Cloning, Morphing, then Tracking Real Emotions

WON-SOOK LEE, TARO GOTO, NADIA MAGNENAT-THALMANN
MIRALab, CUI, University of Geneva, Geneva, Switzerland
E-mail: {wslee, goto, thalmann}@cui.unige.ch
Fax: ++ 41 (0)22 705-7780
Tel: ++ 41 (0)22 705-7766
http://www.miralab.unige.ch

Abstract

We describe the full pipeline from cloning a real person, morphing of several cloned people to the real-time emotion tracking of a person in front of a video camera. As telecommunication is a rising area for virtual reality, there has been the introduction of new standard MPEG-4 for practical and low bandwidth communication in real-time. We focus on the cloning, morphing and tracking methods developed for very efficient and robust communication and visualization, which can be used for a local system or remote system with standard parameters. The advantage in having a 360 ^{o} view of a virtual clone rather than just video streams is a 3D viewing of a virtual emotional face in any angle in any kind of environment. Our system has a big potential for the analysis and understanding of real-emotion, and creating an emotion database. The input for cloning is just two pictures of a person (front and side views) and the individualized 3D face is created by modifying a predefined generic model with detected features on the two photos. The generic model has all function information for animation and shape inside, so that any face from it inherits the functional structure for facial animation. Once several people are virtually cloned, we use a rapid 3D-morphing system, which enables a user to control the shape and skin color in a very intuitive way. It is an integration of 3D shape and 2D skin information. Then any face, (either her/his clone or any other created face from 3D-morphing system), will be used for the visualization of emotion tracking system. With a very simple initialization, the system tracks features like eyes, eyebrows, and a mouth shape on the participant's face. The system successfully tracks any kind of face in real-time, independent of skin color, with or without beard, or glasses.

Keywords: Orthogonal photos, facial animation, texture mapping, morphing, feature tracking, MPEG-4 FAP.

1. INTRODUCTION

Our method focuses on an easy and intuitive method of creating faces and animating them, tracking their true facial expressions, in a virtual world. We design a virtual cloning system starting from photo data and one generic model, then the next step is to generate new population out of cloned face data using 3D-morphing system, which use shape and image morphing, but eliminate many computational intensive steps being profited from 3D-face structure. Our real-time facial feature tracking system allows us to generate and visualize real expressions on either their own clone or any virtual face from a database.

Approaches to the realistic reconstruction of individuals, some of them with a view to animation, include the use of a laser scanner **[1]**, a stereoscopic camera[3], an active light stripper **[2]**, or a video stream **[4]** to reconstruct heads and natural expressions. Modeling has also been done from picture data **[5]**, detecting features, modifying a given generic model and then mapping texture onto it. Not all of these, however, combine sophisticated and reliable shape deformation methods with seamless, high-resolution texture generation and mapping.

Techniques for generating virtual populations have been investigated by DeCarlo et al. **[6]**. However, we are more interested in intuitively controlling the creation of people using 3D-morphing system. The main 2D-morphing methods **[7][8]** have little been exploited in terms combining it with 3D-shape morphing. Nevertheless, this seems a natural approach to better results with less effort. Furthermore, with face cloning methods, many user interactions can be avoided.

In research field of a multimodal interface, recognition of the facial expressions is a very complex and interesting subject where there have been numerous research efforts. Cosatto et al. **[9]** used a sample-based method which needs to make a sample for each person. Kouadi et al. **[10]** used a database and some face markers effectively to process it in real-time. However, the use of markers is not always practical and it is much more attractive to recognize features without them.

Pandzic et al. **[11]** use an algorithm based on edge extraction in real-time, without markers. The presented face tracking method is inspired from this method. The output is converted to MPEG-4 based FAP points **[12][13]**. These feature points are sent to a real-time player that deforms and displays a synthetic 3D animated clone face or morphed face. **Figure 1** shows the whole system of making clones and then tracking the face in real-time.

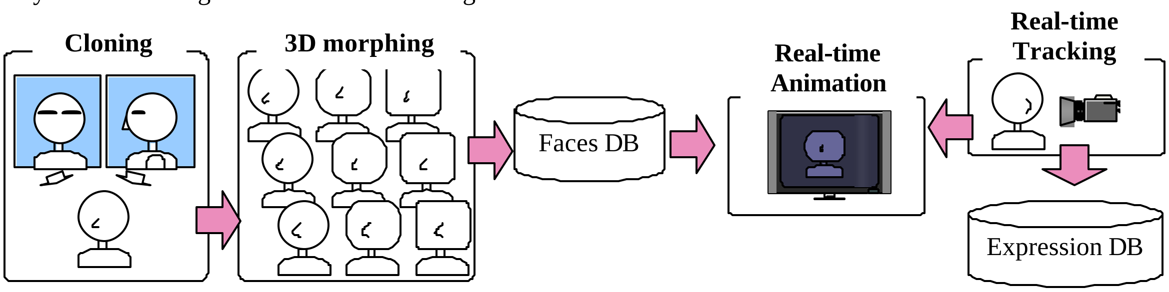
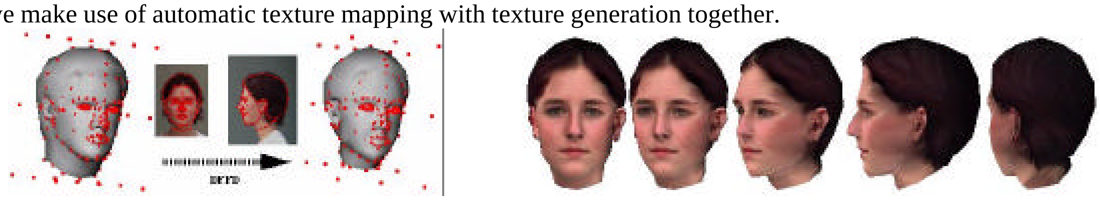


Figure 1: System Overview

2. CLONING

2D photos offer clues to the 3D shape of an object. It is not feasible, however, to consider 3D-points densely distributed on the head. In most cases, we know the location of only a few visible features such as eyes, lips and silhouettes, the key characteristics for recognizing people, which are detected in a semiautomatic way using the structured snake method with some anchor functionality for a subset of feature points. **Figure 2(a)** depicts an orthogonal pair of normalized images, showing the detected features. The two 2D sets of position coordinates, from front and side views, i.e., the *(x, y)* and the *(z, y)* planes, are combined to give a single set of 3D points. The problem is how to deform a generic model, which has more than a thousand points to make an individualized smooth surface. One solution is to use 3D feature points as a set of control points for a deformation. Then the deformation of a surface can be seen as an interpolation of the displacements of the control points. In the Dirichlet-based FFD approach, any point of the surface is expressed relative to a subset of the control points set with the Sibson coordinate system. Therefore DFFD **[14]** is used here to get new geometrical coordinates for a modification of the generic head on which are situated the newly detected feature points. The correct shapes of the eyes and teeth are assured through translation and scaling appropriate to the new head. As shown in **Figure 2(b)**, the result is quite respectable when we consider the input data (pictures from only two views). To improve the realism, we make use of automatic texture mapping with texture generation together.



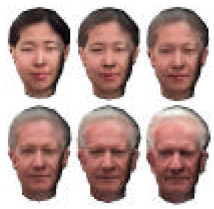
(a) **(b)**

Figure 2: (a) Modification of a generic head according to feature points detected on pictures. Points on a 3D head are control points for DFFD. (b) Snapshots of a reconstructed head in several views.

Texture mapping serves not only to disguise the roughness of shape matching determined by a few feature points, but also to imbue the face with more realistic complexion and tint. If the texture mapping is not correct, the accurate shape is useless in practice. We therefore use information from the set of feature points detected to generate texture in a fully automatically, based on the two views. The main criterion is to obtain the highest resolution possible for most detailed portions. We first connect two pictures along predefined feature lines using geometrical deformations and, to avoid visible boundary effects, a multiresolution technique. We then obtain appropriate texture coordinates for every point on the head using the same image transformation. **Figure 2(b)** shows several views of the head reconstructed.

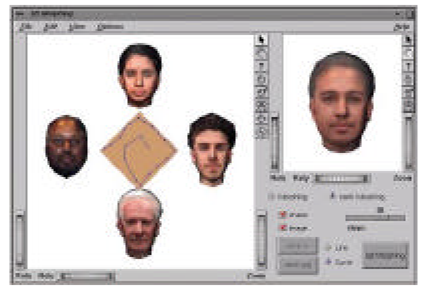
3. 3D MORPHING SYSTEM

Morphing technology has seen a great deal of development, either in only 2D or 3D. In this paper, we vary aspects of both 2D and 3D representations in creating new virtual faces, showing how easy and fast the smooth morphing can be achieved. When we morph one person into another in 3D, we must deal with alteration in both shape and texture. Since every face created through the methods described here shares the same topology, it is relatively easy to vary head shapes in 3D just using a simple linear interpolation of 3D coordinates. Then we utilize 3D information to facilitate 2D morphing among texture images following triangle by triangle the entire 3D head. The resulting head manifests very smooth images without any gaps in the textured parts. This approach for two persons can be very easily extended to several people. **Figure 3(a)(b)** illustrates 3D morphing between two persons. The interface in **Figure 3(c)** shows 3D morphing among several people. It has two windows, the left one for controlling weights among several input faces and the upper right one containing the resulting new face. Since only a barycentric calculation for texture image pixels is required, the calculation is very fast, which makes it possible to see real-time 3D morphing. There is also an option to have the shape and image variation separately to enable more various creations of new faces.



(a)

(b)



(c)

Figure 3: (a) morphing between two persons (b) in another view (c) A dynamic system for creation of new population from several people in intuitive way.

4. FACIAL FEATURE TRACKING SYSTEM

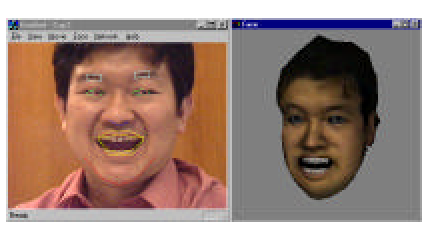
The motivation of the facial feature tracking system is to mo ve the clone in a virtual environment. The system includes a head-mounted video camera and a computer that tracks features. A person puts on the head-mounted camera, and initializes the system first. After the initialization, the system tracks the person's features in real-time. In detail, MPEG-4 Facial Animation Parameters are extracted from a real-time video stream, and they are sent to the virtual environment. These FAPs can be used either to animate the cloned face or to compute the high-level emotions transmitted to the actor. The FAPs take up a very small bandwidth when animating in real-time.

Tracking facial features is a key issue in face analysis. To obtain a real-time tracking, many problems have to be resolved. One important problem lies in the variety of the appearances of individuals, such as skin color, eye color, beards, glasses, etc. The facial features are sometimes not separated by sharp edges, or edges appear at unusual places. This diversity increases the difficulty for the recognition of faces and the tracking of the facial features.

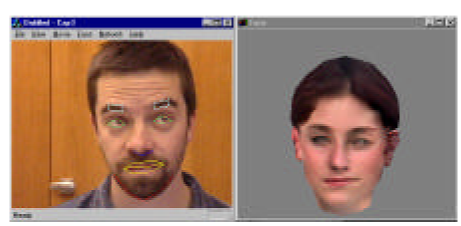
In this application, the facial features and their associated information are set during an initialization phase that will solve the main problem in facial feature differences between people. **Figure 4(a)** shows this initialization. The way to initialize one's face is very easy. The user only moves some feature boxes, like pupil boxes, a mouth box etc, to proper regions. Once the user sets the feature positions, the information around the features, the edge infomation, and the face color information are extracted automatically. Various parameters used during feature tracking are then generated automatically. Those parameters, gathered during initialization phase for every face, contain all the relevant information for tracking the face position and its corresponding facial features without any markers.

The tracking process is separated into two parts: 1) mouth tracking and 2) eye tracking. The edge and the gray level data around the mouth and the eyes are the main data used during tracking. These data is combined with a proper weight, and the best movement from the previous shape is decided as the next shape. Then the shape that is extracted by the image recognition is converted into MPEG-4 based FAPs. It is attractive to obtain an output compatible with the latest standard. This way, the stream of output value can be used on any clone, avatar or even comic characters without any reprogramming. **Figure 4(b)(c)** shows some results. The actor's face movement is used for another one's cloned face movement.

(a)



(b)



(c)

Figure 4: (a) Initialization (b) Feature tracking with his own clone (c) with another virtual face

FAPs are created from a movement of features. Therefore, the neutral face that is used for the initialization is considered as no movement. We use the difference between the initial points and tracked points to make these FAPs. For example, a mouth width was 100 dots and the edge moved 10 dots, movement is shown as 1024*10/100, so that the size of the mouth edge movement was 102. This value does not change when the cloned face is changed to another person. The recognition method for facial expressions in our system does not use any special markers or make-up. It does not need training and only a simple initialization of the system allows it to adapt to a new immediately. Extracted data is transformed into MPEG-4 FAPs that can be used easily for any compatible facial animation system and transmitted to web pages all over the world.

5. CONCLUSIONS

We presented our system covering topics from cloning, creating virtual humans to representation of facial expressions with real-time feature tracking. Just two photographs of a few persons make not only virtual clones of them but also a large number of photo-realistic virtual faces using rapid 3D-morphing method. Through facial tracking method, we create an expression database based on parameters. These three methods are well connected to meet the rising needs of network communication, which is realistic looking, and movement of human with low bandwidth. This approach has potential to be extended for both human face shape and expression analysis.

6. ACKNOWLEDGMENT

The authors would like to thank Laurent Moccozet, Pierre Beylot, Christian Zanardi and Marc Escher. We are grateful to Chris Joslin for proof reading this document. This project is funded by an European project eRENA and Swiss National Research Foundation.

7. REFERENCE

[1] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters, "Realistic Modeling for Facial Animation", In *Computer Graphics (Proc. SIGGRAPH'96)*, pp. 55-62, 1996.

[2] Marc Proesmans, Luc Van Gool. "Reading between the lines - a method for extracting dynamic 3D with texture". In *Proc. of VRST'97*, pp. 95-102, 1997.

[3] http://www.turing.gla.ac.uk/turing/copyrigh.htm

[4] P. Fua, "Face Models from Uncalibrated Video Sequences", In *Proc. CAPTECH'98*, pp. 215-228, 1998.

[5] Kiyoshi Arai, Tsuneya Kurihara, Ken-ichi Anjyo, "Bilinear interpolation for facial expression and metamorphosis in real-time animation", *Visual Computer*, Vol. 12, No. 3, 1996.

[6] Douglas DeCarlo, Dimitris Metaxas and Matthew Stone, "An Anthropometric Face Model using Variational Techniques", In *Computer Graphics (Proc. SIGGRAPH'98)*, pp. 67-74, 1998.

[7] Lee S., Chaw K-Y, Hahn J., Shin SY, "Image morphing using deformation techniques", *Journal Visualization Computer Animation*, 7:3-23, 1996.

[8] Gao P., Sederberg TW, "A work minimization approach to image morphing", *Visual Computer*, International Journal of Computer Graphics, Vol. 14, No. 8/9, pp. 390-400, 1998.

[9] E. Cosatto, H.P. Graf, 'Sample-Based Synthesis of Photo-Realistic Talking Heads', Computer Animation 1998, pp103 – 110

[10] C. Kouadio, P. Poulin, P. Lachapelle, 'Real-Time Facial Animation based upon a Bank of 3D Facial Expressions', CA98, 1999, pp128 – 136

[11] I.S. Pandzic, T.K. Capin, N. Magnenat-Thalmann, D. Thalmann, 'Towards Natural Communication in Networked Collaborative Virtual Environments', FIVE '96, December 1996

[12] SNHC, 'Information Technology – Generic Coding Of Audio – Visual Objects Part 2: Visual', ISO/IEC 14496-2, Final Draft of International Standard, Version of:13, Nov., 1998, ISO/IEC JTC1/SC29/WG11 N2502a, Atlantic City, Oct. 1998

[13] SNHC, 'Text for ISO/IEC FDIS 14496-1 Systems (2nd draft)', ISO/IEC IEC JTC1/SC29/WG11 2501, 1997. ISO/IEC JTC1/SC29/WG11 N2501, Nov.,1998

[14] Moccozet L., Magnenat Thalmann N., "Dirichlet Free-Form Deformations and their Application to Hand Simulation", In *Proc. Computer Animation'97*, IEEE Computer Society, pp.93-102, 1997.