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

Maxime Taron
Charlotte Ghys
Nikos Paragios

Research Report 05-17
December 2005

Ecole Nationale Des Ponts et Chaussées
CERTIS
Centre d'Enseignement et de Recherche en Technologies de l'Information et Systèmes

CERTIS, ENPC,
77455 Marne la Vallee, France,

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

Morphing de Surfaces : Transitions Photo-Réalistes d'Expressions Faciales

Maxime Taron[2]
Charlotte Ghys[2]
Nikos Paragios[1]

[1]MAS, ECP
[2]CERTIS, ENPC, 77455 Marne la Vallee, France, http://www.enpc.fr/certis/

Abstract

Facial animations play a fundamental role in inter-human communication and their reproduction is a keen element of mixed reality systems with a number of applications like human computer interaction, post-production and cinematography. The objective of this paper is to introduce a geometric facial animation mechanism that exploits a fix number of states and is able to execute a subsequent number of expressions. To this end, standard stereo-based techniques are used to reproduce the most characteristic facial expressions in terms of geometry and appearance. A novel free-form-deformation technique based on uncertainties-driven local geometric registration in the space of distance transforms is used to produce a one-to-one mapping between the surfaces and the textures associated with each expression. Standard techniques from image morphing introduce the temporal aspect in the process leading to a promising animation mechanism. Experimental results and comparisons with actual observation demonstrate the potentials of such an approach.

Résumé

Les animations faciales jouent un rôle fondamental dans la communication et leur reproduction est un élément clef dans les systèmes de réalité virtuelle, avec un nombre important d'applications comme l'interaction homme-marchine ou le cinéma. Notre objectif est d'introduire un mécanisme d'animation faciale utilisant un nombre fixé d'états et qui est capable de reproduire les états intermediaires. Pour cela, des techniques de stereo standards sont utilisées pour reproduire les différentes expressions faciales en terme de géométrie et d'apparence. Nous introduisons une nouvelle technique de recalage geometrique basé sur des déformations polynomiales dans l'espace des fonctions implicites. Ceci permet de mettre les surfaces en correspondance en évaluant des informations d'incertitudes liées à la qualité du recalage. Par ailleurs, des techniques classiques de morphing d'images introduisent un aspect temporel et rendent possible le mécanisme d'animation faciale.Des résultats experimentaux comparés à des observations réelles d'animation montrent tout le potentiel de notre approche.

Contents

1 Introduction

Understanding facial expressions is an important aspect of inter-human communication while facial animations are emerging components of human computer interaction systems. Therefore, a substantial amount of effort has been paid in this area in particular from computer graphics with post-production market being the locomotive of such innovation needs.

Simplistic 2D facial animation refers to image morphing techniques [2]. Such techniques are able to interpolate between images corresponding to characteristic facial expressions towards 2D facial animations [17]. Certain constraints are imposed to this end, and particular attention is paid to the mouth region [7]. Recent advances in acquisition as well as on estimation of 3D structure from 2D images [18] have led to photo-realistic 3D face models that capture texture and geometry.

A step further was the use of physics-based models [14], spline surfaces [15], and variational approaches [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 can lead to promising reproduction of facial expressions [23]. More advanced techniques consist of learning statistics between transitions, that are then inherited in the synthesis step. In this paper, we are interested in a similar problem from a different perspective, that is the communication one. State-of-the art communication systems involve audio and image while emerging ones move towards photo-realistic 3D animations with hardware and bandwidth resources being constrained. Therefore, capturing 3D structure/appearance, tracking such structure in images and transferring such information space is a rather unrealistic assumption.

On the other hand, one can assume feasible the reproduction of accurate 3D shape and texture face models for the facial expressions that correspond to human emotions. Once such surfaces have been pre-computed, they can be used along with images (rendering) to create photo-realistic transitions between emotions, that is the objective of this paper. We aim to recover an automated landmark-free method that provides a deformation mechanism leading to a one-to-one correspondence between 3D surfaces that corresponding to human emotions. Two aspects are to be addressed within such an approach; (i) 3D shape recovery from images for a discrete number of states, and (ii) surface registration of structures undergoing topological changes and efficient morphing between these structures.

Shape (3D) reconstruction [8] from images have been a well studies problem, in particular for short base-line binocular camera systems. Based on constraints driven from the epipolar geometry, a number of methods were proposed to recover 3D structure. Simple correlation based techniques, dynamic programming, space curving [13], variational and level set methods [9] as well as combinatorial methods like graph-cuts [12] are some of the state-of-the art methods in the liter-

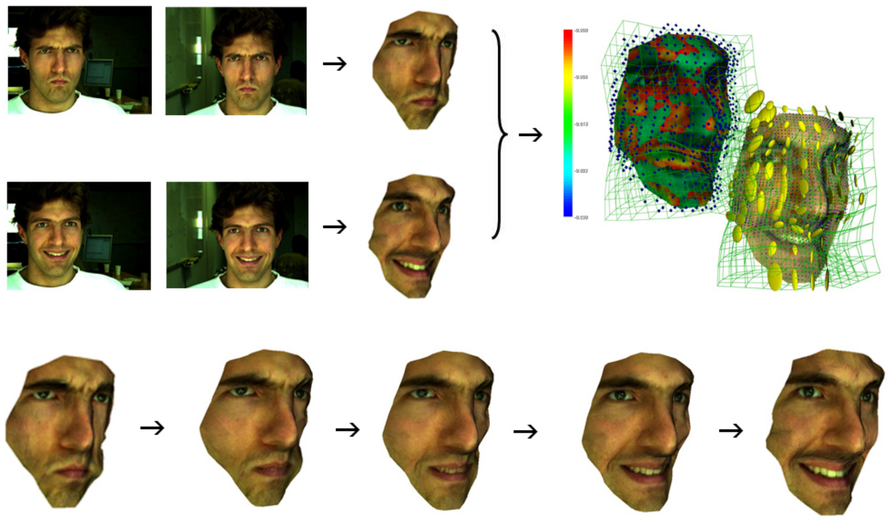


Figure 1: Overview of Uncertainties-driven Surface Morphing towards Photo-realistic Facial Animations.

ature. More accurate methods integrate images with digital projection techniques towards high-resolution data. Within the context of our approach we consider reconstruction and texture mapping for a limited number of expressions that are then used as states to produce animations.

Surface registration is an open problem with particular interest to a number of domains, like medical imaging, computational geometry and graphics. Feature space, nature of transformation, similarity metric and the optimization procedure are the components of any registration technique. Features are often related with the parameterization of the surface and could be sparse cloud points, landmarks representations, representations on orthogonal basis [5], triangulate surfaces [23] or higher order representations like implicit functions [16]. Transformation can be either global, local or could consist of both components [21]. Global have a parametric nature and consist of a limited set of parameters, while local aim to recover a one-to-one correspondence between the elements of the two structures [4] that is an ill-posed problem. Similarity metrics are often based on Euclidean [10] or geodesic distances [1] between surfaces, as well as local and global [22] statistical metrics. Gradient descent, complex conjugate, dynamic programming, simplex methods and simulated annealing are the most popular optimization methods. Such registration methods provide a deformation field that corresponds to the lowest potential of the designed cost function that is insufficient to quantify the quality of the segmentation result.

In this paper, we introduce a novel one-to-one surface morphing technique for registration between facial expressions. In order to account for lack of reconstruction precision, errors as well as address limitations of the discretization/triangulation process, surfaces are represented in an implicit fashion, using distance transforms. Morphing between surfaces is then addressed in this space. 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. Such deformation field is then used in a linear fashion to produce surface-realistic transitions between facial emotions. Registration uncertainties are then used to quantify the outcome of the animation process. The entire process is demonstrated in [Fig. (1)].

The reminder of this paper is organized in the following fashion; in section 2 we briefly review the stereo reconstruction process from pair of images and present static models of human emotions. 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. Discussion and experimental results are part of section 6.

2 Stereo Reconstruction

In this section, we briefly discuss the 3D reconstruction from images [8]. To this end, we first introduce the relation between the two cameras, then we explain rectification and stereo matching and conclude with the 3D reconstruction.

Basic notions from 3D geometry explain such a process. Given a 3D point M, its projection in a stereo system are m and m'. m' is on the projection of m line of sight $[l'_m = Fm]$. l'_m is called epipolar line and F is the fundamental matrix. It encodes the relationship between the two images and all the corresponding points should satisfy: $\left[m'^T Fm = 0\right]$.

Calibration process is about to infer positions of points in one image from positions in the world This is modelled by the projection matrix P such as $[m = PM]$. P can be decomposed as follows : $P = A[Rt]$, where A describes the characteristics of the camera (focal length, location of the image center, real pixel size and distortion of the lens), and $[Rt]$ is a concatenation of a rotation matrix and a translation vector describing the change of world coordinate system.

Once fundamental matrix F is known, it can be used to constrain the correspondence search in one dimension. To simplify and speed up the stereo matching the image are warped so the epipolar lines become scanlines, a process that is known as rectification. Two corresponding points m and m' become :

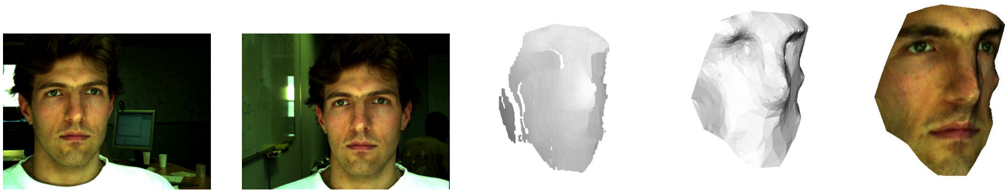
$$m_r = \begin{pmatrix} x \\ y \end{pmatrix}, \quad m'_r = \begin{pmatrix} x + d \\ y \end{pmatrix}$$

where d it the horizontal displacement called disparity. Once epipolar lines have been determined, the stereo problem is simplified to horizontal correspondences. Since the objective of our method is to create surface morphing even for low resolution and quality surfaces, simple normalized correlation is used to determine such correspondences. Towards more precise models, other more efficient techniques can be used like dynamic programming, space curving [13], variational and level set methods [9] as well as combinatorial methods like graph-cuts [12].

Once we have the disparity for each pixel and we know the intrinsic and extrinsic parameters of the camera, we can compute the 3D position of the points by solving the system :

$$m = A[Rt]M = PM, \quad m' = A'[R't']M = P'M.$$

The result of the reconstruction stage is a dense cloud of points. A 3D surface is generated using a simple Delaunay triangulations while one can seek for more advanced mesh generation techniques. In order to capture the texture properties of the different expressions, we use the closest point principle. In an off-line step, for each triangle point we search the closest reconstructed 3D point and we



(1) (2) (3) (4) (5)

Figure 2: (1,2) Stereo Pair, (3) Depth information (4) Low resolution smooth surface representation, (5) Complete model with texture.

associate its texture to the mesh. The texture mapping in the triangle is made by interpolating the texture of the mesh vertices. The entire process (without the calibration part) is shown in [Fig.(2)] where one can see the pair of stereo images, the 3D point cloud, the corresponding mesh and the mesh with the texture mapping.

Such a method is used to reconstruct a number of expressions corresponding to 3D surfaces. Surface morphing consists of finding an appropriate deformation mechanism from one expression to another. Therefore, producing facial animations becomes a registration problem between surfaces where each expression is to be registered to the remaining ones. However a number of limitations are to be addressed; (i) reconstruction errors in particular in areas with limited texture, (ii) low resolution mesh generation due to the use of simple techniques like Delaunay triangulations, (iii) registration between non-regular meshes in terms of volume, parameterization, control points, etc. Implicit representations and distance transforms can naturally cope with most of the above limitations.

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 $\mathcal{D}$ [3]. We consider the positive function ϕ defined on the image domain Ω :

$$\phi_{\mathcal{S}}(x,y,z) = \begin{cases} 0, & (x,y,z) \in \mathcal{S} \\ +\mathcal{D}((x,y,z),\mathcal{S}), & (x,y,z) \in \Omega - \mathcal{S} \end{cases}$$

Where $\mathcal{D}((x,y),\mathcal{S})$ refers to the min distance between the point (x,y,z) and the shape $\mathcal{S}$. Such a space is invariant to a simple transformation T as translation and rotation and can also be modified to account for scale variations :

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

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

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

$$\mathcal{L}(\mathbf{\Theta}, .) : \Omega \to \Omega$$

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 is minimal. Using the implicit representation such a method is equivalent to minimizing :

$$E_0(\mathcal{L}(\mathbf{\Theta})) = \iint_{\mathcal{S}} \rho(\phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta},\mathbf{x}))d\mathbf{s} \tag{1}$$

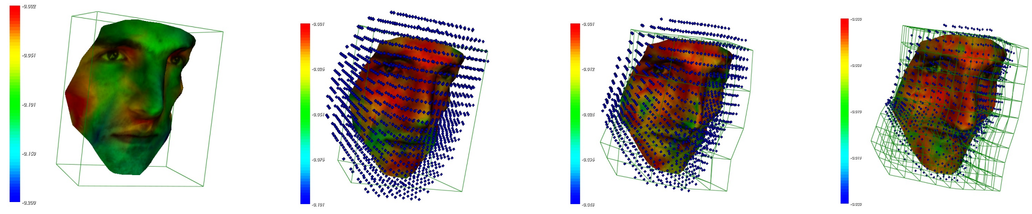
where ρ is a robust estimator and $\phi_{\mathcal{T}}$ the distance transform of the target shape $\mathcal{T}$. In order to prevent the minimization process of such energy to fall into local minima, one can extend registration within a band including numerous isosurfaces. Therefore, a more robust Registration Energy is proposed:

$$E_\alpha(\mathcal{L}(\mathbf{\Theta})) = \iiint_{\Omega} \chi_\alpha(\phi_{\mathcal{S}}(\mathbf{x}))\rho\left(\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathcal{L}(\mathbf{\Theta},\mathbf{x}))\right) d\mathbf{x} \tag{2}$$

where we introduce the indicator function:

$$\chi_\alpha(x) = \begin{cases} 1/(2\alpha) & \text{if } x \in [-\alpha,\alpha] \\ 0 & \text{else} \end{cases}$$

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



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

Figure 3: The colormap indicates the distance to the target shape. (1) Initial position, (2) coarse registration (4) intermediate FFD resolution, (5) Final registration for Morphing Purposes

the scale of the shapes to be registered. It is natural when converging to the optimal solution that α tends to 0. Therefore, we assume a finite number of decreasing set of radii $\{\alpha_0 > ... > \alpha_t > ... > \alpha_n \approx 0\}$ that is equivalent to a scale-space decomposition of the process. On the other hand, if the cardinal of $\mathbf{\Theta}$ is initially too large, there is also a high risk of converging to a local minimum. So, we progressively increase the complexity of the transformation and therefore the size of $\mathbf{\Theta}$ during energy minimization.

Let $\mathbf{\Theta}_{t-1}$ be the parameters defining the transformation $\mathcal{L}_{t-1} = \mathcal{L}(\mathbf{\Theta}_{t-1}, .)$ for which the minimum of energy was reached at scale $t-1$. Also let $\mathcal{S}^{t-1} = \mathcal{L}_{t-1} \circ \mathcal{S}$ the registered shape. The registration between shapes is then equivalent to iteratively minimizing :

$$E_{\alpha_t}(\mathcal{L}(\mathbf{\Theta})) = \iiint_\Omega \chi_{\alpha_t}(\phi_\mathcal{S}(\mathbf{x}))\rho(\phi_{\mathcal{S}^{t-1}}(\mathcal{L}_{t-1}(\mathbf{x})) - \phi_\mathcal{T}(\mathcal{L}(\mathbf{\Theta}, \mathbf{x})))d\mathbf{x}$$

where a correction process is applied when refining scales through the computation of the distance transform for the registered shape $\phi_{\mathcal{S}^{t-1}}()$. Within such a formulation the integration domain is always related to the initial source shape and does not depend on the number of iteration or the size of $\phi_\mathcal{T}$. It may be noticed that both the energy (1) and (2) are equivalent when α_t converges to zero.

Such an objective function can be used in conjunction with multiscale free form deformations as in [11] to address the global to local deformations. Linear splines and Cubic B-spline based free form deformations are an efficient way to model locally smooth transformations on images [19]. 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 of any of control point will induce a local and $\mathcal{C}^2$ field of deformation:

$$\mathcal{L}(\boldsymbol{\Theta}, \mathbf{x}) = \sum_{l=-1}^{2} \sum_{m=-1}^{2} \sum_{n=-1}^{2} B_l(u) B_m(v) B_n(w) (\mathbf{P}_{i+l,j+m,k+n} + \delta \mathbf{P}_{i+l,j+m,k+n})$$

where $\mathbf{x} = (u, v, w)$ and (i, j, k) are chosen so that $\mathbf{P}_{i,j,k}$ is the closest point from $\mathbf{x}$ located in the portion of space defined by $\{x < u, y < v, z < w\}$. B_k is the k^{th} basis function of the cubic B-spline.
This local transformation is a compromise between global and local registration and its parameters consist of the displacement of the control points ($\boldsymbol{\Theta} = \{\delta \mathbf{P}_{L,M,N}\}$). The registration process is initialized using a square box, allowing a rough registration with 24 degrees of freedom. For grids with edge size smaller than five, cubic splines cannot be used. We therefore process this initial step of registration using Linear splines.

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 [24] to control the spatial variations of the displacement:

$$E_{\text{smooth}}(\mathcal{L}(\boldsymbol{\Theta})) = \iiint_{\Omega} \left(|\mathcal{L}_{xx}|^2 + |\mathcal{L}_{yy}|^2 + |\mathcal{L}_{zz}|^2 + 2\,|\mathcal{L}_{xy}|^2 + 2\,|\mathcal{L}_{yz}|^2 + 2\,|\mathcal{L}_{zx}|^2 \right) d\Omega$$

that can be further simplified in the case of the cubic B-spline to the quadratic form $[E_{\text{smooth}}(\mathcal{L}(\boldsymbol{\Theta})) = \boldsymbol{\Theta}^T \mathrm{C} \boldsymbol{\Theta}]$ with C a symmetric matrix.

The objective function $[E_\alpha(\mathcal{L}(\boldsymbol{\Theta})) + w E_{\text{smooth}}(\mathcal{L}(\boldsymbol{\Theta}))]$ 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.

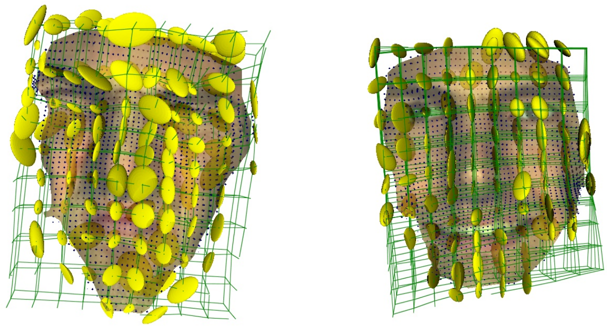


Figure 4: Representation of Uncertainties projected on the registered FFD Grid.

4 Uncertainty estimation on registered shapes

We aim to recover uncertainties on the vector $\mathbf{\Theta}$ in the form of a $[3LMN \times 3LMN]$ covariance matrix by adapting a method initially introduced in [20]. We are considering the quality of the local registration on shapes, that is the zero levelset of the distance transform. Therefore, E_α is formulated in the limit case where α the size of the limited band around the model shape tends to 0.

$$E_0(\mathbf{\Theta}) = \iint_{\mathcal{S}} \phi_{\mathcal{T}}^2(\mathcal{L}(\mathbf{\Theta}; \mathbf{x}))d\mathbf{x} = \iint_{\mathcal{S}} \phi_{\mathcal{T}}^2(\mathbf{x}')d\mathbf{x}$$

where we denote $\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}_i; \mathbf{x})$. Let us consider $\mathbf{q}$ to be the closest point from $\mathbf{x}'$ located on $\mathcal{T}$. As $\phi_{\mathcal{T}}$ is assumed to be a Euclidean distance transform, it also satisfies the condition $[\, \|\nabla\phi_{\mathcal{T}}(\mathbf{x}')\| = 1]$. Therefore one can express the values of ϕ_i at the first order in the neighborhood of $\mathbf{x}'$ in the following manner :

$$\begin{aligned} \phi_{\mathcal{T}}(\mathbf{x}' + \delta\mathbf{x}') &= \phi_{\mathcal{T}}(\mathbf{x}') + \delta\mathbf{x}' \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}') + \circ(\delta\mathbf{x}') \\ &= (\mathbf{x}' + \delta\mathbf{x}' - \mathbf{q}) \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}') + \circ(\delta\mathbf{x}') \end{aligned}$$

This local expression of $\phi_{\mathcal{T}}$ with a dot product reflects the condition that a point to surface distance was adopted. The FFD transformation $\mathcal{L}(\mathbf{\Theta}, .)$ is linear with respect to the vector of parameters $\mathbf{\Theta}$, $\mathbf{x}'$ can be rewritten using a $[3 \times 3LMN]$ matrix:

$$\mathbf{x}' = \mathcal{L}(\mathbf{\Theta}, \mathbf{x}) = \mathbf{x} + \mathcal{X}(\mathbf{x})\mathbf{\Theta}$$

Under the assumption that E_0 is small when reaching the optimum, we can write the classical second order approximation of quadratic energy in the form:

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

Localizing the global minimum of an objective function E is equivalent to finding the major mode of a random variable with density $[\exp(-E_0/\beta)]$. The coefficient

β corresponds to the allowable variation in the energy value around the minimum. 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 $[\Sigma_{\mathbf{\Theta}_i}^{-1} = H_{\mathbf{\Theta}_i}/\beta]$. 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 one can claim that the matrix $H_{\mathbf{\Theta}}$ is not invertible because the registration problem is under-constrained. Then, additional constraints have to be introduced towards the estimation of the covariance matrix of $\mathbf{\Theta}_i$ through the use of an arbitrarily small positive parameter γ :

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

This leads to the covariance matrix for the parameter estimate :

$$\Sigma_{\mathbf{\Theta}} = \beta \left(\iint_{\mathcal{S}} \mathcal{X}(\mathbf{x})^T . \nabla\phi_{\mathcal{T}}(\mathbf{x}') \nabla\phi_{\mathcal{T}}(\mathbf{x}')^T \mathcal{X}(\mathbf{x}) d\mathbf{x} + \gamma \mathbf{I} \right)^{-1} \tag{3}$$

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. Further work will combine 3D surface morphing and 2D texture morphing, uncertainty may therefor be used to assess the time factor on texture mapping.

5 Surface Morphing & Facial Animations

Let us now consider a finite number of facial expressions, like {neutral, happy, sad, fear, anger, surprise, and disgust}. Such expressions have been produced from a number of subjects in front of a low precision binocular stereo system and have been associated with the corresponding geometric information through stereo reconstruction, and surface triangulation. These expressions are shown in [Fig.(5)], are mostly geometric and retain some minimal texture information.

Let S_{exp} be the 3D surface associated with a given expression. Then, using the proposed registration framework we have determined the transitions from one expression to the next in terms of surface representation; that is $\{\mathcal{L}(A, B)\}$, along with uncertainties estimates on the registration process $\{\mathcal{V}(A, B)\}$. Furthermore, information on the duration of the transition has been retained from one expression to the other. Towards animation, two aspects are to be addressed, (i) surface morphing, and (ii) texture morphing.

Surface morphing can be determined in a straightforward fashion from the $\{\mathcal{L}(A, B)\}$ that retains all geometric information needed to transform a given expression A to an expression B. However, a selection is to be made regarding the animation process since the transformation from A to B could be done using either a linear, or a geodesic, or the gradient descent path estimated in the process. The gradient descent path would be a non-uniform path that will first address important discrepancies between surfaces and then will focus on minimal details. Therefore from animation point of view such a strategy is rather unrealistic, should surfaces refer to complete connected entities. Geodesic paths seem to be more prominent direction however, their estimation is challenging and therefore we have decided to adopt a linear interpolation strategy that could be adapted to physiological factors if available (interval between two expressions, etc.). Some results of the surface morphing process are shown in [Fig. (1)].

In terms of texture mapping, what is available is the texture information on the source and the one of the target. Therefore, if we assume that surface correspondences have been properly established, then linear interpolation between the extreme texture values could provide a natural animation from the source to the target as shown in [Fig. (1)]. However, particular attention is to be paid on the distance between the transformed source and the target, an information that is encoded from the uncertainties estimates. In the time being, we investigate how such information can be used to produce more photo-realistic rendering.

5.1 Validation

While explicit validation is possible for the registration aspect of our approach that is not the case for the animation part. The Euclidean distance is used to determine

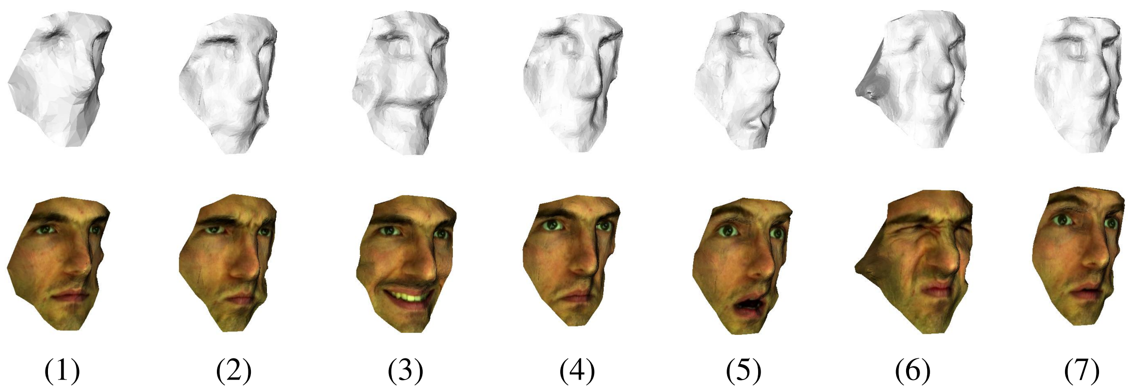


Figure 5: Expressions: (1)neutral, (2) anger, (3) joy, (4)sad, (5) surprise, (6) disgust, and (7) fear.

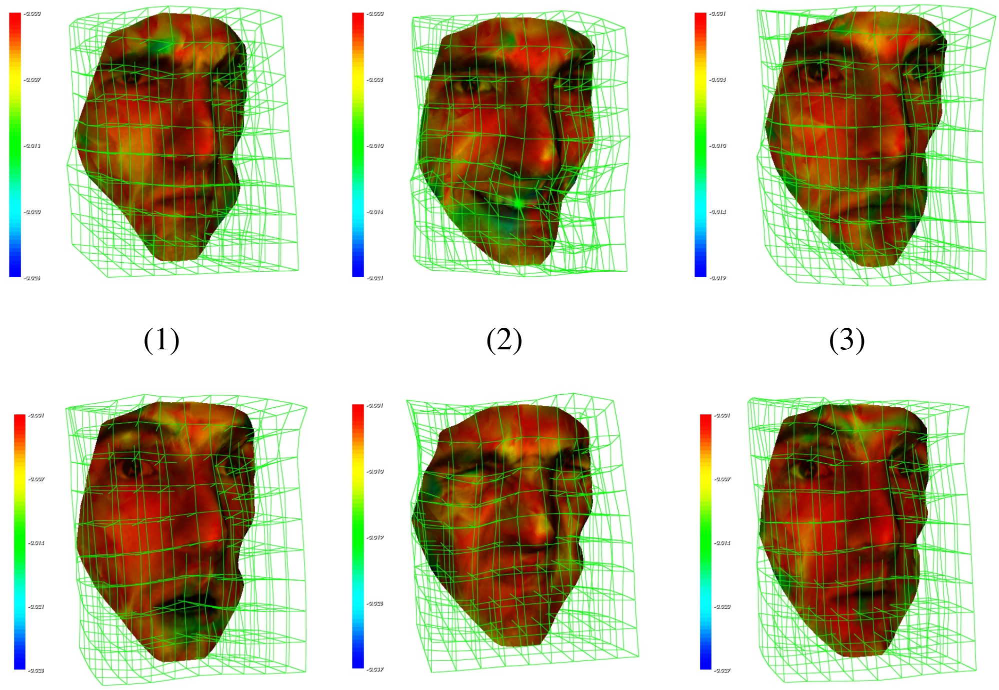
anger	-0.0050720386	1.4020472e-005
joy	-0.0042415187	8.7975668e-006
sadness	-0.0040221638	6.2809795e-006
surprise	-0.0045582722	9.6993954e-006
disgust	-0.0064263316	1.9230867e-005
fear	-0.0048255986	1.0473816e-005

Table 1: Estimation of mean and variance of the distance from the registered shape to the Target

the registration error, and is demonstrated in [Fig. (1)]. One can also see how such error is distributed along the surface for the 6 different expressions assuming the neutral expression to be the origin point. Furthermore, the mean value and the variance of this error is shown in [Tab. (5.1)]. In order to demonstrate the importance of estimating uncertainties, the same error is shown in [Tab. (5.1)] once weighted with the local uncertainties estimates.

6 Discussion

In this paper we have mostly addressed the problem of surface morphing through a free form deformation approach that also provides uncertainties of the registration process. Such a morphing was used to create transitions between facial expressions towards photo-realistic animations. 3D models of expressions were built from low resolution images, and were associated with texture using basic techniques. Registration between these models have provide continuous deformations fields capable of describing the geometric evolution of morphing process. Simple linear interpolation techniques as far texture is concerned were used to produce the intermediate state of the process leading to promising animation results [Fig.



(1) (2) (3)

(4) (5) (6)

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

(6)]. Such a framework can be integrated to the next generation communication systems since it can be adapted to the user expressions, requires limited resources and very low bandwidth for the transmission.

Numerous extensions of the method are under consideration. Introducing visual information on the registration process towards joint surface/visual 3D morphing is the most challenging direction. To this end, appropriate registration criteria that can account for such a multi-modal space are to be considered while preserving the ability of estimating uncertainties. Parallel to that, more precise reconstruction as well as recognition of facial expressions could further improve the performance of the method. To this end, we are willing to construct non-parametric densities that capture the geometry of difference expressions and then use the proposed registration framework with a classifier towards the most reliable expression. Similar concept can be also used in other domains like medical imaging, where 3D registration is of great importance and complete surface/appearance models could lead to better segmentation processes.

References

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

[2] T. Beier and S. Neely. Feature-based image Metamorphosis. In *SIGGRAPH*, pages 35–42, 1992.

[3] G. Borgefors. Distance Transformations in Digital Images. *Computer Vision, Graphics, and Image Processing*, 34:344–371, 1986.

[4] C. Chua and R. Jarvis. 3D Free Form Surface Registration and Object Recognition. *International Journal of Computer Vision*, pages 77–99, 1996.

[5] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[6] D. DeCarlo, D. Metaxas, and M. Stone. An Anthropometric Face Model Using Variational Techniques. In *SIGGRAPH*, pages 67–74, 1998.

[7] T. Ezzat, G. Geiger, and T. Pogio. Trainable Videorealistic Speech Animation. In *SIGGRAPH*, pages 388–398, 2002.

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

[9] O. Faugeras and R. Keriven. Complete Dense Stereovision Using Level Set Methods. In *European Conference on Computer Vision*, pages 379–393, 1998.

[10] S. Granger and X. Pennec. Multi-scale em-icp: A fast and robust approach for surface registration. In *European Conference on Computer Vision*, pages 418–432, 2002.

[11] X. Huang, N. Paragios, and D. Metaxas. Establishing Local Correspondences towards Compact Representations of Anatomical Structures. In *Medical Imaging Copmuting and Computer-Assisted Intervention*, 2003.

[12] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *European Conference on Computer Vision*, volume 3, pages 82–96, 2002.

[13] K. Kutulakos and S. Seitz. A Theory of Shape by Space Carving. *International Journal of Computer Vision*, 38:199–218, 2000.

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

[15] M. Naha and M. Santourens. Animation of a B-Spline Figure. *Visual Computer*, pages 272–276, 1988.

[16] N. Paragios, M. Rousson, and V. Ramesh. Non-Rigid Registration Using Distance Functions. *Computer Vision and Image Understanding*, 2003.

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

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

[19] T. Sederberg and S. Parry. Free-form deformation of solid geometric models. *Proceedings SIGGRAPH '86*, 20:151–160, 1986.

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

[21] D. Terzopoulos and D. Metaxas. Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):703–714, 1991.

[22] P. Viola and W. Wells. Aligment by Maximization of Mutual Information. In *IEEE International Conference in Computer Vision*, pages 16–23, 1995.

[23] 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.

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