Affective Computing and Interaction:

Psychological, Cognitive and Neuroscientific Perspectives

Didem Gökçay
Middle East Technical University, Turkey

Gülsen Yildirim
Middle East Technical University, Turkey

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey · New York

Director of Editorial Content: Kristin Klinger
Director of Book Publications: Julia Mosemann
Acquisitions Editor: Lindsay Johnston
Development Editor: Dave DeRicco
Publishing Assistant: Milan Vracarich Jr.
Typesetter: Milan Vracarich Jr., Casey Conapitski
Production Editor: Jamie Snavely
Cover Design: Lisa Tosheff

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Affective computing and interaction : psychological, cognitive, and neuroscientific perspectives / Didem Gokcay and Gulsen Yildirim, editors.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-61692-892-6 (hardcover) -- ISBN 978-1-61692-894-0 (ebook) 1. Human-computer interaction. 2. Human-machine systems. 3. Affect (Psychology)--Computer simulation. I. Gokcay, Didem, 1966- II. Yildirim, Gulsen, 1978-
QA76.9.H85A485 2011
004.01'9--dc22

2010041639

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 9
Facial Expression Synthesis and Animation

Ioan Buciu
University of Oradea, Romania

Ioan Nafornita
University of Timisioara, Romania

Cornelia Gordan
University of Oradea, Romania

ABSTRACT

Living in a computer era, the synergy between man and machine is a must, as the computers are integrated into our everyday life. The computers are surrounding us but their interfaces are far from being friendly. One possible approach to create a friendlier human-computer interface is to build an emotion-sensitive machine that should be able to recognize a human facial expression with a satisfactory classification rate and, eventually, to synthesize an artificial facial expression onto embodied conversational agents (ECAs), defined as friendly and intelligent user interfaces built to mimic human gestures, speech or facial expressions. Computer scientists working in computer interfaces (HCI) put up impressive efforts to create a fully automatic system capable to identifying and generating photo - realistic human facial expressions through animation. This chapter aims at presenting current state-of-the-art techniques and approaches developed over time to deal with facial expression synthesis and animation. The topic's importance will be further highlighted through modern applications including multimedia applications. The chapter ends up with discussions and open problems.

INTRODUCTION

In human – to – human interaction (HHI) people mostly use their face and gestures to express emotional states. When communicating with each other, people involve both verbal and non-verbal communication ways: speech, facial expression or gestures (nods, winks, etc). As pointed out by Mehrabian (Mehrabian 1968), people express only 7% of the messages through a linguistic language,

DOI: 10.4018/978-1-61692-892-6.ch009

38% through voice, and 55% through facial expressions. Closely related to HHI, Human-computer interaction (HCI) deals with the ways humans communicate with machines. The term is broad and has an interdisciplinary character concerning various scientific fields, such as computer science, computer graphics, image processing, neurophysiology, and psychology. We should note that human-computer interface differs from brain-computer interfaces as the latter describe the communications between brain cells and machine and mainly involves a direct physical link.

Our psychological need of being surrounded by "human–like" machines in terms of their physical appearance and behavior is the main driving force behind the necessity of developing realistic human-computer interfaces. We would like machines acting like us, interpreting our facial expressions or gestures conveyed by emotions and respond accordingly. During the last decade the endeavor of scientists for creating emotion-driven systems was impressive. Although facial expressions represent a prominent way of revealing an emotional state, emotional states may also be expressed, coupled or associated to other modalities, such as gestures, change in the intonation, stress or rhythm of speech, blood pressure, etc. However, within the context of this chapter, we only consider facial expressions whenever we refer to emotion.

A *fully automatic facial expression analyzer* should be able to handle the following tasks (Krinidis, 2003):

1. Detect (and track) the face in a complex scene with random background;
2. Extract relevant facial features;
3. Recognize and classify facial expressions according to some classification rules.
4.

Likewise, a *facial expression synthesizer* should:

1. Create realistic and natural expressions;
2. Operate in real time;
3. Require minimum user interaction in creating the desired expression;
4. Be easily and accurately adaptable to any individual face.

This chapter is focusing on the facial expression synthesis part and animation of synthesized expressions. Once the facial expression is synthesized the facial animation comes next, a task intensively employed in computer graphics applications. For instance, in the film industry, moviemakers try to build virtual human characters that are indistinguishable from the real ones. In the games industry, the designed human characters should be interactive and as realistic as possible. Commercial products are available to be used by users to create realistic looking avatars for chatting rooms, e-mails, greeting cards, or teleconferencing. Face synthesis techniques have also been used for compression of talking head in the video conferencing scenario, such as MPEG – 4 standard (Raouzaiou, 2002).

The purpose of this chapter is to present current state–of–the–art techniques and approaches developed over time concerning facial expression synthesis and animation. These techniques will be elaborated throughout the chapter along with their limitations. The topic's importance is highlighted through modern applications. The chapter ends up with further discussions and open problems.

FACIAL EXPRESSION SYNTHESIS, ANIMATION AND APPLICATIONS

Facial expression synthesis methods roughly may fall under two categories: geometry based approaches where the face geometry is manipulated, and appearance (image) based approaches where face deformation mainly relies on morphing and blending techniques. Geometry based approaches consider the facial geometry extraction and manipulation, and, more precisely, the extraction of geometric parameters of those fiducial points that contribute to face region deformation involved in mimic the expression. Geometry manipulation includes several techniques such as interpolation, parameterization, physics-based modeling (mass-spring, muscle vector, etc.) or finite element method. Appearance based approaches rather rely on representing the face as an array of intensity values that are globally processed.

Facial Expression Descriptors

Each facial expression is the result of several facial muscle activations, so that, while contracted, the facial muscles temporally deform the facial features represented by eyebrows, eyes, nose, mouth and skin texture, leading to changes of the face appearance. When measuring the facial deformation, three characteristics are usually taken into account: location, intensity and dynamics. Location refers to the specific facial region where the facial action that is specific to a particular expression, occurs. The intensity is defined by the magnitude of the geometric deformation of facial features, typically when is fully deformed. We must note that the intensity of a deliberate facial expression can substantially differ from that of a spontaneous one. Generally, a deliberate facial expression possesses an exaggerated intensity. The third characteristic, namely dynamics, conveys information about the temporal evolution of facial expression and the related face surface deformations.

Facial expression measurements and descriptors are required in order to map the facial deformation to a synthetic face. There are several facial expression descriptors proposed in the literature. The most complete system for describing the facial expressions is the so-called Facial Action Coding System (FACS) developed by Ekman and Friesen (Ekman, 1978). The main goal of this system is to encode a comprehensive set of all possible visually distinguishable facial appearances by measuring specific facial muscle movements. The facial motion is decomposed into a set of component actions called Action Units (AUs). However, only 30 AUs out of the full set of 44 are responsible for the anatomical contraction of a specific facial muscle set, while the remaining 14 AUs are referred to as miscellaneous actions (such as blinking, for instance). FACS is intensively employed to analyze and synthesize facial expressions. Examples of several AUs along with their description and muscles involved are depicted in Table 1.

Although FACS seems to be the most popular system used to animate and classify muscle action, it lacks time information, which is important for generating automatic smooth and intensity increasing expression over time. Therefore, AUs should be independently adjusted to the model.

The Facial Animation Parameters (FAP) that is especially dedicated to animate synthetic faces provide another descriptors set. FAP were defined to be compliant to the MPEG-4 standard compression technique and comprises 2 high level parameters (visemes and expressions) and 66 low-level parameters (Raouzaiou, 2002). One major drawback of FAP is the fact that it only expresses discrete expressions. According to Mehrabian (Mehrabian 1996), the emotion is not limited to isolated categories but can be described in terms of combination of three nearly orthogonal dimensions, namely pleasure-displeasure (P), arousal-nonarousal (A) and dominance-submissiveness (D) leading to the PAD emotional space model for describing universal emotions. The PAD parameters are considered high-level facial parameters. Recently, intermediate (middle) level parameters termed Partial Expression Parameters (PEP) were proposed by Zhang et al (Zhang, 2007) to get smoother control over the facial movements. PEP allows high correlations between different FAPs. Aside from those aforementioned descriptors, early works involved the 6 basic limited emotions descriptor set, i.e. angry, happiness, sadness, disgust, surprise and fear. We should note that each basic emotion can be described in terms of FACS

Table 1. Samples of Action Units (AU) in the Facial Action Coding System

AU	Action description	Facial muscle involved
1	Inner Brow Raiser	Frontalis,pars medialis
2	Outer Brow Raiser	Frontalis,pars lateralis
4	Brow Lowerer	Corrugator supercilii, Depressor supercilii
15	Lip Corner Depressor	Depressor anguli oris
20	Lip Stretcher	Risorius
27	Mouth Stretch	Pterygoids, Digastric
s28	Lip Suck	Orbicularis oris
44	Squint	Orbicularis oculi, pars palebralis

Table 2. The six basic expressions and their corresponding combination of AUs

Expression	AUs
Angry	2,4,7,9,10,20,26
Happiness	1,6,12,14
Sadness	1,4,15,23
Disgust	2,4,9,15,17
Surprise	1,2,5,15,16,20,26
Fear	1,2,4,5,15,20,26

through AUs combinations, as drawn in Table 2 (Ekman, 2002).

Geometry Based Manipulations

In the early 80's, Badler and Platt (Blader, 1981) were the first to apply geometric based manipulation strategies to model and simulate the human skin on a muscle model. A set of muscles is attached to a skin mesh. The contraction of the muscle set generates skin mesh deformation. Waters (Waters, 1987) developed and modeled linear and sphincter muscles, where the latter are associated to the lips and eyes region. His early work has been extended by Terzopoulos and Waters (Terzopoulos, 1990) by introducing a third layer (between the skin and muscle layers) corresponding to a fatty tissue. The model provides smoother deformation control over animation. These models are limited in the sense of lacking natural facial expression synthesis. Moreover, fine skin deformations such as temporal wrinkles are difficult to be represented, if not impossible. Hoch et al. (Hoch, 1994) proposed a facial model for facial expression animation based on B-splines surface with 13 × 16 control points. Four action units (AU1 – inner brow raiser, AU2-outer brow raiser, AU4-brow lower, and AU12-lip corner puller) of FACS are chosen for animating the face. Those action units are hard-wired to the model that is adapted to a given input datum consisting of 3D laser-scanner images each comprising 200×200 points determined by hand. The adaptation process is carried out in two phases. First, the model surface is fitted by minimizing the mean square error between the given data points and the surface points. Second, some constraints are considered in order to correctly position the control points in regions where the respective action units apply (i.e. the control points associated with an action unit result in the deformation of the correct region of skin tissue). To look more realistic, the color information of the laser-scanner texture is mapped onto the facial mask. Pighin et al. (Pighin, 2002) have proposed an image-based system which is able to model and animate 3-D face models from images. This technique reconstructs a continuous image function using a set of sample images. Once this function is found, the original samples can be interpolated or extrapolated to produce novel unseen images. The face models are constructed from a set of photographs of a person's face that can be linearly combined to express a wide range of expressions. A texture map is extracted from the photographs using both the face geometry and the camera parameters. Next, a tracking step is accomplished by recovering, for each video frame, the position of the head, its orientation, and the facial expression of a subject. The purpose is to estimate the model parameters through these frames. After capturing multiple views of a subject (with a given facial expression) these photographs are manually marked with a small set of initial corresponding points on the face in the different views (typically, corners of the eyes and mouth, tip of the nose, etc). The 3-D positions of the corresponding points are further used to deform a generic 3D face mesh to fit the face of the particular human subject. Additional corresponding points may be marked to refine the fit. One or more texture maps for the 3D model are finally extracted from the photos. The approach allows either a single view-independent texture map extraction and rendering or the original images can be used to perform view-dependent texture mapping. However, the system is limited, as the authors did not report any results for mapping the expression to different face model.

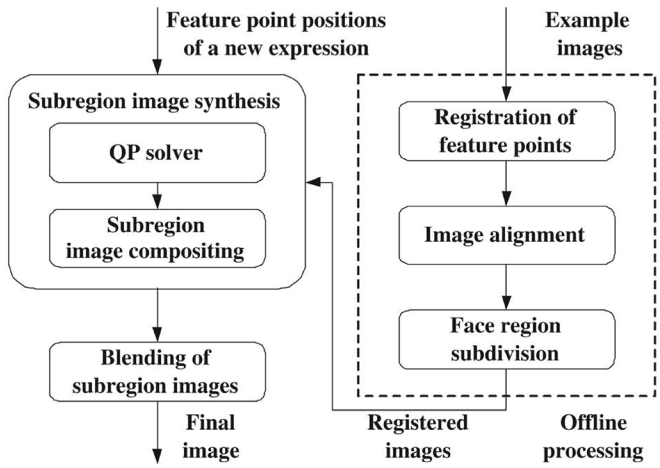
A complex facial model for creating highly realistic facial models and flexible expressions was developed by Zhang et al (Zhang, 2002).

The facial model is based on facial measurements including information about face shape and face texture. Three views of the person's face are acquired by using a non-contact 3D laser range scanner, each producing separate 3D reconstructions of the visible face regions. The geometry and color information of the facial surface is obtained by scanning a subject using a Minolta VIVID 700 Digitizer and the acquired data are registered into a single coordinate system. A geometric face model is created by editing corresponding triangular meshes. The originally generated triangular mesh consists of over 104 triangles most of them being redundant. The process is time consuming however. To reduce computational cost the mesh is adaptively reduced up to 70 percent without sacrificing the visible detail of the facial surface. Based on the triangular mesh, a physically-based face model that incorporates a multi-layer (i.e., epidermal, dermal, hypodermal) facial skin tissue to simulate the real skin is constructed. The model simulates a set of anatomically motivated facial muscle actuators and a rigid skull. A number of 23 major functional facial muscles are selected from FACS to animate facial expressions. The sets of those muscles involved in animation are as follows: 2 frontalis inner, 2 frontalis major, 2 frontalis outer, 2 corrugator supercilliary, 2 orbicularis oculi, 2 zygomaticus minor, 2 zygomaticus majos, 2 nasalis, 1 orbicularis oris, 2 risorius, 2 depressor anguli and 2 mentalis. Various flexible and realistic facial expressions can be generated using principles of Lagrangian mechanics applied to deform the facial surface. Although the model is a complex one simulating several layers, it cannot cope with temporal deformations such as wrinkles. Moreover, the teeth area is not modeled.

Zhang et al (Zhang, 2006) developed geometry-driven approach for facial expression synthesis. For each facial expression a set of feature point positions is provided which is further used to render the model employing an example-based strategy. The overall flowchart is depicted in Figure 1.

134 facial points are manually marked and the resulting marked images are all aligned with a standard image as illustrated in Figure 2. Each face image is next divided into 14 regions where

Figure 1. Flowchart of geometry-driven expression synthesis system proposed by Zhang et al. (Zhang, 2006). With permission from © 2006 IEEE



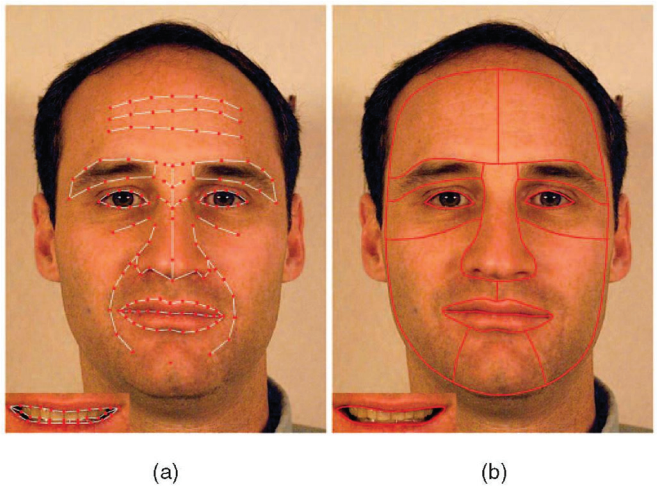
the facial expression synthesis takes place by solving a quadratic programming problem. The animation (motion propagation) is performed by estimating the unknown future location of facial features with the help of a principal components based learning technique. Figure 2 shows the selected feature points and the face subdivision. The final face with synthesized facial expression is obtained by blending those regions. One limitation of the system is the lack of a reliable extrapolation method. Another limitation comes from the blending procedure where image artefacts may become visible. And finally, as many other approaches described in the chapter, the system does not handle out-of-plane head rotations.

Appearance Based Approaches

A self-adaptive mesh is proposed by Yin et al. (Yin, 2001) to compute the deformation of the eyes in a 3D model for eye movement synthesis. Firstly, a Hough transform and deformable template matching combined with color information is used to accurately detect and track the eye features. Once the contours of the iris and eyelids in each frame of the image sequence are extracted the eye features are used to synthesize the real motions of the eye on a 3D facial model. An extended dynamic mesh (EDM) expressed by a nonlinear second-order differential equation is used to create a realistic eye animation. To make the solution of DM equation more stable and accurate, the conventional dynamic mesh method is modified by introducing a so-called "energy-oriented mesh" (EOM) to refine the adaptive meshes. Thus, the eye model adaptation comprises two major steps: 1) coarse adaptation which applies DM method to make the large movement converge quickly to the region of an object followed by a 2) fine adaptation, where the EOM approach is applied to further adjust the mesh obtained after the first step. The movements are obtained according to the energy minimization principle where the meshes are deformed until an equilibrium state is attained. The method was applied to a real face model consisting of a detailed wireframe model with 2954 vertices and 3118 patches, in which there are 120 vertices for each eye. As reported, the facial expressions rather look unrealistic. The adapted wireframe models in successive frames are texture-mapped by using the first frame of the sequence.

Liu et al. (Liu, 2001) proposed a facial expression mapping method based on the so-called

Figure 2. Left: facial points; Right: face subdivisions (Zhang, 2006). With permission from © 2006 IEEE



"expression ratio image" (ERI). Their method is not only able to exhibit facial feature motion but also to capture subtle changes in illumination and appearance (e.g., facial creases and wrinkles) making the face more expressive and the expression more convincing. This approach was one of the first methods capable of mapping one person's facial expression details to a different person's face. ERI involves the presence of four images: A and A * denoting the images of A's neutral face and expression face, respectively, and B and B*, denoting the image of B's neutral face and unknown image of his face with the same expression as A *, respectively. One drawback of this approach is that it requires the expression ratio image from the performer.

Raouzaiou et al (Raouzaiou, 2002) modeled primary facial expressions by using FAPs. They established the so-called face animation tables (FATs) to specify the model vertices that will be spatially deformed for each FAP as well as the deformation magnitude. The FATs value are MPEG-4 compliant, so that an MPEG-4 decoder can receive a face model accompanied by the corresponding FATs to animate synthetic profiles as illustrated in Figure 3. Figure 4 depicts synthetic expressions built using a 3D model of the POSER

Figure 3. Examples for animated profiles corresponding to anger expression. From Raouzaiou et al, (Raouzaiou, 2002) with permission

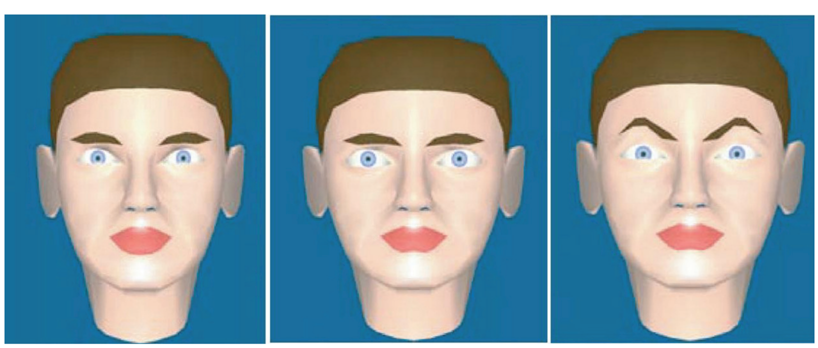
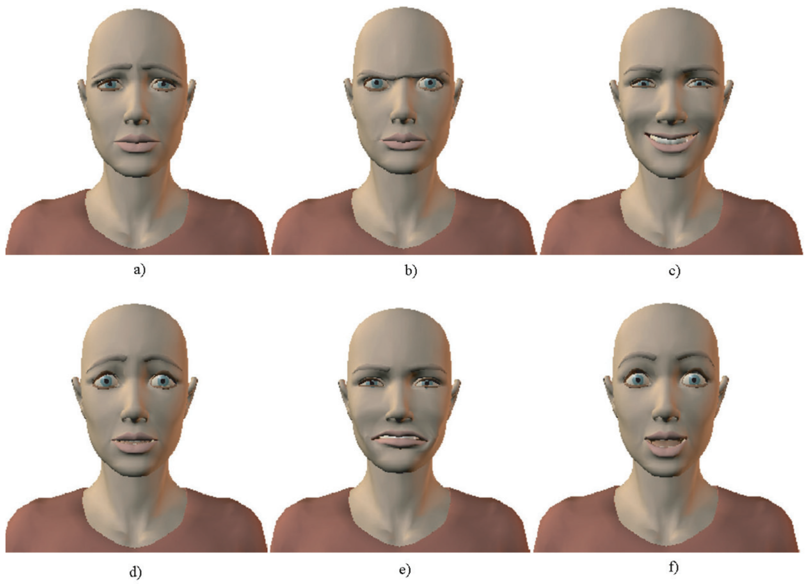


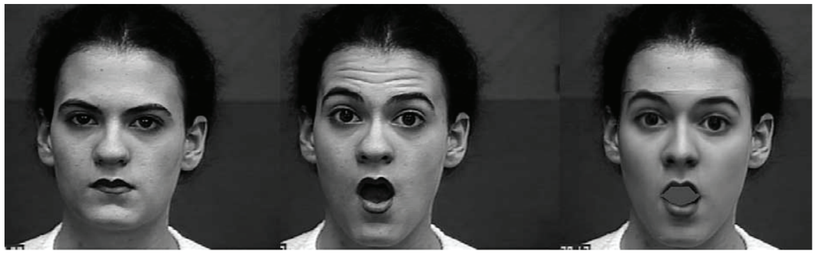
Figure 4. Synthesized archetypal expressions created using the 3D model of the POSER software package: (a) sadness, (b) anger, (c) joy, (d) fear, (e) disgust, and (f) surprise. From Raouzaiou et al, (Raouzaiou, 2002) with permission



software package available at http://poser.smithmicro.com/poser.html.

Wang and Ahuja (Wang, 2003) derived a higher-order singular value decomposition based approach to decompose the facial expression space. The learned expression subspace model is then mapped to a different identity. However, the technique cannot be applied to synthesize expressions of people with unseen facial characteristics/appearance (such as beard) if no similar images exist in the training set for decomposition. A bilinear decomposition technique is proposed by Abboud and Davoine (Abboud 2004) where appearance parameters are encoded through an Active Appearance Model (AAM), which, in turn, relies on Principal Component Analysis (PCA). The shape and texture of a set of training images are modeled with PCA. The method may lead to moderate results when only a limited number of training samples are available. However, the resulting facial expressions rather look blocky and unrealistic, as the synthesis generation accuracy highly depends on the size of the PCA learned space. Tewes et al (Tewes, 2005) used elastic graphs to build a Gabor wavelet based flexible object model (FOM) to synthesize nine different facial expressions for video frames. The model graphs are generated by manually locating the nodes of the graph over facial landmarks in the first frame. The nodes are automatically tracked over time using Gabor wavelets phase information. FOM is constructed as a parameterised model of graph deformation by merging raw data extracted from several video frames using PCA and Neural Gas. Smooth transitions are modeled with Principal Curves. In the training phase four persons are employed to construct the FOM. For testing, a person not contained in the training set is chosen. The corresponding background is discarded and an initial graph is superimposed. The graph is then deformed to generate a specific facial expression. The best matching gesture is

Figure 5. Original(b) and synthesized (c) expression of the subject depicted in (a) using the approach proposed by Ghent and McDonald, (Ghent, 2005). With permission from © 2005 Elsevier

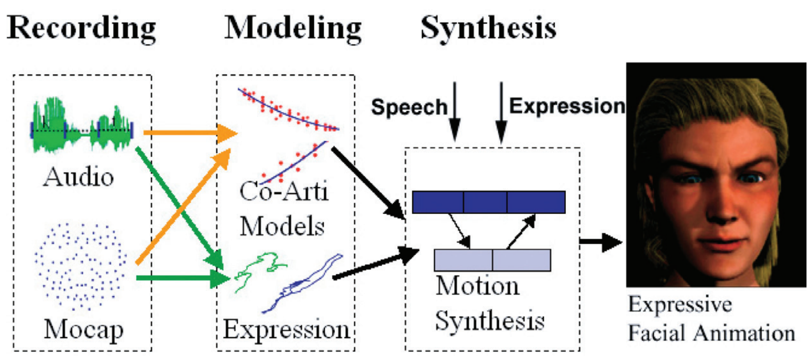


picked up from a set of 9 canonical trained gesture deformations.

Ghent and McDonald (Ghent, 2005) introduce two statistical models named facial expression shape model and facial expression texture model, both derived from FACS. The procedure allows for the generation of a universal mapping function. To map a neutral image of a face to an image of the same subject posing an expression, several radial basis function based neural networks were trained. Figure 5 depicts one person with original and synthesized expression.

Malatesta et al (Malatesta, 2006) extended the work of Raouzaiou et al (Raouzaiou, 2002) by combining MPEG-4 FAPs and action units (AUs) with the help of appraisal theory. The appraisal theory comes from the psychology field and was proposed by Scherer (Scherer, 2001) in order to investigate the connection between the elicitation of an emotion and the response patterning in facial expression. This approach can conduct to intermediate expressions based on sequential checks and derives a cumulative effect on the final combined expression. The theory was directly applied to generate intermediate expressions of hot anger and fear. To generate the cumulative effect the transaction between frames was considered to yield the final expression. Each intermediate expression is derived by the addition of the AUs of the current expression to the AUs of the previous appraisal check. However, as noted by authors, the appraisal method could lead to confusion in the final expression when subsequent expressions are constituted of conflicting animations. For instance, when one intermediate expression includes raised eyebrows ("novelty high" corresponding to anger) and the next intermediate prediction is "goal obstructive" with the predicted facial deformation as lowered eyebrows, those effects would cancel each other. Deng and Neumann (Deng, 2006) proposed a facial expression synthesis system where the animation is controlled by phoneme-isomap space. The isomap framework is introduced for generating low-dimensional manifolds for each phoneme cluster. Given novel-aligned speech input and its emotion features the system automatically generates expressive facial animation by concatenating captured motion data. The best-matched captured motion nodes are found in the database by minimizing a cost function. The method is a data-driven approach and its computational complexity limits the application for real-time expression synthesis. Deng et al (Deng et al, 2006) further extended the work by proposing an approach where learned co-articulation models are concatenated to synthesise neutral visual speech according to novel speech input. A texture-synthesis-based method is next employed to generate novel expression from a phoneme-independent expression eigenspace model that is finally blended with the synthesized neutral visual speech to create the final expressive facial animation. The system overview is depicted in Figure 6.

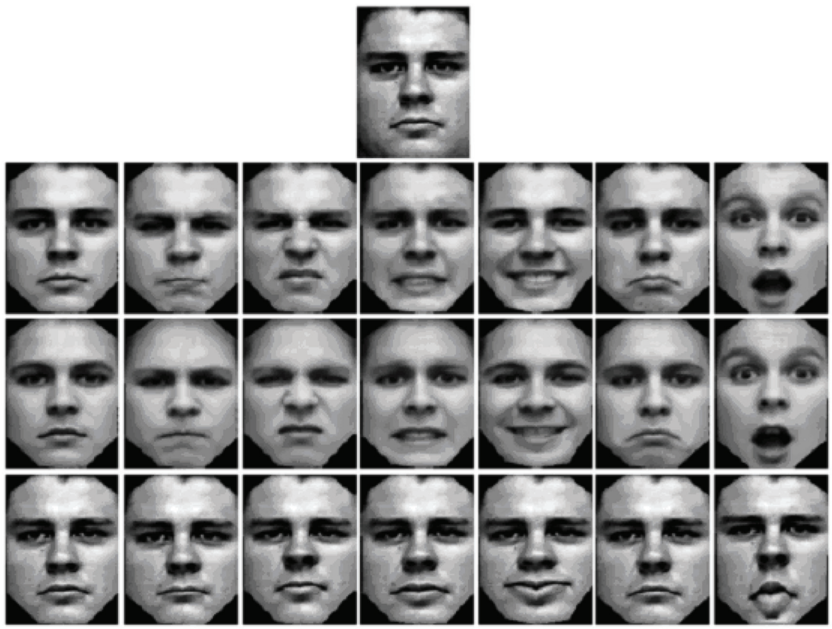
Figure 6. System overview proposed by Deng et al. Here both audio and motion are simultaneously captured as input for the novel model. (Deng et al, 2006). In the figure, Mocap refers to facial motion capture. With permission from © 2006 IEEE



A statistical analysis based approach is proposed by Krinidis and Pitas (Krinidis, 2006) to synthesize facial expressions. The dynamic facial expression model is MPEG-4 compliant, i.e. the statistics is applied to facial animation parameters (FAPs) used by MPEG-4 standard, more precisely to the vectors' displacement. The advantage of this approach is that it permits the analysis of full facial expression animation, starting from a neutral pose to a fully expressive state. When a large face database is available,

Huang and Su (Huang, 2006) propose a facial expression synthesis method based on nonlinear manifold learning for estimating the full facial expression subspace and to create a so-called hallucinated facial expression. The whole process is accomplished in two steps. The first step is related to learning the subspace (manifold) of face images with neutral expression to retrieve the intrinsic parameters of that manifold. The same procedure is applied for face images with happiness expression. The relationship between the two manifolds is learned. The second step concerns the inference issue. More precisely, the parameters of an input face image with neutral expression are obtained and its happy parameters learned from the first step are inferred to reconstruct the happy face image using the happy parameters. The authors reported fairly realistic expressions close to the ground truth. Lee and Elgammal also employed nonlinear manifold decomposition to extract shape and appearance models for facial expression synthesis (Lee, 2006). An empirical kernel mapping is employed for learning low-dimensional nonlinear manifolds that encode facial dynamics from a larger face database. To achieve accurate shape normalized appearance images a thin–plate spline (TPS) warping is used, where every image is warped with its corresponding shape vector into a new image given shape landmark points. Song et al (Song, 2006) proposed an appearance based method where the subtle changes in face deformations and are captured by decomposing the vector field expressed by Helmholtz-Hodge (HH) equations. Different expression states are treated as 3D vector field of the luminance variation. Three sets of feature points, one for the image S of source neutral, one for the source expression $S^{/}$ and the third one T corresponding to the target neutral expression are formed (manually or automatically). The corresponding images are aligned and the motion vector of the feature points between S and $S^{/}$ is computed. A geometric warping on T is next performed. A pixel triangulation on S, $S^{/}$ and T is performed to build the correspondence between the source and target face image 3D vector field. Finally, a HH decomposition based expression mapping is carried out followed by a

Figure 7. Synthesized facial expression images of a new person (Wang, 2008). From left to right: neutral, anger, disgust, fear, happiness, sadness, surprise. First row: original sample face. Second row: the proposed method. Third row: eigentransformation with shape alignment. Fourth row: direct warping of the original face. With permission from © 2008 Elsevier



pixel's luminance update according to the novel values. One advantage of the method is its robustness against illumination variation.

The facial expression dynamics expressed over time is described as discrete-time sequences of random feature vectors by Mana and Pianesi (Mana, 2006). They have employed Hidden Markov Models (HMMs) that are trained on a set of different facial expression appearances with different intensities. The model is next used to generate sequences of feature vectors according to the probability lows that are described by the parameters of the model itself. The vectors are further converted to FAPs according to the MPEG-4 standards. The idea of using manifold for generating facial expression is also embraced by Wang and Wang (Wang, 2008). A person- independent facial expression space is introduced and different subjects with different facial expression intensity are aligned using supervised localized preserving projections. The method not only allows generating basic realistic facial expressions but also mixed expression synthesis. Figure 7 shows the results of the proposed method in comparison with other two approaches.

Sucontphunt et al (Sucontphunt, 2008) developed an interactive system to synthesize 3D facial expressions through 2D portrait manipulation. Pre-recorded facial motion capture database is required for generating fine details. The system exploits the fact that 2D portrait typically relates prominent features of human faces and editing in 2D space is more intuitive than directly working on 3D face meshes. During the editing process the user moves one or a group of 2D control points on the portrait while the rest of the control points are automatically adjusted. The 2D portrait is also used as a query input to reconstruct the corresponding 3D facial expression from the pre-recorded facial motion capture database.

Zhang et al (Zhang 2008) proposed a probabilistic framework based on a coupled Bayesian network (BN) for synthesizing facial expressions through MPEG-4 FAPs while achieving very low bitrate in data transmission. The FAPs and FACS are cast into a dynamic BN while a static BN is used to reconstruct the FAPs along with their intensity. The proposed architecture is suitable for data transmission as 9 bytes / frame can be achieved. Due to the fact that the facial expression is inferred through both spatial and temporal inference the perceptual quality of animation is less affected by the misdetected FAPs when compared to facial expression synthesis using directly the original FAPs, as illustrated in Figure 8.

To capture a large range of variations in expressive appearance, a deep belief network with multiple layers is proposed by Susskind et al (Susskind, 2008). The network is able to learn association specific identities and facial actions in a cleaver way so that, novel combinations of identities and facial actions may be generated by blending them, even for unseen faces (not included in the training process).

The survey of facial expression synthesis we provided herein is not an extensive survey. Rather, we tried to provide a wide spectrum of techniques that we found applicable and promising in this trade. Next, we will try to provide a set of applications relying on the currently available facial expression synthesis technology.

Applications of Facial Expression Synthesis

We categorize the applications of facial expression synthesis in two classes in terms of their physical support: computer generated (cartoon or more

Figure 8. FAP errors (Zhang 2008). The row represents the animation outcome from the Zhang et al. method, while the bottom row shows the results of directly applying the original FAPs. Animation artifacts are visible in this case around the mouth region

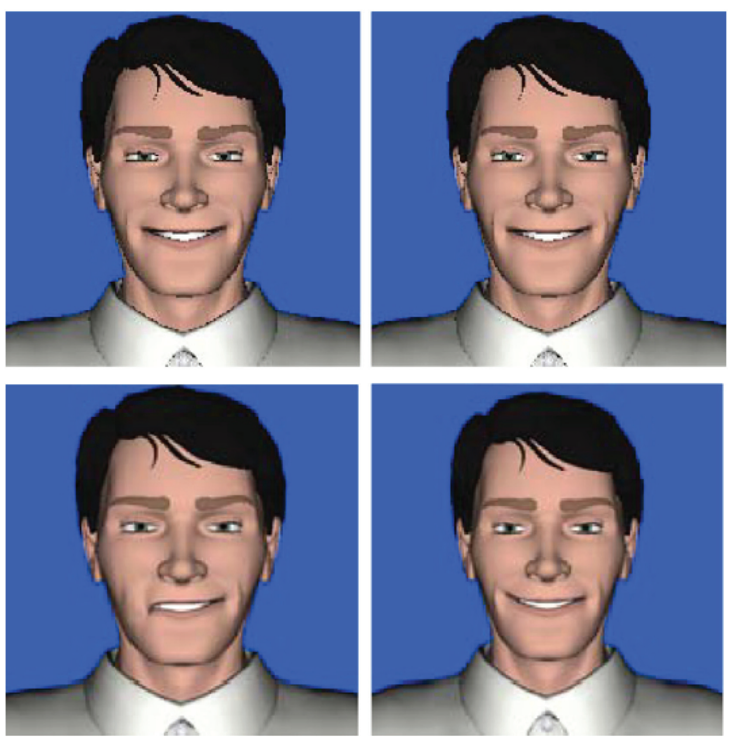
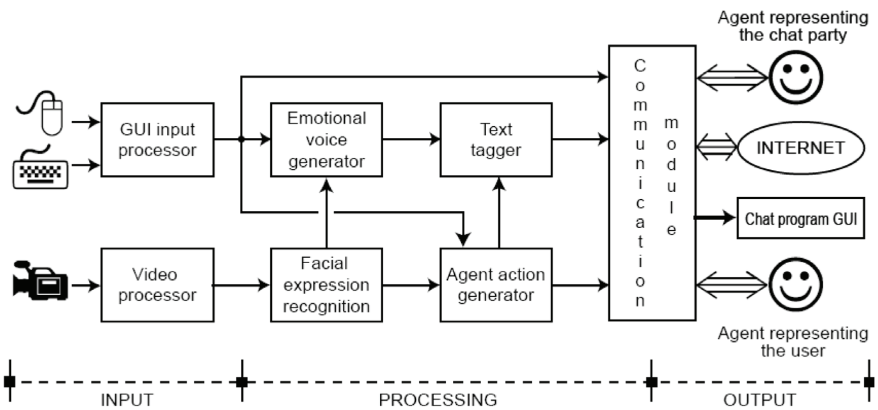


Figure 9. System overview proposed by Chandrasiri et al (Chandrasiri, 2002). With permission from © 2002 IEEE



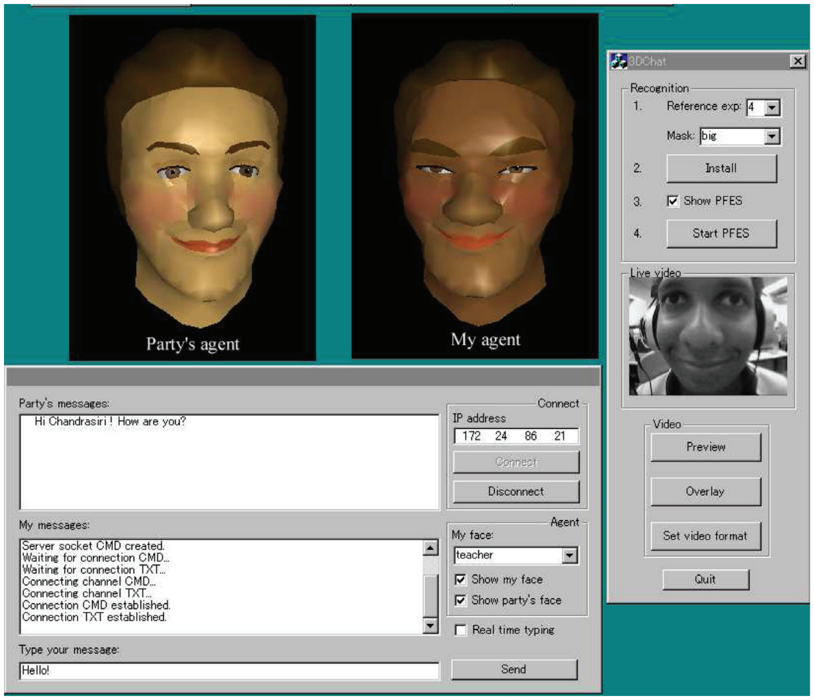
realistic) avatars and social robots. For the first category, applications of facial expression synthesis mainly involve the creation of embodied conversational agents for general communications, either in the form of animated avatars or, more simply, "talking heads". We should note here that, the "talking heads", in their simpler form, do not necessarily imply emotional state characteristics; they may simply imitate mouth region deformation for more or less realistic speech formation. The existing works either deform facial features to imitate the 6 basic expressions or manipulate AUs to animate artificial muscles, while other works carry out the animation process by employing MPEG-4 compliant facial parameters.

Chandrasiri et al (Chandrasiri, 2002) developed a system for Internet chat communication composed of three modules: a real-time facial expression analysis component, a 3D agent with facial expression synthesis and text-to-speech capabilities and a communication module. Figure 9 depicts the system components and their relationship.

The system takes several inputs from the user, such as the current user face with the help of a head-mounted video camera, keyboard and mouse messages. The user face appearance is converted into MPEG-4 compliant FAPs and special tags for text-to-speech synthesis engines are inserted into the chat message using an emotional voice generator module to carry prosody information. The FAPs are also processed by an agent action generator that decides on the appropriate animation intensity command to be sent over to the agent representing the user. The agent actions are replayed by the local agent representation of the user providing some feedback about his behaviour during the chat conversation. Those agent animation commands along with the tagged chat messages are next transmitted to the chart party over the Internet by the communication module. In the same time, the module processes the data coming from other chat parties and passes it to the agents that decode the user emotional state and representing it through facial deformation with synthesized expressions and emotional content voice. Figure 10 illustrates a chat session example.

Choi and Kim (Choi, 2005) proposed a basic framework for integrating the 3D facial animation of an avatar on a PDA via mobile network. In addition, eye-gaze of the user is incorporated using an eye-tracker approach. Basically, the system first detects the facial area within a given image then classifies the expression into 7 emotional weightings. This information along with the eye position of the user is transmitted to the PDA and

Figure 10. A chat session using the system proposed by Chandrasiri et al (Chandrasiri, 2002). Both agents, the chat window setup and live video window are depicted. With permission from © 2002 IEEE

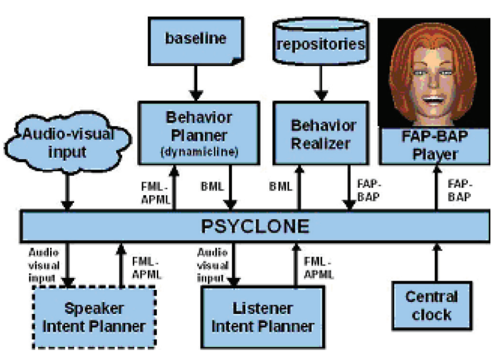


used for non-photorealistic facial expression animation for the PDA avatar. On the other hand, Albrecht et al. (Albrecht, 2005) created photo-realistic animations of a talking head capable of expressing a continuum of shades of emotion. The proposed system is able to generate facial expressions with various intensities and also mixed emotions by blending basic emotions. Zhang et al. (Zhang, 2007) proposed PEP to depict the facial expression movements for a Chinese talking avatar, acting as mid-level parameters between FAPs and PAD. The PAD-PEP mapping model and PEP-FAP translation model are then implemented to translate the PAD parameters to PEP parameters and then to FAP parameters for facial expression synthesis. The PAD values are calculated using the method described in (Mehrabian 1996).

Most recently, the work of Catherine Pelachaud and her colleagues greatly contributed to the development of virtual agents expressing emotions through facial expressions, gaze, body movement, gesture and voice (Pelachaud, 2009). The dynamics of each facial deformation is defined by three parameters: the expression intensity (magnitude) is controlled through a spatial extent; the expression duration (onset, apex and offset) is controlled by a temporal parameter; the deformation acceleration is also controlled. Although impressive, the model is limited to the six basic emotions. Niewiadomski et al. (Niewiadomski et, 2009a) proposed an algorithm for generating multimodal sequential expressions where the expressions enable the recognition of other affective states that are not prototypical expression of the six basic emotions (such as relief, for instance). In their second paper, Niewiadomski et al. (Niewiadomski et, 2009b) evaluate the recognition of expression representing 8 emotional states that are generated over a ECA face. They have reported high recognition rate (a maximum of 93% for angry and a minimum of 41% of embarrassment, often confused with anxiety and tension), suggesting a reliable model in expressing distinctive emotions. A conversational agent embodied in a 3D face named Greta, that tries to achieve a believable behaviour while interacting with a user was developed by Niewiadomski et al. (Niewiadomski et, 2009c). Greta's architecture is depicted in Figure 11. The model is not limited to facial expression generation, it also communicates through gestures, gaze, head or torso movements. User's audio and video information is captured for generating proper model behaviour. An FML-APML XML – language is used to specify the agent's communicative intentions (emotions). The Listener Intent Planner module generates the listener's communicative intentions, while the speaker's communicative intentions are intended to be preformed by the Speaker Intent Planner. The Behaviour Planner module receives as input the agent's communicative intensions (written as FNL-AMPL) and generates a list of signals using a BML tag each corresponding to a given modality: head, torso, face gaze, gesture or speech. Furthermore, signalling through the Behaviour Realizer generates the MPEG4 FAP-BAS files which, in turn, results in an animation through a dedicated player. The modules are synchronized using a central clock and a Psyclone messaging system.

An interesting application of facial expression application is proposed by Vogiatzis et al. (Vogiatzis, 2008) who mapped synthetic facial expressions into an INDIGO robot (http://www.ics.forth.gr/indigo/) that serves as a guide for a museum. The robot's affective component is described in terms of three issues, i.e., personality, emotional state and the mood, while the facial expression is described by different components such as lip-synchronized speech, face idle emotions, emotional expressions, conversational expressions and look-at behaviour. The facial expression synthesis information is converted into MPEG-4 FAPs.

Figure 11. Greta's architecture



Another robot application is developed by Ge et al (Ge, 2008) who mapped 2D real face image appearance expressing the 6 basic emotions into an expressive robotic face. Unlike the robot proposed by Vogiatzis et al., where facial expressions are displayed on a monitor, the Ge's robot head is composed of moving parts imitating the facial components such as eyebrows, eyes, eyelids and lips. Those parts are moved by servomotors according to the desired expression. The robot head has 16 degrees of freedom. The robot first detects and recognizes the facial expression of a human then imitates it. The input consists of two images, one corresponding to the neutral one and one corresponding to the desired expression. When the human changes his facial expression the difference between the two images is derived and the displacement and velocity information is extracted. This information is further multiplied by a weight vector to reach the desired animation effect on a robot head. The weight vector is added on the face plane of the robot head in its neutral state (default state) forcing the facial parts to move and simulate the human's expression. Figure 12 depicts two examples for "happy" and "surprise" expression, respectively. This application resembles learning by imitation which is observed in chimpanzees and humans through the mirror neuron and pre-motor systems found in frontal lobe.

The work of Cynthia Breazel and her colleagues (Breazel et al., 2009) is, perhaps, the most prominent in the field of emotional robots. Her work focuses in creating social robots that interact with

Figure 12. Human expression imitation by a robot (Ge, 2008). Detected keyframes associated to the video (left), the recognized expression (middle) and the corresponding response of the robot. With permission from © 2008 IEEE

humans and imitate their emotional states. The first robot called Kismet (Breazel, 2002) is augmented with expression performing capabilities. Kismet acts like babies to encourage exaggerated expressions of emotion in people. Her most recent robot Leonardo combines art with state of the art research in social intelligent robots.

RELATED PROJECTS AND SOFTWARES

The importance of the human-computer interaction and, in particular, the development of intelligent and expressive human-like machines is emphasized through large EU projects, such as SEMAINE (http://www.semaine-project.eu/), HUMAINE (http://emotion-research.net/), or dedicated workshops (http://www.enterface.net/). Open source or facial expression oriented freely distributed software tools include Xface (http://xface.itc.it/index.htm) for creating MPEG-4 and keyframe based 3D talking heads or Greta (http://www.tsi.enst.fr/~pelachau/Greta/).

DISCUSSIONS AND CONCLUSIONS

In this chapter we have tried to review the most prominent state-of-the-art techniques for facial expression synthesis and animation prospectively. Table 3 summarizes the approaches described in this Chapter along with their shortcomings and descriptors used.

Applications of this topic to real life, including embodied conversational agents enhanced or humanoid robots with affective components are also briefly discussed.

From the synthesis point of view, both geometry and appearance approaches have their own specific limitations. The problem with the geometry based (and more specifically the physically) approaches is the difficulty of generating natural looking expressions obtained for subtle skin deformation such as wrinkles and furrows. The main advantage of these approaches is that they are, generally, subject independent. On the other hand, because the appearance approaches rely on morphing and warping techniques, modeling requires a large database. These approaches (especially PCA and manifold based techniques) find emotional subspaces using subjects from the database, thus, mapping one expression from a subject in the database to another unseen (not included in the database) subject, sometimes leading to unsatisfactory results.

Apart from the approach dependent drawbacks, all current methods share (more or less) issues that are insufficiently tackled:

- *Open mouth issue*. This issue appears for certain expression, such as happiness or surprise and is more specific to appearance approaches. When going from neutral to surprise, teeth appear, and, consequently, they must be modeled as well. A few works accurately addressed this issue, most works ignoring this aspect (see below).
- *Frontal pose*. Most approaches were developed and tested only to subjects with

Table 3. Summary of facial expression synthesis techniques presented in the chapter

Method	Descriptors and characteristics	Shortcomings
(Blader, 1981)	Set of muscles attached to a skin mesh	Lack of natural expressions
(Waters, 1987)	Sphincter muscles	Does not address fine details as temporal wrinkles
(Terzopoulos, 1990)	Set of muscles attached to a skin mesh plus a third layer (fat tissue)	Does not address fine details as temporal wrinkles
(Hoch, 1994)	B-spline and 4 AUs	Lack of natural expressions
(Pighin, 2002)	3D Face mesh	It requires a set of images and it is face-dependent.
(Zhang, 2002)	FACS	Does not address wrinkles and teeth model
(Zhang, 2006)	134 facial points	Blending artefacts, lack of a reliable extrapolation method
(Yin, 2001)	Extended dynamic mesh (2954 vertices and 3118 patches)	Unrealistic facial expressions
(Liu, 2001)	Expression ratio image	It requires the expression ratio image from the performer.
(Raouzaiou, 2002)	FAPs and FATs	Limited facial deformation
(Wang, 2003)	Statistical method based on singular value decomposition of a training data	Not applicable for images not included in the training data.
(Abboud 2004)	Active Appearance Model	Not applicable for images not included in the training data.
(Tewes, 2005)	Gabor wavelet based flexible object model derived from graph nodes (landmarks)	Its accuracy highly depends on the landmarks location.
(Ghent, 2005)	Texture and shape model derived from FACS	Unrealistic facial expressions
(Malatesta, 2006)	FAPs + AUs	It generates confusion when subsequent expressions are constituted of conflicting animations.
(Deng, 2006)	Eigenspace model	Not applicable for images not included in the training data.
(Krinidis, 2006)	FAPs	Not applicable for images not included in the training data.
(Huang, 2006)	Manifold (learning) model	Large database requirement
(Lee, 2006)	Nonlinear manifold (learning) model	Large database requirement
(Mana, 2006)	HMM and FAPs	It highly depends on the training set
(Wang, 2008)	Manifold (learning) model	Large database requirement
(Sucontphunt, 2008)	3D face meshes	Computational expensive and requires extensive user interaction
(Zhang 2008)	FAPs, FACS and Bayesian Network	Large database requirement
(Susskind, 2008)	Deep Belief Nework	Large database requirement

frontal pose, although some works had dealt with 3D scanned images. Modeling facial expressions along with rotated faces is much difficult and requires high computational load, making the process unattractive for the current real-time facial expression synthesis models.

- *Illumination problem*. This is a common issue to any face related analysis. The reflection of human skin is approximately specular when the angle between the view direction and lighting direction is around 90^0. The light reflection should be treated seriously, as slight illumination variation (in its direction or intensity) may lead to failure in accurate synthesis of facial expression, particularly when the morphing step is employed. While other face analysis topics, such as face recognition for instance, addressed this issue, not much work was devoted in the literature for facial expression synthesis. Approaches borrowed from the face recognition field (including image gradients or illumination cone model) may help in accurately synthesizing facial expression under uncontrolled illumination variation.
- *Smooth deformation of fine geometric details*. Geometric details are important for human perception and measure the emotion intensity. However, they are difficult to be synthesized and to smoothly deform them is even harder so that the expression is realistically generated.
- *The own-race bias*. This term refers to the tendency of people to be more accurate in perceiving and recognizing differences amongst the faces of their own race than those pertaining to other racial groups. For the particular case of facial expression, several studies indicated major differences in facial expression perception of different race groups. For instance, comparing British and Japanese people asked to decipher non-verbal relationships in photographs, Kito and Lee (Kito, 2006) have found that Japanese participants were generally better at understanding subtle information provided by facial expression to decipher interpersonal relationships. In another study, Elfenbein and Amady (Elfenbein, 2002) brought evidence that the emotion recognition accuracy was higher when emotions were expressed and perceived by participants from the same "cultural" group. The own race bias concept highly relates to the fourth major and generic objective of the expression synthesizer, i.e., easy and accurate model adaptation to individual faces. Although important for universal facial expression mapping, this issue was not addressed, to our knowledge, by scientists involved in creation of facial expression synthesis models.
- *Generating distinctive (non-ambiguous) synthetic facial expression.* This is a common issue with the human ability of accurately discriminate between some particular emotions. Humans often confuse, for instance, fear and surprise due to the fact that they are perceptually similar and engage common muscle configuration and AUs. *Fear* and *surprise* share 5 common AUs out of 7 (see Table 2). Similar findings were reported when facial expressions have been classified automatically (Buciu et al, 2003). The difficulty of accurately recognizing these two emotions by both humans and machines, and, moreover, the almost common corresponding AUs set consequently leads to problems in generating distinctive associated synthetic facial expression for *fear* and *surprise*.

The first four issues above have been recently addressed by the Digital Emily Project that is a collaboration between the facial animation company Image Metrics and the Graphics Laboratory at the University of Southern California's Institute for Creative Technologies (see http://gl.ict.usc.edu/Research/DigitalEmily/ and Alexander et al (Alexander et al, 2009). The common work employs latest generation techniques in high-resolution face scanning, character rigging, video-based animation and compositing (blending). An actress

Figure 13. High resolution 3D geometry from fourteen of the thirty-three facial scans of the Emily actress's face posing several facial expressions (Alexander et al, 2009). With permission from © 2009 IEEE

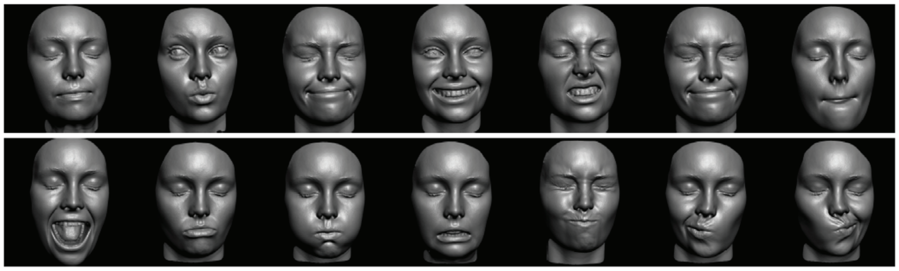
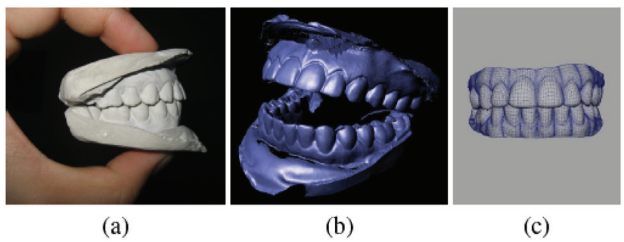


Figure 14. (a) A plaster cast of Emily actress's upper and lower teeth. (b) Tesulting merged 3D model before remeshing. (c) Remeshed model (Alexander et al, 2009). With permission from © 2009 IEEE



is first filmed on a studio set speaking emotive lines. The light variation is captured as a high dynamic range light probe image while the face of the actress is three-dimensional scanned in thirty-three facial expressions down to the level of skin pores and fine wrinkles. Animated eyes and teeth was added to the model and a semi-automatic video-based facial animation system was used to animate the 3Dface do to match the performance seen in the original video. The final face is also illuminated using the acquired reflectance maps with a complex skin translucency shading algorithm. The resulting model was generally accepted as being a real face. Figure 13 illustrates several high-resolution scans of the actress' face posing several facial expressions, while Figure 14 depicts the process of generating the model's teeth. It should be noticed that each resulting face mesh contains approximately three million polygons while the artificial teeth mesh is generated using 600.000 polygons, both processes requiring high computational load and resources. Despite being highly detailed and accurate, the expression scans require some pre-processing before blending into the final model. Mesh artifacts around teeth and eye regions may occur due to irregular edges with poor triangulation.

The applications involving humanoid robots also have limitations. On one hand, emotional robots with higher degree of freedom have to be built. The higher the degree of freedom, the more natural expression is achieved. On the other hand, robots with high degree of freedom may suffer from the uncanny valley issue, defined as the repulsion among humans caused by "too" human-like robots (MacDorman, 2006), leading to creepy robot faces. A fundamental question still remains when mapping a synthetic expression either to an affective robot or a computer generated avatar: what is the minimal level of display in order to generate an emphatic response and avoid thus the frightening effect when exaggerated (deformation) expressions are posed ? How social intelligent robots plan appropriate actions according to stimuli is another challenge. The issue is complex and involves advanced learning strategies that have to be addressed and implemented in the future.

The ultimate goal of human-computer interaction concerning the development of friendly human interfaces is the necessity of having real-time photo-realistic expression synthesis for unseen images. Although current implementations are close to achieving this goal (see Digital Emily Project), to date this complex requirement is still not fully accomplished, in spite of great endeavor provided by a large number of computer scientists and researchers. In addition, the human-computer interface is not limited to realistic facial expression synthesis, but also realistic gestures, eye gazes or natural lip movements, presenting important issues to be carefully addressed.

REFERENCES

Abboud, B., & Davoine, F. (2004) Appearance factorization based facial expression recognition and synthesis. In *IEEE International Conference on Pattern Recognition*, 958 – 965.

Albrecht, I., Schröder, M., Haber, J., & Seidel, H.-P. (2005). Mixed feelings: expression of non-basic emotions in a muscle-based talking head. *Virtual Reality (Waltham Cross)*, *8*(4), 201–212. doi:10.1007/s10055-005-0153-5

Alexander, O., Rogers, M., Lambeth, W., Chiang, M., & Debevec, P. (2009), Creating a Photoreal Digital Actor: The Digital Emily Project, *IEEE European Conference on Visual Media Production* (CVMP).

Badler, N., & Platt, S. (1981). Animating facial expressions. *Computer Graphics*, 245–252.

Breazeal, C. (2002). *Designing Sociable Robots*. MIT Press.

Breazeal, C., Gray, J., & Berlin, M. (2009). An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research*, *28*(5), 656–680. doi:10.1177/0278364909102796

Buciu, I., Kotropoulos, C., & Pitas, I. (2003) ICA and Gabor representations for facial expression recognition. *IEEE International Conference on Image Processing (ICIP 2003),* September 14-17,Barcelona, Spain, pp. 1054 – 1057.

Chandrasiri, N. P., Barakonyi, I., Naemura, T., Ishizuka, M., & Harashima, H. (2002) Communication over the Internet using a 3D agent with real-time facial expression analysis, synthesis and text to speech capabilities. In *Proceeding of the 8th International Conference on Communication Systems*, 01, 480 – 484.

Choi, S.-M., & Kim, Y.-G. (2005) An affective user interface based on facial expression recognition and eye – gaze tracking. *First International Conference on Affective Computing & Intelligent Interaction*, 3784, 907 – 914.

Deng, Z., & Neumann, U. (2006) eFASE: expressive facial animation synthesis and editing with phoneme-isomap controls. *Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (LNCS, 251 – 260).,New York: Springer.

Deng, Z., Neumann, U., Lewis, J. P., Kin, T.-Y., & Bulut, M. (2006). Expressive facial animation synthesis by learning speech coarticulation and expression spaces. *IEEE Transactions on Visualization and Computer Graphics*, *12*(6), 1523–1534. doi:10.1109/TVCG.2006.90

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System. Consulting Psychologists Press*. Alto, CA: Palo.

Ekman, P., Friesen, W. V., & Hager, J. C.,(2002). *Facial Action Coding System*. The Manual.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–235. doi:10.1037/0033-2909.128.2.203

Ge, S. S., Wang, C., & Hang, C. C. (2008). A Facial Expression Imitation System in Human Robot Interaction, *The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 213 – 218.

Ghent, J., & McDonald, J. (2005). Photo-realistic facial expression synthesis. *Image and Vision Computing*, *23*(12), 1041–1050. doi:10.1016/j.imavis.2005.06.011

Hoch, M., Fleischmann, G., & Girod, B. (1994). Modeling and animation of facial expressions based on B-splines. *The Visual Computer*, *11*, 87–95. doi:10.1007/BF01889979

Huang, L., & Su, C. (2006). Facial expression synthesis using manifold learning and belief propagation. *Soft Computing - A Fusion of Foundations*. *Methodologies and Applications*, *10*(12), 1193–1200.

Kito, T., & Lee, B. (2004). Interpersonal perception in Japanese and British observers. *Perception*, *33*, 957–974. doi:10.1068/p3471

Krinidis, S., Buciu, I., & Pitas, I. (2003*). Facial expression analysis and synthesis: A survey*. Proceedings of HCI International, 10th Int. Conference on Human-Computer Interaction, pp. 1432 – 1433, June 22-27, Crete, Greece.

Krinidis, S., & Pitas, I. (2006) Facial expression synthesis through facial expressions statistical analysis. In *Proc. 2006 European Signal Processing Conference*.

Lee, C.-S., & Elgammal, A. (2006) Nonlinear shape and appearance models for facial expression analysis and synthesis. *International Conference on Pattern Recognition*, 1, 497 – 502.

Liu, Z., Shan, Y., & Zhang, Z. (2001). Expressive expression mapping with ratio images. *In International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 271 – 276.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*.

Malatesta, L., Raouzaiou, A., Karpouzis, K., & Kollias, S. (2006). MPEG-4 facial expression synthesis based on appraisal theory, *3rd IFIP Conference* on *Artificial Intelligence Applications and Innovations*, 378 – 384.

Mana, N., & Pianesi, F. (2006) HMM-based synthesis of emotional facial expressions during speech in synthetic talking heads. In Proc. International Conference on Multimodal Interfaces, 380 – 387.

Mehrabian, A. (1968). Communication without words. *Psychology Today*, *2*(4), 53–56.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology (New Brunswick, N.J.)*, *14*, 261–292. doi:10.1007/BF02686918

Niewiadomski, R., Bevacqua, E., Maurizio, M., & Pelachaud, C., (2009c) Greta: an interactive expressive ECA system. *AAMAS* (2), pp. 1399-1400

Niewiadomski, R., Hyniewska, S., & Pelachaud, C. (2009a) Modeling emotional expressions as sequences of Behaviors, International conference on Intelligent virtual agents IVA'09, Amsterdam.

Niewiadomski, R., Hyniewska, S., & Pelachaud, C. (2009b) Evaluation of Multimodal Sequential Expressions of Emotions in ECA, International conference on Affective Computing & Intelligent Interaction ACII'09, Amsterdam.

Pelachaud, C., (2009) Modelling Multimodal Expression of Emotion in a Virtual Agent, *Philosophical Transactions of Royal Society B Biological Science*, B, 364, pp. 3539-3548.

Pighin, F., Szeliski, R., & Salesin, D., H. (2002). Modeling and animating realistic faces from images. *International Journal of Computer Vision*, *50*(2), 143–169. doi:10.1023/A:1020393915769

Raouzaiou, A., Tsapatsoulis, N., Karpouzis, K., & Kollias, S. (2002). Parameterized Facial Expression Synthesis Based on MPEG-4. *EURASIP Journal on Applied Signal Processing*, (10): 1021–1038. doi:10.1155/S1110865702206149

Scherer, K. R. (2001) Appraisal considered as a process of multilevel sequential checking. In Scherer, K.R., Schorr, A., & Johnstone, T., (Eds) *Appraisal Processes in Emotion: Theory Methods*, Research. Oxford, New York: Oxford University Press, 92-129.

Song, M., Wang, H., Bu, J., Chen, C., & Liu, Z. (2006) Subtle facial expression modeling with vector field decomposition. In *IEEE International Conference on Image Processing*, 2101 – 2104.

Sucontphunt, T., Mo, Z., Neumann, U., & Deng, Z. (2008) Interactive 3D facial expression posing through 2D portrait manipulation. *Proceeding of Graphics Interface*, 71(10 – 12), 177 – 184.

Susskind, J. M., Hinton, G. E., Movellan, J. R., & Anderson, A. K. (2008). Generating Facial Expressions with Deep Belief Nets. In Kordic, V. (Ed.), *Affective Computing, Emotion Modelling, Synthesis and Recognition*. ARS Publishers.

Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J. R., & Anderson, A. K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, *45*(1), 152–162. doi:10.1016/j.neuropsychologia.2006.05.001

Terzopoulos, D., & Waters, K. (1990). Physically-based facial modeling and animation. *Journal of Visualization and Computer Animation*, *1*(4), 73–80.

Tewes, A., Würtz, R. P., & Von der Malsburg (2005) A flexible object model for recognising and synthesising facial expressions. In Takeo Kanade, Nalini Ratha, & Anil Jain (eds.), *Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication*,(LNCS, 81-90) Springer.

Vogiatzis, D., Spyropoulos, C., Konstantopoulos, S., Karkaletsis, V., Kasap, Z., Matheson, C., & Deroo, O. (2008) An affective robot guide to museums. *In Proc. 4th International Workshop on Human-Computer Conversation*.

Wang, H., & Ahuja, N. (2003) Facial expression decomposition. In *IEEE International Conference on Computer Vision*, 958 – 965.

Wang, H. & Wang, K. (2008) Affective interaction based on person independent facial expression space. *Neurocomputing, Special Issue for Vision Research*, 71(10 – 12), 1889 – 1901.

Waters, K. (1987). A muscle model for animating three – dimensional facial expression. *Computer Graphics*, *22*(4), 17–24. doi:10.1145/37402.37405

Yin, L., Basu, A., Bernögger, S., & Pinz, A. (2001). Synthesizing realistic facial animations using energy minimization for model-based coding. *Pattern Recognition*, *34*(11), 2201–2213. doi:10.1016/S0031-3203(00)00139-4

Yongmian, Z. Q. J. Z. Z. B. Y. (2008). Dynamic Facial Expression Analysis and Synthesis With MPEG-4 Facial Animation Parameters. *IEEE Trans. on Circuits and Systems for Video Technology*, *18*(10), 1383–1396. doi:10.1109/TCSVT.2008.928887

Zhang, Q., Liu, Z., Guo, B., Terzopoulos, D., & Shum, H.-Y. (2006). Geometry – driven photo-realistic facial expression shyntesis. *IEEE Trans. on Visualization and Computer Graphics*, *12*(1), 48–60. doi:10.1109/TVCG.2006.9

Zhang, S., Wu, Z., Meng, H. M., & Cai, L. (2007) Facial expression synthesis using PAD emotional parameters for a chinese expressive avatar. *In Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction,* 4738, 24 – 35.

Zhang, Y., & Prakash, E., C., & Sung, E. (2002). Constructing a realistic face model of an individual for expression animation. *International Journal of Information Technology*, *8*(2), 10–25.

KEY TERMS AND DEFINITIONS

Embodied Conversational Agents (ECAs): Friendly and intelligent user interfaces built to mimic human gestures, speech or facial expressions.

Facial Action Unit (AUs): A facial action unit is an objective description of facial signals in terms of component motions.

Facial Action Coding System (FACS): A system developed to encode a comprehensive set of all possible visually distinguishable facial appearances by measuring specific facial muscle movements.

Facial Action Parameters (FAPs): Geometrical facial descriptors.

Face Animation Tables (FATs): MPEG-4 compliant values of model vertices used to deform the face geometry.

PAD Parameters: High-level parameters defined as pleasure-displeasure (P), arousal-nonarousal (A) and dominance-submissiveness (D) use to express a continuous facial space model for describing universal emotions.

Partial Expression Parameters (PEP): middle level facial parameters describing the facial motion.