Cloning Your Own Face With A Desktop Camera

Zhengyou Zhang, Zicheng Liu, Denis Adler, Michael Cohen, Erik Hanson, Ying Shan
Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
E-mail:zhang@microsoft.com http://research.microsoft.com/~zhang/

Abstract: We have developed an easy and cost-effective system that constructs textured 3D animated face models from videos with minimal user interaction. Our system first takes, with an ordinary video camera, images of a face of a person sitting in front of the camera turning the head from one side to the other. After five manual clicks on two images to tell the system where the eye corners, nose top and mouth corners are, the system automatically generates a realistic looking 3D human head model and the constructed model can be animated immediately (different poses, facial expressions and talking). A user, with a PC and a video camera, can use our system to generate his/her face model in a few minutes. The face model can then be imported in his/her favorite game, and the user sees themselves and their friends take part in the game they are playing. We will demonstrate the system on a laptop computer *live* at the conference, and participants can try it to model their own faces.

Keywords: face modeling, facial animation, geometric modeling, 3D computer vision

Animated face models are essential to computer games, film making, online chat, virtual presence, video conferencing, etc. Generating realistic 3D human face models and facial animations has been a persistent challenge in computer graphics. So far, the most popular commercially available tools have utilized laser scanners. Not only are these scanners expensive, the data are usually quite noisy, requiring hand touchup and manual registration prior to animating the model. The goal of our system is to allow an untrained user with a PC and an ordinary camera to create and instantly animate his/her face model in no more than a few minutes.

The equipment of our system include a computer and a video camera. We assume the intrinsic camera parameters have been calibrated, a reasonable assumption given the simplicity of calibration procedures [3].

The first stage is data capture. The user simply turns his/her head from left all the way to the right, or vice versa. The user then selects an approximately frontal view, and the original video is split into two sequences. The selected frontal image and its successive image are called the *base images* in the sequel. The user then locates 5 markers in each of the two base images. The 5 markers correspond to the two inner eye corners, nose tip, and two mouth corners.

The next processing stage computes the face mesh geometry and the head pose with respect to the camera frame using the two base images and markers as input. This is done through the following steps [1]:

Preprocessing: compute automatically a mask image to locate the approximate area of head motion.

Feature matching and motion determination: a robust technique based on least-median-squares is used to match points of interest and simultaneously determine the head motion across images [4].

Reconstruction: Reconstruct matched points in 3D space.

Fitting: the parameters which define the face mesh geometry are estimated through fitting face mesh to 3D reconstructed points and also silhouettes.

The third stage determines the head motions in the video sequences, based on the same matching technique mentioned above [4]. Both the head motions and the face geometry can be further refined using a recently developed technique called *model-based bundle-adjustment* [2].

The final stage blends all the images to generate a facial texture map [1]. This is now possible because the face regions are now registered by the head motions estimated in the previous stage.

The accompanying 5-minutes video, available at `ftp://ftp.research.microsoft.com/Users/zhang/FaceModeling.mpg`, shows an older version of the system. The reader is referred to [1, 2] for technical details.

References

[1] Z. Liu, Z. Zhang, C. Jacobs, and M. Cohen. Rapid Modeling of Animated Faces From Video. In *Proc. The Third International Conference on Visual Computing* (Visual 2000), pages 58-67, Sep. 2000, Mexico City. Also available as MSR technical report from http://research.microsoft.com/~zhang/Papers/TR00-11.pdf.

[2] Y. Shan, Z. Liu, and Z. Zhang. Model-Based Bundle Adjustment with Application to Face Modeling. In *Proc. International Conference on Computer Vision* (ICCV'01), Vancouver, Canada, July 2001. (in this volume)

[3] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[4] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, Oct. 1995.