A BIOMECHANICAL MODEL FOR IMAGE-BASED ESTIMATION OF 3D FACE DEFORMATIONS

Ilse Ravyse , Hichem Sahli

Vrije Universiteit Brussel, Department ETRO,
Audio Visual Signal Processing (AVSP)
Pleinlaan 2, 1050 Brussel
{icravyse, hsahli}@etro.vub.ac.be

ABSTRACT

To reproduce a face motion from an image sequence, natural motion parameters provide a semantical and efficient way of representation. State-of-the art techniques for 3D face motion estimation employ a limited set of predefined key-shapes of face structure, and thereby restrict the possible face motion which can cause distortions. We propose a new approach in which such distortions are avoided by augmenting a 3D structural surface face model with a physical motion model originating from continuum mechanics. Implementation with a displacement-based FEM does not only describe, but also explain the motion of the face's skin tissue caused by the muscle force parameterized actuations. The correct usage of the motion model and the mapping of the 3D scene flow to the 2D optical flow, allow posing the 3D deformation estimation as an inverse problem, for which a solution has been obtained using numerical solvers.

Index Terms— Image motion analysis, Finite element methods, Modeling, Biomechanics, Multimedia communication

1. INTRODUCTION

The communication of non-verbal face gestures is used in a wide range of applications as tele-presence and surveillance, and as an element in the upcoming new media in games and intuitive user-interfaces with virtual actors. This paper addresses analysis techniques which allow to replace the raw video recording of a person by a high level semantical representation (to transmit and to store efficiently) and to animate a face model according to the face appearance.

Facial expressions are 3-dimensional as they are produced by 3D deformation of the skin, the face shape (due to articulation) as well as the head movements (3D rigid motion of the head). Extracting these 3D deformation parameters from 2D image-sequences is an ill-posed problem, due to the perspective projection. A-priori knowledge of the face surface must be used to solve the problem [1]. The estimation of the motion of a human face is made by constraining the projection of the motion of the 3D face (human head) model to the measured optical flow (apparent motion) [2]. For geometrical motion models (as Candide [2]) the deformations are predefined with disadvantage that the 'variation modes' which are dropped might prevent some expressions. Also the active appearance models [3] cannot be deformed to other variations than those they have been trained for. In spring-based muscle models [4, 5] an explicit setup of the muscle topology and the extend functions for muscle control do limit the nonrigid motion capabilities [6]. Given the usage of the continuum mechanical Finite Element Model (FEM) for biological objects in surgery simulation [7], an elastic thin shell FEM [8] can provide the natural motions of the skin of a human face. The estimation of the external forces by FEM to constrain motion is currently seen as state-of-the-art in computer vision and graphics applications to control nonrigid motions [9, 10].

In this paper we present an innovative solution to the nonrigid face deformation estimation problem, where the estimated natural nonrigid motion parameters will be those of the muscle actuations on the face applied by a FEM. The motion estimation problem and a two step solution is formulated in section 2. The results of solving the inverse non-rigid motion estimation problem with the proposed implementations are given in section 3. In Section 4 some conclusions are drawn.

2. 3D DEFORMATION ESTIMATION

The problem of estimating the non-rigid motion of the face can be formulated as: estimating the 3D flow, constrained by a physically-based face motion model, from a 2D image sequence. We argue that by coupling the flow estimation with the physical model of the skin and muscles, the 3D deformations of the face model will appear more natural.

The 3D velocity field that describes the motion of a deformable surface, imaged by a sensor, is called scene flow,

The research of this paper was partially sponsored by the *Instituut voor aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen* (IWT) as a specialization grant.

$\mathbf{W}$ and is related to the (2D) optical flow $\mathbf{u}$ as :

$$\mathbf{u} = \frac{d\mathbf{x}}{dt} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} \frac{d\mathbf{X}}{dt} \triangleq \mathbf{J_X W} \approx \mathbf{J_X}[\mathbf{Q} - \mathbf{X}] \quad (1)$$

where $\mathbf{X} = [x, y, z]^T$ the starting point of the 3D flow vector $\mathbf{W}$ and $\mathbf{Q}$ its ending point, and $\mathbf{x}$ the projection of $\mathbf{X}$ on the 2D image. The Jacobian of the scene flow projection, is $\mathbf{J_X} = \frac{f}{z}\begin{pmatrix} 1 & 0 & -\frac{x}{z} \\ 0 & -1 & \frac{y}{z} \end{pmatrix}$ where f is the focal length of the perspective imaging system.

Estimating scene flow from a 2D image sequence is a mathematical inherently ill-posed inverse problem [11]. Denote by $\widetilde{\mathbf{u}}(\mathbf{W})$ the *modeled optical flow*. It correspond to the forward model, being the projection of a 3D scene flow (of a known 3D motion model) on the image plane using Eq.(1). The 3D face motion is then estimated by comparing the modeled optical flow to the measured optical flow $\mathbf{u}$. We will thus select the parameters for which $\widetilde{\mathbf{u}}(\mathbf{W})$ can most likely resemble $\mathbf{u}$ in the least-squares sense following the same formulation of the SfM-problem of [12]:

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}}{\arg\min} \left[\left(\sum_{k=1}^{m} \|\mathbf{u}_k - \widetilde{\mathbf{u}}(\mathbf{W}_k)\|^2 \right) + \psi \right] \quad (2)$$

where $\widehat{\mathbf{W}}$ is the estimated 3D face deformation; m is the number of visible face model vertices used in the estimation; $\mathbf{u}_k(u_k, v_k)$ is the measured optical flow at image location $\mathbf{x}_k$; $\mathbf{W}_k$ is the face scene flow at the point $\mathbf{X}_k$ (corresponding to $\mathbf{x}_k$); $\widetilde{\mathbf{u}}$ is the predicted (modeled) optical flow corresponding to the scene flow $\mathbf{W}$. For solving Eq.(2) one needs a regularization term ψ and/or constraints on the motion model parameters. Here, we use the regularization on the image [1] by 'measuring' a regularized optical flow field [13]. The remaining ill-conditioning of the problem will be solved via regularization on the 3D surface level [1], imposed by face shape modeling, as it will be defined in sections 2.3.2 and 2.3.3.

2.1. Separability of Pose and Shape

The *a-priori* scene knowledge for 3D motion estimation from an image sequence, may be given under the form of 2 sets of information which are closely related to each other, namely the *a-priori* structure and the motion model knowledge. Forchheimer [14] presented this as part of the geometrical face motion model. For physics-based models, Terzopoulos and Witkin [15] suggested that the mechanical equations of the motion can be decomposed in a rigid and a nonrigid part. The decomposition of the 3D scene flow $\mathbf{W}$ into a rigid motion field $\mathbf{V}$ and a non-rigid motion field $\mathbf{U}$ is formulated as $\mathbf{W} = \mathbf{V} + \mathbf{U}$. As long as the changes in depth of the considered point is not too large, the optical flow $\mathbf{u}$ can also be decomposed:

$$\mathbf{u} \approx \mathbf{w} + \mathbf{v} \quad (3)$$

where $\mathbf{v}$ is the apparent motion that moves the structure rigidly to a virtual state, and $\mathbf{w}$ the motion that handles the deformation. Hence, the face motion estimation can be also decomposed into two steps, namely pose and shape estimation.

2.2. Pose estimation problem

We define the bulk motion field $\widetilde{\mathbf{v}}$ as the part of the 2D motion field $\mathbf{u}$ that absorbs all parts of the instantaneous 2D motion which follows a rigid motion model in the corresponding real 3D scene. The *modeled optical flow* is then defined by:

$$\widetilde{\mathbf{u}}(\mathbf{W}_k) \triangleq \widetilde{\mathbf{v}}_k(\boldsymbol{\omega}, \mathbf{t}) = \mathbf{J}_{\mathbf{X}_k}\left[\mathbf{Q}_k(\boldsymbol{\omega}, \mathbf{t}) - \mathbf{X}_k\right] \quad (4)$$
$$\text{with} \quad \mathbf{Q}_i(\boldsymbol{\omega}, \mathbf{t}) = \mathbf{X}_k + \boldsymbol{\omega} \times \mathbf{X}_k + \mathbf{t} \quad (5)$$

The objective function in Eq.(2) is linear in the 3D rigid motion parameters ($\boldsymbol{\omega}$ and $\mathbf{t}$). Therefore the estimation problem can be solved efficiently using the *entire* dense estimated optical flow field. Note that, (i) the parameter m in Eq.(2) corresponds to the image size, and (ii) no regularization is required ($\psi = 0$) since there are only 6 parameters to be estimated.

2.3. Shape estimation problem

The part of the flow that was not taken into account by the pose estimation, i.e. the part which was not absorbed by the estimated bulk flow $\widetilde{\mathbf{v}}$, is the nonrigid motion. As the nonrigid flow is not directly accessible from the image measurement, Eq.(3) provides it as $\mathbf{u} - \widetilde{\mathbf{v}}$. Accordingly, the Jacobian, in Eq.(1), for the non-rigid optical flow is evaluated at $\widehat{\mathbf{Q}}_k$, being the virtual point's position after the pose estimation. Finally, the *modeled optical flow*, for the non-rigid motion $\mathbf{U}$ parameterized by the force parameters $\bar{\mathbf{F}}$, is then defined by

$$\widetilde{\mathbf{u}}(\mathbf{W}_k) \triangleq \widetilde{\mathbf{w}}_k(\bar{\mathbf{F}}) = \mathbf{J}_{\widehat{\mathbf{Q}}_k}\left[\mathbf{U}(\bar{\mathbf{F}}, \widehat{\mathbf{Q}}_k) - \widehat{\mathbf{Q}}_k\right] - \widehat{\mathbf{v}}_k \quad (6)$$
$$\text{with} \quad \widehat{\mathbf{v}}_k = \mathbf{J}_{\mathbf{X}_k}\left[\widehat{\mathbf{Q}}_k - \mathbf{X}_k\right]$$
$$\text{and} \quad \widehat{\mathbf{Q}}_k = \mathbf{X}_k + \widehat{\boldsymbol{\omega}} \times \mathbf{X}_k + \widehat{\mathbf{t}}$$

The nonrigid deformation estimation becomes a force estimation problem formulated as follows.

2.3.1. The FEM Face Model

A face model with homogeneous soft tissue skin material parameters and a 3D bended surface geometry [16] is used to obtain an accurate motion model of the skin deformation in the human face. We use a displacement-based FEM in which the face muscles are modeled as external constraint forces. The interaction of the physical muscles with the soft skin tissue indeed imposes a distributed stress on the skin layer that on its turn causes the face surface motion. For the numerical determination of the displacements, FEM approximates well the continuum equilibrium equation (between force and displacement) as a non-linear relationship at the wire-frame model's node points or vertices $\widehat{\mathbf{Q}}_k$:

$$\mathbf{K}(\widehat{\mathbf{Q}}, \bar{\mathbf{U}})\bar{\mathbf{U}} = \bar{\mathbf{F}} \Rightarrow \mathbf{U}_k = \mathbf{C}(\bar{\mathbf{F}}, \widehat{\mathbf{Q}}) \quad (7)$$

where $\mathbf{K}$ is a stiffness matrix dependent on the object's material and geometry, $\bar{\mathbf{U}}$ is the tissue displacement field of the face wire-frame vertices, $\bar{\mathbf{F}}$ is the distributed muscle force field of the vertices; The inversion of this nonlinear system is made by a nonlinear multivariate function $\mathbf{C}$ that relates tissue displacement field $\mathbf{U}_k$ of the vertex $\mathbf{X}_k$ to $\bar{\mathbf{F}}$. More details are to be found in [17]. This model parameterizes the tissue displacement and its projection by a 3D force field in Eq.(6).

2.3.2. Geometrical constraint

The 3D force field lies on a plane tangential to the model, motivated by the muscle anatomy of a face in which muscle fibres run approximately parallel to the skin. Equivalently, at the vertex point $\mathbf{X}_k$ the force $\mathbf{F}_k$ is perpendicular to the model normal $\mathbf{n}_k$:

$$\psi_1 = \alpha \sum_{k=1}^{m} \|\mathbf{F}_k \cdot \mathbf{n}_k\|^2 \tag{8}$$

where α is a penalty parameter to fine-tune the constraint.

2.3.3. Smoothness constraint

The 3D force field varies smoothly in space by imposing Tikhonov regularization.

$$\psi_2 = \beta \sum_{k=1}^{m} \left(\|\nabla f_k\|^2 + \|\nabla g_k\|^2 + \|\nabla h_k\|^2 \right) \tag{9}$$

where β is a penalty parameter to fine-tune the constraint; and the components of a force $\mathbf{F}$ at a vertex $\mathbf{X}_k$ are defined as (f_k, g_k, h_k) of which the variation in the spatial directions x,y and z are suppressed, as long as the other model terms allow this. The muscle anatomy indeed shows that every muscle has an extend range and does not operate solely on one face surface point. The surrounding tissue thus receives a weakened stress from a nearby muscle, giving rise to a smoothing of the muscle force field.

2.4. Numerical Implementation

For estimating the 3D facial deformations from image sequence, the following general algorithm is applied: (i) *Initialization.* Position the deformable mesh in the 3D world (by a least-squares estimation of the camera parameter and the model scaling) and adapts its natural features to those of the image by a scene calibration procedure [17, 18]. (ii) *Pose estimation.* Solve the linear least-squares problem, given by Eq.(2) and Eq.(4), using as inputs the dense optical flow, yield the rigid transform parameters $\boldsymbol{\omega}$ and $\mathbf{t}$. (iii) *Shape estimation.* Solve the non-linear least-squares problem, given by Eq.(2), Eq. (6) and Eq.(7), with the above defined geometrical and smoothness constraints. Note that, at each iteration of this least-squares problem the non-linear function $\mathbf{C}$, in Eq.(7), is evaluated by solving the mechanical displacement-based finite element problem with the current estimate of $\bar{\mathbf{F}}$ as input.

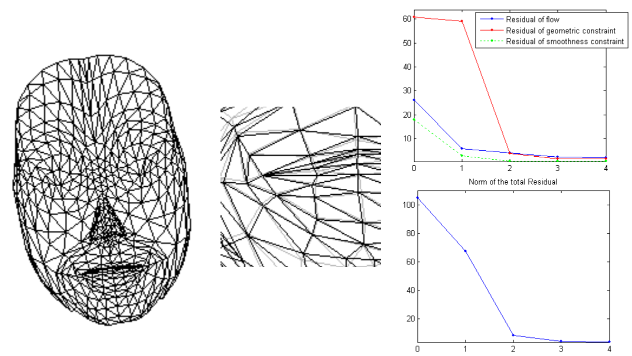
3. EXPERIMENTS

The proposed method can be validated by visualizing (rendering) the new position of the face model, estimated by applying the virtual bulk motion parameters and the force parameters in the full face motion model, using the following equation:

$$\widehat{\mathbf{X}}_k' = \widehat{\boldsymbol{\omega}} \times \mathbf{X}_k + \widehat{\mathbf{t}} + \mathbf{C}(\widehat{\mathbf{F}}, \widehat{\mathbf{Q}}) \qquad k = 1, \ldots, m \tag{10}$$

3.1. Comparison with Ground truth

Having in mind a video-phony type of application, we have used a FEM model that is built upon a generic geometric model, namely the *Waters* model [4], for which we know that it is suitable for animation. A two frame synthetic sequence has been created: the model has been positioned in the 3D world and perspectively projected on the first image frame, as shown in Figure 1a ; in second frame (Fig. 1b-grey) the right mouth corner has been moved sidewards tangentially to the model (obtaining a 'half-laughing' expression) and projected.



(a) (b) (c)

Fig. 1. (a) first synthetic image frame of the *Waters* sequence, (b) estimated motion (black) versus the original deformation (grey); (c) nonrigid motion estimation residuals

Figure 1b(-black) depicts the results of the estimation overlayed on the second frame. The visual difference between the results (black) and the ground truth (grey) is small. The considered accuracy lies indeed somewhat in the middle between character animation and surgical simulation: we need a natural motion tracking of the major features of the face, but still intend to obtain semantical parameters for animation. In Figure 1c the residuals of each term in Eq.(2), and the total residual are displayed for each iteration of the solver. The least-squares method converges: at the beginning of the iterations the residual drops dramatically and is followed by the flat tail in only a few iterations. Looking at the separate residual terms (top of Fig. 1c) the geometric and the smoothness constraint are less imposed after iteration 2. The regularization is thus gradually reduced, which makes the proposed nonrigid model fits better the ill-posed estimation problem.

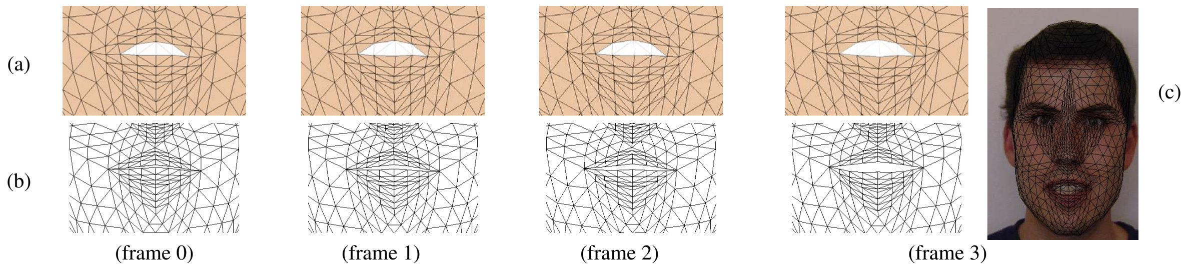


Fig. 2. Face model position on frames 0 to 3 of the *Peter* sequence [2], obtained by (a,c) the proposed nonrigid motion estimation and (b) the geometric motion estimate of P. Eisert [2]

3.2. Real Face Sequence Analysis

In Figure 2 we compare the estimated natural motion parameters using the proposed approach, to the results of P. Eisert [2]. The *Peter* sequence do not exhibit rigid motion in the examined frames, only the nonrigid mouth opening. The mouth opening is well visible on the mesh, but by closely examining the face input frames (Fig. 2c), we can see that the lips do not really open as a line. In fact, our mechanical motion model uses more parameters (namely the forces on each vertex, for these results we refer to [17]) than the one of Eisert (who is applying geometric mouth opening parameters). Therefore the shape details can be better grasped by the proposed natural motion model, and the mouth gesture is followed more precisely. The overlay of the projected estimated model on the face images is given in Fig. 2c. The deformation of lips in the 3D face model follow well those of the real person.

4. CONCLUSION

We proposed a novel 3D face motion estimation method in which pose and shape are separately handled. The face shape deformation estimation is formulated as a non-linear least-squares optimization problem in which regularization parameters originate from the mechanical properties of a face. Within this approach a physically-based FEM face model simulates the deformation of a face in a natural, realistic and accurate way. By employing distributed muscle forces as parameters of the model, the quantitative (predefined) parametric description of a geometrical shape changes, has been replaced by a qualitative semantical (natural) one. The proposed extraction of the motion parameters provides displacement estimates of the 3D face wire-frame model, which accurately reproduce the gestures of the recorded person. This non-rigid motion estimation of the face tissue is a case-study to relate 3D mechanical continuum motion equations to image measurement and can be extended to deal with other natural deformable objects.

5. REFERENCES

[1] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade, 'Three-dimensional scene flow', *IEEE PAMI*, 27(3), pp. 137–154, Mar. 2005.

[2] P. Eisert, 'Very Low Bit-Rate Video Coding Using 3-D Models', Ph.D., Universitat Erlangen, Germany, 2000,

[3] I. Matthews and S. Baker, 'Active appearance models revisited', *IJCV*, 60(2), pp. 135 – 164, Nov. 2004.

[4] D. Terzopoulos and K. Waters, 'Analysis and synthesis of facial image sequences using physical and anatomical models', *IEEE PAMI*, 15(6), pp. 569–579, Jun. 1993.

[5] I. Essa, S. Basu, T. Darrell, and A. Pentland, 'Modeling, tracking, and interactive animation of faces and heads using input from video', in *Computer Animation '96, Geneva, Switzerland*, Jun. 1996.

[6] S. Gibson and B. Mirtich, 'A survey of deformable modeling in computer graphics', Tech. Rep. TR-97-19, MERL, Nov. 1997.

[7] G. Picinbono, H. Delingette, and N. Ayache, 'Realtime large displacement elasticity for surgery simulation: Non-linear tensor-mass model', *miccai*, 2000.

[8] Y. Bascar and D. Weichert, 'Nonlinear continuum mechanics of solids', Springer-Verlag: Berlin, 2000, ISBN 3-540-66601-X.

[9] L.V. Tsap, D.B. Goldgof, and S. Sarkar, 'Nonrigid motion analysis based on dynamic refinement of finite element models', *IEEE PAMI*, 22(5), pp. 526–543, May 2000.

[10] J.P. Gourret, N. Magnenat-Thalmann, , and D. Thalmann, 'Simulation of object and human skin deformations in a grasping task', *Proc. ACM SIGGRAPH*, pp. 21–30, 1989.

[11] H. Spies, B. Jähne, and J. L. Barron, 'Range flow estimation', *CVIU*, 85(3), pp. 209–231, Mar. 2002.

[12] M. Harville, A. Rahimi, T. Darrell, G. Gordon, and J. Woodfill, '3d pose tracking with linear depth and brightness constraints', in *IEEE ICCV*, pp. 206–213, 1999.

[13] Lixin Yang, 'Structure and Motion Analysis : Variational Methods and Related PDE Models', Ph.D., VUB-ETRO, Brussels, Belgium, Jul. 2004.

[14] H. Li, A. Lundmark, and R. Forchheimer, 'Image sequence coding at very low bitrates: A review', *IEEE Image Processing*, 3(5), pp. 589–605, Sep. 1994.

[15] D. Terzopoulos and A. Witkin, 'Physically-based models with rigid and deformable components', *IEEE Computer Graphics and Applications*, 8(6), pp. 41–51, Nov. 1988.

[16] E. Ventsel and T. Krauthammer, 'Thin Plates and Shells', , 2001, ISBN 0-8247-0575-0.

[17] I. Ravyse, 'Facial Analysis and Synthesis', Ph.D., VUB-ETRO, Brussels, Belgium, May 2006.

[18] I. Ravyse and H. Sahli, 'Facial Analysis and Synthesis Scheme', *Proc. Advanced Concepts for Intelligent Vision Systems (ACIVS)*, LNCS 4178, pp. 810–820, 2006.