A Real-Time Face Tracking System based on Morphable 3D Model Fitting

Kazuhiro HIWADA
Toshiba Corporation[*]

Atsuto MAKI
Kyoto University[†]

Akiko NAKASHIMA
Toshiba Corporation[‡]

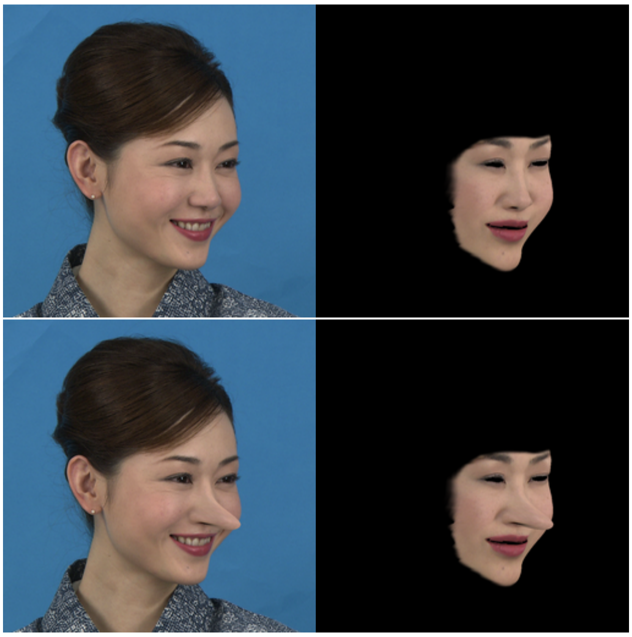

Figure 1: Mimicking. Clockwise from top left, an input frame in a video, a fitted CG mask, a geometrically morphed CG mask that has a longer nose, and a mimicked frame with the longer nose overlaid.

1. OVERVIEW

We demonstrate a real-time face tracking system, which works based on morphable 3D model fitting. The technical details are partly presented in [2]. Given a target user's individual morphable 3D model in the form of a combination of 3D linear bases, our system is capable of tracking the rigid and non-rigid motion of the target in live video. Furthermore, thanks to the tracking accuracy of our system, we also perform real-time mimicking, partial modification of the input image, such as shown in **Figure** 1.

For our purpose of real-time tracking, we take an analysis-by-synthesis approach that has also been taken in [3, 4] for different applications. That is, we first synthesize the image of a target face using the individual morphable 3D model of the target and then minimize the error between the synthesized image and the input image in each frame. The essence of our demo is that we introduce a coordinate-oriented error minimization [2] by adapting a linear solution [1] for estimating the tracking parameters. In order to measure the error and minimize it, we employ a coarse-to-fine optimization technique (See **Figure** 2). Only the motion parameters, **R** and **t**, are computed using the result of coarse matching whereas we also determine the morphing coefficients, **c**, using fine matching. Moreover, we demonstrate that the proposed framework alleviates the complication in treating the illumination variability on the object surface which may be caused for instance by the motion of the object itself.

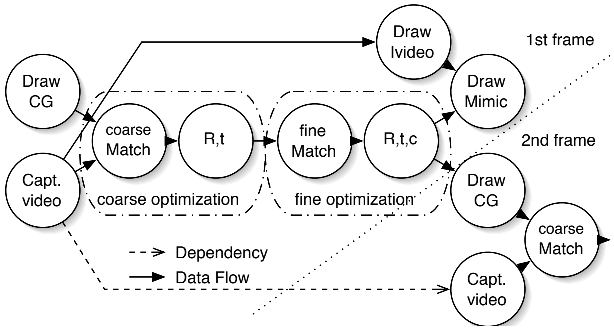

Figure 2: The *coarse-to-fine* fitting procedure. In order to speed up the fitting procedure, the optimization is executed in two steps, *coarse* and *fine*.

2. SPECIFICATION OF THE SYSTEM

Our demonstration system consists of a camera and a macintosh PC. The PC has 2 CPUs (PowerPC 1.25GHz) and a GeForce4 GPU. Capturing 640×480 images using Sony VX2000 DV-Camera, we cut them down to 512×480 gray-scale images for fitting. The fitting procedure is mapped into 2 posix threads (See **Figure** 3). Our system can execute the fitting procedure and display a result 15 times per second.

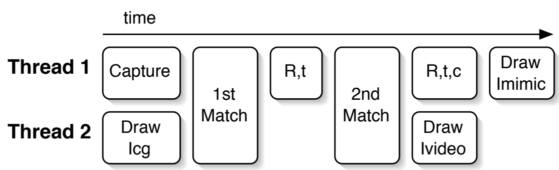

Figure 3: A thread mapping of real-time fitting procedure.

[*]Broadband System LSI Project, System LSI Division, Semiconductor Company. e-mail: kazuhiro.hiwada@toshiba.co.jp
[†]Academic Center for Computing and Media Studies. e-mail: maki@media.kyoto-u.ac.jp
[‡]Multimedia Laboratory, Corporate Research & Development Center. e-mail: akiko.nakashima@toshiba.co.jp

3. DEMO EXAMPLES

Figure 4 shows some results sampled in our real-time model fitting. Tracking is done through 300 frames. Not only the motion but also the flexion is properly obtained to the extent that the continuous operation keeps track of ordinary movement of pose and surface over time. **Figure** 5 also shows similar results but with presence of partial occlusion. It illustrates situations where the tracking is stable even when some portion of the face is occluded.

Figure 4: Real-time fitting result.

Figure 5: Real-time fitting under partial occlusion.

Figure 6 illustrates the resulting images of mimicking. In the figure, only the part of surface around the nose is rendered whereas the fitted structure of the target woman is geometrically morphed to mimic input video in order to have a longer nose. This is an example that shows such a partial manipulation of a 3D model is possible within the proposed framework.

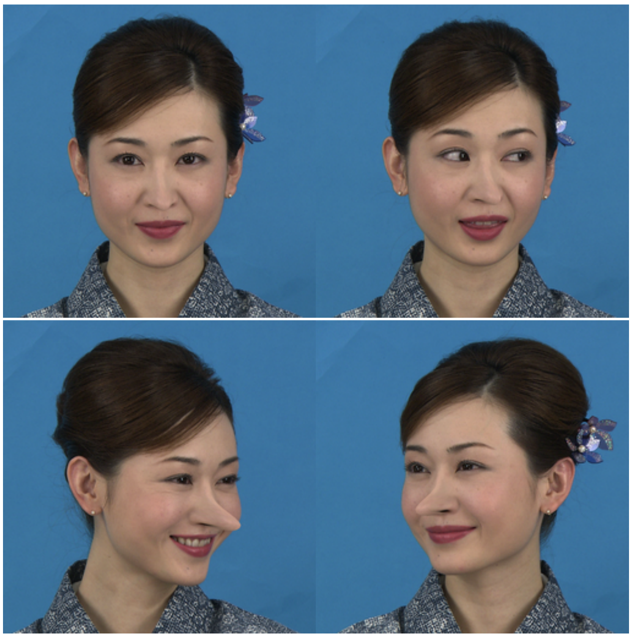

Figure 6: Making the nose longer. Illumination effect is taken into account.

4. SUMMARY

We realize a real-time tracking of a non-rigid human face with high accuracy, thanks to the key advances in:

- the framework, which is the combination of analysis-by-synthesis approach, use of a 3D morphable model which consists of a small but sufficient number of morph bases for expressing an individual facial structure, and use of the linear error minimization method,
- reducing computational time by the coarse-to-fine technique and parallel computation,
- the modular design that allows various techniques for estimating lighting environment in order to deal with the illumination variability on the target surface.

We further plan to demonstrate the robustness of our fitting procedure against partial occlusions and/or illumination variability.

5. REFERENCES

[1] M. Brand and R. Bhotika. Flexible flow for 3d nonrigid tracking and shape recovery. In *Proc. CVPR*, pages 315–322. IEEE, 2001.

[2] K. Hiwada, A. Maki, and A. Nakashima. Mimicking video : Real-time morphable 3d model fitting. In *VRST'03*. ACM, 2003.

[3] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salestin. Synthesizing realistic facial expressions from photographs. In *SIGGRAPH 98*, pages 75–84. ACM, 1998.

[4] S. Romdhami, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *ECCV'02*, 2002.