Estimation of Rigid and Non-Rigid Facial Motion Using Anatomical Face Model

Alper Yilmaz Khurram Shafique Mubarak Shah
Department of Computer Science
University of Central Florida, Orlando FL-32816, USA
{yilmaz, khurram, shah}@cs.ucf.edu

Abstract

We present a model-based approach to recover the rigid and non-rigid facial motion parameters in video sequences. Our face model is based on anatomically motivated muscle actuator controls to model the articulated non-rigid motion of a human face. The model is capable of generating a variety of facial expressions by using a small number of muscle actuator controls. We estimate rigid and non-rigid parameters in two steps. First, we use a multi-resolution scheme to recover the global 3D rotation and translation by linear least square minimization. Then, we estimate the muscle actuator controls using the Levenberg-Marquardt minimization technique applied to a function, which is constrained by both optical flow and the dynamics of the deformable model. We present the results of our system on both real and synthetic images.

1. Introduction

A realistic analysis of 3D facial motion requires the recovery of both 3D global motion vectors and local motion parameters. One of the fundamental problems in this regard is the description of local motion parameters. Most of the current systems designed to solve this problem use "Facial Action Coding System", FACS [10] for describing non-rigid facial motions. Despite its wide use, FACS has the drawback of lacking the expressive power to describe different variations of possible facial expressions [11].

In this paper, we propose a system that can capture both rigid and non-rigid motions of a face. Our approach uses a realistic parameterized muscle model proposed in [1], which overcomes the limitations of the FACS and provides realistic generation of facial expressions as compared to the other physical models. The muscle model is motivated by the non-rigid motion of face, the physics of facial muscles and the skin. We use the face model described in [2], which is composed of 850 polygons and 18 synthetic muscles. First we conform the generic face model onto the face using deformations based on anthropometrical statistics [3][4]. Then we recover the motion parameters in two steps: (1) we use a multi-resolution (pyramid) scheme to recover global 3D rotation and translations; (2) we estimate the contraction of muscles, which are constrained by the optical flow equation, by using the Levenberg Marquardt method for three disjoint regions. These regions are defined such that the muscles in one region do not affect other regions.

The organization of the rest of the paper is as follows. In next section, a brief review of previous research is presented. We describe our model in section 3. The detailed analysis of global and local motion estimation is presented in sections 4 and 5 respectively. The implementation details are given in section 6. Finally, we will conclude by demonstrating experimental results.

2. Previous Research

Approaches for analyzing and synthesizing rigid and non-rigid human face motion differ by the choice of model selection and description of expressions.

Terzopoulos and Waters [1] used an anatomical model for describing the face. Their model encodes specialized knowledge about facial expressions, anatomical structure of the muscle and histology of biomechanics. Their algorithm proceeds by deforming the conformed face mesh for generating synthesized facial expressions using nine manually initialized snakes for frontal views. Use of active contour limits recovering the facial expressions in the presence of occlusions due to out of plane rotations.

DeCarlo and Metaxas [5] defined a partial 3D face model with parametric representation to facilitate the motion due to facial expressions. They have constrained their system by the optical flow equation, anthropometric statistics and the edge information for overcoming optical flow problems on boundaries. Use of parameters for generating facial expressions from partial face model limit ability to generate deformations other than predefined ones.

In [7], Koch described a parametric model for 3D shapes where the parameters are approximated by an analysis through synthesis approach. Similar to [5], the system updates the shape and rigid motion information iteratively by minimizing the reconstruction error in 2D using spatiotemporal derivatives. In contrast to [1] and [5], this approach isn't capable of estimating local motion.

Huang and Goldgof [8] proposed a generic adaptive model, where the number of nodes of the mesh is updated

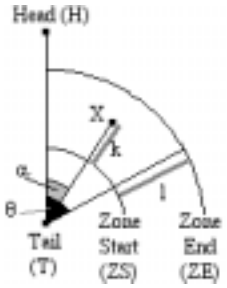

Figure 1: Muscle Model.

using external forces. Their model differs from previous models by its ability to adapt its structure according to objects' non-rigid motion by adjusting both the number of number of nodes and the individual vertex motion for minimizing the error.

Similar to [5], Li et al. described a rigid and non-rigid model motion using expression parameters assuming perspective projection and using the optical flow constraint equation [9]. In contrast to [5], non-rigid motion parameters are modeled using the affine motion model, which gives them more flexibility to generate different expressions. A synthesis feedback is used to reduce the error accumulated due to motion estimation in tracking.

Our approach is partly motivated by the research conducted by [1], [5] and [9]. In contrast to [1], while utilizing the muscles contraction parameters as our local deformation model, we are using the optical flow constraint similar to [5]. Our model differs from [5] in two ways. First, we are using muscle parameters to define local motion, which is more a natural way as compared to using predefined facial actions; second, we are following a two step methodology to calculate rigid and non-rigid motions which minimize the error due to higher dimension of the parameter space.

3. Anatomic Model

Local non-rigid motion of the face is caused by both the contractions of the muscles surrounding the face and the rotation of the jaw for opening the mouth. In [2], to simulate the behavior of muscle, Waters formalized structure of muscle as given by Figure 1. Every muscle is defined by its head (H), tail (T), zone start (ZS), zone end (ZE), and muscle zone (θ). Given a 3D vertex X, we can calculate the angle α, between the vectors |XT| and |HT|, and the distance of the point to the muscle zone start, *k*.

The mathematical representation of the structure shown in Figure 1 is formulated as

$$[X' \quad Y' \quad Z']^T = [X \quad Y \quad Z]^T + [a_i \quad b_i \quad c_i]^T \lambda_i \qquad (1)$$

where subscript i means i^{th} muscle, λ_i is the muscle contraction value and a_i, b_i and c_i are given by

$$a_i = (X - T_{ix})\Psi_i \qquad b_i = (Y - T_{iy})\Psi_i \qquad c_i = (Z - T_{iz})\Psi_i \qquad (2)$$

where

(a) (b)

Figure 2: (a) Wire frame model (left); muscles superimposed (middle); texture map (right), (b) Division of face into three regions.

$$\Psi_i = \left(1 - \frac{\cos\alpha_i}{\cos\theta_i}\right) \times \left(\cos\frac{k_i\pi}{2l_i} \times \delta(x - (|ZE_i| - |XT_i|))\right) \times \delta(x - (|\theta_i| - |\alpha_i|)) \qquad (3)$$

In equation 3, in contrast to original representation of Waters, we have used sigmoid function,

$$\delta(x) = (1 + e^{-x})^{-1} \qquad (4)$$

to approximate the step function with a continuous function for computing valid derivatives. Given a 3D vertex, the muscle contraction parameter λ_i is the only variable in the equation, i.e. equation 1 can be rewritten as

$$X' = X + f_i(\lambda_i) \qquad (5)$$

We model the effect of each muscle contraction on a single vertex by linear combination of flow vectors,

$$[X' \quad Y' \quad Z']^T = [X \quad Y \quad Z]^T + \sum_{i=1}^{k} \lambda_i [a_i \quad b_i \quad c_i]^T \qquad (6)$$

where *k* is the number of muscles.

For opening the jaw, instead of muscle actions, we rotate the jaw vertices of the face mesh by angle β around x-axis, $X'=R_\beta X$. The effect of jaw rotation on the jaw vertices, which were moved due to muscle contractions, is an additive factor, and it has no effect on other vertices. Following this observation, a preliminary division of the face will be the jaw region and the rest of face. The new location of the jaw vertex due to jaw rotation is,

$$X' = R_\beta X + \sum_{i=1}^{k} f_i(\lambda_i) \qquad (7)$$

For the other region, we simply set β=0, so that the rotation matrix is the identity matrix, which will result in equation 5. The model, muscles superimposed and texture map from the Claire sequence is shown in figure 2a. We only use the texture of head in our system because the texture of the neck is not always available due to clothing.

4. Global Motion Estimation

Object motion in 3D is defined in terms of rotational and translational velocities, $X' = RX + T$. Since frame-to-frame rotation is small, the rotation matrix, R, can be approximated using Euler angles. We apply perspective projection to project 3D space to a 2D image plane. It can easily be shown that the 2D optical flow is given by

$$u = f\left(\frac{V_1}{Z} + \Omega_2\right) - \frac{V_3}{Z}x - \Omega_3 y - \frac{\Omega_1}{f}xy + \frac{\Omega_2}{f}x^2 \qquad (8)$$

$$v = f\left(\frac{V_2}{Z} - \Omega_1\right) + \Omega_3 x - \frac{V_3}{Z}y + \frac{\Omega_2}{f}xy - \frac{\Omega_1}{f}y^2 \qquad (9)$$

where u and v are the flow vectors in the x and y directions respectively, Ω_1, Ω_2 and Ω_3 are the rotational velocities and V_1, V_2 and V_3 are the translational velocities. We use optical flow constraint $f_x u+f_y v+f_t=0$ where f_x and f_y are spatial derivatives and f_t is temporal derivative. We estimate the 3D rotational and translational velocities using a linear least squares fit by substituting equations 8 and 9 into optical flow constraint equation.

5. Local Motion Estimation

Once the global motion is compensated, we estimate the non-rigid deformations. We use equation 7 along with the optical flow constraint to derive the minimization scheme for recovering the non-rigid motion. Using the perspective projection, 2D flow vectors in image plane is

$$u = \dot{x} = \frac{f}{Z^2}\left(Z\dot{X} - X\dot{Z}\right) = \frac{f}{Z}\dot{X} - \frac{x}{Z}\dot{Z} \tag{10}$$

$$v = \dot{y} = \frac{f}{Z^2}\left(Z\dot{Y} - Y\dot{Z}\right) = \frac{f}{Z}\dot{Y} - \frac{y}{Z}\dot{Z} \tag{11}$$

Combining equation 7 with equation 10 will result in the 2D optical flow in the x direction and it is given by

$$u = \frac{1}{Z}\sum_{i=1}^{k} a_i'\lambda_i - \frac{xy}{f}\sin\beta - x\cos\beta + x \tag{12}$$

where $a_i'=fa_i-xc_i$ and f is focal length. Similarly, equations 7 and 11 will give 2D optical flow in the y direction,

$$v = \frac{1}{Z}\sum_{i=1}^{k} b_i'\lambda_i - f\sin\beta - \frac{y^2}{f}\sin\beta \tag{13}$$

where $b_i'=fb_i-yc_i$.

Substituting equations 12 and 13 to optical flow equation results in the error functional,

$$\sum e^2 = \sum\left[\frac{1}{Z}\sum_{i=1}^{k}(a_i' f_x + b_i' f_y)\lambda_i - \frac{xyf_x + f_y\left(f^2+y^2\right)}{f}\sin\beta + xf_x(1-\cos\beta) + f_t\right]^2 \tag{14}$$

We use Levenberg-Marquardt for finding the unknown muscle contraction parameters λ_i, and jaw rotation angle β that minimizes the nonlinear function of equation 14. The Jacobian is given by

$$\frac{\partial e}{\partial \lambda_i} = f_x\frac{1}{Z}\left(fa_i - xc_i\right) + f_y\frac{1}{Z}\left(fb_i - yc_i\right) \tag{15}$$

$$\frac{\partial e}{\partial \beta} = f_x\left[-\frac{xy}{f}\cos\beta + x\sin\beta\right] + f_y\left[-f\cos\beta - \frac{y^2}{f}\cos\beta\right] \tag{16}$$

In order to improve the estimation, and using the fact that effect of muscles is confined to small regions, we divide face into 3 regions as shown in Figure 2b, and perform Levenberg Marquardt separately in these regions.

6. Algorithm

Our algorithm executes in two steps: first step calculates the global motion and the second step calculates the local motion. Let R^k, T^k be the 3D rotation and translation for frame k. For k=0 R^0 is identity and T^0 is 0.

First step (Global motion estimation):

1. Create multi resolution representation (pyramid) of frame k and k+1.

For Each Level (From Coarse to Fine):

2. Estimate rigid motion parameters using least squares, by solving the equations from optical flow constraint.
3. Synthesize the transformed face and calculate error.
4. Repeat steps 2 and 3 until residual error is minimized

Second step (Local motion estimation): Transform the mesh using global motion parameters and pyramid representation for transformed face. At each level perform

1. Solve muscle unknowns using Levenberg-Marquardt.
2. Synthesize the transformed face, recalculate residual.
3. Repeat steps 1 and 2 until residual error is minimized

7. Results:

We applied our method to both real and synthetic images. Synthetic images were obtained by performing 3D rotations and translations on the texture map. For real images, we used the "Claire" sequence, which consists of 80 frames. The texture map, we used for all experiments is the same as shown in Figure 2a.

We have designed two sets of experiments for qualitative evaluation of our system. First set of experiments evaluates the performance of the system for estimation of global motion. Figure 3 shows the estimation accuracy of the synthetic and real images.

The second set of experiments deals with both global and local motion estimation. For each frame of the Claire sequence, we compute the global motion and local motion. For global motion, we use 2 levels of pyramids and 16 iterations per pyramid. Local motion is estimated by Levenberg-Marquardt method and iterations are terminated when the change of residual error per iteration becomes less than a threshold.

The convergence of global estimation is shown graphically in Figure 4, where the normalized residual error value is plotted against iterations for different frames. Some of the frames from the sequence along with the synthesized frames are shown in Figure 5. Figure 6 shows the normalized residual errors for Levenberg Marquardt iteration in one of the frames for three regions. Note that the error always decreases or remains constant. This is because L.M method does not update the parameters if error increases at some iteration.

8. Conclusions

A method is proposed for generating synthesized face images from a video sequence using the anatomic structure of the face. The approach is based on the estimation of global and local motion separately using different minimization schemes constrained by optical flow. Local motion is based on the muscle actuator

control values and jaw rotation, which is a natural approach compared to widely used FACS. The method is shown to produce reasonable results to obtain both global and local motion parameters.

References

[1] D. Terzopoulos, K. Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models," *IEEE Trans. on PAMI*: 13(6), 1993.

[2] K. Waters, "A muscle model for animating three-dimensional facial expression," *Computer Graphics*: 22(4), pp. 17-24,1987.

[3] L. Farkas. Anthropometry of the Head and Face, Raven Press, 1994.

[4] D. DeCarlo, D. Metaxas, M.Stone, " An Anthropometric Face Model using Variational Techniques," *Proc. SIGGRAPH*, pp. 67-74, 1998.

[5] D. DeCarlo, D. Metaxas, "Combining Information using Hard Constraints," *CVPR*, pp. 132-188, 1999.

[6] Y. Yacoob, L. Davis, "Computing Spatio-Temporal Representations of Human Faces," *CVPR*, pp. 70-75, 1994.

[7] R. Koch, "Dynamic 3D Scene Analysis through Synthesis Feedback Control," *IEEE Trans. PAMI*: 13(6), June 1993.

[8] W.C. Huang, D.B. Goldgof, "Adaptive-Size Meshes for Rigid and Non-Rigid Shape Analysis," *IEEE Trans. on PAMI*: 15(6), 1993.

[9] H. Li, P. Roivainen, R. Forchheimer, "3D Motion Estimation in Model-Based Facial Image Coding," *IEEE Trans. on PAMI*: 15(6), 1993.

[10] P. Ekman, W.V. Friesen, *Facial Action Coding System.* PaloAlto, Calif: Consulting Psychologists Press, Inc., 1978.

[11] A. Essa, A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," *IEEE Trans. on PAMI*: 19(7), 1997.

Figure 3: Global motion estimation for synthetic (top row) and real (bottom row) images. Images from left to right: initial, to be recovered, resynthesized.

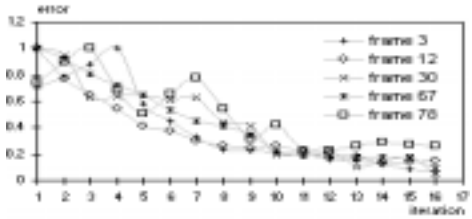

Figure 5: Residual Error in each iteration of Global motion estimation plotted for different frames.

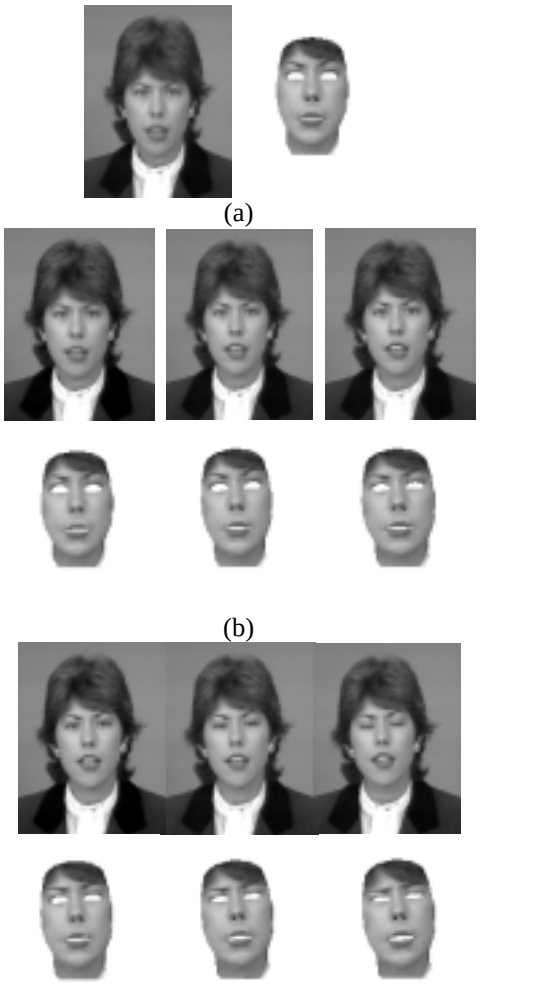

Figure 6: Claire sequence: (a) initial frame and texture map; (b) first row: frames 1, 2, 3; second row: resynthesized faces using recovered facial motion; (c) first row: frames 4, 5, 6; second row: resynthesized face using recovered facial motion.

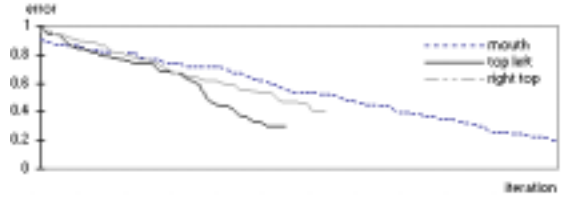

Figure 7: Convergence of Levenberg-Marquardt method for different regions of head.