Crowd Systems and Motion Synthesis

CGI Techniques Background Research

Derek Gokstorp
i7632019

Abstract

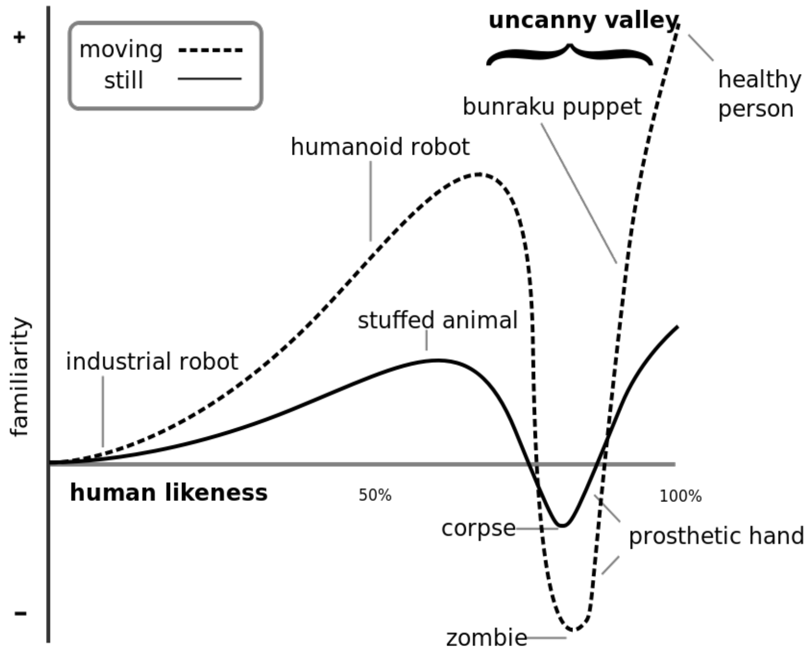
Crowds and flocks are used heavily in computer animation in film. The systems and software used to set up crowds are often extremely complex and very difficult and time consuming to use. There have been a myriad of approaches to crowd simulation in the past. This paper will show that using a variety of these approaches in concert it is possible to get fairly realistic motion from a limited variety of motion clips. Used in conjunction with using a rule based crowd system it is possible to get an easy to use and direct-able rudimentary crowd tool.

Introduction

Historically crowds were created with ‘extras’; basically lay people paid to portray the crowd. They give extremely realistic results as in films like: *The Wizard of Oz* (1939), *The Ten Commandments* (1956), *Ben Hur* (1959), *Sparticus* (1960), *The Last Emperor* (1987), *Stargate* (1994), and *Braveheart* (1995). However this can be extremely expensive, very time consuming and sometimes even impossible due to practicality (think fully cg environments, or areas sensitive to large groups of people) or financial constraints. Serious research into implementing cg crowds is a fairly new field, and wasn’t a feasible research topic until about 30 years ago.

Background Research

Mori (1970) described the ‘uncanny valley’. At the time Mori was doing early research into robotics and decided to write about an ‘intuition’ or ‘feeling’ he had while looking at wax dolls and robotic hands according to an interview he gave to Kageki (2012). The basic principle behind the uncanny valley is that as an artificial entity’s appearance becomes more and more like a true human likeness, a human observer’s affinity response will become higher and higher. Until the likeness starts to approach a high degree of likeness to a human and then there is a sharp valley followed by a rise again (see below – note: ‘familiarity’ on the y axis should actually read ‘affinity’ this graphic was from a very old translation). Note the exaggeration of the effect with motion.



Mori's Uncanny Valley.

At the time the research was virtually unnoticed, however, today it is important in many fields – especially cg.

Hogan (2003) discusses the fact that stories have been an important part of human history (from very early to present day), and has been shown to affect our emotional states. Green (2004), argues that if the right preconditions exist while and after listening to and interpreting a story narrative transport and narrative effect can occur. This field has mostly been taken for granted in terms of the movie industry, but heavily researched in marketing in order to influence our decisions on what we buy and what we will believe in commercials according to Keen (2007). While film may not necessarily want you to buy something (other than itself perhaps), in order for it to be effective you must buy *into* the story itself.

Recognition is related to affinity and there is research from Kandel et al. (2000), Pratt et al. (2010) Bidet-Ildei et al. (2006) and Goodale & Milner (1992) that indicates that the ability to recognize natural motion lays in the posterior temporal sulcus and may be innate. But more importantly, activation of this area of the brain can also *impose* covert attention. This makes sense from an evolutionary standpoint – if you see a tiger, bear, or another person wielding a club moving toward you – it would be to your advantage to take notice and pay attention to it! It is important to understand that having characters that unintentionally creep into the uncanny valley can very quickly kill the audience's suspension of disbelief and effectively destroy narrative transport and narrative effect. This is why natural looking motion is so important in the field of visual effects.

Reynolds (1987) first proposed a simple system to describe the motion of flocks, schools and herds of what he described as bird-oids or boids (a play on the New York pronunciation of the word ‘birds’ as well) meant to represent any type of flocking herding or schooling thing. The system composed of the boids being ‘aware’ of the other boids (and their motion) around itself and adjusting its own direction accordingly based on simple rules. Simple simulations using this method are mesmerizing to watch.

There has been a plethora of research into different techniques of generating crowd-like motion: cellular automata, by Blue & Adler (1998), social force model, Helbing & Molnar (1995), behaviouristic models, Robin et al. (2009) etc. There are some inherent problems with modeling crowds this way for film for two main reasons:

1. These are simulations, which can be time consuming to author and create and not necessarily directable or editable (removing one character can change the entire simulation).

2. These types of models rarely take into consideration the motion data that is available for individual characters and if they do they are extremely slow and do not consider how to transition realistically from one motion to the next.

Additionally there has been some excellent research into motion synthesis in terms of blending clips together:

Jehee & Yong (1999), build on earlier work by Lee et al. (1997). They propose using a multilevel b-spline to manipulate a curve (instead of a surface). Using constraints and an inverse kinematic solver they calculate keyframes for deriving the displacement map, and then apply the multilevel b-spline to do the derivation.

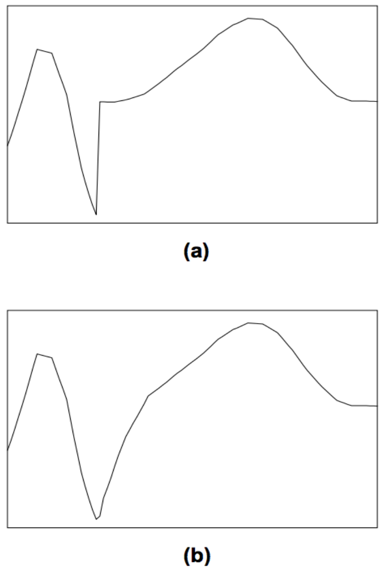


Figure 1: *Graphs of a joint angle over time: (a) two animations are concatenated resulting in a discontinuity, (b) the same animations are blended with a fall-off.*

Mizuguchi et al. (2001) propose a much simpler linear blend between poses – Figure 1 to the left shows the combined channels with and without linear blending. They go on to discuss an interactive transition editor that would allow an animator or technical director to paramatize the transition(s) as well as interactively preview them. Although initially proposed for interactive games – this could easily be applied to transitions for film.

.

Kovar et al (2002), go into more detail giving an excellent outline of the conditions that should be met when considering transition points. They point out that although motion is normally defined by a root position, and joint angle representations comparing these directly are not very useful in terms of motion transition. Global position is irrelevant as the character can be translated to be in a similar position as any other character as well as rotated around the vertical axis. Velocities and accelerations can also be used in conjunction with the relative orientations of joints to determine optimal transition points in the data.

Wang and Bodenheimer (2004) give methods for determining a visually appealing length for motion transitions. Interestingly, they also do informal testing to see how people respond to the three methods they use with fairly promising results. Impressively, they are also the only study surveyed here to show video of the resultant motions. The method that showed most promise first calculates the cost for transitioning from frame *i* to frame *j* with blend length *b*, the equation for calculating this is below:

$$D_{f_i f_j} = \sum_{t=1}^{b} \frac{d_{f_i f_j t}}{b}.$$

Here $d_{f_i f_j t}$ is the difference between two corresponding frames which is equal to:

$$d_{f_i f_j t} = \sum_{k=1}^{m} w_k \left\| \log \left(q_{j-b+t,k}^{-1} q_{i+t-1,k} \right) \right\|^2$$

m is the number of non-global joints in the skeleton and

$$q_{i,k}, q_{j,k}$$

represent orientations (expressed as quarternions) of joint *k* at frames *i* and *j*. The weighted term w_k described in Wang and Bodenheimer (2003). The log-norm term is the geodesic norm in quarternion space. And global degrees of freedom were blended using Rose et al. (1998).

Industry

Since the late 1990's and early 2000's digital crowd shots have become more and more commonplace in film. Industrial Light and Magic's (ILM) film The Phantom Menace was one of the first films to have large-scale digital effects crowds. According to Lucasfilm Ltc (unknown), and Stansweet & Vilmur (2007), during the pod race on Tatooine, ILM special effects artists hundreds of coloured 'q-tips' as crowd characters. They then had fans blowing on them to create the illusion of movement. Since the shot was from such a far distance it is impossible to see the characters close enough to distinguish it. Although these were not digital effects (special effects are done in camera – whereas digital effects are computer generated), they help to illustrate that far crowd shots need not have realistic motion. The same cannot be said for mid-range and close to camera shots.

Crowd shots (especially human crowd shots) are much more commonplace in films today and the techniques used to generate them have become much more complex. Some of the largest and memorable crowd scenes have been the hundreds of thousands of orcs and elves in the *Lord of the Rings* trilogy. A software system that was created to deliver these shots was invented and named Massive. Massive is an acronym that stands for Multiple Agent Simulation System in Virtual Environment and is effectively an artificial intelligence system. Rankin (2014) states:

> Massive is virtually synonymous with CG crowds and by default many facilities simply buy Massive and figure out how to make it work within (or more likely alongside) their existing pipeline.

Massive has several failings when it comes to creating direct-able crowds. Massive was created with the idea that each crowd individual or agent should be 'intelligent', capable of being aware of it's environment and other agents and re-acting to them. Rankin (2014), describes the characters in *Lord of the* Rings as having over 200 motion capture clips and having a 'brain' to navigate the resultant motion graph. Because it is a simulation – if a single element or agent is changed the entire shot needs to be re-simmed and re-rendered. He goes on to say:

> Another problem with this approach is that it produces motion that looks like a video game for two main reasons:
>
> 1) In order to be able to build a reasonably efficient motion graph, you have to have all the actions begin and end in one of a limited number of poses, which is unnatural for the performer and the repetition becomes subtly apparent to the viewer
>
> 2) The motion graph approach involves transitioning from one clip to another as they pass through those aforementioned common poses, which inevitably have different speed and acceleration characteristics causing a non-physical change in momentum. Because the clips are so short, in the duration of a given shot, most crowd characters are going to perform a transition and all of those subtly wrong momentum changes account for the recognizable "Massive look" of so many crowd shots in major films
>
> Finally, these so-called intelligent agents are not naturally equipped to follow the sort of instructions or timings that a director wants. I typically contrast Massive agents to extras on set who are not hired for their intelligence, but rather their obedience. And worse, like humans, these agents ended up using less than 10% of their brains because once we finally got into shot production we found that only a small subset of the actions were regularly used and indeed much of the motion capture and brain logic never came into play at all.
>
> The Massive paradigm is also one that favours introducing diversity primarily through procedural variation and as such there was only one mocap performer captured for each class of agent. So in LOTR, every orc with a 2-handed axe is actually the same orc, with slightly randomized overall scale and leg length and every elf with bow and arrow is actually the same elf. And so on. The nature of the project and the types of characters involved was on our side, though. There were many different types of weapons that the orcs used, so their army could be a jumble of a dozen different character types and the repetition was hidden. Conversely, the elves were intended to be a regimented and homogeneous army, so the repetition was beneficial. But such amenable requirements cannot be relied upon in general.
>
> Procedural variation also has the effect of degrading the physicality of the motion capture because in reality movements don't scale linearly

A.L.I.C.E is proprietary program used in the industry at the Moving Picture Company (MPC), of it Rankin (2014) states:

> Having evaluated Massive, MPC initially set out to reverse engineer it, but with a team of programmers (including one with a doctorate in biomechanics) working alongside a team of crowd TDs, we quickly found more efficient ways to do many of the things that Massive did poorly.
>
> One of the fundamental improvements in the ALICE engine is that the motion blending algorithm is able to automatically find the best transition zones between two clips, minimizing the unnatural changes in momentum that occur in Massive and removing both the need to pass through repeated poses and the manual process of setting in and out points on every clip that Massive requires.
>
> In parallel with this, we made a change to the way we planned and directed motion capture. Recognizing that in most shots no character will perform more than two actions, we captured all the actions that should naturally transition from one to the next, together. So having reduced the issues with clip transitions, we also reduced the number of clip transitions that would appear on screen.
>
> This approach also lent itself to having incredibly lightweight brains, because each agent would only have enough logic to control the one or two mocap clips it would need to execute in a given shot. It might sound like this requires more shot-specific brain building, but in fact, with reuse of logic across agents and shots, it results in considerably less overall brain building than the exhaustive Massive approach.
>
> …
>
> One of the casualties of this intensive era of development was what we called “pose caches” that were used extensively on *Kingdom of Heaven*. A pose cache was a very efficient way to populate a standing or sitting crowd from a library of clips of similar behaviour cached at the origin. With a pose cache you wouldn't need any brain logic at all, just a layout of positions and orientations that would be substituted for clothed characters at render time. By the time we were working on *Prince Caspian*, standing crowds were being populated more in the fashion of Massive, albeit with much lighter weight logic to control them.
>
> We expanded on the idea of avoiding clip transitions by choreographing and capturing entire fight sequences between pairs and larger groups of opponents. These fight vignettes could be placed by hand in the foreground and mid-ground of shots, while it was possible to extract loops from them to distribute randomly in the background where a convincing melee is just a chaotic blur of swinging limbs and weapons.

References:

Ben Hur, 1959. [film, DVD]. Directed by William Wyler. USA: Loew's Inc.

Bidet-Ildei, C., Orlianguet, J.P., Sokolov, A.N. & Pavlova, M., (2006). *Perception of elliptic biological motion* in *Perception* 35(8) 1137-1147.

Blue, V.J., & Adler J.L., (2001). *Cellular Automata microsimulation for modeling bi-directional pedestrian walkways* in *Transportation Research Part B: Methodological* 35, 293-312.

Braveheart, 1995. [film, DVD]. Directed by Mel Gibson. USA: Paramount Pictures.

Goodale, M. A. & Milner, A. D., (1992). *Separate Visual Pathways for Perception and Action* in *Trends Neuroscience 15 (1): 20-5.*

Green, M. (2004) *Transportation into Narrative Worlds: the Role of Prior Knowledge and Perceived Realism* in *Discourse Processes,* 38 247-266.

Helbing, D. & Molnar, P., (1995). *Social Force Model for Pedestrian Dynamics* in *Physical Review – Part E* 51, 4282-4286.

Hogan, P.C., (2003). *The Mind and it's Stories: Narrative Universals and Human Emotion,* Cambridge: Cambridge University Press.

Jehee, L., & Yong, S. S., (1999). *A hierarchical approach to interactive motion editing for human-like figures* in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 39–48.

Kandel, E. R., Jessel, T. M., & Sanes, J. R. (2000). "Chapter 27: Central Visual Pathways", *Principles of Neural Science* (4 ed.), New York: McGraw-Hill.

Kageki, N. (2012) *An Uncanny mind: Masahiro Mori on the Uncanny Valley and Beyond* [online].

Keen, S., (2007). *Empathy and the Novel,* New York: University Press 34.

Kovar, L., Gleicher, M. & Pighin, F., (2002). *Motion graphs* in *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 473–482.

Lee, S., Wolberg, G., & Shin, S.Y., (1997). *Scattered data interpolation with multilevel b-splines* in *IEEE Transactions on Visualization and Computer Graphics* 3(3):228-244.

Lucasfilm Ltd, unknown. *Star Wars: Episode I The Phantom Menace* [online]. Lucasfilm Ltd. Available from: http://lucasfilm.com/star-wars-episode-1-the-phantom-menace [Accessed 14 November 2014].

Mar, R.A., Oatley, K., Hirsh, J., dela Paz, J. & Peterson, J.B., (2006). *RE: Bookworms versus Nerds: Exposure to fiction vs non-fiction, divergent association with social ability and the simulation of fictional social worlds* in *Journal of Research in Personality,* 40 649-712.

Mar, R.A., Kelley, W.M., Heatherton, T.F. & Macrae, C.N., (2007). *Detecting agency from the biological motion of veridical versus animated agents* in *Social Cognitive and Affective Neuroscience,* 2 199-205.

Mizuguchi, M., Buchanan, J., & Calvert, T. (2001). *Data Driven Motion Transitions for Interactive games* in *Proceedings of Eurographics Short Presentations.*

Mori, M. (1970) *The Uncanny Valley* trans. MacDorman, K. F. & Kageki, N. (2012) *IEEE Robotics & Automation Magazine* 19:2: 98–100.

Pratt, J., Radulescu, P. V., Guo, R. M. & Abrams, R. A., (2010). *It's Alive! Animate Motion Captures Visual Attention* in *Psychological Science* (11) 21, 1724-1730.

Pool, P.M., (1986). *The Incredible Machine*. Washington D.C.: The National Geographic Society.

Rankin, O., (2014). pers.comm., 12 November.

Reynolds, C. W. (1987) *Flocks, Herds, and Schools: A Distributed Behavioral Model*, in *Computer Graphics*, 21 (4) (SIGGRAPH '87 Conference Proceedings) 25-34.

Robin, T., Antonini, G., Bierlaire, & M., Cruz, J., (2009). *Specification, Estimation and Validation of a Pedestrian Walking Behaviour Model* in *Transportation Research – Part B* 43, 36-56.

Sparticus, 1960. [film, DVD]. Directed by Stanley Kubrick. USA: Universal Pictures.

Stansweet, S. & Vilmur, P., (2007). *The Star Wars Vault: Thirty Years of Treasures From the Lucasfilm Archives.* New York: Simon and Schuster.

Stargate, 1994. [film, DVD]. Directed by Roland Emmerich. USA, France: Metro-Goldwyn-Mayer, Pathé.

The Last Emperor, 1987. [film, DVD]. Directed by Bernardo Bertolucci. China, United Kingdom, Italy: Columbia Pictures Entertainment, Inc.

The Wizard of Oz, 1939. [film, DVD]. Directed by Victor Fleming. USA: Metro-Goldwyn-Mayer.

The Ten Commandments, 1956. [film, DVD]. Directed by Cecil B. DeMille. USA: Paramount Pictures.

Wang, J., & Bodenheimer, B., (2004). *Computing the duration of motion transitions: an empirical approach* in *SCA '04 Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer Animation* 335-344.