Uncertainties-driven Surface Morphing: The case of Photo-realistic Transitions between Facial Expressions*

Maxime TARON [1,2] Charlotte Ghys [1,3] Nikos PARAGIOS [1]

[1] MAS, Ecole Centrale Paris, 92295 Chatenay Malabry, France http://vision.mas.ecp.fr

[2] CERTIS, Ecole des Ponts et Chaussées, 77455 Marne la Vallee, France

[3] France Télécom R&D, TECH/IRIS/VIA 22307 Lannion, France

E-mail: {maxime.taron ; charlotte.ghys ; nikos.paragios}@ecp.fr

Abstract

Reproduction of facial animation play a fundamental role in applications requiring human-computer interactions The objective of this paper is to introduce a geometric mechanism that exploits a fix number of states and is able to execute a subsequent number of transitions between facial expressions. Standard stereo-based techniques are used to reproduce the geometry and appearance of the most characteristic facial expressions. A novel free-form-deformation technique based on uncertainty driven local geometric registration in the space of distance transforms is used to produce a one-to-one mapping between the surfaces and the associated textures. Standard techniques from image morphing introduce the temporal aspect in the process. Experimental results and comparisons with actual observation demonstrate the potentials of such an approach.

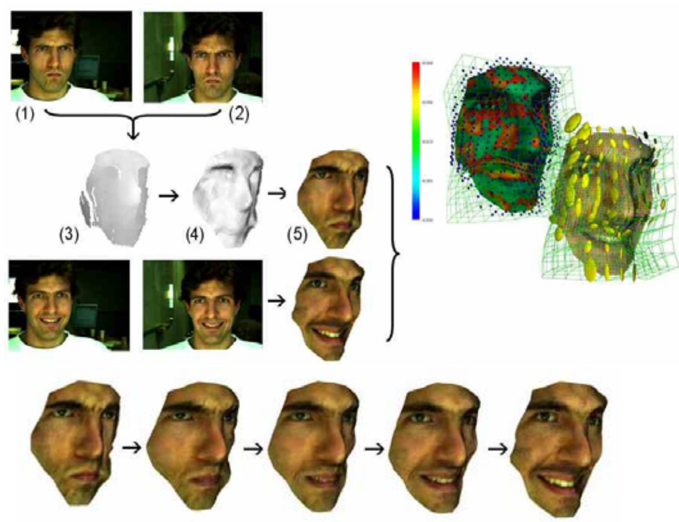


Figure 1. Overview of Uncertainties-driven Surface Morphing towards Photo-realistic Facial Animations (From Anger to Joy).

1 Introduction

In the past fifteen years, particular efforts have been made in the area of human-computer interaction systems with for instance applications of Facial animation.

Simplistic 2D facial animation computing transition between characteristic expressions refers to image morphing techniques [8]. Then, acquisition of 3D structure from 2D images have led to photo-realistic 3D face models that capture texture and geometry [9].

A step further was the use of physics-based models [6] along with images to simulate facial expressions. Such methods rely on high resolution 3D shape data and models that when combined with tracking components leads to reproduction of facial expressions [14]. In this paper, we are interested in a similar problem with the perspective of future applications to communication devices. The bandwidth resource being constrained, transmitting 3D structure in real-time is rather unrealistic. However, one can assume feasible the reproduction of 3D shape and texture face models for the expressions that correspond to human emotions. Therefore the objective of the paper is the recovery of a landmark-free method providing a deformation mechanism.

Shape (3D) reconstruction [3] from images have been a well studies problem in particular for short base-line binocular camera systems. Number of methods were proposed from simple correlation based techniques, to more accurate one like level-set methods or graph-cuts [5]. Within our approach we consider reconstruction and texture mapping for a limited number of expressions that are then used as key states to produce animations.

Surface registration is an open problem with application

*Research work supported by the Imaging & Vizualisation Dept. Siemens Corporate Research (Princeton USA) the authors also aknowledge the MMI Group from Delft University (Netherland) for fruitful discussion regarding face modelling.

to various domains. It requires first the choice of a feature space (point clouds, landmarks...) to model the shapes. Their representation in orthogonal basis [2], using triangulated surfaces [14] or higher order representations on implicit functions [7]. Transformation can be either global or local, based on a set of parameters or parameter free aiming at a perfect match of structures. Similarity metrics between surfaces are often based on Euclidean or geodesic distances [1], as well as statistical metrics [13]. This set of choice usually leads to cost function being minimized using various methods like gradient descent, dynamic programming, simplex methods or simulated annealing. This is however insufficient to quantify the quality of the segmentation result.

In this paper, we account for the lack of reconstruction precision and address limitations of the discretization/triangulation process by using distance transforms. A multi-resolution Free Form Deformation procedure is used to recover the most promising registration field along with uncertainties measures that characterize the quality of the obtained alignment map. This field is used in a linear fashion to produce surface-realistic transitions between facial emotions. The entire process is demonstrated in [Fig. (1)].

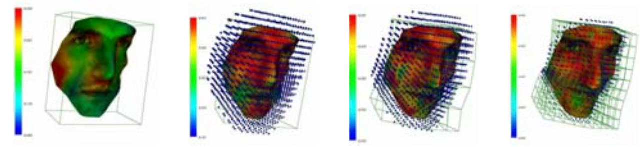
In section 2 we briefly review the stereo reconstruction process. Registration between surfaces with uncertainties is presented in sections 3 and 4, while in sections 5 texture mapping and morphing between end-states are presented along with experimental results.

2 Stereo Reconstruction

Let the projection of a 3D point M, with a stereo system be m and m'. Considering an image point m, the spatial line of sight going through m projects on the other image as an epipolar line containing m'. It can easily be shown that there exists a 'fundamental' matrix verifying $m.F.m^T$. Where F is estimated during the calibration step of the two cameras and encodes the relationship between the two images.

Once F is known, it is used to constrain the correspondence search in one dimension. To simplify and speed up the stereo matching, images are warped so the epipolar lines become scanlines. This is the rectification step. Therefore the stereo problem for retrieving corresponding points m and m' simplifies to a horizontal correspondence search, and the value $d = x_{m'} - x_m$ is called disparity.

Since the objective of our method is to create surface morphing even for low resolution and quality surfaces, normalized correlation is used to determine such correspondences. Once we know the disparity for each pixel and the intrinsic and extrinsic parameters of the camera, the 3D position of every point is computed resulting in a a dense cloud of points [3] out of which a smooth 3D mesh is constructed.



(1) (2) (3) (4)

Figure 2. The colormap indicates the distance to the target shape. (1) Initial position, (2) coarse registration (4) intermediate FFD resolution, (5) Final registration

The entire process (without the calibration part) is shown in [Fig.(1)] where one can see the pair of stereo images (1,2), the 3D point cloud (3), the corresponding mesh (4) and the textured mesh (5). Such a method is used to reconstruct a number of expressions corresponding to various 3D surfaces. Surface morphing consists of finding an appropriate deformation mechanism from one expression to another. Therefore, producing facial animations turns into a registration problem between surfaces where each expression is to be registered to the remaining ones.

3 Registration through Implicit Polynomials

In the present framework, a 3D shape $\mathcal{S}$ is embedded in a higher dimensional space through the use of Euclidean distance transform $\mathcal{D}$.The positive distance transform of the shape $\mathcal{S}$ is denoted $\phi_{\mathcal{S}}$. Such a space is invariant to a simple transformation T as translation and rotation and can also be modified to account for scale variations :

$$\mathcal{S}_2 = T \circ \mathcal{S}_1 \Rightarrow \phi_{\mathcal{S}_1} = \phi_{\mathcal{S}_2} \circ T.$$

In the most general case an explicit relation between the distance function of the source and the target is not present.

Now consider a smooth diffeomorphism $\mathcal{L}$ defined on the domain Ω and depending upon a vector of parameters $\mathbf{\Theta} \in \mathbb{R}^n$.

Standard point-based registration consists of applying $\mathcal{L}$ to the source shape $\mathcal{S}$ and minimizing the integral defined on $\mathcal{S}$ such that some metric error between the transformed source and the target $\mathcal{T}$ is minimal. In order to prevent the minimization process of such energy to fall into local minima, one can extend registration within a band including numerous isosurfaces of the source distance transform. Therefore, a robust quadratic registration energy is proposed:

$$E_\alpha(\mathcal{L}(\mathbf{\Theta})) = \iiint_\Omega \mathbb{1}_\alpha(\phi_{\mathcal{S}}(\mathbf{x}))\left(\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x}))\right)^2 d\mathbf{x}$$

where $\mathbb{1}_\alpha$ is the indicator function of the segment $[-\alpha/2, \alpha/2]$.

This Energy is minimized through the calculation of variations. Within such a process the selection of the parameter α is crucial, since to some extent, it refers to the scale of the shapes to be registered. Therefore minimization is performed for a decreasing set of α values while the degrees of freedom of the transformation $\mathcal{L}$ increases.

Such an objective function can be used in conjunction with multiscale Free Form Deformations (FFD) to address the global to local deformations [4]. FFD is an efficient way to model smooth transformations on images [10]. Deformations of shapes (and their implicit representation $\phi_{\mathcal{S}}$) are recovered by evolving a square control lattice $\mathbf{P}$ that is overlaid on the initial distance transform structure. Let us consider the control lattice points $\{\mathbf{P}_{L,M,N}\}$ defining the initial regular grid. The displacement $\delta\mathbf{P}$ of any control point will induce a local and $\mathcal{C}^2$ field of deformation:

$$\mathcal{L}(\mathbf{\Theta}, \mathbf{x}) = \sum_{i,j,k} K_i(\mathbf{x}_x) K_j(\mathbf{x}_y) K_k(\mathbf{x}_z)(\mathbf{P}_{i,j,k} + \delta\mathbf{P}_{i,j,k})$$

where along the x axis, K_i is the i^{th} cubic B-spline basis function. It has a compact support $[(\mathbf{P}_{i-2})_x, (\mathbf{P}_{i+2})_x]$. This local transformation is a compromise between global and local registration and its parameters consist of the displacement of the control points ($\mathbf{\Theta} = \{\delta\mathbf{P}_{L,M,N}\} \in \mathbb{R}^{3LMN}$). The registration process is initialized using an affine transformation. Then, when refining the FFD lattice the new transformation is initialized with the output of the former minimization process. To recover a smooth transformation and avoid folding when increasing complexity of FFD grid, we adopt a regularization term motivated by the thin plate energy functional [15] to control the spatial variations of the displacement:

$$E_{\text{smooth}}(\mathcal{L}(\mathbf{\Theta})) = \iiint_\Omega \|\mathcal{H}(\mathcal{L}_x)\|_F^2 + \|\mathcal{H}(\mathcal{L}_y)\|_F^2 + \|\mathcal{H}(\mathcal{L}_z)\|_F^2$$

with $\mathcal{H}$ the Hessian and $\|.\|_F$ the Frobenius norm. This can be further simplified in the case of the cubic B-spline to the quadratic form $[E_{\text{smooth}}(\mathcal{L}(\mathbf{\Theta})) = \mathbf{\Theta}^T \mathrm{C} \mathbf{\Theta}]$ with C a symmetric matrix.

The objective function $[E_\alpha(\mathcal{L}(\mathbf{\Theta})) + wE_{\text{smooth}}(\mathcal{L}(\mathbf{\Theta}))]$ is optimized using a standard gradient descent method. The coherence of distance transform allow to perform a very sparse discretization of the space when minimizing the Energy and leads to exceptional fast results.

4 Uncertainty estimation on registered shapes

We aim to recover uncertainties on the quality of the local registration vector $\mathbf{\Theta}$ in the form of a $[3LMN \times 3LMN]$ covariance matrix. We adapt a method introduced in [11] and that was previously considered in 2D in [12]. E_α is reformulated in the limit case where α the size of the limited band around the model shape tends to 0.

$$E_0(\mathbf{\Theta}) = \iint_{\mathcal{S}} \phi_{\mathcal{T}}^2(\mathcal{L}(\mathbf{\Theta}; \mathbf{x})) d\mathbf{x} = \iint_{\mathcal{S}} \phi_{\mathcal{T}}^2(\mathbf{x} + \mathcal{X}(\mathbf{x})\mathbf{\Theta}) d\mathbf{x}$$

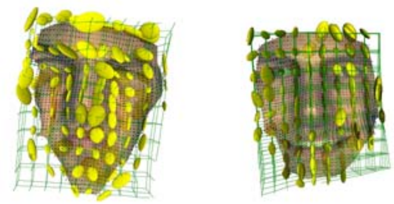


Figure 3. Representation of Uncertainties projected on the registered FFD Grid.

where the FFD has been rewritten in a linear fashion. If we denote q the closest point from $\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}; \mathbf{x})$ located on $\mathcal{T}$. When using the distance condition $[\, \|\nabla\phi_{\mathcal{T}}(\mathbf{x}')\| = 1]$ and assuming that E_0 is small when reaching the minimal, a second order approximation of the Energy could be written :

$$E_0(\mathbf{\Theta}) = \iint_{\mathcal{S}} \left[(\mathbf{x} + \mathcal{X}(\mathbf{x})\mathbf{\Theta} - \mathbf{q}) \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}')\right]^2 .$$

Localizing the global minimum of an objective function E is equivalent to finding the major mode of a random variable with density $[\exp(-E_0/\beta)]$. In the present case of a quadratic energy (and therefore Gaussian random variable), the covariance and the Hessian of the energy are directly related by $[\Sigma_{\mathbf{\Theta}_i}^{-1} = H_{\mathbf{\Theta}_i}/\beta]$ with β a scaling coefficient. This leads to the following expression for the covariance :

$$\Sigma_{\mathbf{\Theta}}^{-1} = \frac{1}{\beta} \iint_{\mathcal{S}} \mathcal{X}(\mathbf{x})^T . \nabla\phi_{\mathcal{T}}(\mathbf{x}') . \nabla\phi_i(\mathbf{x}')^T . \mathcal{X}(\mathbf{x}) d\mathbf{x}$$

In the most general case, the Hessian is not invertible. Additional constraint are added to the Energy acting like an extra regularization term and leading to the covariance matrix for the parameter estimate :

$$\Sigma_{\mathbf{\Theta}} = \beta \left(\iint_{\mathcal{S}} \mathcal{X}(\mathbf{x})^T . \nabla\phi_{\mathcal{T}}(\mathbf{x}') \nabla\phi_{\mathcal{T}}(\mathbf{x}')^T \mathcal{X}(\mathbf{x}) d\mathbf{x} + \gamma \mathbf{I} \right)^{-1}$$

Initially imagined to compute a statistical learning in the space of deformations, these uncertainty also reveals of higher interest in the case of surface registration. Indeed, it may be inferred that areas with high uncertainty on surface registration will also present errors on texture mapping. In the present article however, 3D registration and 2D Texture mapping are uncorrelated and therefore do not allow to take these additional information into account.

5 Surface Morphing & Facial Animations

Recording the face of a subjects with a low precision binocular stereo system, the subsequent set of expressions has been reconstructed leading to a triangulated and textured surface : S ={neutral reference, anger, joy, sadness, surprise, disgust and fear}. The registration framework allows the computation of the deformation field from the

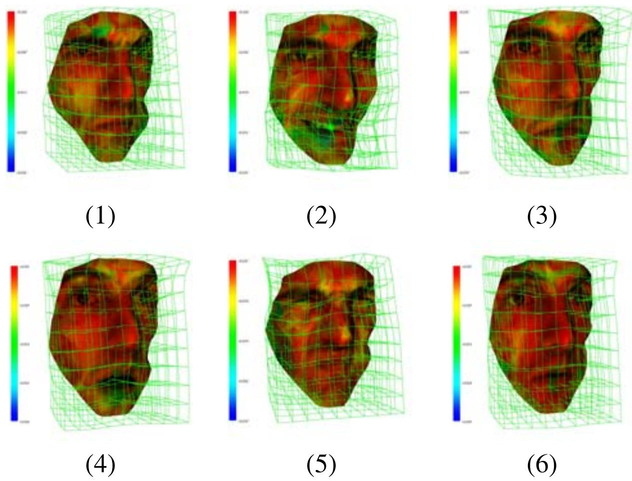


Figure 4. Expressions: (1) anger, (2) joy, (3)sad, (4) surprise, (5) disgust, and (6) fear.

'neutral reference' to any of these shapes $\{\mathcal{L}_{S_i}\}$ along with uncertainties estimates $\{\mathcal{V}_{S_i}\}$. Transition from one expression to the next is therefore characterized with the deformation $\{\mathcal{L}_{S_{i+1}} - \mathcal{L}_{S_i}\}$. Information on the duration of the transition has also been retained from the orginal recording.

Let's first adress surface morphing. $\{\mathcal{L}_{S_{i+1}} - \mathcal{L}_{S_i}\}$ retains all geometric information needed to transform surfaces. However, the transformation S_i to S_{i+1} could be done using a linear, or a geodesic transition, or following the gradient descent path of the registration process. This last choice would be highly non uniform and therefore unrealistic. Geodesic paths is a more prominent direction however, their estimation is challenging and beyond the scope of the article. So we focused on a linear interpolation strategy adapted to physiological factors (varying the speed of transitions) as shown on [Fig. (1)].

Texture information is available on the source and target. Assuming a proper surface transition between states, a linear interpolation between the extreme texture values provide a natural animation process as shown in [Fig. (1)]. However, particular attention is to be paid on the distance between the transformed source and the target, such information is encoded from the uncertainties estimates and we investigate how this will be used to produce more photorealistic rendering.

Explicit validation is possible for the registration aspect of our approach with the use of Euclidean distance quantifying the error as shown on [Fig. (4)] for 6 different expressions registered starting from neutral, all shapes beeing scaled with size 1.

6 Discussion

The problem of surface morphing was addressed through the use of FFD Registration providing uncertainties estimations. Photo-realistic transitions between expressions were achieved on low-resolution images with a small set of data (The FFD-field and textures). If adapted to the user expressions through a learning step, such framework could be adapted to the next generation communication systems.

The present method could be extended with the introduction of a joint surface/texture registration leading to a 1 step 3D face morphing. An appropriate criteria that account for a multi-modal space has to be introduced and will also reinforce the meaning of uncertainties. Parallel to that, precise reconstruction and recognition of facial expressions are investigated.

References

[1] M.-F. Beg, M. Miller, A. Trouve, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV*, pages 139–157, 2005.

[2] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *CVIU*, 61:38–59, 1995.

[3] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[4] X. Huang, N. Paragios, and D. Metaxas. Establishing Local Correspondences towards Compact Representations of Anatomical Structures. In *MICCAI*, 2003.

[5] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *ECCV*, pages 3:82–96, 2002.

[6] Y. Lee, D. Terzopoulos, and K. Walters. Realistic Modeling for Facial Animation. In *SIGGRAPH*, pages 55–62, 1995.

[7] N. Paragios, M. Rousson, and V. Ramesh. Non-Rigid Registration Using Distance Functions. *CVIU*, pages 89:142–165, 2003.

[8] F. Parke and K. Waters. *Computer Facial Animation*. AK Peters, 1996.

[9] F. Pighin, J. Hecker, D. Lischinski, R. Sziliski, and D. Salesin. Synthesiazing Realistic Facial Expressions from Photographs. In *SIGGRAPH*, pages 75–84, 1998.

[10] T. Sederberg and S. Parry. Free-form deformation of solid geometric models. volume 20, pages 151–160, 1986.

[11] C. Stewart, C.-L. Tsai, and B. Roysam. The dual bootstrap ICP algorithm with application to retinal image registration. *IEEE Trans. Med. Img.*, 22:1379–1394, 2003.

[12] M. Taron, N. Paragios, and M.-P. Jolly. Uncertainty-driven non-parametric knowledge-based segmentation: The corpus callosum case. In *VLSM*, pages 198–207, 2005.

[13] P. Viola and W. Wells. Aligment by Maximization of Mutual Information. In *IEEE ICCV*, pages 16–23, 1995.

[14] Y. Wang, X. Huang, C.-H. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High Resolution Acquisition, Learning and Transfer of Dynamic 3-D Facial Expressions. In *EUROGRAPHICS*, pages 677–686, 2004.

[15] Z. Xie and G. Farin. Deformation with hierarchical b-splines. *Mathematical Methods for Curves and Surfaces: Oslo 2000*, pages 545–554, 2001.