Visual Comput (2008) 24: 13–30
DOI 10.1007/s00371-007-0175-y

ORIGINAL ARTICLE

Nikolaos Ersotelos
Feng Dong

Building highly realistic facial modeling and animation: a survey

N. Ersotelos (✉) · F. Dong
Department of Information Systems and Computing, Brunel University, Uxbridge, UB8 3PH, UK
Nikolaos.Ersotelos@brunel.ac.uk

Abstract This paper provides a comprehensive survey on the techniques for human facial modeling and animation. The survey is carried out from two different perspectives: facial modeling, which concerns how to produce 3D face models, and facial animation, which regards how to synthesize dynamic facial expressions. To generate an individual face model, we can either perform individualization of a generic model or combine face models from an existing face collection. With respect to facial animation, we have further categorized the techniques into simulation-based, performance-driven and shape blend-based approaches. The strength and weakness of these techniques within each category are discussed, alongside with the applications of these techniques to various exploitations. In addition, a brief historical review of the technique evolution is provided. Limitations and future trend are discussed. Conclusions are drawn at the end of the paper.

Keywords Facial modeling · Facial expression and animation · Generic model adaptation · Morphable modeling · Pseudo muscles · Performance-driven animation · Blend shapes

1 Introduction

Facial modeling and animation have been a research challenge and focus for many years. They play the most substantial role in depicting human characters. The recent advance in facial animation that allows us to produce a rich set of stunning effects on synthetic humans has already brought profound impact on the industry. Meanwhile, within the computer graphics community, new efforts are still emerging and research interests in synthesizing high quality facial animation show no sign of abating.

Examples of early work in facial modeling and animation include [27] and [3]. These works have generated very simple and artificial looking face models and expressions (e.g., models with connected vertical and horizontal lines in [3]). Moreover, the control of facial animation in these works involved a complicated parameterization process and hence appeared to be difficult for untrained users.

However, since the appearance of these pioneer works, significant progress has been materialized by the researchers from the computer graphics community, which have developed a large number of techniques to generate high quality face models and highly realistic facial expressions. However, despite this progress, the existing computer synthesized human facial animation still requires costly resources and sometimes involves considerable manual labors. Furthermore, the outcomes are not yet completely realistic. Therefore, at the current stage, solutions are cost effective, but fully realistic facial animation is still not entirely available.

This survey aims at providing a comprehensive survey for the existing techniques in the area of facial modeling and animation, giving analysis to the strength and weakness for a wide range of techniques. Here we pay special attention to more recent techniques, which allow the production of highly realistic results, as compared to other surveys given in the past [24, 26]. In particular, the analysis on the latest techniques allows us to look into their suitability to different applications and foresee the future research trend in this area. In addition, the survey also provides a historical view on the evolution of these techniques.

The survey is carried out from two perspectives:

- Face modeling, where we introduce the techniques for producing high quality face models up to the latest.
- Facial animation, where we concentrate on the techniques that allow facial animations with high realism.

Facial modeling and facial animation are two strongly interrelated issues. In fact, generating realistic facial animation often involves modeling techniques, for example, to build multiple face layers such as in [20], or to carry out deformation on the face models for desired facial expressions. Therefore, the quality of facial animation is determined by both the employed methods of facial modeling and facial animation. Such a relationship is demonstrated further in our discussions in the rest of the paper.

Figure 1 outlines the structure of the techniques that are covered by this survey. The details of these techniques are given in the following sections.

This paper is structured as follows: Sect. 2 provides a historical view on the technical evolution; Sect. 3 describes two main approaches for face modeling, generic model individualization (Sect. 3.1) and example-based face modeling (Sect. 3.2). Section 4 introduces three main approaches for facial animation, including simulation-based approach (Sect. 4.1), performance-driven animation (Sect. 4.2) and blend shape-based approach (Sect. 4.3). Section 5 provides analysis to their strength and weakness from an application perspective; Sect. 6 gives the limitation of the current techniques and the future research trend. Finally, the conclusion is given in Sect. 7.

2 Brief history

Great interest has been received in computer simulation of human faces and their movements during last few decades. An early example of success was the facial action coding system (FACS), which was introduced by Ekman and Friesen in 1978 to describe primitive facial activities. Early work on computer facial modeling and animation dated back to 1970s, during which the first 3D facial animation was created by Parke [27]. This was followed by a few landmark works in 1980s, which included the deformable face model from [3], the classic work on facial animation using pseudo muscles from [42].

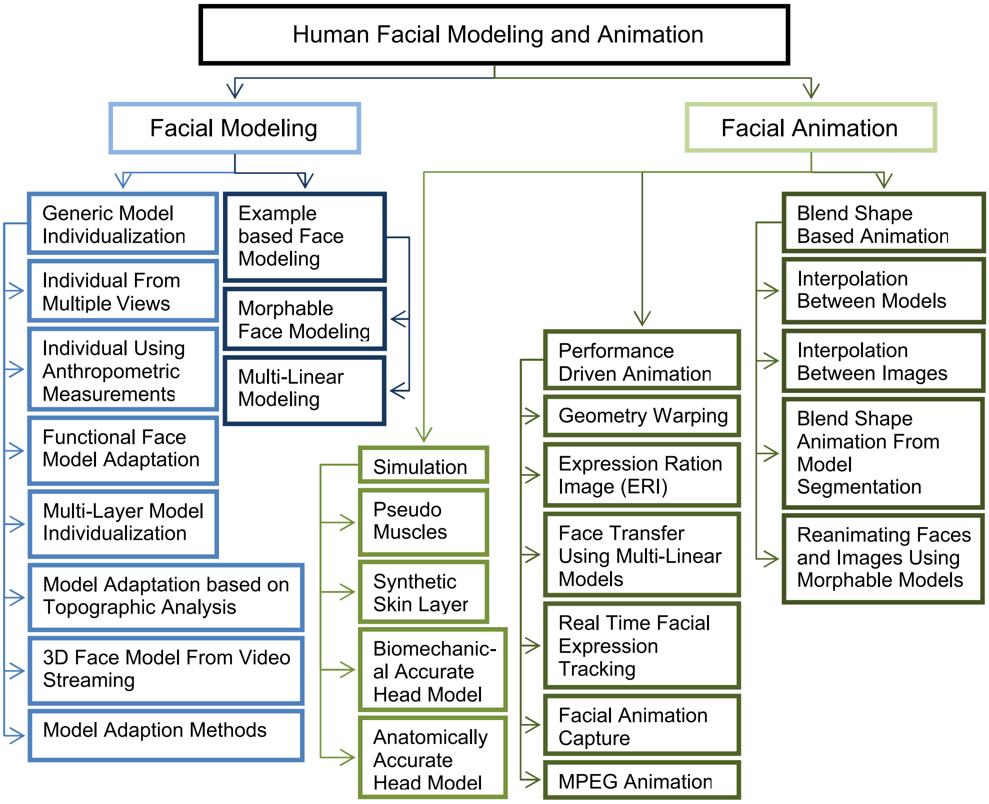

Fig. 1. Categorization of facial modeling and animation methods

Subsequently, a much larger number of significant works were published during the 1990s. Apart from those inherited the pseudo-muscle-based approach, such as in [38], many researchers strived to target high quality face models by adopting natural face measures for the modeling, such as anthropometrics, [8]. Meanwhile, people also started to use scanned 3D models from scanners, which created face models with a great deal of detail and hence allowed them to largely overcame "cartoon styled" artifacts, [4]. To produce facial animation on the scanned models, pseudo muscles needed to be located, [20]. In addition, aiming at a natural looking facial animation, cameras started to be employed to capture facial movements, such as in [12, 31], which constituted the early stage of performance-driven facial animation.

More recently, with the rapid advance of hardware, the performance-driven approach has played a more significant role. For example, high quality facial expression with fine details was created from a set of existing photos using different techniques, as in [5, 14, 23, 46, 47]. Meanwhile, video analysis on human faces was applied to capture facial movements and hence improve the realism of synthesized facial animation, such as in [6, 40].

On the other side, with the aim of improving the comprehension on the mechanism of facial movements, anatomy-based high quality face models have been built in recent years. This has included the multi-layer model from [15], which simulated a head model with skin, muscle, skeleton, etc, and the volumetric model of a human head from [37], which covered a range of head structures including muscles, tissues etc.

However, despite the recent technical advance within this area, the currently available techniques are still not able to meet the requirement from many envisaged applications. To compromise with limited computation resources, considerable amount of simplification has to be made in anatomy-based face models. Also, due to the limitation from the current image and video analysis, the performance-driven approach has to involve a large number of equipments and the accuracy is still subjective to further enhancement. Therefore, facial modeling and animation is still an on-going research issue, and there is a long way before the satisfactory completion of the technology.

3 Face modeling

The aim of face modeling is to generate realistic face models with high visual fidelity. The basic way to represent the shape of a face is to use a triangle mesh. A more complex model involves multiple layers which mimic the anatomical structure of a face.

With the advance of data capturing hardware, many 3D face models have been created via using 3D laser scanner. On the other side, over the last few years, researchers have proposed and developed various techniques on producing quality face models. These techniques can be divided into two categories:

- Generic model individualization, which is based on the idea of creating a face model for a specific subject by carrying out feature-based deformation to a generic model – more details are given in Sect. 3.1.
- Example-based face modeling. This is to create a face model with desired facial features through the linear combinations of an existing face model collection – more details are given in Sect. 3.2.

3.1 Generic model individualization

Generic model individualization generates a facial model for a specific individual through the deformation of a generic model. This is also named as model adaptation. Given the positions of some selected facial features from an individual face, such as eyes corners, mouth and nose positions, the adaptation generates the model for the individual by aligning the corresponding facial features of the generic model towards these given feature positions.

As the basic inputs of the generic model individualization approach, the features the individual face can be positioned in different ways. For example:

1) In [31], the feature positions are given manually through a number of multi-viewed photographs. For more details, see Sect. 3.1.1.
2) In [8] anthropometric measurements are used to describe faces with particular features. For more details, see Sect. 3.1.2, 3.1.3 and 3.1.4.
3) In [45] the individual face is described by a frontal view image, and image analysis is used to detect the face features. For more details, see Sect. 3.1.5.
4) Face features are given by data tracking of video streams using multiple cameras in [49] For more details, see Sect. 3.1.6.

Notably, all the model individualization methods involve deforming a generic model to an individual model. The techniques involved in this process decide the efficiency and effectiveness of the individualization. We will discuss these techniques in Sect. 3.1.7.

3.1.1 Individualization from multiple views

Generating face model constitutes an important step in the work of Pighin et al. [31] which presents a method of generating facial expressions from photographs. To create a model for an individual subject, the method adapts a generic face model to the subject, which is portrayed by a number of photographs (images) taken from different view angles. To assist this adaptation process, features points such as eye corners, nose tip, mouth corners are manually identified in these images – see Fig. 2. A scattered data interpolation technique is used to deform the

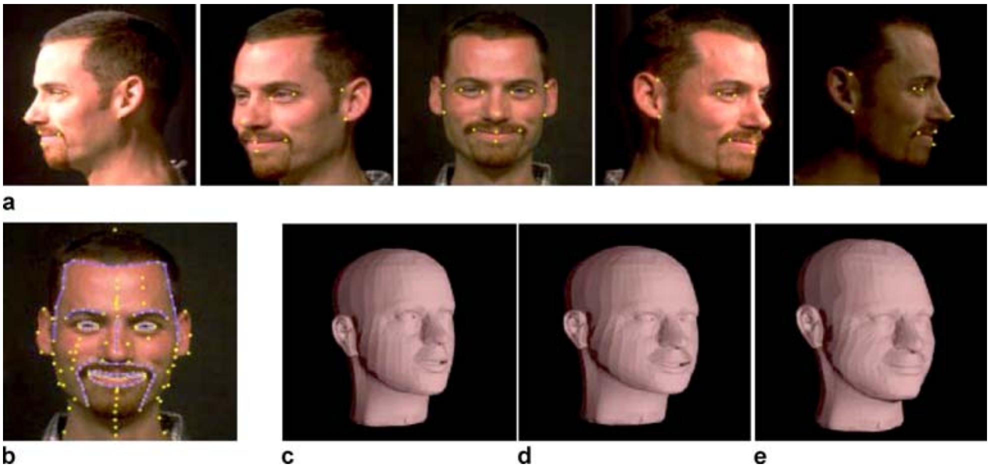

Fig. 2a–e. Model-fitting process: **a** a set of input images with marked feature points, **b** facial features annotated using a set of curves, **c** generic face geometry (shaded surface rendering), **d** face adapted to initial 13 feature points (after pose estimation) **e** face after 99 additional correspondences have been given [31]

generic model to fit into the feature points identified from the images.

Furthermore, the work of [31] goes beyond face modeling towards facial animation. To do this, the above modeling process has to be repeated for several different facial expressions of the same subject. This creates a facial model for each facial expression. Then a 3D shape morphing technique is applied to these facial models to generate transactions in between these models to create facial animations. This issue will be further discussed Sect. 3.3.

However, this proposed method of face modeling involves considerably large amount of manual work. For example, a large number of facial feature points have to be given manually in the multiple viewed images. This is extremely inconvenient and time consuming. Therefore, this algorithm only can work offline.

Lee et al. [21] perform generic face model individualization using photographs from two views: a front view and a side view. To facilitate this, a generic model is also divided into a number of feature regions around each identified feature, and the individualization is based on the facial features obtained from the feature detection, which is carried out in two successive steps: a global matching, which is used to find locations of facial features using statistical data from the images, and a detailed matching, which recognizes the shape for each facial feature using a specific method designed for the feature.

The global feature matching achieves a high accuracy rate. Among the faces that are tested, which cover a wide range of human races, ages, hair colors, and both genders, the authors find the proper position of 201 features out of 203. However, they do not attempt to handle faces with other accessories, such as glasses.

The detailed matching, which employs multi-resolution edge detection methods, achieves different success rate for different features. For example, the forehead recognition is over 90%, the eyes and mouth extraction are around 80%, and the nose identification reaches about 70%.

After the feature detections from the two input views, the features points from these 2D views are combined into 3D points, based on which the deformation of the generic model takes place. The deformation starts from a global transformation to align the 3D points with the generic model, followed by Dirichlet freeform deformation to match the shape of the generic model against the feature points.

3.1.2 Individualization using anthropometric measurements

To create models that match different individuals, DeCarlo et al. [8] propose a facial modeling approach based on facial anthropometric measurements. These measurements are used as the fundamental elements to describe and generate a wide range of geometrical 3D head models.

Anthropometry is a biological science of measuring human body. More specifically, it stores statistical measurements of human body parts in libraries. These libraries contain data which characterizes human bodies by gender, color and age. Data from the anthropometry studies is used in many applications such as plastic surgery planning, human-factors analysis, and 3D human head construction, etc.

The anthropometric measurement – see Fig. 3 involved in [8] includes around 130 feature points and their relative distances, describing the characteristics of a human face. These feature points are based on the Farkas [10] system. Given such a measurement of an individual face, the algorithm generates a static facial model using variational modeling. Variational modeling is an optimization method

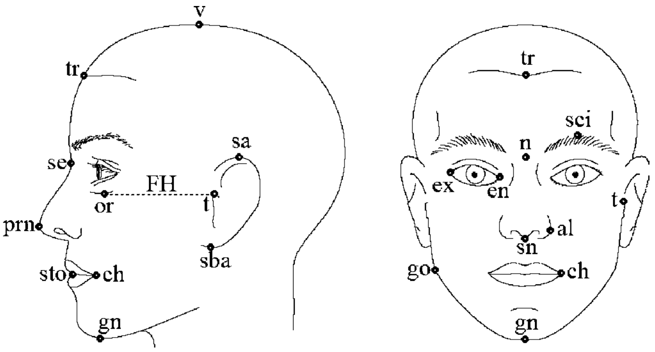

Fig. 3. Anthropometric landmarks on the face [8]

which generates a face model constrained by the features described by the anthropometric measures.

A major disadvantage of using anthropometric measurement as in DeCarlo et al. [8] is that the anthropometric measurements only provide a statistical description of a human face and hence cannot offer more specific human face features such as hooked nose or double chin. Moreover, the employed optimization process involves considerably large computation resource.

3.1.3 Functional face model adaptation

Zhang et al. [48] present an algorithm which allows us to adapt a functional generic face model to an individual face model. The generic model is equipped with a number of pseudo muscles in order to support facial animation – see Sect. 4.1 for more details of pseudo muscle-based facial animation.

The process starts with the specification of a small set of anthropometric landmarks on the 2D images of both the generic and scanned model. The 3D positions of the landmarks are recovered automatically by using a projection mapping approach. A global adaptation is then carried out to adapt the size, position and orientation of the generic model towards the scanned model, referring to a series of measurements based on the recovered 3D landmarks. Following the global adaptation, a local adaptation deforms the generic model to fit all of its vertices to the scanned model. Meanwhile, the underlying muscle structure of the generic model is automatically adapted as well, such that the reconstructed model not only resembles the individual face in shape and color but also allows facial animation from the adapted pseudo muscles.

3.1.4 Multi-layer model individualization

To continue from [8], Kähler et al. [15] present their technique to generate animatable 3D face models. Compared to the models in [8], this work is featured by generating head models with multiple layers to simulate anatomical head structures including skin, muscle, skull, mass-spring, and other separated components (e.g., eyes, teeth, tongue), etc – see Fig. 4. For each model, up to 24 major muscles are used for facial expressions and speech articulations. The skin and muscles are attached to the skull via a mass-spring system. These head models allow real-time animation based on the simulation of facial muscles and elastic skin prosperities.

To facilitate the modeling for individuals, a generic model is provided with the above five layers. Similar to [8], some landmarks are taken from a standard set of the anthropometric literature, which are small dots placed on the model to define the features. These tagged anthropometrically meaningful landmarks allow us to fit the generic model to scanned 3D face models, creating a wide variety of animated face models. Moreover, by using the anthropometric measurements to simulate the growth of a human head, the technique is capable of generating animatable human heads at different ages.

As an extension from this method, Kahler et al. apply the technique to scanned real skull data. As a result, they are able to reconstruct expressive faces from the skull data in an application which is named as "reanimating the dead" [16].

3.1.5 Model adaptation-based on topographic analysis

Besides the use of anthropometric landmark data as introduced in the previous sections, Yin & Weiss [45] suggest the use of topographic representation to give facial features for face model individualization from a generic model. To generate the topographic representation, an

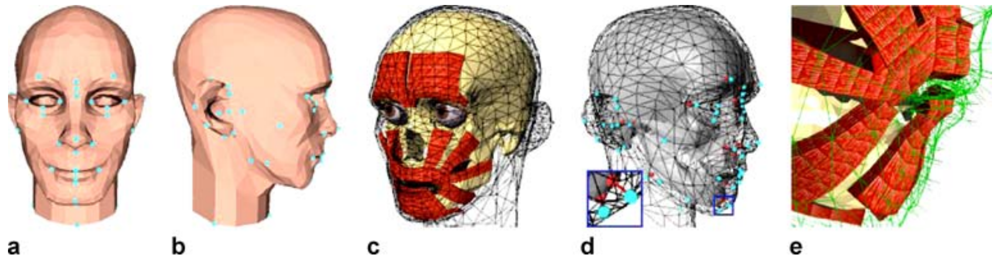

Fig. 4a–e. The reference head: **a** head geometry with landmarks, front view; **b** side view; **c** skull and facial components; **d** skull landmarks related to subset of skin landmarks; **e** facial detail showing spring mesh connecting skin and muscles [15]

input face model is needed. Then a topographic analysis treats the face image as a topographic terrain, and labels each of its pixel as one of the topographic labels, including peak, ridge, saddle, hill, float, ravine, or pit. The hill-labeled pixels are further divided into convex, concave, saddle or slope hill. Following the topographic analysis, the individualization carries out an optimization using the criteria measured by these labels. This optimization adapts a generic model to an individual model.

3.1.6 3D face model from video streaming

The method from Zhang et al. [49] presents an approach of generating facial models and expressions by using multiple video cameras. Video cameras are used to capture model data via stereo matching. A fitting process is then employed to fit a face template to the captured data.

The data capture hardware consists of 6 synchronized video streams (4 monochrome and 2 colors), running at 60 fps (frames per second). This provides a stereo system, which uses three of the cameras to capture from the left side while the other three cameras are from the right. A spacetime stereo algorithm is employed during this data capture, which calculates a time sequence of depth maps using stereo matching.

Then, a template fitting and tracking process is used to fit a face template to the depth maps in the first frame and then perform tracking through the whole sequence. The result of this fitting gives high quality meshes with vertex correspondence over the time. The output of the data capture gives high-resolution 3D meshes at 20 fps sequence, capturing the face geometry, color and motions.

Once acquired, this sequence of models can be interactively manipulated to create expressions using a technique which is named “faceIK” by the authors. More details about the facial animation are given in Sect. 4.3.1.

A main disadvantages of this algorithm is that it is quite resource demanding as it involves special equipments to capture the face geometry and motion. Also, the work is only presented for the animation of a single face without considering variations of different individuals.

3.1.7 Model adaptation methods

All of the above generic model individualization methods involve model adaptation, a technique which deforms a generic model to an individual (target) model. Basically, this is a scatter data interpolation problem, which drives the movement for each vertex on the generic model towards the target model, given only a sparse set of feature point positions as input.

A common approach to solve such a problem is to construct and minimize an interpolation function based on radial basis functions [15, 31]. To further improve the quality of the model adaptation, Kahler et al. [15] develop an automatic procedure which allows us to refine the adaptation by using model subdivision without requiring users to input a large number of dense feature points.

In addition, in the work of DeCarlo et al. [8] employs variational modeling for the model adaptation. It works on B-spline face models. The anthropometric measures are used as linear and non-linear surface constraints and subsequently surface fairing with these constraints are applied. This generates smooth surface models which match the anthropometric features.

Zhang et al. [48] deforms a generic model to an individual model through global and local adaptation. The global adaptation involves size and orientation adaptation based on some landmarks that are semi-automatically identified, while the local adaptation includes moving the model vertices locally to fit into the individual model. Essentially, the global adaptation is a process of face pose recovery and model scaling, and the local adaptation provides small adjustments for the vertices positions according to the local geometry.

In [45], second-order differential equations are used to define the adaptation of the generic model. Within these equations, topographic features are used to define the internal and external forces. The external force drives the deformation of the model, while the internal force maintains the shape of the model during the deformation. The result of the deformation matches the generic model to a single frontal view of an individual face.

In [49], the specific adaptation problem is to fit a generic model to a depth map captured from the videos by using Gauss–Newton optimization. The depth map represents the face geometry by a depth value h at each point (x, y). Similar to [45], the fitting metric has two terms, a depth matching, which measures the difference in depth between the generic and target model, and a regularization term, which maintains a good shape for generic model after the deformation.

3.2 Example-based face modeling

The techniques introduced in this section concern creating face model through the combination of existing models. Such methods require support from a collection of face models. Given desired facial features, for example, by using a face photograph as in [4], optimization method is used to find the right combination coefficients. The linear combination of the collected face models with these optimized coefficients provides a close match between the synthetic model and the desired facial features. Noticeably, linear interpolation is also used to create facial animations from a number of example expressions, such as in [31, 46, 47]. These will be discussed in Sect. 3.

3.2.1 Morphable face modeling

Blanz & Vetter [4] present a face modeling technique named as morphable modeling. A distinct strength of this modeling approach is that it allows generate a wide variety of face modeling with minimal user input. More specifically, a face model can be created from a single photograph supplied by users and the created model matches the facial features portrayed by the photograph.

The method requires an example set of 3D face models. Morphable face modeling is based on transforming the shapes and textures of these example face models into a vector space representation. The shape and texture of a new face model is represented by a linear combination of these transformed vectors. Moreover, this method allows face manipulations using complex parameters, such as gender, fullness of a face and distinctiveness etc.

However, the implementation of the morphable face modeling is not straightforward, since it requires a large collection of 3D face models within which the dense point to point correspondence between the models have to be established. Also, the algorithm is time consuming – it takes tens of minutes to acquire the geometry of a face from a photograph.

Further, based on the morphable face modeling technique, Blanz et al. [5] propose a face exchange technique, which allows the replacement of an existing face in a target photograph with a new face. To do this, only a single image of the new face is required. By making use of the morphable face modeling, one can create a 3D face model for the new face as a combination of the existing face model collection. By employing optimization method, one renders the model with proper illuminations and postures. This allows the rendered result to fit into the target photograph and thence replace the existing face.

Remarkably, the morphable face modeling has been utilized in the work of [13], in which 3D face model is tracked from a real-time video sequence. It has been found that 3D morphable modeling is extremely well suited to the task of fitting a 3D model to a target video in real time.

3.2.2 Multi-linear modeling

Multi-linear face modeling is another way to create a desired face model from existing face model examples [40]. Similar to the morphable modeling in [4, 5], the multi-linear face modeling requires careful pre-processing of the collected examples to set up full vertex to vertex correspondence between the examples. Then, these examples are organized in the form of a data tensor, which encodes model variations in terms of different attributes, such as identity, expression and viseme. This allows independent variation of each of these attributes. By using the organized data tensor, an arbitrary face model with desired facial expression can be modeled as a linear combination of these examples.

In fact, the multi-linear face modeling was proposed in [40] to make face models for face transfer. The face transfer allows mapping video recorded performance of an individual face to the facial animation of another one. More details of face transfer will be discussed in Sect. 4.2.3.

3.3 Discussion

Comparing between the generic model individualization (GMI) approach and the example-based face modeling (EFM) approach, we can see that a number of strength and weakness of these two approaches:

- GMI only needs one generic model, without involving the support from a face model collection as in EFM. Further, it is necessary that the face models in the collection required by EFM are registered to each other with vertex-to-vertex correspondence. Such a collection might not be accessible for many potential users. As a conclusion, GMI is suitable for applications which have no access to large model collections.
- GMI requires a challenging process of providing facial features. This involves either considerable amount of manual work, or highly cost equipments & vision techniques, as in [49]. In contrast, given a face model selection, EFM allows us to generate face models from one single face image without requiring the identification of facial features. Further, the techniques in EFM such as [4] also allow us to recover face models with facial expressions from single pictures.
- To our best knowledge, EFM only works for face models with single layers, while researcher have used GMI to develop multi-layered anatomical-based models. Extension of EFM to accommodate multi-layered models can be a difficult challenge as potentially it needs to involve multi-layered model collection. Also, the optimization method required in EFM may also be more complicated to deal with multi-layer models.

Table 1 provides a summary for these two approaches and their typical examples.

Table 1. Comparison between GMI and EFM

	Examples	Strength	Weakness
GMI	[31] [21] [8] [15] [49] [48] [45]	1. Only require a generic model 2. Works for models with multiple layers	1. Need to identify facial features 2. Need considerable user inputs
EFM	[5] [4] [40]	1. Do not need to identify facial features 2. Only need a single face image input 3. Allow generate facial expressions	1. Need support of a registered face collection 2. Difficult to generate multi-layer models

4 Facial expression and animation

Driven by the desire of improving visual realism of facial animation and creating naturally looking facial expressions, great effort has been made from the graphics community, which can be categorized as follows:

- Simulation-based approach, which employs simulation methods to generate synthetic facial movements by mimicking the contraction of facial muscles. More details are given in Sect. 4.1.
- Performance-driven animation tries to learn facial expressions from recorded videos or captured face movements and subsequently makes synthetic facial expressions by applying them to a face model. More details are given in Sect. 4.2.
- The blend shape-based approach creates new facial expressions of a face from the linear combination of collected expression examples of the same subject (face). More details are given in Sect. 4.3.

4.1 Simulation-based approach

The motivation of the simulation-based approach is to create synthetic facial expressions by simulating facial muscle actions on a face model. This requires us to define the functionality and locations for a number of pseudo muscles on the face model. The functionality of a pseudo muscle is defined in terms of its influence on the face model, which depends on the employed simulation method. The overall synthetic facial expression is determined by the combination of the pseudo muscle contractions [42]. Further, initiated by the idea of using pseudo muscle simulation, a number of multi-layer models have been developed to simulate the anatomical structure of human face, including skull, muscle, soft tissue, skin etc. [20, 37, 38]. This greatly improves the visual realism of the synthetic expressions.

4.1.1 Pseudo muscles

The paper from Waters [42] presents a classic work in synthesizing facial animation using pseudo muscles. The movement (extraction) of each pseudo muscle is defined to link to a particular area of the face model. For example, if a facial expression with an open mouth is wanted, only the muscles associated to mouth areas need to be adjusted. This can potentially avoid the facial distortion created by [3], in which a stretched point on the mesh can take effect to the whole mesh shape. In general, the pseudo muscles influence either the upper or lower face. The upper facial muscles are responsible for changing the appearance of the eyebrows, the upper and lower lids of the eyes, while the lower facial muscles determine the appearance of the chin, ears, lips, and the areas around the eyes and the neck.

Further, in [42], pseudo muscles are classified into small groups according to their functionalities. The outcome of this classification is also known as action units (AU), which defines the pseudo muscle actions during various facial movements. In other words, the AU for a specific facial expression (e.g., smile) tells us which muscles need to be activated to synthesize this facial expression. This physics-based simulation greatly reduces the amount of work that must be input by an animator, as he/she can directly specify required facial expressions by controlling the movements of the AUs.

4.1.2 Synthetic skin layers

To improve upon the basic pseudo muscle-based facial animation as presented in [42], the work from Terzopoulos & Waters [38] proposes facial animation by contracting pseudo (synthetic) muscles embedded in an anatomically motivated face skin model.

More specifically, the face model is composed of three synthetic skin layers which are made of spring-mass. The physical simulation propagates the muscle forces through the physics-based synthetic skin thereby deforms the skin to produce facial expressions. This is a combination of the pseudo muscle approach with the anatomy-based facial modeling, which significantly improves the realism of synthetic facial expressions compared to the earlier techniques [27, 42].

This approach is also amenable to improvement through the use of more sophisticated biomechanical models and more accurate numerical simulation methods. This is, of course, subject to an increase in computational expense.

4.1.3 Biomechanical skin model

Lee et al. [20] addresses the challenge of automatically creating individual facial models with highly realistic facial expressions based on pseudo muscles. This method allows us to adapt a well-structured generic face model to an individual face model acquired by a 3D Cyberware scanner in a highly automated manner. This adaptation process is similar to the deformation techniques introduced in Sect. 3.1. The outcomes of the algorithm are functional facial models which allow significant amount of facial details and high quality facial animation. The algorithm is applicable to a wide range of individuals.

To allow facial animation, the generic geometric model used by the algorithm consists of a number of different layers. More specifically, on top of the face model, five different layers are used, including the epidermis, dermis, sub-cutaneous connective tissue, fascia and the muscles. Each layer is described as a triangle deformable tissue, which is connected respectively with all the other layers – see Fig. 5. Such a skin tissue modeling is identified as "physically-based modeling of human facial tissue". This

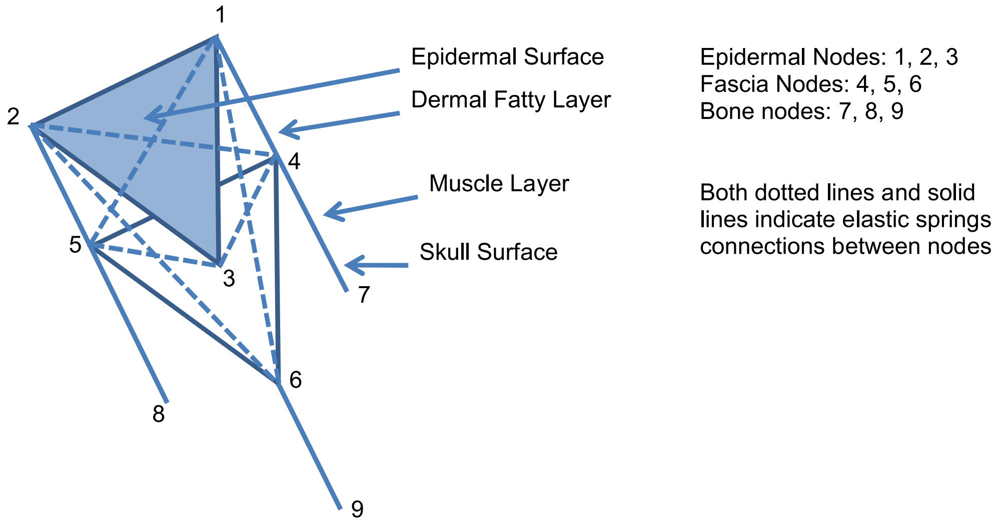

Fig. 5. Triangle deformable tissue [26]

biomechanical face skin model is claimed to be more accurate than those used in the previous methods as in [38].

However, after the creation of a functional facial model, considerable experience is required from users to generate desired facial expressions. For example, to animate the face model, the user has to pull in/out or even to move the elastic angles of the triangles. The disadvantages of that process are the complexities and difficulties to implement a reliable facial expression because it expects from the user a very good consideration of transforming the triangles. Moreover, it is difficult to model the detailed skin deformations, such as expression wrinkles, and consequently the results tend to be less realistic.

4.1.4 Anatomically accurate head model

The recent work from [37] has been claimed as the modeling of a living male subject with high anatomical accuracy. This is a volumetric modeling based on the data acquired from the subject, covering a range of structures including facial musculature, passive tissues and underlying skeletal structures. The rigid articulated cranium and jaw consist of about 30 000 surface triangles, while the flesh is modeled in the form of 850 000 tetrahedral, out of which 370 000 are simulated. 32 facial muscles are included. The data is captured using laser and MRI scans. As claimed by the authors, this model was constructed within a two-month period from 5 undergraduate students – see Fig. 6 for an illustration of the model.

Animating such a complex structure is a highly non-linear process which involves controllable an-isotropic muscle activations based on fiber directions. To allow for the animation of such a complex structure, the authors propose a performance-based method, which is capable of automatically determining muscle activations by tracking a sparse set of surface landmarks on a performer. Then the resulting animation is obtained by using a non-linear finite element method. Once the controls are reconstructed, the model can be subject to many applications, such as interaction with external objects, dynamic simulation to capture ballistic motion. The facial expressions can be edited in the activation space. The results are claimed to be not only visually plausibly but also anatomically accurate.

Noticeably, the simulation-based approach has also been applied to simulate human body structures and movements. A typical example appears in Wilhelms & Gelder [43], which performs body modeling with multi-layers including skeleton, muscles and skin. All these three components are connected in order to create the animation movements between the skin, skeleton and muscles. That process is presented as one of the first succeeded techniques for constructing a 3D body for animating purpose. However, according to Allen Van Gelder [11] by assigning the same stiffness to all springs it appears that

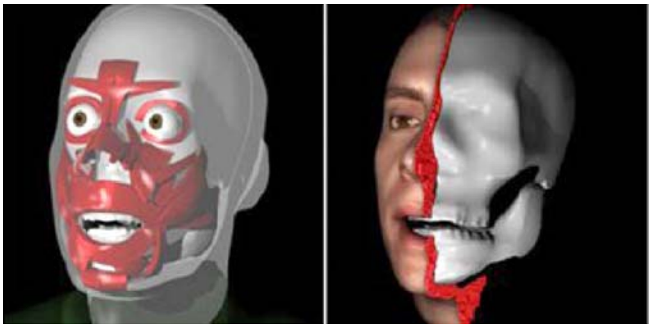

Fig. 6. Anatomically accurate head model [37]

the system fails to simulate a uniform elastic membrane, for equilibrium calculations.

4.2 Performance-driven animation

A simple performance-driven animation is geometry warping based, in which prominent facial features are tracked through a sequential of facial expression images and subsequently their movement is copied to a new face to synthesize facial expression [22, 29, 44]. On top of this, a more delicate algorithm also involves the use of expression ratio image (ERI) in order to create fine details such as face wrinkles [23]. Other examples of performance-driven facial animation include: live facial performance capture [12]; face transfer using multi-linear models [40]; vision-based facial animation control [6].

4.2.1 Geometry warping-based method

Geometry warping-based method gives the simplest form of performance driven animation. The input of a geometry warping-based method consists of two photographs from the same person as source images – the first one is a neutral face while the second one gives a facial expression, plus a target neural face. By manually or automatically locating the facial features, such as eyes, mouth, nose in these images, the algorithm calculates the movements of these features during the facial movement using difference vectors and subsequently applies the difference vectors to the target neural face. This transfers the facial expression from the source image to the target face.

4.2.2 Expression ratio image (ERI)

Normally, facial animation that generates from the geometric deformation of facial models lacks fine details that often appear in real human facial expressions, such as creases and wrinkles. These fine details are normally captured by illumination change during the facial movement. To address the challenge of presenting realistic facial expression, Liu et al. [23] present an image-based approach, which employs the expression ratio image (ERI). ERI allows the capture of the illumination change caused by fine facial details from existing face images. The work is inspired by the research in presenting and transferring lighting and illuminations between images [7, 25, 36].

In fact, a large amount of tiny visible details of a facial expression comes from the change of surface normal on the face, which gives rise to the change of illuminations under a fixed lighting environment. The idea of using ERI comes from the observation that such an illumination change can be extracted in a skin-color independent manner. An ERI is defined as a ratio image which captures the illumination change from example facial expression images. When applying such a ratio image to a neural face, all the fine details of the facial expression from the example images can be preserved – see Fig. 7 as an example.

However, this method works under the assumption that the illumination change only happens when the facial expression changes and the source and target images have similar lighting conditions. If one of the above conditions cannot be held, the ERI-based approach is not able to create a high quality output.

4.2.3 Face transfer using multi-linear models

The work presented by Vlasic et al. [40] allows mapping facial movements from a recorded video to a target face. Such facial movements include visemes (speech-related mouth articulations), facial expressions and head pose. The authors have termed this technique as “Face Transfer”. Face transfer extracts the facial movements of an individual subject and subsequently applies them to a target face – see Fig. 8 as an example.

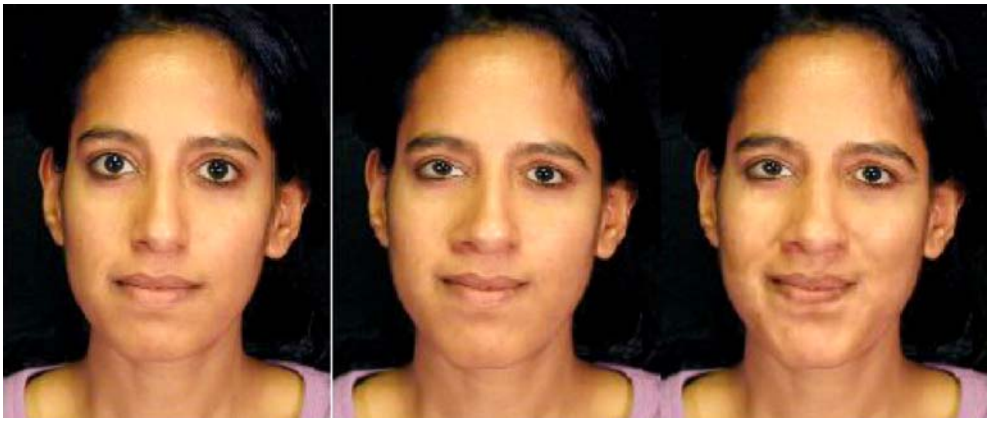

Fig. 7. Expression ratio image approach. *First image* is with a neutral expression (input image). The *second image* is after the geometric warping and the *third one* with ERI. The wrinkles in the *third image* are the result using the ERI [23]

As introduced in Sect. 3.2.2, face transfer is based on a multilinear modeling of 3D face meshes. Multilinear models consist of a collection of face meshes and their estimated variations in terms of different attributes, such as their sizes, identifies and expressions. The modeling allows each of the attributes to be adjusted separately, which facilitates the editing of the facial performance. In principle, if the face collection is sufficiently large, the multilinear modeling allows us to generate any face with any expressions and any visemes. However, as a proof of concept, the collection used in [40] only offers very limited size.

Continuously, the multilinear model can be linked to an optical flow-based tracker. The tracker estimates performance parameters and detailed 3D geometry from video recordings. The mapping from the performance parameters back to the 3D shape is then calculated. Arbitrarily mixture of pose, identity, expressions, and visemes from two or more videos are allowed.

An advantage of the face transfer technique is that it neither requires performers to wear visible facial markers or to be recorded by special face-scanning equipment. Hence, it provides an easy-to-use facial animation system from the users point of view.

4.2.4 Real time facial expression tracking

Chai et al. [6] demonstrate vision-based control of 3D facial animation by presenting a system which allows us to extract animation control parameters from a video and subsequently apply these parameters to a 3D face model to create high quality facial animation. By using this method, users are able to control the animation using recorded face actions. It is expected that low quality videos is sufficient for such a facial animation control.

The input of the algorithm is a single video stream recording the user's facial movement, a preprocessed motion capture database, and a 3D head model captured by laser scanner.

The system has four components. First, video analysis, which tracks the head position in the video, identifies a small number of features and subsequently gives expression control and head pose parameters. Second, preprocessed motion capture data contain a set of head motion & expression data. The head motion and facial deformations are automatically decoupled in the motion capture data during the pre-processing. Third, expression control and animation transform noisy and low resolution control parameters into high quality motions. Fourth, expression retargeting applies the synthesized motion to animate 3D high resolution models.

The expression tracking step involves tracking 19 2D features on the face: one for each point of the upper and lower lip, one for each mouth corner, two for each eye brow, four for each eye and three for the nose. The expression control parameters generated in the first step includes the parameters to describe the movement of the features such as the mouth, nose, eye and eyebrow of the actor in the video. These noisy and low resolution parameters are converted into high resolutions by using an example-based motion synthesis method, which compares the low resolution parameters with the motion data captured in the second step, and subsequently synthesizes a proper high resolution motion. Finally, to map the synthesized motion to a target 3D model, an efficient expression cloning technique is used, which pre-computes a number of bases for the facial deformation of the target model, and then blends them to create run-time facial expressions according to the synthesized motion.

4.2.5 Facial animation capture

Guenter et al. [12] create a system for capturing human facial expression from a live performance and replaying it in a highly realistic manner. The system allows capture the geometry, color and shading information of a face expression and subsequently re-display it using a high quality deformable polygon model decorated by a dynamic texture map.

A 3D scan is used to capture the geometry of the actor's face. To acquire the live performance of the face, multiply cameras (6 calibrated cameras) are used to record the face movement from different positions simultaneously. To help the data capture, a large number of sample points (182 points) are placed on the face. Their positions are reconstructed through the recorded videos from the multiple cameras, which are then used to distort the face geometry in order to create desired facial expressions.

Together with the tracking of the geometric data, the multiple video cameras also capture multiple high reso-

Fig. 8. With the multilinear model, the expression, and the viseme from the *second* and *third left*, respectively, can be transferred to a neural subject face (the *far left*). The *far right* gives the result [40]

lution video images. These images are used to generate a dynamic texture map for the face model which gives a vivid appearance during the re-display of the facial expression. To generate the texture map, the images from the cameras are weighted differently according to their positions.

4.2.6 Principal component-based facial motion capture and analysis

Kshirsagar et al. [19] present their techniques for facial motion capture using principal component analysis (PCA). Through the PCA, a number of principal facial movements are identified from the captured data, which provide insight into the mechanism of general facial movement. This is subsequently used to control the synthesis of new facial expressions.

6 cameras and 27 markers are used for the capture. These markers are compatible to MPEG-4 feature point locations. The work is primarily aimed at speech animation. During the recording of a speech performance, the movements of 14 markers are extracted after removing global head movement.

By applying PCA to the captured data, 6 principal face movements are obtained, which cover the major mouth actions during speech, including open mouth, puckered lips, parted lips, etc. Then, based on the observation that suppressing or enhancing certain principal face movements can generate satisfactory results during the practice of facial expression mixture, various controls are applied to the synthesis of by assigning different weights to different principal movements.

4.3 Blend shape-based approach

The blend shape-based approach creates a desired facial expression through the combination from a set of existing examples. This bears similar idea as the face modeling method presented in Sect. 3.2, which also employs linear combination from a number of exiting face models. The combination can be linear interpolation applied either to images [46, 47], or to face models [31, 49], or can be morphable modeling based, such as in [4].

4.3.1 Interpolation between models

Given a number face models with different facial expressions, a straightforward idea is to generate facial expressions in-between by using linear interpolation. This has been used frequently in many applications of facial animation. For example, the work from [31] and [49] both generate facial expression following the reconstruction of face models with various expressions.

Noticeably, the work from [32] uses linearly combined 3D face models to recover face position and facial expression from an input video sequence. This recovery process takes place in each frame of the video, and it employs a continuous optimization method in order to find the best matched model at each frame. The 3D model, which is used for the fitting, is based on the linear combination of a set of face models with different facial expressions. This face model set is generated using the technique presented in [31].

However, an obvious disadvantage of this approach is that only facial expressions in between existing examples can be created. Therefore, the technique requires considerable large number of facial expression examples.

Also, linear interpolation does not possess high accuracy and hence is not a prefect solution for generating in between expressions. Remarkably, the recent work from [49] overcomes the weakness of using linear interpolation by presenting a technique named "faceIK". Essentially, faceIK is an inverse kinematics technique, which blends the models to generate different facial expressions under user-specified controls. Moreover, the authors also present a new representation name "face graph", which encodes the dynamics of the face sequence and can be traversed to create desired facial animations.

4.3.2 Interpolation between images

The technique presented in Zhang et al. [46, 47] allows us to generate high quality facial expression with significant details such as wrinkles, given a set of example images of different facial expressions. This technique can also be applied to 3D models.

To use these example images, geometry positions of the feature points in the example images need to be identified. Then, a photorealistic facial expression can be obtained from a convex combination of the example expressions based on these positions. Since this technique makes use of high quality expression examples, it can generate photorealistic and natural looking expressions with fine facial details.

Further, to overcome the challenges of automatically recovering feature points from the images, the authors develop a technique to infer missing feature points from the tracked face by using an example-based approach. This allows us to be less demanding on the feature point recovery and tracking technique. In other words, the system should still be well-functioned even if the number of tracked feature points is fewer than what the system requires.

4.3.3 Blend shape animation from model segmentation

To generate high quality facial expression for 3D face models using blend shape methods, Joshi et al. [14] present a method which segments the face models into small regions. The shape blend animation based on this segmentation allows us to handle specific part of a face without affecting other irrelevant parts and hence preserve significant amount of complexity of human expressions. The face model segmentation method presented in [14] is claimed to be automatic and physically motivated.

4.3.4 Reanimating faces and images using morphable models

Continuing from their previous work [4], Blanz et al. [4] present a technique that allows us to change facial expressions in existing images and videos. The morphable modeling method, which was used previously for face modeling (see Sect. 3.2.1), is extended here to cover facial expressions. To achieve this, instead of only collecting neutral faces as in [4], 35 laser scans of facial expressions are also captured and stored in the face model collections. Morphable modeling using such a face model collection allows 3D reconstruction of non-neutral face models, i.e., it generates face models with expressions. The reconstructed model can be adjusted to change its facial expression and rendered back to its original image or video.

The advantage of using this morphable modeling-based approach is that it can work on any face shown in a single image/video, without requiring example expression data from that particular person.

4.4 Discussion

Comparing the three major facial animation approaches as mentioned above, we can conclude their strength and weakness as follows:

- The simulation-based approach only needs to involve computing resources. Basically, it employs and simplifies a number of biomechanical muscle equations for producing visually correct facial movements. The performance-driven approach, which requires the capture of live facial expressions, involves considerable high cost computer vision equipments for data capturing. Given the state of the art of computer vision, the performance of such a data capture process may not be easy to improve rapidly in the near future. For the blend shape approach, a collection of face models with different facial expression is required. Again, obtaining and accessing such a face collection can be not easy from an ordinary user point of view.
- The performance-driven approach has the most potential for achieving more visual realism. In fact, the performance of the simulation-based approach is always limited by the underlying biomechanical simulation method. Unfortunately, most of the biomechanical simulations, including those claimed with high accuracy as in [37], take a great simplification from the real world model. In contrast, the methodology of the performance-driven approach, which learns facial expression from real performance, offers a mechanism to be close to the real world animation as much as possible.
- The shape blend approach is only capable of creating animation in-between the existing examples, which is a great limitation. Also, its performance heavily relies on the quality of the example models, as well as the employed interpolation methods.
- In principle, the simulation-based approach bears great potential in medical applications. It can work in conjunction with anatomical and medical data, medical simulation, which provides high quality medical models. Such a simulation can potentially provide guidelines for medical professionals in their practice.

Table 2 gives a summary of comparison of the performance between these three major approaches.

Table 2. Comparison between the simulation, performance-driven and shape blend approach

	Examples	Strength	Weakness
Simulation-based approach	[42] [38] [20] [37]	1. Only require computing resources 2. Great potential in medical applications	Artificial-looking facial expression
Performance-driven approach	[44] [23] [40] [6] [12] [19]	Great potential to achieve visual realism equipments	Need to involve motion capture
Shape blend approach	[31] [49] [46, 47] [14] [5]	Easy implementation	Need high quality facial expression examples

5 Applications

Facial modeling and animation faces a wide range of demands from industry nowadays. This section will discuss how the techniques described within this paper meet these demands, providing their suitability towards different areas of exploitation. Here, we will focus on major applications including movie industry, computer games, medicine and telecommunication. Of course, it is well understood that the actual potential of facial modeling and animation goes beyond these limited number of applications discussed here. In addition, as a typical example of applying facial animation to multimedia, we also introduce facial animation in MPEG-4 in Sect. 5.5.

5.1 Movie industry

The movie industry has received huge benefit from the advance of facial modeling and animation techniques. The increasing number of computer-made movies, such as Toy Story, Shrek, Monsters Inc, Monster House, King Kong, has demonstrated that the current techniques are well suitable to create cartoon styled movies. To this end, traditional techniques, such as geometric face modeling and simulation-based facial animation [20, 38, 42], are quite competent for the job.

However, major work still needs to be done to achieve results with high realism. Due to its post-processing nature, movie production usually allows considerably long processing in order to achieve the desired quality. Hence, methods involving high cost resource are permitted, such as large face models with significant details captured from 3D scanners, facial motion captured using costly equipments, etc. However, given the state of the art, there is still a long way to go before we are able to completely remove artificial looking from computer synthesized faces and produce outcomes that are indistinguishable with those from real faces.

5.2 Computer games

The nature of computer games implies that rapid response speed and, in many occasions, the power of real-time processing is essential. Given the fact that a large number of computer games take place in hand-held devices nowadays, many of them employ face models with moderate quality and artificial looking facial movement in order to compromise with the speed issue.

Highly accurate computation for facial animation, such as anatomy-based simulation or high degree of shape blending (interpolation), is normally not necessary for computer games. To maintain a good balance between the speed and image quality, preprocessed or captured facial movement data can be stored so as to take off the burden from real time computing. Although this only supports a limited number of facial expressions, it should be suitable in the context of computer game design at present stage as many current games only involve limited facial expressions.

Hence, given the limitation of the current techniques and computing power, there is an immense barrier to overcome before we achieve completely freeform and realistic facial animation in computer games.

5.3 Medicine

Computer-based simulation can help medical society to enhance their knowledge to the underlying problem of facial movement and expressions by creating insight views for facial anatomy and structures. The aim of such an application is to achieve highly accurate simulation rather than just to gain a stunning looking. While early effort on pseudo muscle-based simulation [42] did not provide required precision, recent work on anatomy-based simulation [37] makes a significant step towards this direction. Noticeably, this application also overlaps with the research in medical visualization, which allows us to display the data of human anatomy in a highly accurate manner.

5.4 Telecommunication

The capability of displaying a talking face is always a desirable feature in telecommunication. Here the challenge is to reduce the time lag while the display quality is improved. Due to bandwidth limit, transmitting high resolution face images is usually practically prohibitive. A feasible solution is to store high-resolution generic face models locally, which can be adapted to different individuals involved in the communication. Given these models, only movements of the models need to be transmitted during the communication, which cost less bandwidth. However, as we introduced above, a fully automatic model adaptation is still not available, hence manual identification of facial features is still necessary. Further, completely traceless, fast and reliable facial performance tracking is still a challenge. Therefore, large progress is still required before we are allowed to view each other vividly via computer facial animation techniques during a telecommunication.

5.5 MPEG-4 facial animation

A superb feature of MPEG-4 is the support of integration of natural and synthetic scenes through an object-based audiovisual representation. It was also the first time to standardize the tools supporting 3D facial animation. To achieve high quality face modeling and animation outcomes, three types of facial data are specified [1] – see Fig. 9:

1) *Facial animation parameters (FAPs).* are designed to allow the animation of a 3D facial model, reproducing

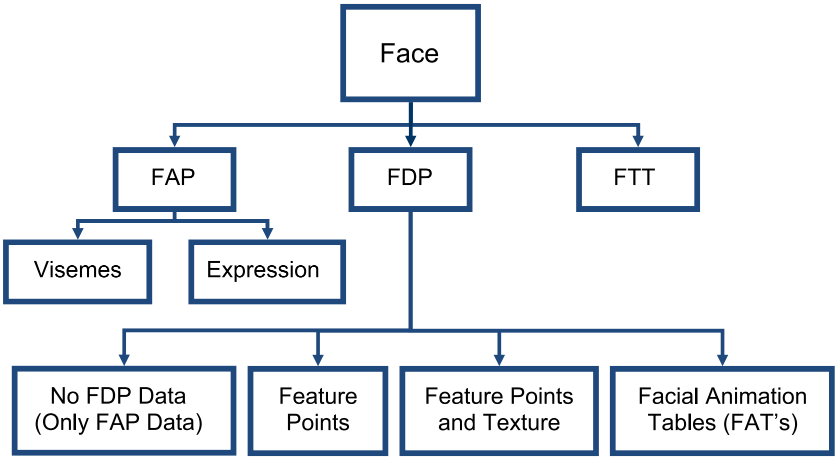

Fig. 9. MPEG-4 facial animation data

facial movements, expressions, emotions and speech pronunciation. The FAP set includes 68 parameters, 66 of which are low-level parameters related to the movements of lips, jaw, eyes, mouth, cheek, nose, etc., and the rest two are high-level parameters related to expressions and visemes. By using the FAPs, one can generate different expressions, including joy, sadness, anger, fear, disgust and surprise. Visemes are the visual analog to phonemes and allow the efficient rendering of visemes for better speech pronunciation.

2) *Facial Definition Parameters (FDPs).* FDPs allow us to configure a 3D facial model, which is to be animated by means of FAPs as described above. This can be done by either adapting a previously available model or by sending a completely new model alongside with the information about how to perform its animation. A MPEG file can potentially carry A) no FDP data, or B) feature points, which constitute a set of 3D features represented by their coordinates, or C) feature points plus textures, which include additional texture information for the model, or D) facial animation tables (FATs), which are designed to define full animation control of a complete new model.
3) *FAP Interpolation Table (FIT).* FIT allows interpolation for the FAPs. This facilitates the definition of facial animation as only a small set of FAPs needs to be included in the MPEG file to describe a dynamic facial animation. This small set of FAP is used to determine the values of other FAPs during the animation using the interpolation based on FIT.

6 Limitations and future trend

Rapid progress has been made in facial modeling and animation in recent years. Comparing recent results with those from early years, we can achieve much more vivid and realistic looking faces synthesized from computers. However, on the other side, great limitations still exist for future improvement. Here, we provide a brief summary on the major limitations of the current technology and discuss how to address these challenges in future:

1) Although we can capture extremely high quality face models with great deal of fine details through 3D scanning, how to efficiently represent and effectively manipulate such a large model has become a critical issue. These manipulations are important if we wish to modify or apply facial animation to the model.
2) While the simulation-based facial modeling and animation provides valuable insights to human facial structure and its movement, the current computational models have made significant simplification and hence lack sufficient accuracy. One potential approach to improve the accuracy is to combine facial animation with Medical Visualization technology, which offers truthful display of facial structures through captured medical volumetric data (CT & MRI).
3) The performance-driven approach for facial animation is greatly limited by current computer vision technology. Due to these limitations, particular arrangements, such as face marks and fixed camera positions, need to be made during facial movement capture, and subsequently complex data processing and analysis is required. However, frequently, many of these captures still fail to provide convincing results.
4) Currently, most of the face-rendering techniques require at least two photographs in order to allow multiple views of a face. However, in many applications users desire to create multiple views from a single input image. The existing techniques which allow us to do so [4] need support from a considerably large face model database, in which the models have to be registered. This poses as a great limitation.

Correspondingly, in future, more work will be carried out in conjunction with anatomy and medical visualization to enhance the quality and accuracy of face models to meet the demands from medical science. On the other side, although the performance-driven approach has proved its great potential through its initial results, its future depends much on the progress of motion capture techniques in computer vision, including software analysis and hardware performance. In addition, further efforts will be needed to develop more robust techniques for face rendering from a single input of a face image.

The techniques covered in this survey are mainly designed to create face shapes and facial movements with respect to time. To carry out a comprehensive simulation to real face, combining these techniques with the modeling and animation of other associated components is necessary, such as hair, eyes, and tongue. In particular, hair modeling has been consistently challenging in computer graphics, due to its complicated and dynamic nature. Although already being adopted by many commercial software packages, hair modeling is only at its infant stage. At present, computer-synthesized results are still generally artificial looking and have large difference with real hair, especially with respect to hair animation. For more details of hair modeling, please see a recent survey paper [41].

7 Conclusion

This paper has presented a comprehensive survey for the techniques of human facial modeling and animation. We have carried out the survey from two perspectives: facial model and facial animation.

Facial modeling techniques concern how to make 3D face models. This can be further categorized into the generic individualization approach and the example-based face modeling approach. The generic individualization approach is less resource demanding – it only requires one generic model, while the example-based face modeling approach needs a considerably large model collections. On the other hand, the example-based face modeling can create a face model from one frontal face image without identifying face features, while the generic individualization approach needs considerable input to identify facial features.

For facial animation, it can be done by either the simulation-based approach, or the performance-driven approach, or the shape blend approach. The simulation-based approach only requires computing resources, while the performance-driven approach involves considerable motion capture equipments. However, the simulation-based approach only provides artificial looking animations, while the performance-driven approach has the potential to be more realistic. The performance of the shape blend approach depends largely on the existing facial expression examples. Although it is easy to implement, it only allows synthetic expression in between the existing examples.

Facial modeling and animation face many demands from various applications, including movie industry, computer games, medicine and telecommunication. Given the current state of the art of the technology, great effort is required before we are able to completely remove artificial looking from computer synthesized faces and produce outcomes that are indistinguishable from those from real faces. This is particularly true in computer game manufacture where the capability for real time computation is often significant.

Future computer facial animation depends largely on the progress of computer vision technology, based on which we can potentially produce vivid facial animation via the capture of facial movements. On the other hand, from the perspective of medical science, computer simulation helps to enhance the knowledge to the underlying problem of facial movement and expressions. In future, more work will be needed in conjunction with anatomy and medical visualization to enhance the quality and accuracy of face models.

References

1. Abrantes, G.A., Pereira, F.: MPEG-4 facial animation technology: survey, implementation, and results. IEEE Trans. Circuits Syst. Video Technol. **9**(2), 290–305 (1999)
2. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. ACM Trans. Graph. **22**(3), 587–594 (2003)
3. Badler, N., Platt, S.: Animating facial expressions. ACM SIGGRAPH Comput. Graph. **15**(3), 245–252 (1981)
4. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, pp. 187–194. ACM SIGGRAPH, New York (1999) (URL: http://www.kyb.tuebingen.mpg.de/bu/people/volker/)
5. Blanz, V., Scherbaum, K., Vetter, T., Seidel, H.-P.: Exchanging faces in images. Comput. Graph. Forum **23**(3), 669–676 (2004)
6. Chai, J.-X., Xiao, J., Hodgins, J.: Vision-based control of 3D facial animation. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 193–206. EUROGRAPHICS, San Diego (2003) (URL: http://faculty.cs.tamu.edu/jchai/projects/face-animation/)
7. Debevec, P.E.: Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 189–198. ACM SIGGRAPH, New York (1998)
8. DeCarlo, D., Metaxas, D., Stone, M.: An anthropometric face model using variational techniques. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 67–74. ACM SIGGRAPH, New York (1998)
9. Du, Y., Lin, X.: Emotional facial expression model building. Pattern Recognit. Lett. **24**, 2923–2934 (2003)
10. Farkas, L.: Anthropometry of the Head and Face, 2nd edn. Raven Press, New York (1994)

11. Gelder, A.V.: Approximate simulation of elastic membranes by triangulated spring meshes. J. Graph. Tools **3**(2), 21–42 (1998)
12. Guenter, B., Grimm, C., Wood, D., Malvar, H., Pighin, F.: Making faces. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 55–66. ACM SIGGRAPH, Boston, MA (1998)
13. Hiwada, K., Maki, A., Nakashima, A.: Mimicking video: real-time morphable 3D model fitting. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, pp. 132–139. ACM SIGGRAPH, Osaka (2003)
14. Joshi, P., Tien, W.C., Desbrun, M., Pighin, F.: Learning controls for blend shape based realistic facial animation. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 187–192. EUROGRAPHICS, San Diego (2003)
15. Kähler, K., Haber, J., Yamauchi, H., Seidel, H.-P.: Head shop: generating animated head models with anatomical structure. In: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 55–63. EUROGRAPHICS, San Antonio, TX (2002) (URL http://www.mpi-inf.mpg.de/∼kaehler/slides/sca02-headshop_files/v3_document.htm)
16. Kähler, K., Haber, J., Seidel, H.-P.: Reanimating the dead: reconstruction of expressive faces from skull data. ACM Trans. Graph. **22**(3), 554–561 (2003)
17. Kalra, P., Garchery, S., Kshirsagar, S.: Facial deformation models. In: Magnenat-Thalmann, N., Thalmann, D. (eds.) Handbook of Virtual Humans, chap. 6. John Wiley & Sons, West Sussex, England (2004)
18. Koch, A.: Structured design implementation – a strategy for implementing regular data paths on FPGAs. In: Proceedings of the 1996 ACM 4th International Symposium on Field Programmable Gate Arrays, pp. 151–157. ACM SIGDA, Monterey, CA (1996)
19. Kshirsagar, S., Egges, A., Garchery, S.: Expressive speech animation and facial communication. In: Magnenat-Thalmann, N., Thalmann, D. (eds.) Handbook of Virtual Humans, chap. 10. John Wiley & Sons, West Sussex, England (2004)
20. Lee, Y., Terzopoulos, D., Waters, K.: Realistic modeling for facial animation. In: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, pp. 55–62. ACM SIGGRAPH, New York (1995)
21. Lee, W., Goto, T., Kshirsagar, S., Molet, T.: Face cloning and face motion capture. In: Magnenat-Thalmann, N., Thalmann, D. (eds.) Handbook of Virtual Humans, chap. 2. John Wiley & Sons, West Sussex, England (2004)
22. Litwinowicz, P., Williams, L.: Animating images with drawings. In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 409–412. ACM SIGGRAPH, New York (1994)
23. Liu, Z., Shan, Y., Zhang, Z.: Expressive expression mapping with ratio images. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 271–276. ACM SIGGRAPH, New York (2001)
24. Magnenat-Thalmann, N., Thalmann, D.: Handbook of Virtual Humans. John Wiley & Sons, West Sussex, England (2004)
25. Marschner, S.R., Greenberg, D.P.: Inverse lighting for photography. In: IST/SID 5th Colort Imaging Conference, pp. 262–265. IS&T, Scottsdale (1997) (URL: http://www.graphics.cornell.edu/pubs/1997/MG97.html)
26. Noh, J., Neumann, U.: A Survey of Facial Modeling and Animation Techniques. USC Technical Report 99-705, Integrated Media Systems Center, University of Southern California (1998)
27. Parke, F.: Computer generated animation of faces. In: Proceedings of the ACM Annual Conference, pp. 451–457. ACM, Boston, MA (1972)
28. Parke, F.I.: A Parametric Model for Human Faces. PhD Thesis, University of Utah, Salt Lake City, UTEC-CSc-75-047, USA (1974)
29. Parke, F.I., Waters, K.: Computer Facial Animation. AK Peters, Wellesley, MA (1996)
30. Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K.: Digital photography with flash and no-flash image pairs. ACM Trans. Graph. **23**(3), 664–672 (2004)
31. Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., Salesin, D.H.: Synthesizing realistic facial expressions from photographs. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 75–84. ACM SIGGRAPH, New York (1998)
32. Pighin, F., Szeliski, R., Salesin, D.: Resynthesizing facial animation through 3D model-based tracking. In: Proceedings of the 7th IEEE International Conference on Computer Vision, vol. 1, pp. 143–150. IEEE Computer Society, Los Alamitos, CA, USA (1999)
33. Pratscher, M., Coleman, P., Laszlo, J., Singh, K.: Outside-in anatomy based character rigging. In: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 329–338. Eurographics, Los Angeles (2005)
34. Seo, H., Magnenat-Thalmann, N.: An automatic modeling of human bodies from sizing parameters. In: Proceedings of the 2003 Symposium on Interactive 3D Graphics, pp. 19–26. ACM SIGGRAPH, Monterey, CA (2003)
35. Seo, H., Cordier, F., Magnenat-Thalmann, N.: Synthesizing animatable body models with parameterized shape modifications. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 120–125. Eurographics, San Diego (2003)
36. Shashua, A., Riklin-Raviv, T.: The quotient image: class-based re-rendering and recognition with varying illuminations. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 129–139 (2001)
37. Sifakis, E., Neverov, I., Fedkiw, R.: Automatic determination of facial muscle activations from sparse motion capture marker data. ACM Trans. Graph. **24**(3), 417–425 (2005)
38. Terzopoulos, D., Waters, K.: Physically-based facial modeling, analysis, and animation. Vis. Comput. Animation **1**, 73–80 (1990)
39. Tu, P.-H., Lin, I.-C., Yeh, J.-S., Liang, R.-H., Ouhyoung, M.: Surface detail capturing for realistic facial animation. J. Comput. Sci. Technol. **19**(5), 618–625 (2004)
40. Vlasic, D., Brand, M., Pfister, H., Popović, J.: Face transfer with multilinear models. ACM Trans. Graph. **24**(3), 426–433 (2005)
41. Ward, K., Bertails, F., Kim, T.Y., Marschner, S.R., Cani, M.P., Lin, M.C.: A survey on hair modeling: styling, simulation, and rendering. IEEE Trans. Vis. Comput. Graph. **13**(2), 213–234 (2007) (URL: http://www.cs.unc.edu/∼wardk/research.html)
42. Waters, K.: A muscle model for animating three-dimensional facial expression. Comput. Graph. **22**(4), 17–24 (1987)
43. Wilhelms, J., Gelder, A.V.: Anatomically based modeling. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 173–180. ACM Press/Addison-Wesley Publishing Co., New York (1997)
44. Williams, L.: Performance driven facial animation. Comput. Graph. **24**(4), 235–242 (1990)
45. Yin, L., Weiss, K.: Generating 3D views of facial expressions from frontal face video based on topographic analysis. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 360–363. ACM SIGMULTIMEDIA, New York (2004)
46. Zhang, Q., Liu, Z., Guo, B., Shum, H.: Geometry-driven photorealistic facial expression synthesis. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 177–186. Eurographics, San Diego (2003)
47. Zhang, Q., Liu, Z., Guo, B., Terzopoulos, D., Shum, H.: Geometry-driven photorealistic facial expression synthesis. IEEE Trans. Vis. Comput. Graph. **12**(1), 48–60 (2006)
48. Zhang, Y., Sim, T., Tan, C.L.: Rapid modeling of 3D faces for animation using an efficient adaptation algorithm. In: Proceedings of the 2nd International

Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 173–181. ACM SIGGRAPH, Singapore (2004)

49. Zhang, L., Snavely, N., Curless, B., Seitz, S.M.: Spacetime faces: high resolution capture for modeling and animation. ACM Trans. Graph. **23**(3), 548–558 (2004) (URL: http://grail.cs.washington.edu/projects/stfaces)

NIKOLAOS ERSOTELOS is a PhD student in computer science at the Department of Information Systems and Computing of Brunel University. His research targets on generating new algorithms for constructing new modeling and rendering techniques for facial synthesized expressions.
He was awarded the BSc degree ('99) in music technology from Hertfordshire University, UK. In 2005 he was awarded the MSc degree in media production and distribution with distinction from Lancaster University, UK.

FENG DONG is a lecturer in computer graphics at the Department of Information Systems and Computing, Brunel University, UK. His research interests include fundamental computer graphics algorithms, medical visualization, volume rendering, human modeling, and virtual reality. Dong received a PhD in computer science from Zhejiang University, China.
He is a member of the UK Virtual Reality Special Interest Group (VRSIG).