Generating 3D Views of Facial Expressions From Frontal Face Video Based on Topographic Analysis

Lijun Yin* Kenny Weiss
Department of Computer Science
State University of New York at Binghamton
Binghamton, NY 13902

ABSTRACT

In this paper, we report our newly developed 3D face modeling system with arbitrary expressions in a high level of detail using the topographic analysis and mesh instantiation process. Given a sequence of images of facial expressions at frontal views, we automatically generate 3D expressions at arbitrary views. Our face modeling system consists of two major components: facial surface representation using topographic analysis and generic model individualization based on labeled surface features and surface curvatures. The realism of the generated individual model is demonstrated through 3D views of facial expressions in videos. This work targets the accurate modeling of face and face expression for human computer interaction and 3D face recognition.

Categories and Subject Descriptors: I.4.7 [Image Processing and Computer Vision] Feature Measurement - Feature Representation; I.5.1 [Pattern Recognition]Applications; I.3.7 [Computer Graphics]Three-Dimensional Graphics and Realism

General Terms: Algorithms

Keywords:facial expression; face modeling; feature analysis

1. INTRODUCTION

Accurate 3D face representation and modeling could help improve the recognition of 3D face and 3D facial expressions. This research attempts to model the human face based on a frontal view image in a high level of accuracy. We developed a novel face modeling system using an explicit face surface representation, the so-called topographic representation, and a generic model individualization process. Most existing work for 3D face modeling utilized non-expressive (i.e., neutral) faces as a source to construct the 3D face surface. Model-based coding tracks facial expressions using 2D facial motion parameters [6], which are inherently incapable of generating accurate 3D expressions in arbitrary views. Several successful approaches have been developed relying on either morphable model database [1] or multiple photographs [5]. The recent work reported in [3] used extended Active Appearance Model [2] to synthesize 3D virtual views of a facial expression for recognition, given a near frontal view video input. However, the virtual view 3D model is a 3D avatar other than the subject shown in the video. In other words, the 3D view of facial expressions is not individualized. In order to create an accurate 3D expression model, we not only track the face motion and generate the animation in its original view, but also create a 3D arbitrary view with the model instantiation through the video sequence. In this paper, we give a new representation of a face by a so-called "topographic map". Tracing the behavior of some features (e.g., eyes, nose, mouth, chin, and wrinkles) across expression sequences can reveal precious information about the nature of the underlying physical process. The investigation of the relationship between the surface and the topographic labels will make the face representation establish on an intuitive higher level. Our face modeling system consists of two major components: facial surface representation using topographic analysis and generic model individualization based on the labeled surface features and the surface curvatures. The system composition is illustrated in Figure 1, which will be described in following sections.

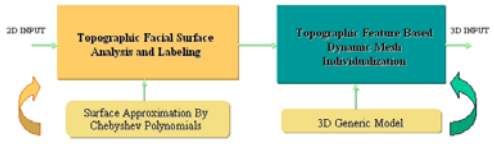


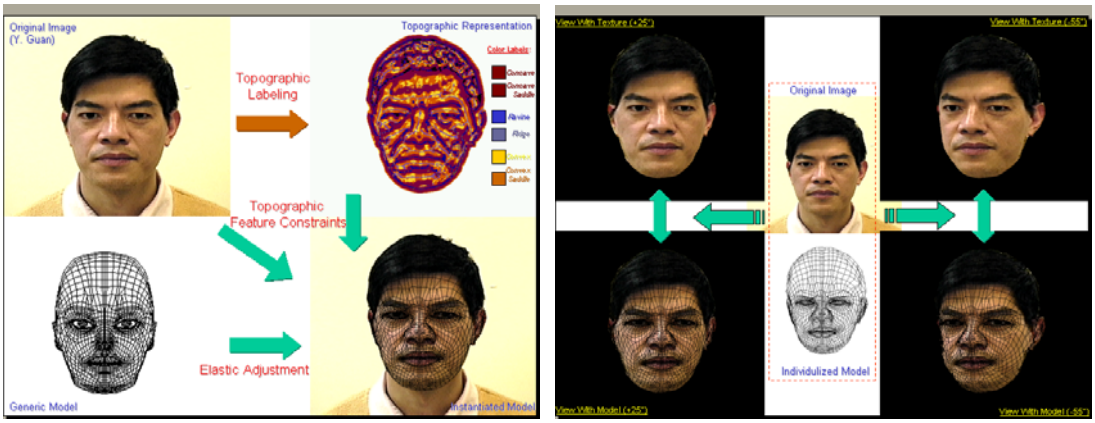
Figure 1: Composition of face modeling system

2. TOPOGRAPHIC REPRESENTATION

In order to deform a generic model to a facial surface, we need to make an explicit representation for the facial structure. Light intensity variations on an image are caused by an object's surface orientation and its reflectance. In visual perception, exactly the same visual interpretation and understanding of a pictured scene occurs no matter how the imaging condition is or whether the image is enhanced or not. The only difference is that the enhanced image has more contrast, is nicer to look at, and is understood more quickly by the human visual system. This fact suggests that edge-based detection methods cannot be expected to have the robustness associated with human visual perception because they are inherently incapable of invariance under monotonic transformations. However, the topographic categories peak,

*lijun@cs.binghamton.edu. Acknowledgment: This material is based upon the work supported by the National Science Foundation under grant No. IIS-0414029.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'04, October 10-16, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-893-8/04/0010 ...$5.00.



(a) (b)

Figure 2: Face modeling at work: (a) Topographic map creation and model adaptation; (b) Created 3D model in different views.

pit, ridge, valley, saddle, flat, and hill do have the required invariance, they can reveal the three-dimensional intrinsic surface of the object [4]. For this reason, we believe the topographic labeling is promising for our purpose. Topographic analysis treats the face image as a topographic terrain surface. Each pixel is assigned one of the topographic label peak, ridge, saddle, hill, flat, ravine, or pit Hill-labeled pixels can be further specified as one of the labels convex hill, concave hill, saddle hill or slope hill. We furthermore distinguish saddle hills as *concave saddle hill, convex saddle hill*, distinguish saddle as *ridge saddle* and *ravine saddle*. The different composition of these basic primitives will give a fundamental representation of different skin surface details. The primitive label classification approach is determined by the estimation of the first-order and second-order directional derivatives. The gradient vector is $\nabla f = (\partial f/\partial x, \partial f/\partial y)$, the second directional derivatives are calculated to form the *Hessian matrix* [7]. The eigenvalues (λ_1 and λ_2) of the Hessian are the values of the extrema of the second directional derivative, and their associated eigenvectors (ω_1 and ω_2) are the directions in which the second directional derivatives have greatest magnitude, their directions are orthogonal to each other.

$$H\omega_1 = \lambda_1\omega_1, H\omega_2 = \lambda_2\omega_2 \qquad (1)$$

The feature labeling is based on the values of λ_1, λ_2 and ∇f. For example, a pixel is labeled as a peak if the values in this pixel satisfied the following condition: $\lambda_1 < 0$, $\lambda_2 < 0$ and $\nabla f = 0$. Although we classified twelve categories of topographic labels, six of them are sufficient to describe the characteristics of a face. Figure 3 shows the six primitive features and their classification rule in our face modeling system.

There is a direct relation between λ_1 and λ_2 and the curvature K_i in the direction ω_1 and ω_2:

$$K_i = \frac{1}{\sqrt{1+\nabla f \cdot \nabla f}} \cdot \frac{\lambda_i}{1+(\nabla f \cdot \omega_i)^2}, i = 1, 2. \qquad (2)$$

Since a temporal skin "wave" is associated with the movement of the expression, the skin surface with a certain expression at a different time will have different shape, resulting in the different label changes. The curvature of the skin shape follows an expression's change while the expression occurs from the initiation stage to the release stage, for

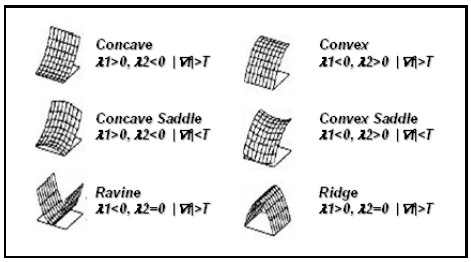


Figure 3: Classification rule for six typical topographic features

example, from a convex hill to a convex saddle hill. This dynamic labeling procedure produces a topographic label "map", which represents the skin surface and is changed along with the skin tissue movement. Because the curvature of a temporal skin shape reflects the relative depth information, it can be used as the external force to drive the dynamic mesh to form the skin details on the model (which will be described in the next section).

Figure 2(a) shows one example of the topographic map which is extracted from the face image "Guan". As it illustrated, the hill features locate in the facial skin surface, the ridge and ravine features lie in the facial organ areas and wrinkle areas.

3. 3D MODEL INDIVIDUALIZATION

The model instantiation process is decomposed into a global part and a local part. The global adaptation should estimate the scale, position and orientation of the observed face and fit the model accordingly by an affine transformation (see details in [8] for our previous work). The local adaptation scheme intends to deform the face model into the non-rigid face area.

According to the principle of minimum potential energy, of all possible kinematically admissible displacement configurations that an elastic body can take up, the configuration which satisfies equilibrium makes the total potential energy assume a minimum value [6]. This implies that for meshes associated with the image observations, if we let nodes move by successive steps based on the principle of minimum potential energy, and reduce strain energy at each step, finally

we should obtain a fine adaptation on this image. This motivates us to develop an energy oriented mesh to minimize the energy in each adaptation step. Extending our previous work by using the energy-oriented mesh method [8], we apply this dynamic mesh onto the facial areas, which have been labeled by the topographic features, with not only the 2-D external force (*e.g.* topographic gradient) but also the depth (*e.g.* topographic curvature) for model deformation.

The nodes of a dynamic mesh are mobile observers or sampling sites and they distribute themselves over the image data so as to represent the face with sufficient accuracy. Clearly, it is beneficial to concentrate on those nodes of a facial mesh where they will do the most good — in highly curved areas of the facial surface, especially in articulate regions around the nose, eyes, mouth and the wrinkles around these areas. We take the model as a dynamic structure, in which the elastic meshes are constructed from nodes connected by springs. The external forces of the nodes are used to link the dynamic mesh to the observed face image data. The motion for the dynamic node system is formulated to a second-order differential equation [6].

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} + \gamma_i \frac{d\mathbf{x}_i}{dt} + \mathbf{g}_i = \mathbf{f}_i; i = 1, \ldots, N \qquad (3)$$

where $\mathbf{x}_i$ is the position of node i, m_i is a point mass of node i, γ_i is the damping coefficient dissipating kinetic energy in the mesh through friction, $\mathbf{f}_i$ is the external force acting on node i, $\mathbf{g}_i$ is the internal force on node i due to the springs connected to neighboring nodes j. $\mathbf{g}_i$ is determined by the topographic gradient (I_i); $\mathbf{f}_i$ is determined by the topographic curvature K_i and the type of topographic labels $(L_i = 1, ..., 6)$. $\mathbf{d}_j$ is a 3D vector of the spring displacement.

$$\mathbf{g}_i = \sum_{j \in M_i} (k_2 I_i + k_1)\mathbf{d}_j, \qquad \mathbf{f}_i = \sum_{j \in M_i} (L_i \cdot K_i)\mathbf{d}_j \qquad (4)$$

Differing from the traditional method [6], we fit the generic model onto the topographic surface of a face with a 3D external force for simulating the skin "wave" deformation: (1) In the image plane, an external force is exerted by the gradient of the topographic surface; (2) In the direction that is perpendicular to the image plane, a vertical spring with one end attached to node i and the other end able to slide along the topographic surface. We apply image data (*e.g.*, surface curvatures) as external force which deflect (or pull) the mesh perpendicular to the image plane so that its 3-D shape becomes consistent with the face surface. In addition to external forces applied, the fitting process incorporates a feedback procedure which automatically adjusts spring parameters according to the topographic features that the spring covers.

We use six types of topographic labels: ridge, ravine, convex hill, convex saddle hill, concave hill and concave saddle hill. The ridge and ravine labels signify the key features location (e.g., eyebrows, eyes, nose, nose-bridge and mouth). Hill labels represent the facial surface features which connect ridge or ravine regions. It makes sense to increase the stiffness and the external force of springs in areas close to the ravine and ridge lines and boundaries of different regions. As a result, the mesh can distribute itself in both salient feature areas and facial surface "wave" areas. This dynamic modeling process creates a "tight" fitting model. Figure 2(b) shows the result of model adaptation constrained by six types of topographic regions of Figure 2(a).

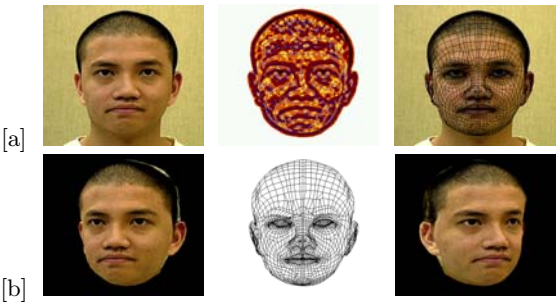


Figure 4: Face modeling example ("Alau") : Row: [a] from left: Original face; Topographic map; Adapted model; Row [b] from left: Different views of the created model with without texture.

4. EXPERIMENT

The accurate representation of 3D face is demonstrated through video images. 840 face images in 7 video sequences have been modeled by our system. Figure 4 shows one example of the modeling process. The face model contains 3708 vertices, and 4117 meshes. Figure 5 shows a sample video to demonstrate the realism of the 3D face creation with multiple view expressions. The running time for each frame is 1.07 seconds on Pentium-IV 3GHz PC. Note that our current implementation on topographic labeling requires that each face image has uniform background so that we can easily remove the background without labeling. The sihluette of face region is extracted by using the existing active tracking algorithm using multiple-frame fusion method [8]. The face contour shows a consistent concave or concave saddle region, therefore the head region can be reliably distinguished from other regions such as neck and clothes.

5. CONCLUSION

Topographic labeling provides us with an efficient face surface representation. The limitation of this work is that the frontal face image is required since the "depth" information is extracted based on the face structure of the front view. Our future work will conduct an improvement on the robustness to the lighting condition and image noise in low resolution, and address the computation issue for a real time application of human computer interaction and face expression recognition.

6. REFERENCES

[1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In SIGGRAPH99, p187-194.

[2] T. F. Cootes and C. Taylor. Statistical models of appearance for computer vision. In Technical Report, University of Manchester, Manchester, UK, 2001.

[3] L. Zalewski, et al. Synthesis and recognition of facial expressions in virtual 3d views. IEEE 6th Inter. Conf. on Automatic Face and Gesture Recognition, 2004.

[4] R. Haralick and et al. The topographic primal sketch. The Int. J of Robotics Research, 2(2):50–72, 1983.

[5] F. Pighin and *et al.* Synthesizing realistic facial expressions from photographs. SIGGRAPH'98, Orlando,FL., 1998.

[6] D. Terzopoulos. Analysis & synthesis of facial image sequences using physical & anatomical models. IEEE Trans. PAMI, 1993.

[7] O. Trier and A. Jain, et. al. Recognition of digits in hydrographic maps: binary vs topographic analysis. IEEE Trans. PAMI, 19(4), 1997.

[8] L. Yin, et al. Generating realistic facial expressions with wrinkles for model based coding. CVIU, 84(11):201-240, 2001.

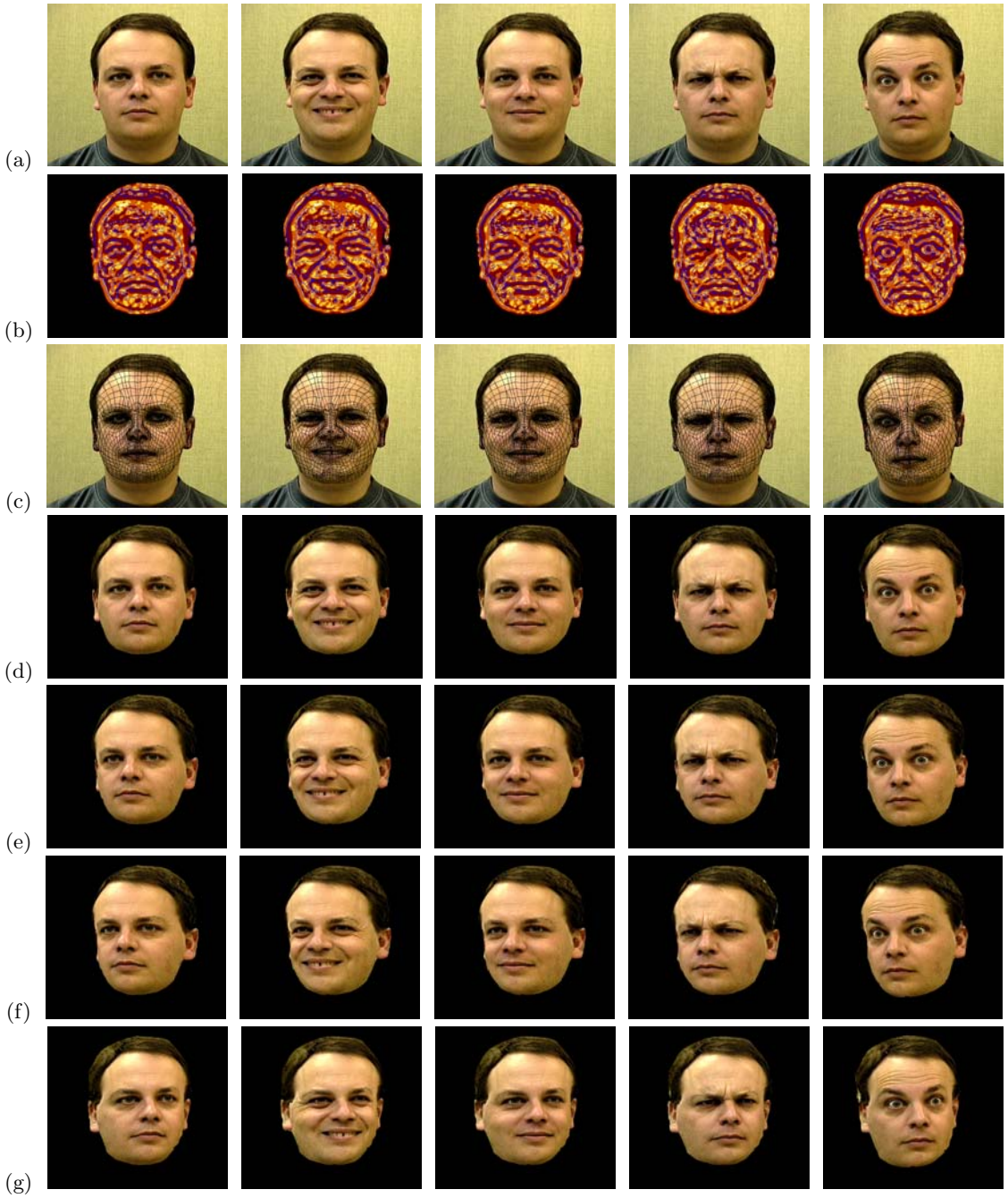


Figure 5: Creation of 3D face expressions in multiple views (Video Frame 1, 30, 70, 92, 127.) (a) original video in frontal view; (b) Topographic labels on face regions (six categories of primitive labels); (c) Model instantiation onto the front view face sequence; (d) Individualized 3D model and with texture mapped in front view; (e)-(f): left view of the 3D expression model sequence; (g): right view of the 3D expression model sequence.