Face Animation by Real Time Feature Tracking

Xiaozhou Wei[†], Zhiwei Zhu[‡], Lijun Yin[†,*] Qiang Ji[‡]
[†]Department of Computer Science, SUNY at Binghamton, Binghamton, NY 13902
[‡]Department of ECSE, Rensselaer Polytechnic Institute, Troy, NY 12180

1 Introduction

Tracking limited facial features in video and reproducing its expression remotely could save transmission bandwidth significantly, however, a number of issues need to be accounted including the real time face feature tracking under a variety of imaging conditions, and the real time face modelling and animation using limited number of feature parameters. Thus we developed a noval system to create face animation with twenty-two tracked features [1] .

2 Related Works

The appearance-driven tracking and the model based tracking approach are widely applied in feature extraction [Tomasi and Kanade 1991]. To produce a realistic facial expression for avatar communication, physical based method [Terzopoulos 1993] using physical based animation and elastically deformable model, image based model fitting use a synthesis by analysis strategy, while modification-based method cloning face expression with motion data mapping from source to target [Noh and Neumann 2001].

3 Face Tracking

In order to obtain the motion parameters, first an active sensing system is built with an IR sensitive camera and two concentric rings of IR LEDs. Eye tracking result is further improved by a mean shift eye tracker. Ratio of pupil ellipse axes is used to characterize the percentage of eye closure. Twenty-two facial features around eyes and mouth are selected for tracking. Multi-scale and orientation Gable wavelet is used to represent features. With the assumption of smooth motion of each feature, Kalman filer is used for tracking. A measurement model used in Kalman filter is further defined as: $\mathbf{O_t} = H\mathbf{S_t} + \mathbf{U_t}$ where H is the measurement matrix and $\mathbf{U_t}$ models measurement uncertainty. Based on above two models and some initial conditions (e.g., detected pupil position), the state vector $\mathbf{S_{t+1}}$ along with its covariance matrix can be updated, and therefore the position prediction of each feature ($\vec{P} = (x,y)^T$) can be obtained. Finally, by combining the head motion with the Kalman filtering, we can obtain the accurate feature location in the current frame. Based on detected eyes and certain anthropometric statistics, 3D face pose is then tracked in 3D space. The detected face normal vector is perpendicular to the face plane, which is represented by three angles of orientation $(\alpha_{roll}, \beta_{pitch}, \gamma_{yaw})$.

4 Animation By Dynamic Inference

To transfer the source motion parameters (MP) to target animation parameters (AP), first the individual 3D face avatar is firstly created by our face modelling algorithm based on two views of an individual face. We developed an algorithm to reliably identify features on the profile such as chin tip, mouth, nose tip, and nose bridge. After the process of static face modelling, regions like mouth, nose, eye, eyeballs, and eyebrows are then marked and the areas to be influenced by the feature points are defined. In order to obtain APs, we need to process source MPs by a series of transformations. The motion of each feature point on a face can be modelled as a local skin deformation plus a global head affine transformation. Given a feature point $\vec{v}_o$ with the front view of a source subject, its corresponding point ($\vec{v}_p$) after the local deformation due to expressions and global motion due to the head pose change is formulated as follows: $\vec{v}_p = T \cdot R \cdot \vec{v}_f, \quad \vec{v}_f = \delta\vec{v} + \vec{v}_o$ where T, R are translation and rotation, $\vec{v}_o$ is a feature point with neutral expression of the front view. $\delta\vec{v}$ is the deformation applied on $\vec{v}_o$, which is also the approximation of APs. The animation data obtained is further adapted to apply on avatar model. Except the 22 feature control feature points, the 3D animation vector V of non-feature vertex i (denoted as V^i) can be derived by the following equation:

$$V^i = \sum_{j=1}^{N} (\omega(d_{i,j}) \cdot V_k^j) \qquad (1)$$

where $d_{i,j}$ is the distance between vertex i and vertex j, $d_{i,j}, j = 1,2,\ldots,N, (e.g.\ N = 3)$ have been arranged in increasing order. $\omega(d_{i,j})$ is a weight function which will have a large output of weight value for a small input of $d_{i,j}$.

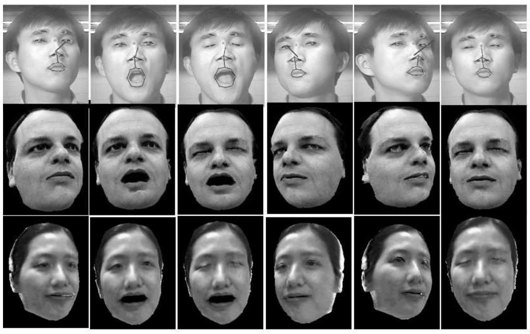

Figure 1: Tracked Video Generated Animation Sequences

5 Result

The active tracking and animation system works in a fully automatic fashion without any human intervention, and it works for any person. Experiment result in Figure 1 shows the good visual quality on created expressions with movements of eye, eyebrow and mouth, eye open/closure and head pose change. From the output of the tracking system, the 3D face model can be animated in the speed of 33 frames per second. The data rate for animating each frame is 64 bytes per frame. Given the animation speed from 15 frames/second to 30 frames/second, the data transmission rate ranges from 7.5K bits/s to 15K bits/s. Video clip demonstration can be found at: http://www.cs.binghamton.edu/˜xwei/avatar.avi. Our future work will further develop the face instantiation algorithm for making the model be individualized in real time.

Reference

Jun-yong Noh, Ulrich Neumann, *Expression Cloning,* SIGGRAPH01, pp277-288.

D. Terzopoulosa and K. Waters, ”Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models.” *IEEE Trans. PAMI*, 15(6), 1993

C. Tomasi and T. Kanade, ”Detection and Tracking of point features”. *Carnegie Mellon University Technical Report*, (CMU-CS-91-132), April 1991

[1]This material is based upon work supported in part by the National Science Foundation under grant No. IIS-0414029 and the AFOSR. Corresponding author: lijun@cs.binghamton.edu