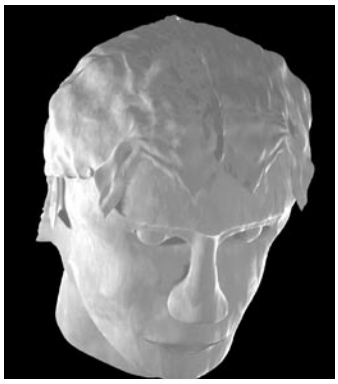
Computer facial animation




Computer facial animation is primarily an area of computer graphics that encapsulates models and techniques for generating and animating images of the human head and face. Due to its subject and output type, it is also related to many other scientific and artistic fields from psychology to traditional animation. The importance of human faces in verbal and non-verbal communication and advances in computer graphics hardware and software have caused considerable scientific, technological, and artistic interest in computer facial animation.

Although the development of computer graphics methods for facial animation started in the early 1970s, major achievements in this field are more recent and have taken place since the late 1980s.

Computer facial animation includes a variety of techniques from morphing to three -dimensional modelling and rendering. It has become well-known and popular through animated feature films and computer games but its applications include many more areas such as communication, education, scientific simulation, and agent-based systems (for example, online customer service representatives).

rendering
the process of generating an image from a model by means of computer programs

History

Human facial expressions have been the subject of scientific investigation for more than one hundred years. The study of facial movements and expressions started from a biological point of view. After some older investigations, i.e. by John Bulwer in late 1640s, Charles Darwin's book The Expression of the Emotions in Men and Animals can be considered a major departure for modern research in behavioural biology.

More recently, one of the most important attempts to describe facial activities (movements) was the Facial Action Coding System (FACS). Introduced by Ekman and Friesen in 1978, FACS defines 64 basic facial Action Units (AUs). A major group of these Action Units represent primitive movements of facial muscles in actions such as raising brows, winking, and talking. Eight AUs are for rigid three-dimensional head movements, i.e. turning and tilting left and right and going up, down, forward and backward. FACS

morphing
a special effect in motion pictures and animations that changes (or morphs) one image into another through a seamless transition

keyframe
or a key frame in animation and film making is a drawing which defines the starting and ending points of any smooth transition texture - a bitmap image applied to a surface in computer graphics

has been successfully used for describing desired movements of synthetic faces and also in tracking facial activities.

Computer based facial expression modelling and animation is not a new endeavour. The earliest work with computer based facial representation was done in the early 1970s. The first three-dimensional facial animation was created by Parke in 1972. In 1973, Gillenson developed an interactive system to assemble and edit line drawn facial images. And in 1974, Parke developed a parameterized three-dimensional facial model.

The early 1980s saw the development of the first physically-based muscle-controlled face model by Platt and the development of techniques for facial caricatures by Brennan. In 1985, the short animated film Tony de Peltrie was a landmark for facial animation; for the first time computer facial expression and speech animation were a fundamental part of telling the story.

The late 1980s saw the development of a new muscle-based model by Waters, the development of an abstract muscle action model by Magnenat-Thalmann and colleagues, and approaches to automatic speech synchronization by Lewis and by Hill. The 1990s saw increasing activity in the development of facial animation techniques and the use of computer facial animation as a key storytelling component as illustrated in animated films such as Toy Story, Antz, Shrek, and Monsters, Inc, and computer games such as Sims. Casper (1995) is a milestone in this period, being the first movie with a lead actor produced exclusively using digital facial animation (Toy Story was released later the same year).

computer vision
the science and technology of machines that see

The sophistication of the films increased after 2000. In The Matrix Reloaded and Matrix Revolutions dense optical flow from several high-definition cameras was used to capture realistic facial movement at every point on the face. Polar Express used a large Vicon system to capture upward of 150 points. Although these systems are automated, a large amount of manual clean-up effort is still needed to make the data usable. Another milestone in facial animation was reached by The Lord of the Rings where a character specific shape base system was developed. Mark Sagar pioneered the use of FACS in entertainment facial animation, and FACS based systems developed by Sagar were used on Monster House, King Kong, and other films.

Techniques

2D Animation

Two-dimensional facial animation is commonly based upon the transformation of images, including both images from still photography and sequences of video. Image morphing is a technique which allows in-between transitional images to be generated

between a pair of target still images or between frames from sequences of video. These morphing techniques usually consist of a combination of a geometric deformation technique, which aligns the target images, and a cross-fade, which creates the smooth transition in the image texture. An early example of image morphing can be seen in Michael Jackson's video for Black And White. In 1997 Ezzat and Poggio working at the MIT Center for Biological and Computational Learning created a system called Miketalk, which morphs between image keyframes, representing visemes, to create speech animation.

Another form of animation from images consists of concatenating together sequences captured from video. In 1997 Bregler et al. described a technique called video-rewrite, where existing footage of an actor is cut into segments corresponding to phonetic units which are blended together to create new animations of a speaker. Video -rewrite uses computer vision techniques to automatically track lip movements in video and these features are used in the alignment and blending of the extracted phonetic units. This animation technique only generates animations of the lower part of the face, these are then composited with video of the original actor to produce the final animation.

computer vision
a branch of artificial intelligence that deals with computer processing of images from the real world

3D Animation

Three-dimensional head models provide the most powerful means of generating computer facial animation. One of the earliest works on computerized head models for graphics and animation was done by Parke. The model was a mesh of 3D points controlled by a set of conformation and expression parameters. The former group controls the relative location of facial feature points such as eye and lip corners. Changing these parameters can re-shape a base model to create new heads. The latter group of parameters (expression) are facial actions that can be performed on a face, such as stretching lips or closing eyes. This model was extended by other researchers to include more facial features and add more flexibility. Different methods for initializing such "generic" models based on individual (3D or 2D) data have been proposed and successfully implemented. The parameterized models are effective due to the use of limited parameters, associated with the main facial feature points. The MPEG-4 standard defines a minimum set of parameters for facial animation.

Animation is done by changing parameters over time. Facial animation is approached in different ways. Traditional techniques include:

1. shapes/morph targets,
2. bones/cages,
3. skeleton-muscle systems,
4. motion capture on points on the face and

5. knowledge based solver deformations.

1. Shape based systems offer a fast playback as well as a high degree of fidelity of expressions. The technique involves modelling portions of the face mesh to approximate expressions and visemes and then blending the different sub meshes, known as morph targets or shapes. Perhaps the most accomplished character using this technique was Gollum, from The Lord of the Rings. Drawbacks of this technique are that they involve intensive manual labor, are specific to each character and must be animated by slider parameter tables.

2. 'Envelope Bones' or 'Cages' are commonly used in games. They produce simple and fast models, but are not prone to portray subtlety.

3. Skeletal Muscle systems, physically-based head models form another approach in modelling the head and face. Here the physical and anatomical characteristics of bones, tissues, and skin are simulated to provide a realistic appearance (e.g. spring-like elasticity). Such methods can be very powerful for creating realism but the complexity of facial structures make them computationally expensive and difficult to create. Considering the effectiveness of parameterized models for communicative purposes (as explained in the next section), it may be argued that physically-based models are not a very efficient choice in many applications. This does not deny the advantages of physically-based models or the fact that they can even be used within the context of parameterized models to provide local details when needed. Waters, Terzopoulos, Kahler, and Seidel (among others) have developed physically-based facial animation systems.

4. Motion capture uses cameras placed around a subject. The subject is generally fitted either with reflectors (passive motion capture) or sources (active motion capture) that precisely determine the subject's position in space. The data recorded by the cameras is then digitized and converted into a three-dimensional computer model of the subject. Until recently, the size of the detectors/sources used by motion capture systems made the technology inappropriate for facial capture. However, miniaturization and other advancements have made motion capture a viable tool for computer facial animation. Facial motion capture was used extensively in Polar Express, where hundreds of motion points were captured. This film was very accomplished and while it attempted to recreate realism, it was criticised for having fallen in the 'uncanny valley', the realm where animation realism is sufficient for human recognition but fails to convey the emotional message. The main difficulties of motion capture are the quality of the data which may include vibration as well as the retargeting of the geometry of the points.

5. Deformation Solver Face Robot.

alignment
the adjustment of an object in relation to other objects, or a static orientation of some object or set of objects in relation to others

Speech Animation

Speech is usually treated in a different way to the animation of facial expressions; this is because simple keyframe-based approaches to animation typically provide a poor approximation to real speech dynamics. Often visemes are used to represent the key poses in observed speech (i.e. the position of the lips, jaw and tongue when producing a particular phoneme); however, there is a great deal of variation in the realisation of visemes during the production of natural speech. The source of this variation is termed coarticulation, which is the influence of surrounding visemes upon the current viseme (i.e. the effect of context.) To account for coarticulation, current systems either explicitly take into account context when blending viseme keyframes or use longer units such as diphone, triphone, syllable or even word and sentence-length units.

One of the most common approaches to speech animation is the use of dominance functions introduced by Cohen and Massaro. Each dominance function represents the influence over time that a viseme has on a speech utterance. Typically the influence will be greatest at the center of the viseme and will degrade with distance from the viseme center. Dominance functions are blended together to generate a speech trajectory in much the same way that spline basis functions are blended together to generate a curve. The shape of each dominance function will be different according to both which viseme it represents and which aspect of the face is being controlled (e.g. lip width, jaw rotation etc.) This approach to computer-generated speech animation can be seen in the Baldi talking head.

Other models of speech use basis units which include context (e.g. diphones, triphones etc.) instead of visemes. As the basis units already incorporate the variation of each viseme according to context and to some degree the dynamics of each viseme, no model of coarticulation is required. Speech is simply generated by selecting appropriate units from a database and blending the units together. This is similar to concatenative techniques in audio speech synthesis. The disadvantage to these models is that a large amount of captured data is required to produce natural results, and whilst longer units produce more natural results, the size of database required expands with the average length of each unit.

Finally, some models directly generate speech animations from audio. These systems typically use hidden Markov models or neural nets to transform audio parameters into a stream of control parameters for a facial model.

motion capture
a technique of recording the actions of human actors and using that information to animate digital character models in 3D animation

texture
a bitmap image applied to a surface in computer graphics