A MINIMALIST APPROACH TO FACIAL RECONSTRUCTION

M. GOTLA AND Z. HUANG
Department of Computer Science, School of Computing
National University of Singapore, Singapore 119260
E-mail: meherzad, huangzy@comp.nus.edu.sg

We propose a minimalist approach to 3D facial reconstruction. By applying cognitive and anatomical heuristics, we show that a realistic face model can be reconstructed from a generic face model using only two real facial images and minimal user interaction. There is no need for camera calibration. The method is very suitable to rapidly customize face models used as avatars in VR walk-through applications.

1 Introduction

Much work has already been done in the area of facial reconstruction [14, 17, 22, 10, 8, 11, 3, 12, 4, 6, *and* 16]. An excellent summary of the techniques used in facial reconstruction and animation can be found in [15]. However, most work emphasizes photo realistic quality that requires extensive input, a large number of sample points, and intensive user interaction.

Cognitive psychologists have demonstrated, through phenomena such as the differential inversion and the caricature advantage (Figure 2)[21], that humans appear to pay special attention to contours of facial features such as the eyes, lips, etc.(Figure 1). These are believed to form the dimensions of face space [7]. The dimensions of the face space are not formally known. "Common-looking" faces are closer to a generic face, which forms the face-space origin, while distinctive faces that are easily recognized are sparse points away from the origin. These two observed phenomena seem to corroborate this theory and provide for us a direction for our solution.

Work of cognitive psychologists inspires our solution:

1. As demonstrated by the differential inversion effect, humans process some parts of the face better than others. Obvious indicators like the eyes, mouth etc. should therefore be dealt with in greater detail.

2. The fact that humans can recognize people from just pencil sketches implies that complex texture information, while helpful, is not essential in recognizing faces. It appears that humans seem to pay special attention to the contours of facial features. Therefore, deforming the contours of a

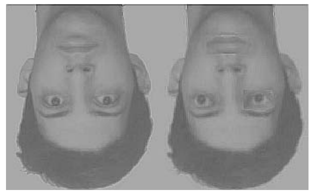

Figure 1. The eyes and mouth of the image on the right have been inverted. At a first glance, both images appear the same and everything seems normal. However, if the page is turned upside down, the grotesqueness of the image on the right becomes apparent immediately.

Figure 2. The caricatured images in the third row are more recognizable than the veridical images in the second row. Reproduced with permission of the publisher.

generic face model could be a good way to reconstruct the face.

3. In addition, the caricature advantage suggests that we could potentially increase the cognitive appeal of our output by exaggerating deformations from a generic face, possibly by a low to medium degree.

We propose a minimalist approach to 3D facial reconstruction. Our goal is to minimize the quantity and quality requirement of the input images and human interaction in the reconstruction process while still maintaining quality of result that is good enough for human perception. The main idea is to apply cognitive and anatomical heuristics to the algorithm. We focus only on those aspects of the face which are cognitively significant. In addition, our solution's feature extraction component takes advantage of anatomical expertise in the form of redundancies in the human face to help to trace the features of interest. This expertise is primarily in the form of rules that all human faces conform to and have been compiled in [15].

From our knowledge, all reconstruction frameworks using images need camera calibration. It is a very tedious process though significant progress has been made on techniques of self-calibration in computer vision society [18].

We can avoid the camera calibration process. We do not try to reconstruct face features precisely. Instead, we only want to get them approximately so that the resulting face is good enough for human perception. We found that we do not need the exact camera parameters. In our reconstruction solution, only two facial images from two special camera poses are used as input - a front and a profile view. Thus, we can estimate the camera position and orientation. From our experiments the reconstruction results are still acceptable, even when these two images are not taken exactly from the front and side views.

The most relevant work to ours is [12]. We follow the same trend towards automatic reconstruction. The major difference is that we introduce cognitive and anatomical heuristics into processes of feature extraction and deformation. By doing this, we can segment the extraction and deformation to multi-processes with each sub-process working on its own important feature of face model. The accuracy and efficiency of algorithm have been improved.

2 The Solution

Our solution consists of three steps - feature extraction, 3D deformation, and texture mapping (Figure 3). We apply cognitive and anatomical heuristics in feature extraction and deformation. Texture mapping is used to enhance the final results.

2.1 Feature Extraction

The function of this process is to obtain cognitively significant features represented by contours of the face space dimensions - e.g., the eyes, lips, etc. We do this by the technique of Hierarchical Deformable Templates (HDTs) which is an extension of the snake method introduced by [9]. An HDT represents hierarchically encoded anatomical expertise to help finding the cognitively significant features. Essentially, this involves taking a pre-defined face tem-

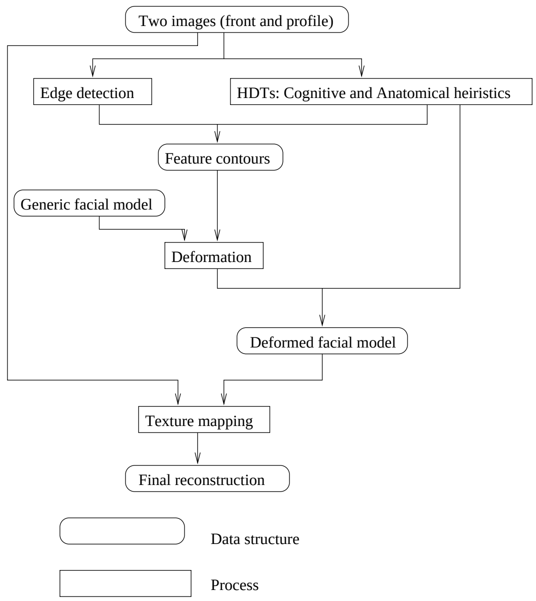

Figure 3. Function model of reconstruction process.

plate containing contours of a generic face model and allowing these contours to be deformed by facial images in a constrained fashion. These constraints are represented in a hierarchical manner and codify the anatomical expertise required to constrain the search space for individual contours.

An HDT consists of the following:

1. Child templates which are either templates or snakes (base case).
2. Approximate relative locations of these templates
3. Approximate relative size of these templates
4. Rules that its children obey.
5. References to the feature-rings of 3D generic face model.

A snake is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges. The name “snake”comes from the appearance of these splines

Figure 4. The illustration of snake behavior.

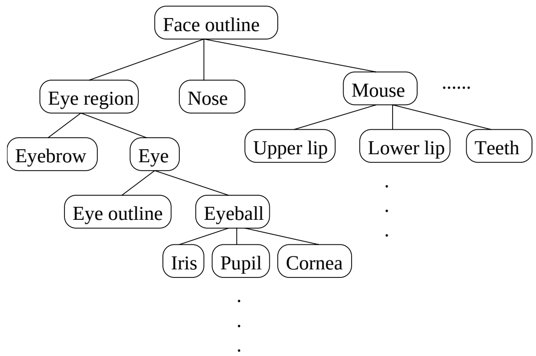

Figure 5. A top-down extraction of facial features.

to "slither" during the energy-minimization stage. The energy-minimization concept is borrowed from the classical physics. In mechanics, a ball released from a height will tend to lose energy and fall to a lower potential. In thermodynamics, a heated object will have a tendency to lose energy to its surroundings. These concepts transform readily to image processing when one views an image as a "terrain" of gradients with sharp edges acting as "cliffs". Edges can then simply be found by letting objects "fall off" these cliffs (Figure 4).

A top-down approach is used to extract the facial contours (Figure 5). First a face outline template finds the approximate location of the face. It then enforces approximate feature locations for the eyes, for instance, which are the child nodes of the face HDT. Finally the third level HDTs enforce their own set of constraints on its own children, e.g., the eye HDT enforces its set of constraints on the eye-ball (pupil, cornea, and iris) and eyelid.

For each image, we use a separate HDT to extract its contours in 2D. The results of this process are then used to deform the generic 3D model in next step.

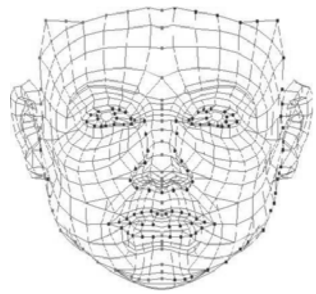

Figure 6. The front view feature-rings in a generic face model. The feature-rings are shown with highlighted points.

2.2 3D Deformation

A common technique in reconstructing the 3D structure of the face is to deform a generic 3D face instead of recreating the model from scratch (e.g. using Cyberware scanners [2]). A very general approach is Free Form Deformation (FFD) introduced by [20] that instead of deforming the object directly, the object is embedded in a space, i.e., FFD block, that is then deformed. The recent progress is the Dirichlet Free-Form Deformation (DFFD) that the FFD block is generalized by using the Voronoi diagrams [13].

Most prior work of facial reconstruction deforms the generic face using structural information derived from photogrammetric techniques. We introduce cognitive and anatomical heuristics to deformation in our approach. We pre-define a feature-ring of vertices for each facial feature on the generic face model corresponding to those extracted by HDTs in the extraction phase (see highlighted vertices in Figure 6). These features represent the most cognitively significant parts of the face. Each of these features is then deformed using the same snake-based technique to match the derived contours. The rest of the face is deformed by interpolation based on their locations relative to the feature vertices of the feature-ring.

We perform a two-stage deformation on the generic face, once for each image. We first take the front view and deform the features that reside on the XY plane with contours obtained from the front view image. After this, we deform the features that lie on the YZ plane using the contours obtained from the profile image.

Figure 7. The process to derive a texture map from two input images.

2.3 Texture Mapping

To provide additional realism, we texture-map the two original images onto the deformed model. We blend these two images to get a texture map (Figure 7). We use a very simple inverse-distance blending function - i.e. on the final texture map, the contribution of the front image decreases in the direction away from the center while that of the profile image increases correspondingly. After creating the texture map in this manner, we project the texture map to the 3D model by using a cylinder.

3 Sample Results

A sample result is shown on the next page. Input is shown in Figure 8: two real facial images and a generic 3D face model. In Figure 9, the reconstruction results of the top row are obtained using only one image - the front view of the face. The results in bottom row are obtained using two images - front and profile. The 3D features are better reconstructed using two images. But if only the front view image is available for a face, e.g., from a web page, ID or passport photo, the 3D-reconstruction result is still very distinguishable.

However, because we use a very simple blending function for texture mapping as described in Section 2.3, sometimes the derived texture map has discontinuity (see the texture map in the top right image of Figure 10). This results in artifacts in the reconstructed 3D face.

4 Implementation

Our prototype system is implemented on PC, Windows NT 4.0 using Microsoft Visual C++ 5.0, CosmoPlayer VRML plug-in, Microsoft Vision SDK 1.0, ImageMagick DLLs, and OpenGL.

We choose the (Virtual Reality Modeling Language) VRML 2.0 standard to represent our 3D objects because it is a de-facto standard. We use Mi-

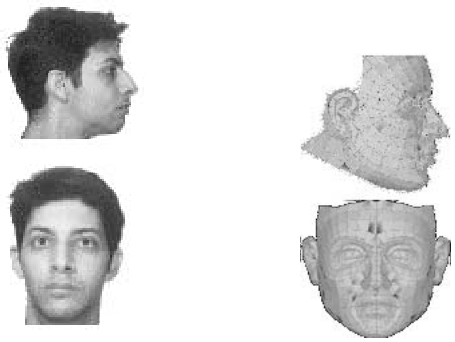

Figure 8. Two images of the first author's face (left) used as input and a generic face model (right).

Figure 9. The reconstruction results using one image (top) and two images (bottom) after texture mapping.

Figure 10. The reconstruction result using two images, Copyright 1995 University of Bern[5].

Figure 11. The reconstruction result using only one portrait of the second author.

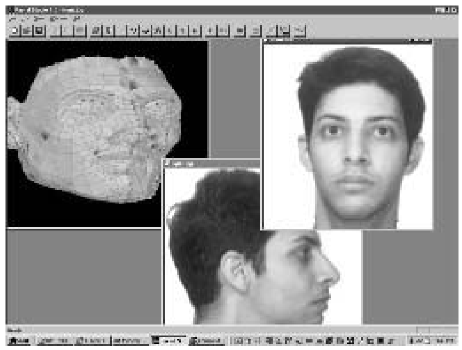

Figure 12. Some sample snapshots of the user interface.

crosoft's Vision SDK 1.0 together with image .dlls from ImageMagick. It allows us to perform low-level pixel operations in images.

The goal of our solution is to produce photo-realistic 3D face models without user interaction. The interface, in an ideal case, would therefore be a simple command line specifying details such as the generic face model and the images. However, in order to produce better results, we have an interface to allow limited manual intervention to aid the system's extraction process (Figure 12).

5 Conclusion

We have presented a minimalist approach to 3D facial reconstruction by introducing cognitive and anatomical heuristics into feature extraction and 3D deformation process. The results have shown this method can rapidly customize a generic face model using only one or two images of a real face.

6 Future Work

To achieve full automation of reconstruction is our final goal. At present, user interaction is required under some circumstances when we do not obtain optimal results from the automatic feature extraction process, e.g., noisy images. Therefore, we are trying to make this process more robust. At present, we only make use of edge-gradient information and anatomical heuristics. We are exploring methods of introducing a probabilistic framework over multiple criteria. It will encode more useful target information into HDTs.

At present, our texture-blending produces artifacts because of errors in our rough estimate (front and profile) of facial poses. We are in the process of improving this by applying anatomical heuristics into the blending function.

7 Acknowledgment

This work was supported partly by Academic Research Grant (RP3982704) of National University of Singapore.

References

1. Akimoto, T., Seunaga, Y., and Wallace, R. S., *Automatic Creation of 3D Facial Models.* IEEE Computer Graphics and Applications, Vol. 13, No. 5, pp. 16-22.
2. Cyberware Laboratory, *4020/RGB 3D Scanner with Color Digitizer.* Cyberware Laboratory, Inc, Monterey, California, 1990.
3. Essa, I., Basu, S., Darrell, T., and Pentland, A., *Tracking and Interactive Animation of Faces and Heads Using Input from Video.* Computer Animation '96, 1996, pp. 68-79.
4. Fua, P. and Miccio, C., *From Regular Images to Animated Heads: A Least Squares Approach.* ECCV Conference, Freiburg, Germany, May 1998.
5. *ftp://iamftp.unibe.ch/pub/Images/FaceImages/README.*
6. Guenter, B., Grimm, C., Wood, D., Malvar, H., and Pighin, F., *Making Faces.* In the proceedings of SIGGRAPH, 1998, pp. 55-66.
7. Johnston, R.A. and Ellis, H. D., *The development of face recognition.* In Cognitive and Computational Aspects of Face Recognition : Explorations in Face Space. Routledge, 1995.
8. Kalra, P. and Magnenat Thalmann, N., *Modeling of Vascular Expressions in Facial Animation.* Computer Animation '94, pp. 50-58.
9. Kass, M., Witkin, A., and Terzopoulos, D., *Snakes: Active Contour Models.* In the proceedings of the First International Conference on Computer

Vision, 1987, pp. 259-268.
10. Kurihara, T. and Arai, K. A., *Transformation Method for Modeling and Animation of Human Face from Photographs.* Computer Animation 91, pp. 45-58.
11. Lee, Y., Terzopoulos, D., and Waters, K., *Realistic Modeling for Facial Animation.* In the proceedings of SIGGRAPH, 1995, pp. 55-62.
12. Lee, W. S., Kalra, P., and Magnenat Thalmann, N., *Model Based Face Reconstruction For Animation.* In the proceedings of MMM'97, pp. 323-338.
13. Moccozet, L. and Magnenat Thalmann, N., *Dirichlet Free-Form Deformation and their Application to Hand Simulation.* Computer Animation'97, pp. 60-70.
14. Parke, F. I., *A Model for Human Faces that allows Speech Synchronized Animation.* Computer and Graphics, Pergamon Press, Vol. 1, No. 1, 1975. pp. 1-4.
15. Parke, F. I. and Waters, K., *Computer Facial Animation.* AK Peters, Wellesley, Massachussetts, 1996.
16. Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., and Salesin, D., *Synthesizing Facial Expressions from Photographs.* In the proceedings of SIGGRAPH, 1998, pp. 75-84.
17. Platt, S. and Badler, N., *Animating Facial Expressions.* In the proceedings of SIGGRAPH, 1991, pp. 245-252.
18. Pollefeys, M., Koch, R., and Van Gool, L., *Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters.* Proc. ICCV'98 (international Conference on Computer Vision), Bombay, 1998, pp. 90-95.
19. Prosmans, M. and Van Gool, L., *Getting Facial Features and Gestures in 3D.* In the Proceedings of the NATO Advanced Study Institute on Face Recognition : From Theory to Applications, Stirling, Scotland, UK, June 23-July 4, 1997.
20. Sederburg, T. W., *Free-Form Deformation of Solid Geometric Models.* In the proceedings of SIGGRAPH, 1986, pp. 151-160.
21. Stevenage, S. V., *Expertise and the Caricature Advantage.* In Cognitive and Computational Aspects of Face Recognition : Explorations in Face Space. Routledge, 1995, pp. 24-46.
22. Waters, K., *A Muscle Model for Animating Three-Dimensional Facial Expression.* In the proceedings of SIGGRAPH, 1987, pp. 17-24.