

Facial expression analysis and synthesis: A survey

Stelios Krinidis, Ioan Buciu and Ioannis Pitas

Department of Informatics
Aristotle University of Thessaloniki
Box 451, 54006 Thessaloniki, Greece
e-mail: pitas@zeus.csd.auth.gr

Abstract

Human facial expression analysis and synthesis play a central role in the social context, being investigated by many psychologists over the time. Moreover, an increased interest has been shown in developing an automated facial expression analyzer capable of recognizing, classifying and then synthesizing human expressions on synthetic “talking heads”. This paper surveys the state of the art in this area.

1 Introduction

The analysis and synthesis of facial expressions have challenged many researchers not only from the field of psychology but also from the field of computer science. It is well known among psychologists that the social context is dominated by language. However, the language alone is insufficient when it comes to successful social interaction. As a consequence, the nonverbal communication systems, including facial expressions, have kept an increased attention from the psychological perspective.

The analysis and synthesis of facial expressions is not limited only to psychology. Many efforts have been made by the computer scientists working in computer interface (HCI) to automatically identify and generate realistic human facial expressions. A fully automatic facial expression analyser should be able to cope with the following tasks: 1) detect the face in a scene; 2) extract the facial expression features; 3) recognize and classify facial expressions according to some classification rules. Likewise, a facial expression synthesizer should have as an ultimate goal the: 1) creation of realistic expressions; 2) operation in real time; 3) automation as much as possible; 4) easy adaptation to individual faces.

The survey proceeds as follows. Section 2 introduces the facial expressions analysis, whereas facial expressions synthesis is presented in Section 3. Conclusions are drawn in Section 4.

2 Facial Expression Analysis

There are three steps concerning an automatic facial expression analysis: face detection (tracking), facial feature extraction and facial expression classification. We shall focus our presentation to the classification issues.

2.1 Facial Expression Classification

The classification is done according to certain facial action coding schemes. Two classes of approaches can be found in this regard:

- ***Spatio-temporal approaches.***

In (Essa & Pentland, 1999) we find a method to extract the spatio-temporal motion energy representation of facial motion for an expression. Hidden Markov Model (HMM) is another approach that has been used to accomplish facial expression classification. (Otsuka & Ohya, 1998) used HMM to match a temporal sequence of the 15 dimensional feature vector to the models of six basic facial expressions. (Lien, Kanade, Cohn & Li, 1998) facial expression features to HMMs that classify it into FACS action units. An extension of the pseudo 2-D HMMs that is called pseudo 3-D HMMs (P3DHMMs) is applied to dynamic facial expression recognition by (Muller, Eickeler & Rigoll, 1999). (Hoey, 2001) presents a hierarchical dynamic Bayesian network that models the visual events at each of a number of temporally abstract levels. A level in the network consists of a mixture of Markov chains (MMC).

- ***Spatial approaches.***

Neural networks systems are often used for facial expression recognition and are used either directly on facial images or combined with principal component analysis (PCA), independent component analysis (ICA) or Gabor wavelets filter. (Rosenblum, Yacoob & Davis, 1996) use a system of networks where the complexity of recognition facial expressions is divided into three layers of decomposition. The first layer identifies expression of emotion; the second layer identifies motion of three facial features while the third layer recovers motion directions. (Fasel, 2002) developed a system based on convolutional neural networks in order to allow for an increased invariance with regard to translation and scale changes. He uses multi-scale simple feature extractor layers in combination with weight sharing feature extraction layers. Another neural network is used by (Dailey, Cottrell, Padgett & Adolphs, 2002). The data are processed in three steps: first the image is filtered by applying a grid of overlapping 2-D Gabor filters; the second step is to perform dimensionality reduction by applying PCA; the reduced data is fed into a neural network containing 6 outputs one for each of the six basic emotions. Support Vector Machines is another approach used to tackle facial actions classification employed by (Schulze, Scheffler & Omlin, 2002).

3 Facial Expression Synthesis

In this Section, facial expressions synthesis approaches are summarized. Facial modeling and expressions synthesis research falls into two major categories, those based on geometric manipulations and those based on image manipulations.

3.1 Geometric Manipulations

- ***Interpolation*** techniques offer an intuitive approach to facial expressions synthesis. Typically, an interpolation function specifies smooth motion between two key-frames at extreme positions, over a normalized time interval. Although, interpolations are fast and easily generate primitive facial expressions, their ability to create a wide range of realistic facial configurations is severely restricted (Pighin, Hecker, Lischinski, Szeliski & Salesin, 1998).

- ***Parameterisation*** techniques for facial expressions (Parke & Waters, 1996) overcome some of the limitations and restrictions of simple interpolations. Ideally parameterizations specify any possible face and expression by a combination of independent parameter values. Unlike interpolation techniques, parameterizations allow explicit control of specific facial configurations. Combinations of parameters provide a large range of facial expressions with relatively low

computational costs. Nevertheless, there is no systematic way to arbitrate between two conflicting parameters, hence parameterization rarely produces natural human expressions. Another limitation is that the choice of the parameter set depends on the facial mesh topology and, therefore, a complete generic parameterization is not possible. Furthermore, tedious manual tuning is required to set parameter values, and even after that, unrealistic motion or configurations may result.

- **Physics-based muscle models** fail into three categories: mass spring systems, vector representations, and layered spring meshes.

- **Mass-spring** methods propagate muscle forces in an elastic spring mesh that models skin deformation. Forces applied to elastic meshes through muscle arcs generate realistic facial expressions.

- **The vector muscle** method models the action of muscles upon skin exploiting a delineated deformation field. The positioning of vector muscles into anatomically correct positions can be a daunting task. No automatic way of placing muscles beneath a generic or person-specific mesh is reported. They are widely used because of their compact representation and independence of the facial mesh structure.

- **Layered spring mesh muscles** models detailed anatomical structure and dynamics of the human face (Terzopoulos & Waters, 1990). This model achieves great realism. However, simulating volumetric deformations with 3D lattices requires extensive computation.

- **The Finite Element Method (FEM)** is a numerical approach to approximating the physics of an arbitrary complex object. An object is decomposed into area or volume elements, each endowed with physical parameters. The dynamic element relationships are computed by integrating the piecewise components over the entire object (Nebel, 2001). They are quite computationally intensive methods.

- **Pseudo or simulated muscle** modeling produces realistic results by approximating human anatomy, but it is difficult to consider the exact modeling and parameter tuning needed to simulate a specific human facial structure. Muscle forces are simulated in the form of splines, wires, or free form deformations:

- **Free form deformation (FFD)** deforms volumetric objects by manipulating control points arranged in a 3D lattice. FFDs can deform many types of surface primitives, including polygons; quadric, parametric, and implicit surfaces; and solid models. On the other hand, rational FFD (RFFD) (Kara, Mangili, Thalmann & Thalmann, 1992) incorporates weight factors for each control point. Dirichlet free form deformations (DFFD) (Lee, Kara & Thalmann, 1997) is another free-form variation.

- **Spline pseudo muscle** models supports smooth and flexible deformations, allowing localized deformation on the surface, reducing the computational complexity.

- **Wire** curves together with a collection of domain curves provide an implicit modeling primitive (Singh & Fiume, 1998).

- **Wrinkles:** There are two types of wrinkles, temporary wrinkles that appear for a short time in expressions, and permanent wrinkles that form over time as permanent facial features. Wrinkles and creases are modeled by physically based modeling with plasticity or viscosity, and texture techniques like bump mapping:

- **Bump mapping** produces perturbations of the surface normals that alter the shading of a surface, either by defining wrinkle functions, or using morphing. Bump mapping was difficult to compute in real time.

- **Physically based wrinkle** models use the plastic-visco-elastic properties of the facial skin and permanent skin aging effects (Wu, Thalmann & Thalmann, 1994). Viscosity is responsible for time dependent deformation, while plasticity is for non-

invertible permanent deformation that occurs when an applied force goes beyond a threshold.

- Other wrinkles approaches are spline segments and time-varying homotopy based on homotopy sweep technique (Moubaraki, Tanaka, Kitamura, Ohya & Kishimo, 1994).

3.2 Image Manipulations

- **2D and 3D morphing** effects a metamorphosis between two target images or models.

A 2D image morph consists of a warp between corresponding points in the target images and a simultaneous cross dissolve. Typically, the correspondences are manually selected to suit the needs of the application. The 2D and 3D morphing methods can produce realistic facial expressions, but they share similar limitations with the interpolation approaches.

- **Vascular expressions:** Realistic face modeling and animation demand not only face deformation, but also skin color changes that depend on the emotional state of the person. Not much research is reported on this subject. It can be achieved by changing the color of all the polygons during strong emotion (Karla & Thalmann, 1994), or using texture mapping.

- **Model fitting:** Some techniques have been developed, in order to fit a predefined wireframe on a face. This face model can be of any level of detail, but it is usually limited in locating a rough outline of the face and a set of facial movements. Such model-fitting algorithms on faces derived from images are described in (Ahlberg, 2001).

- **Expressions using tracking:** The difficulties of other methods in achieving life-like facial expressions led to the performance driven approach where tracked human actors control the expressions. Real time video processing allows interactive animations where the actors observe the animations they create with their motions and expressions. Accurate tracking of feature points or edges is important to maintain a consistent and life-like quality of expression.

4 Conclusions

In this paper, we describe and survey the issues associated with facial expressions analysis and synthesis. Analysis of human facial expressions consists of three steps: face detection (tracking), facial feature extraction and facial classification. Automatically classification of human facial expressions constitutes the best step of an automatic facial expression analysis system. This task performed according to certain facial action coding schemes, using either spatio-temporal or spatial approaches. Generation of facial expressions can be summarized as follows. First, an individual specific model is obtained and fitted into the prearranged prototype mesh. Second, the constructed individual facial model is deformed to produce facial expression. Wrinkles and vascular effects are also considered for added realism. The goal of research related to the synthesis, achieving realism in real time in automated way, has not been reached yet.

References

- Ahlberg, J. (2001), Real-time facial feature tracking using an active model with fast image warping, in *Proceedings of the International Workshop on Very Low Bitrate Video*, pp. 39-43.
- Dailey, M. N., Cottrell, W. C., Padgett, C., & Adolphs, R. (2002). EMPATH: A Neural Network that Categorizes Facial Expressions. *Journal of Cognitive Science*, 14(8), 1158-1173.
- Essa, I., & Pentland, A. (1999). Coding, Analysis, Interpretation, Recognition of Facial Expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7), 757-763.

- Fasel, B. (2002). Multiscale Facial Expression Recognition using Convolutional Neural Networks. *In Proc. of the third Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP'2002)*.
- Hoey, J. (2001). Hierarchical unsupervised learning of facial expression categories. *IEEE Workshop on Detection and Recognition of Events in Video (EVENT'01)*, 92-99.
- Karla, P., Mangili, A., Thalmann, N.M., and Thalmann, D. (1992), Simulation of facial muscle actions based on rational free form deformations, *in Computer Graphics Forum*, pp. 59-69.
- Karla, P., and Magnenat-Thalmann, N. (1994), Modeling of vascular expressions in facial animation, *in Computer Animation*, pp. 50-58.
- Lee, W.S., Karla, P., and Magnenat-Thalmann, N. (1997), Model based face reconstruction for animation, *in Proceedings of Multimedia Modeling*, pp. 323-338.
- Lien, J. J. J., Kanade, T., Cohn, J. F., & Li C. C. (1998). Detection, Tracking, and Classification of Action Units in Facial Expression. *In Journal of robotics and Autonomous Systems*, 324-329.
- Moubaraki, L., Tanaka, H., Kitamura, Y., Ohya, J., and Kishino, F. (1994), Homotopy-based 3D animation of facial expression, *Tech. Rep. IE-94-37*, IEICE.
- Muller, S., Eickeler, S., & Rigoll G. (1999). Pseudo 3D HMMs for Image Sequence Recognition. *In IEEE Proc. Inter. Conf. on Image Processing*, 237-241.
- Nebel, J.C. (2001), Soft tissue modeling from 3D scanned data, *in Kluwer*, pp. 85-97.
- Otsuka, T., & Ohya, J. (1998). Spotting Segments Displaying Facial Expression from Image Sequences Using HMM. *In Proc. Inter. Conf. on Automatic Face and Gesture Recognition*, 442-447.
- Parke, F.I., and Waters, K. (1996), Computer Facial Animation, *A.K.Peters Ltd*.
- Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., and Salesin, D.H. (1998), Synthesizing realistic facial expressions from photographs, *in Proceedings of SIGGRAPH*, pp. 75-84.
- Rosenblum, M., Yacoob, Y., & Davis, L. (1996). Human expression recognition from motion using a radial basis function network architecture. *IEEE Trans. on Neural Networks*, 7(5), 1121-1138.
- Schulze, M., Scheffler, K., & Omlin, C. W. (2002). Recognizing Facial Actions with Support Vector Machines. *In Proc. PRASA*, 93-96.
- Singh, K. and Fiume, E. (1998), Wires: A geometric deformation technique, *in Proceedings of SIGGRAPH*, pp. 405-414.
- Terzopoulos, D. and Waters, K. (1990), Physically-based facial modeling analysis and animation, *Visualization and Computer Animation*, 1(4), pp. 73-80.
- Wu, Y., Magnenat-Thalmann, N., and Thalmann, D. (1994), A plastic-visco-elastic model for wrinkles in facial animation and skin aging, *in Proceedings of Pacific Graphics*.