FACE MODELING FOR RECOGNITION

Rein-Lien Hsu and Anil K. Jain

Dept. of Computer Science & Engineering, Michigan State University, MI 48824
Email: {hsureinl, jain}@cse.msu.edu

ABSTRACT

3D Human face models have been widely used in applications such as facial animation, video compression/coding, augmented reality, head tracking, facial expression recognition, human action recognition, and face recognition. Modeling human faces provides a potential solution to identifying faces with variations in illumination, pose, and facial expression. We propose a method of modeling human faces based on a generic face model (a triangular mesh model) and individual facial measurements containing both shape and texture information. The modeling method adapts a generic face model to the given facial features, extracted from registered range and color images, in a global-to-local fashion. It iteratively moves the vertices of the mesh model to smoothen the non-feature areas, and uses the 2.5D active contours to refine feature boundaries. The resultant face model has been shown to be visually similar to the true face. Initial results show that the constructed model is quite useful for recognizing nonfrontal views.

1. INTRODUCTION

Current trend in face recognition is to use 3D face model explicitly. As an object-centered representation of human faces, 3D face models are used to overcome the large amount of variations present in human face images. These variations, which include extra-subject variations (individual appearance) and intra-subject variations (3D head pose movement, facial expression, lighting, and aging) are still the primary challenges in face recognition. However, the three major recognition algorithms [12] merely use viewer-centered representations of human faces: (i) a PCA-based algorithm; (ii) a LFA-based (local feature analysis) algorithm; and (iii) a dynamic-link-architecture-based paradigm.

Researchers in computer graphics have been interested in modeling human faces/heads for facial animation. We briefly review three major approaches to modeling human faces. DeCarlo et al. [5] use the anthropometric measurements to generate a general face model. This approach starts with manually-constructed B-spline surfaces and then applies surface fitting and constraint optimization to these surfaces. In the second approach, facial measurements are directly acquired from 3D digitizers or structured light range sensors. Water's [14] face model is a well-known instance. A morphable model [3] was constructed from a linear combination of eigenshapes and a linear combination of eigentextures, based on laser scans of 200 subjects. The third approach, in which models are reconstructed from photographs, only requires low-cost and passive input devices (video cameras). For instance, Chen and Medioni [4] build face models from a pair of stereo images. However, currently it is still difficult to extract sufficient information about the facial geometry only from 2D images. This difficulty is the reason why Guenter et al. [7] utilize a large number of fiducial points to capture 3D face geometry for photorealistic animation. Even though we can obtain dense 3D measurements from high-cost 3D digitizers, it still takes too much time to scan every face. Hence, advanced modeling methods which incorporate some prior knowledge of facial geometry are needed. Reinders et al. [13] use a fairly *coarse* wire-frame model, compared to Water's model, to do model adaptation for image coding. Lee et al. [10] modify a generic model from two orthogonal pictures (frontal and side views), or from range data for animation. Lengagne et al. [11] and Fua [6] fit a range animation model to uncalibrated videos using bundle-adjustment and least-squares fitting, given five manually-selected features points and initial camera positions. Zhang [15] deforms a generic mesh model to an individual's face based on two images, each of which contains five manually-picked markers.

We propose a face modeling method which adapts an existing generic face model (a priori knowledge of human face) to an individual's facial measurements. Our goal is to employ the learned 3D model to verify the presence of an individual in a face image database/video, based on the estimates of head pose and illumination.

2. FACE MODELING

We model an individual face starting with a generic face model, instead of extracting isosurfaces directly from facial measurements (range data or disparity maps), which are often noisy (e.g., near ears and nose) as well as time-consuming, and usually generates equal-size triangles. Our

0-7803-6725-1/01/$10.00 ©2001 IEEE

modeling process aligns the generic model using facial measurements in a global-to-local way so that feature points/ regions that are crucial for recognition are fitted to the individual's facial geometry.

2.1. Generic face model

We choose the Water's animation model [14], which contains 256 vertices and 441 facets for one half of the face. The use of triangular meshes is suitable for the free-form shapes like faces and the model captures most of the facial features that are needed for face recognition. Figure 1 shows the frontal and a side view of the model, and features such as eyes, nose, mouth, face border, and chin. There are openings at both the eyes and the mouth.

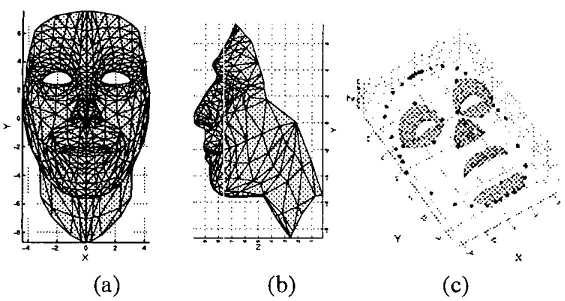


Fig. 1. 3D triangular-mesh model and its feature components: (a) frontal view; (b) side view; (c) feature components.

2.2. Facial measurements

Facial measurements include information about face shape and face texture. 3D shape information can be derived from a stereo pair combined with shape from shading, a sequence of frames in a video, or obtained directly from range data. The range database of human faces used here [1] was acquired using a Minolta Vivid 700 digitizer. It generates a registered 200 × 200 range map and a 400 × 400 color image. Figure 2 shows a range map and a color image of a frontal view, and the texture-mapped appearance. The locations of face and facial features such as eyes and mouth in the color texture image can be obtained by our face detection algorithm [8]. The corners of eyes, mouth, and nose can be easily obtained based on the locations of detected eyes and mouth. Figure 3 shows the detected feature points.

2.3. Model construction

Our face modeling process consists of global alignment and local adaptation. Global alignment brings the generic model and facial measurements into the same coordinate system. Based on the head pose and face size, the generic model is translated and scaled to fit the facial measurements. Figure 4 shows the global alignment results in two different modes. Local adaptation consists of *local alignment* and *local feature refinement*. Local alignment involves translating and scaling several model features, such as eyes, nose, mouth, and chin to fit the extracted facial features. Local

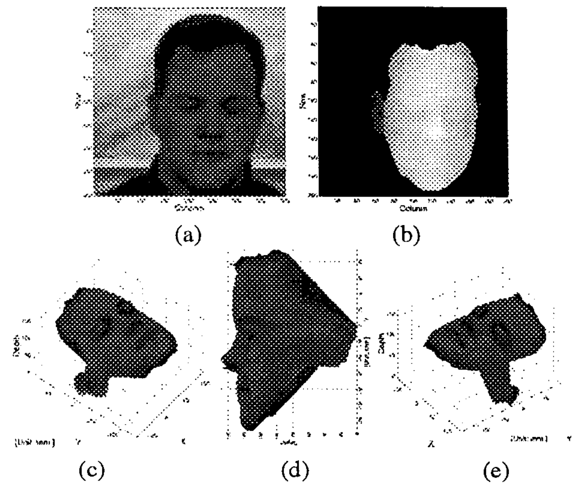


Fig. 2. Range data of a face: (a) color texture; (b) range map; and with texture mapping of (c) a left view; (d) a profile view; (e) a right view.

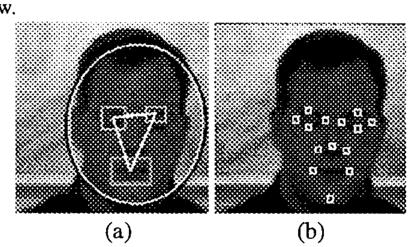


Fig. 3. Facial features overlaid on the color image: (a) obtained from face detection; (b) generated for face modeling.

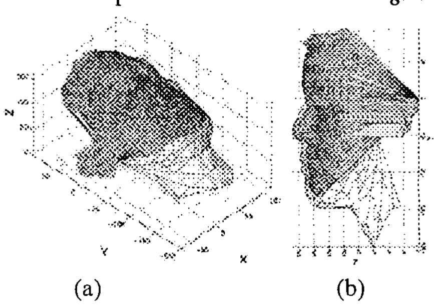


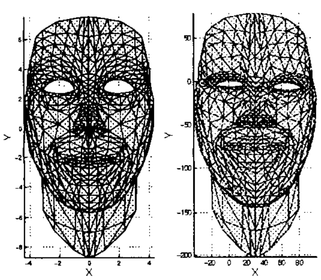
Fig. 4. Global alignment from the generic model (red) to the facial measurements (blue): the target mesh is plotted in (a) for a hidden line removal mode for a frontal view; (b) for a see-through mode for a profile view.

feature refinement makes use of *displacement propagation* and *2.5D active contours* to smoothen the face model and to refine local features. Both the alignment and the refinement of each feature (shown in Fig. 1(c)) is followed by displacement (of model vertices) propagation, in order to blend features in the face model.

Displacement propagation inside a triangular mesh mimics the transmission of message packets in computer networks. Let N_i be the number of vertices that are connected to a vertex i, J_i be the set of all the indices of vertices that are connected to a vertex i, w_i be the sum of weights from all vertices that are connected to vertex i, and d_{ij} be the Euclidean distance between a vertex V_i and a vertex V_j. ΔV_j is the displacement of vertex V_j, and α is a decay factor, which can be determined by the face size and the size of active facial feature in each coordinate. Eq.(1) computes the contribution of vertex V_j to the displacement of vertex V_i.

$$\Delta V_{ij} = \begin{cases} \Delta V_j \cdot \frac{w_i - d_{ij}}{w_i \cdot (N_i - 1)} \cdot e^{-\alpha d_{ij}}, \\ \quad N_i > 1, \ w_i = \sum_{j \in J_i} d_{ij} \\ \Delta V_j \cdot \ e^{-\alpha d_{ij}}, \\ \quad N_i = 1, \ j \in J_i \, . \end{cases} \quad (1)$$

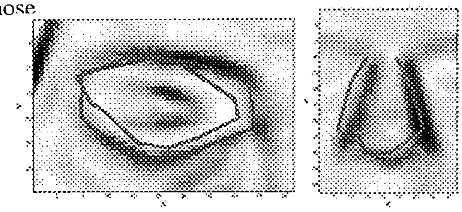
The total displacement ΔV_i of V_i can be obtained by summing up all the displacement contributed from its neighbor vertices. The displacement will decay during propagation and it continues for a few iterations, which is determined by the number of edge connections from the current feature to the nearest neighbor feature. In the future implementation, we will include symmetric property of a face and facial topology in computing this displacement. Figure 5 shows the results of local alignment for a frontal view.



(a) (b)

Fig. 5. Local feature alignment and displacement propagation shown for frontal views: (a) the generic model; (b) the model adapted to eyes, nose, mouth, and chin.

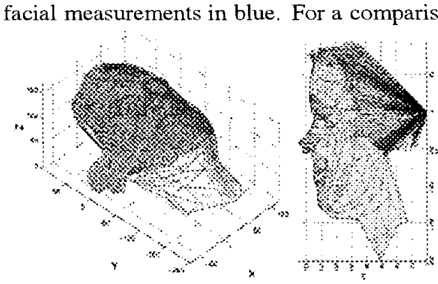
Local feature refinement follows local alignment to further adapt the results of alignment to an individual face by using 2.5D active contours (snakes). We modify Amini et al.'s [2] 2D snakes for our 3D active contours on boundaries of facial features. Hence, the crucial initial contours for fitting the snakes are known in our generic face model. Another important point for fitting snakes is to find appropriate external energy maps that contain local maximum/minimum at the boundaries of facial features. For the face and the nose, the external energy is computed by the maximum magnitude of vertical and horizontal gradients from range measurements. These two facial features have steeper borders than others. For features such as eyes and the mouth, the external energy is obtained by the product of the magnitude of the luminance gradient and the squared luminance. Figure 6 shows the results of local refinement for the left eye and nose.



(a) (b)

Fig. 6. Boundary alignment: initial (blue) and refined (red) contours overlaid on the energy maps for (a) left eye and (b) nose.

Although our displacement propagation smoothes non-feature skin regions in local adaptation, they can be further updated if a dense range map is available. Figure 7 shows the overlay of the final adapted face model in red and the target facial measurements in blue. For a comparison with



(a) (b)

Fig. 7. The adapted model (red) overlapping the target measurements (blue): the adapted model plotted (a) in 3D; (b) with colored facets at a profile view.

Fig. 2, Fig. 8 shows the texture-mapped face model. We further use a face recognition algorithm [9] to demonstrate the use of 3D model. The training database contains 504 image from 28 subjects and 15 images generated from our 3D face model, shown in Fig. 9. All the 10 test images were correctly matched.

3. CONCLUSIONS AND FUTURE WORK

Face modeling plays a crucial role in face recognition systems. We have proposed a method to adapt a generic face

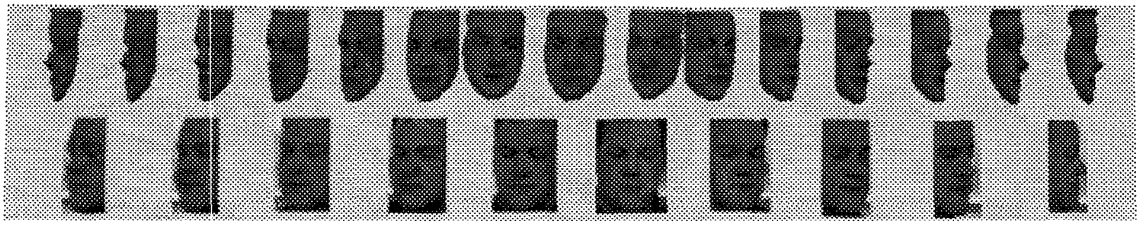


Fig. 9. Face matching: the first row shows the 15 training images generated from the 3D model; the second shows 10 test images captured from a CCD camera.

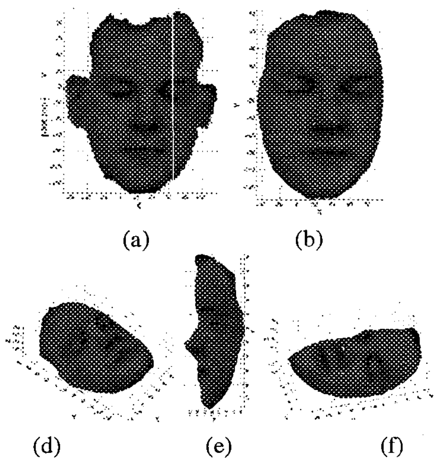


Fig. 8. The texture-mapped (a) input range image; adapted mesh model (b) from a frontal view; (d) from a left view; (e) from a profile view; (f) from a right view.

model to input facial features in a global-to-local fashion. The model adaptation first aligns the generic model globally, and then aligns and refines each facial feature locally using displacement (of model vertices) propagation and active contours associated with facial features. The final texture mapped model is visually similar to the original face. Initial matching experiments based on the 3D face model show encouraging results. Our goal is to extend this work to recognize faces in videos.

4. ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Patrick Flynn for providing the range datasets [1], and Dr. Wey-Shiuan Hwang for providing us his face recognition software [9].

5. REFERENCES

[1] Range databases: <http://sampl.eng.ohio-state.edu/~sampl /data/3DDB/RID/minolta/faceimages.0300/>

[2] A.A. Amini, T.E. Weymouth, and R.C. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Trans. PAMI*, vol. 12, pp. 855-867, Sept. 1990.

[3] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *SIGGRAPH Conf. Prof.*, pp. 187-194, 1999.

[4] Q. Chen and G. Medioni, "Building human hace models from two images," *IEEE 2nd Workshop Multimedia Signal Processing*, pp. 117-122, Dec. 1998.

[5] D. DeCarlo, D. Metaxas, and M. Stone, "An anthropometric face model using variational techniques," *SIGGRAPH Conf. Proc.*, pp. 67-74, Jul. 1998.

[6] P. Fua, "Using model-driven bundle-adjustment to model heads from raw video sequences," *Int'l Conf. Computer Vision*, pp. 46-53, Sept. 1999.

[7] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin, "Making faces," *SIGGRAPH Conf.*, pp. 55-66, Jul. 1998.

[8] R.-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," *Tech. Report MSU-CSE-01-7*, Michigan State Univ., Mar. 2001.

[9] W.-.S. Hwang and J. Weng, "Hierarchical discriminant regression," *IEEE Trans. PAMI*, vol. 22, pp. 1277-1293, Nov. 2000.

[10] W. Lee and N. Magnenat-Thalmann, "Fast head modeling for animation," *Image and Vision Computing (IVC)*, pp. 355-364, vol. 18, no. 4, Mar. 2000.

[11] R. Lengagne, P. Fua, and O. Monga, "3D stereo reconstruction of human faces driven by differential constraints," *IVC*, vol. 18, no.4, pp. 337-343, Mar. 2000.

[12] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *IVC*, vol. 16, no. 5, pp. 295-306, Mar. 1998.

[13] M.J.T. Reinders, P.J.L. van Beek, B. Sankur, and J.C.A. van der Lubbe, "Facial feature localization and adaptation of a generic face model for model-based coding," *Signal Processing: Image Comm.*, vol. 7, no. 1, pp. 57-74, Mar. 1995.

[14] F.I. Parke and K. Waters, "Appendix 1: Three-dimensional muscle model facial animation," *Computer Facial Animation*, A.K. Peters, Sept. 1996.

[15] Z. Zhang, "Image-based modeling of objects and human faces," *Proc. SPIE*, vol. 4309, Jan. 2001.