

A Functional Spatio-Temporal Model for Geometric Shape Analysis

Lara Fontanella - Luigi Ippoliti¹ and Pasquale Valentini
Department of Quantitative Methods and Economic Theory

Abstract: A functional spatio-temporal model for shape is investigated for objects represented by landmark data. The model describes a time-varying deformation of the ambient space in which the objects of interest lie. By using bases functions defined by principal warps in space and time facilitates the model specification and the fitting of the data represented by Procrustes tangent coordinates. The fitted model can be interpreted either just in terms of the finite set of landmarks at the finite set of data times, or in terms of a deformation of space which varies continuously through time. The method is illustrated on a set of data representing 4 different facial expressions.

Keywords: Shape analysis, spatio-temporal models, procrustes tangent coordinates, principal warps, bending energy matrix, splines.

1. Introduction

Spatio-temporal modelling has largely been developed through applications in geostatistics, hydrology and meteorology. More recent activities in the area include environment monitoring, tracking, functional MRI, health data and facial analysis. Motivated by these applications, the field has adopted various modelling strategies, essentially depending on the underlying objective of the analysis and the scale and type of data.

In this paper we are interested in applications in shape analysis in which the data are represented by a set of landmarks, and the objective is to construct a spatio-temporal model which is able to describe the time-varying deformation of the shape space obtained from the landmark coordinates. For simplicity, we assume that the data take the form of a regular array in space-time. That is, we have one-dimensional observations

$$y_{ij}, i = 1, \dots, M, j = 1, \dots, T$$

at sites $\mathbf{x}_i \in \mathcal{R}^d$ and at times $t_j \in \mathcal{R}$. Typically, we consider objects in $d = 2$ dimensions but the methodology extends to higher dimensions. Thus, the objective is to model the data as

$$y_{ij} = z(\mathbf{x}_i, t_j) + e_{ij}, \tag{1}$$

where $\{z(\mathbf{x}, t)\}$ is a deterministic space-time process and e_{ij} an error term. We usually assume the error terms e_{ij} are independent, identically distributed. Once a smooth process $\hat{z}(\mathbf{x}, t)$ has been fitted, it can be used for interpolation.

¹Address of correspondence: University G. d'Annunzio, Chieti-Pescara, DMQTE, Viale Pindaro 42, 65127 Pescara - ITALY. e-mail: ippoliti@unich.it

A key-property of much spatio-temporal data is that observations at nearby sites and times will tend to be similar to one another. This underlying smoothness characteristic of the space-time process $z(\mathbf{x}, t)$ can be captured by using a finite-dimensional space of deterministic drift functions, or using autocorrelation to make nearby values similar. In principle, both of these approaches can be applied in space and/or time. For the type of data we are considering the former approach is preferred and thus the space-time model results from a tensor product of drift in times and drift in space. The outline of the paper is as follows. In Section 2 we describe how to represent the data in a form suitable for fitting the space-time model. In Section 3 we describe the modelling strategy. Specifically, roughness penalties for directions of variation in space and for rate of variation in time are constructed and combined to give a joint space-time penalty. Also, a transformation of the data is given which facilitates the fitting of these smoothing models. Section 4 discusses a parametric version of the model and illustrates its application on a set of landmarks representing the dynamic of four different facial expressions. Finally, Section 5 concludes the paper with a discussion.

2. The Data

Suppose landmark data are available on different individuals at a common set of times, taking the form of a 4-way array $\{v_{ikmj}\}$ where:

$l = 1, \dots, N$ labels different individuals; $k = 1, \dots, K$ labels different landmarks;
 $s = 1, \dots, d$ labels different coordinates; $j = 1, \dots, T$ labels different times $t_1, \dots, t_T$.

It is convenient to represent these data as a collection $\{\mathbf{v}_{lj}\}$ of $K \times d$ matrices. Then, since the shape of an object determines its coordinates only up to a similarity transformation, it is necessary to reduce the data to just the shape information. We do this using Procrustes tangent coordinates about a centered and scaled mean configuration $\boldsymbol{\mu}$, say (Dryden and Mardia, 1998). In general, a convenient choice for $\boldsymbol{\mu}$ is the Generalised Procrustes estimate based on all NT configurations, but the exact choice does not matter.

Let $\{\mathbf{y}_{lj}\}$ denote the $K \times d$ (centered not Helmertized) matrix of Procrustes tangent coordinates of the data $\{\mathbf{v}_{lj}\}$. Next, assuming all N individuals are i.i.d., we take a sample average of the Procrustes coordinates to get averaged data $\bar{\mathbf{y}}_j$. It is this form of the data to which we wish to fit our spatio-temporal model. For this purpose it is convenient to rewrite the data as a $M \times T$ matrix $\mathbf{Y}$, say, with $M = Kd$ and the j th column of $\mathbf{Y}$ defined by stacking the $d = 2$ columns of $\bar{\mathbf{y}}_j$ on top of one another.

3. The Drift-Drift model

In this section we describe our spatio-temporal model which involves drift functions in space $\mathbf{x}_i \in \mathcal{R}^d$ and time $t \in \mathcal{R}$. There are three ingredients in this model:

- (a) $\mathcal{F}$: a p -dimensional vector space of functions $\mathcal{R}^d \rightarrow \mathcal{R}$, specifying possible directions of shape variation. Let $\{f_\alpha(\mathbf{x}) : \alpha = 1, \dots, p\}$ denote a basis.

- (b) $\mathcal{G}$: a q -dimensional vector space of functions $\mathcal{R} \rightarrow \mathcal{R}$, specifying possible rates of variation in time. Let $\{g_\beta(t) : \beta = 1, \dots, q\}$ denote a basis.
- (c) r : a rank, $r > 0$ representing complexity of the model.

Both $\mathcal{F}$ and $\mathcal{G}$ will usually include the constant function to accommodate an intercept term. From these ingredients we form a deterministic spatio-temporal process of the form

$$z(\mathbf{x}, t) = \sum_{\alpha=1}^p \sum_{\beta=1}^q a_{\alpha\beta} f_\alpha(\mathbf{x}) g_\beta(t),$$

where the $p \times q$ matrix of coefficients $\mathbf{A} = (a_{\alpha\beta})$ has rank r . Let the matrix $\mathbf{F}$, of dimension $N \times p$ denote the values of the spatial basis functions at sites (landmarks) $\mathbf{x}_i$, $i = 1, \dots, N$. Similarly, let the matrix $\mathbf{G}$, of dimension $T \times q$ denote the values of the temporal basis functions at the times t_j , $j = 1, \dots, T$. Thus, the model (1) takes the form

$$\mathbf{Y} = \mathbf{FAG}' + \mathbf{E}, \tag{2}$$

where $\mathbf{Y} = (y_{ij})$ and $\mathbf{E} = (e_{ij})$. If the basis functions are chosen so that $\mathbf{F}$ and $\mathbf{G}$ have orthonormal columns, then $\mathbf{A}$ can be estimated using the dominant r components in a singular value decomposition of $\mathbf{F}'\mathbf{Y}\mathbf{G}$ (Kent et al., 2001), though if an intercept term is separated out estimation can be a bit more involved.

3.1 Choices of Drift Functions

There are various choices for the space of drift functions and some examples include:

- (i) *Polynomials*. Low order polynomials form the simplest choice. These are very simple functions and do not depend on the choice of sites (for $\mathcal{F}$) or on the choice of times (for $\mathcal{G}$). However, especially for high order, they have the disadvantage to be rather wild in their oscillatory behaviour and grow rapidly as $|\mathbf{x}| \rightarrow \infty$ or $t \rightarrow \infty$.
- (ii) *Trigonometric functions*. Fourier expansions can be very useful for data which are periodic in space.
- (iii) *Principal Kriging Functions or Principal Splines*. These are sets of functions in space or time adapted to the locations of the sites or times, respectively, and are arranged from coarse scale to fine scale in terms of variation (Bookstein, 1991; Mardia et al., 1998). These functions have some advantages over polynomials. First they grow less quickly than polynomials outside the domain of the data and so, in general, are more stable for extrapolation. Second, they are adaptive to the given arrangement of times or sites, so that they can represent more detailed behaviour where the times or sites are most dense.

Next, we describe how to construct and use the principal kriging functions in our model. For simplicity we largely focus on functions of time but similar a construction also holds in space. Essentially, there are two main components in the construction. First, is a vector space of functions $\mathcal{G}_0$ of dimension $q_0 \geq 0$ called the "null space", and which will form a subspace of $\mathcal{G}$. Second, is a "potential" function $\tau(t)$, say, which is conditionally positive

definite with respect to $\mathcal{G}_0$. That is, for all distinct times $t^{(h)}$, $h = 1, \dots, h_0$, and all vectors of coefficients $\delta = (\delta_h, h = 1, \dots, h_0) \neq 0$,

$$\sum_{h_1, h_2=1}^{h_0} \delta_{h_1} \delta_{h_2} \tau(t^{(h_1)} - t^{(h_2)}) > 0 \quad (3)$$

whenever $\sum_{h=1}^{h_0} \delta_h g(t^{(h)}) = 0$ for all $g \in \mathcal{G}_0$. Given the data times $t_j, j = 1, \dots, T$, define the $T \times T$ matrix, $\mathbf{P}$, by $\mathbf{P} = (\tau(t_i - t_j))$. Next, define an $T \times q_0$ "drift" matrix $\mathbf{U}$ by $u_{j\beta} = g_{0\beta}(t_j)$, where $\{g_{0\beta}(t_j), \beta = 1, \dots, q_0\}$ is a basis of functions in $\mathcal{G}_0$. It is assumed that the data times are suitably spaced so that this matrix is of full rank q_0 . These matrices can be combined into an $(T + q_0) \times (T + q_0)$ matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{P} & \mathbf{U} \\ \mathbf{U}' & \mathbf{0} \end{bmatrix}$$

with partitioned inverse

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{L} \end{bmatrix}, \text{ say,}$$

which, in particular, defines the matrices $\mathbf{B}$ and $\mathbf{C}$. Denote the eigenvectors of $\mathbf{B}$ by $\xi_k, k = 1, \dots, K$ with the eigenvalues written in nondecreasing order. The first q_0 eigenvalues of $\mathbf{B}$ are 0 and the corresponding eigenvectors are given by the q_0 columns of $\mathbf{U}$. Finally, let $\tau_0(t)$ denote the vector function of t with the j th element $\tau_0(t)_j = \tau(t - t_j), j = 1, \dots, T$, and let $\mathbf{u}_0(t)$ denote the vector function of t with β th component $g_{0\beta}(t), \beta = 1, \dots, q_0$. The k th principal kriging function is defined by

$$g_k^{PKF}(t) = \xi_k' [\mathbf{B}\tau_0(t) + \mathbf{C}\mathbf{u}_0(t)].$$

It can be shown that $g_k^{PKF}(t)$ is an interpolating function with $g_k^{PKF}(t_j) = (\xi_k)_j$. The vector space $\mathcal{G}$ is defined to be the span of $g_\beta^{PKF}(t), 1 \leq \beta \leq q$, where q is a specified dimension, $q_0 \leq q \leq T$. For further details see, for example, Mardia et al. (1996).

Different possible choices for $\tau(t)$ are available; in general, any valid covariance function for which any null space of functions $\mathcal{G}_0$ will suffice is a possible choice. In this paper, we consider $\tau(t) = |t|^3$, which is conditionally positive definite whenever $\mathcal{G}_0$ contains linear functions, $\text{span}(1, t)$. In this case, it turns out that the k th principal kriging function is an interpolating cubic spline which minimises the penalty function

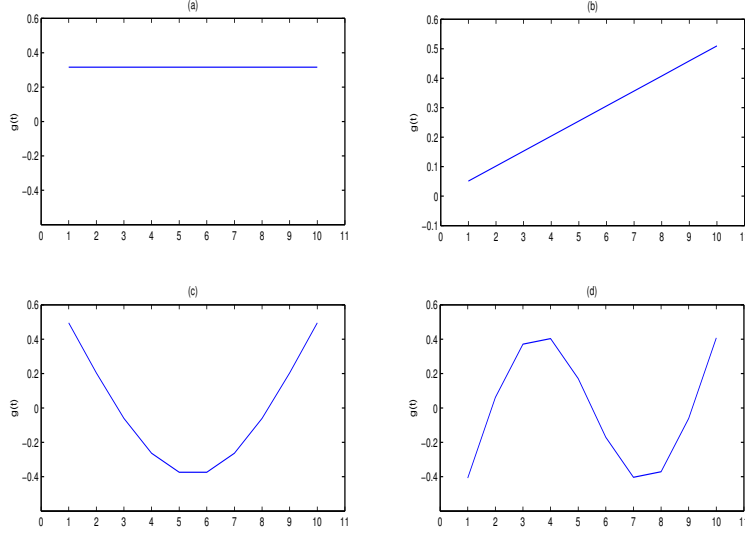
$$\Phi(g) = \int_{-\infty}^{\infty} \left(\frac{d^2 g(t)}{dt^2} \right)^2 dt,$$

subject to the constraints $g_k(t_j) = (\xi_k)_j, j = 1, \dots, T$. This example motivates the alternative name "principal spline" for a principal kriging function. As an example, a plot of some principal splines for $T = 10$ equally spaced time points is given in Figure 1.

Note how the functions appear to mimic the qualitative behaviour of the successive polynomials in t . However, also note that the principal splines grow only linearly as $|t| \rightarrow \infty$, more slowly than the quadratic and higher order polynomials.

Principal splines are also easily dealt with for two-dimensional data at landmarks $\mathbf{x}_i, i = 1, \dots, N$. In this case, it is common to use the thin-plate spline penalty

Figure 1: First four principal cubic splines for $T=10$ equally spaced time points: (a) and (b) are the constant and linear functions corresponding to the null space $\mathcal{G}_0$; (c) and (d) are analogous to a quadratic and cubic function.



$$\Phi(f) = \int \left\{ \left(\frac{\partial^2 f}{\partial x[1]^2} \right)^2 + \left(\frac{\partial^2 f}{\partial x[1]x[2]} \right)^2 + \left(\frac{\partial^2 f}{\partial x[2]^2} \right)^2 \right\} d\mathbf{x},$$

where the integral is over $\mathcal{R}^2$ and $\mathbf{x} = (x[1], x[2])$ denotes the components of $\mathbf{x}$. The above construction of principal splines can be carried out with little change in this setting as well. Replace the potential function by $\tau(\mathbf{x}) = |\mathbf{x}|^2 \log |\mathbf{x}|$, $\mathbf{x} \in \mathcal{R}^2$, and the null space by $\mathcal{F}_0 = \text{span}(1, x[1], x[2])$. Principal splines in $\mathcal{R}^2$ were introduced by Bookstein (1989) in the context of deformations in shape analysis, and he used the term "principal warps". The matrix $\mathbf{B}$ instead, was called the "bending energy matrix".

3.2 Model Details

In this section we describe how to obtain the matrices $\mathbf{F}$ and $\mathbf{G}$ in equation (2). Let $t_1, \dots, t_T$ denote the common set of times at which the data are observed. Associated with these times is the $T \times T$ symmetric positive semidefinite bending energy matrix $\mathbf{B}$ of rank $T - 2$. For our modelling purposes it is useful to combine all the eigenvectors (principle warp vectors) of $\mathbf{B}$ (excluding the constant vector) into the $T \times (T - 1)$ column orthonormal matrix $\mathbf{G}$. Denote the corresponding eigenvalues of $\mathbf{B}$ by β_j , $j = 1, \dots, T - 1$, with $\beta_1 = 0$, arranged in nondecreasing order. The columns of $\mathbf{G}$ roughly correspond to the effects of orthogonal polynomials at the data times.

A similar construction can be carried out in space for the $\mathbf{F}$ matrix. In particular, let $\{\mathbf{x}_1, \dots, \mathbf{x}_K\}$ denote a fixed collection of landmarks in $\mathcal{R}^2$. In our application we shall use the K rows of $\boldsymbol{\mu}$. Then, a $K \times K$ bending energy matrix can be constructed with analogous properties to the above setting, though this time of rank $K - 3$. Then, since a deformation can be constructed from a pair of functions $\mathcal{R}^2 \rightarrow \mathcal{R}$, we combine together two copies of the eigenvectors. After removing 4 degrees of freedom for the constraints in

Procrustes tangent space, we are left with the $2K \times (2K - 4)$ column orthonormal matrix $\mathbf{F}$. The first two columns of $\mathbf{F}$ represent linear functions orthogonal to the similarity transformations - i.e. the Bookstein's uniform components; see Bookstein (1995) and Mardia (1995, Section 7). The remaining columns of $\mathbf{F}$ come in pairs

$$\begin{bmatrix} \boldsymbol{\xi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\xi} \end{bmatrix},$$

where $\boldsymbol{\xi}$ is the $K \times 1$ eigenvector of the bending energy matrix for the thin-plate spline. Also, let $\alpha_{2j-1} = \alpha_{2j}$, $j = 1, \dots, K - 2$ denote the corresponding eigenvalues in nondecreasing order, each listed twice. Note $\alpha_1 = \alpha_2 = 0$.

Thus, the $(2K - 4) \times (T - 1)$ matrix of coefficients, $\mathbf{A} = \{a_{ij}\}$, in equation (2) is obtained as $\mathbf{A} = \mathbf{F}'\mathbf{Y}\mathbf{G}$. As shown in Section 4, this matrix is of key importance for specifying different models.

Finally, if the eigenvectors in $\mathbf{F}$ and $\mathbf{G}$ are interpolated to yield a pair of space-time functions $(\Psi_1(x[1], t), \Psi(x[2], t))$ then the penalty function

$$\Phi(\Psi_1, \Psi_2) = \sum_{i=1}^2 \int \left\{ \left(\frac{\partial^4 \Psi_i}{\partial x[1]^2 \partial t^2} \right)^2 + 2 \left(\frac{\partial^4 \Psi_i}{\partial x[1] \partial x[2] \partial t^2} \right)^2 + \left(\frac{\partial^4 \Psi_i}{\partial x[2]^2 \partial t^2} \right)^2 \right\} d\mathbf{x} dt,$$

where the integral is over $\mathbf{x} \in \mathcal{R}^2$, $t \in \mathcal{R}$, reduces to

$$\sum_{i=1}^{2K-4} \sum_{j=1}^{T-1} a_{ij}^2 \alpha_i \beta_j,$$

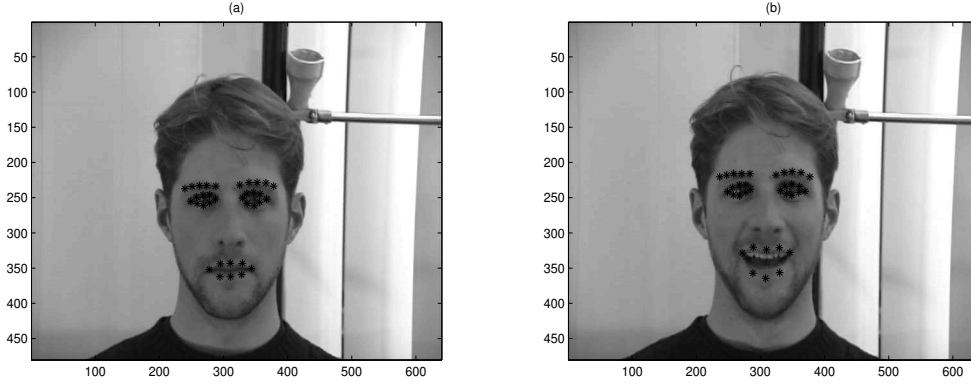
which depends just on the coefficients and the eigenvalues.

4. Modelling Facial Expressions

In this section we consider the FG-NET Database with Facial Expressions and Emotions from the Technical University Munich. This is an image database containing face images showing a number of subjects performing the six different basic emotions defined by Eckman and Friesen (1971). Specifically, the database contains material gathered from 18 different individuals, each performing all six desired actions three times. Additionally three sequences doing no expressions at all are recorded. All together this gives an amount of 399 sequences. Depending on the kind of emotion, a single recorded sequence can take up to several seconds. From each recorded sequence we have extracted 5 frames summarising the dynamic of the expression and on each of them we have manually placed a set of 34 landmarks. For our modelling purposes we have considered here four different expressions, namely: disgust, happiness, sadness and surprise. There is thus complete information on $N = 18$ subjects at $T = 5$ times on $K = 34$ landmarks. As an example Figure 2 shows, for a specific subject, two frames representing the first (t_1 - neutral) and last (t_7 - surprise) frames. For the purposes of this paper we shall ignore any differences between the individual subjects. Furthermore, we also ignore changes in size and limit attention only to changes in the shape of the object.

For this data set, let $\mathbf{A} = \mathbf{F}'\mathbf{Y}\mathbf{G}$ denote the 64×4 doubly rotated version of the Procrustes tangent matrix for the mean data. The rows of $\mathbf{A}$ label the eigenvectors in space and the

Figure 2: Frames at point times t_1 and t_7 . The frames represent: (a) neutral and (b) surprise expressions respectively.



columns label the eigenvectors in time. The mean shape μ is obtained as the generalized Procrustes estimate based on the $T = 5$ configurations of the neutral expression such that the model provides insights on its deformation.

Some important models and their application to the expression data are described below and a selection of fitted models is plotted in Figures 3-6 where each "o" represents a landmark of the the Procrustes mean shape μ , and each closed circle represents the position of a landmark at the initial time.

1 Full Rank. These models can be specified in terms of the nonzero entries in the estimated $\mathbf{A}$. For example, the notation $[1 : J, 1 : L]$ will mean that the block of entries specified by the first J rows and first L columns will be allowed to be nonzero.

$\text{FULL} = [1:64, 1:4]$. In this case there is no data reduction (other than the averaging over individuals in the first place). For this model, the residual sum of squares (RSS) is zero. Figure 3-6(a).

$\text{LIN-LIN} = [1:2, 1]$. This is the simplest possible model, linear in space (a two-dimensional subspace) and linear in time (one-dimensional). Thus the this model consists only of a constant variation in a single direction arising from a linear transformation in two dimensions. This model captures some of the main features of the data, but fails to capture the curvature of the paths, and does not fit well in general at several landmarks.

$\text{FULL-LIN} = [1:64, 1]$. In this case there is full flexibility in the spatial variation but the variation in time is still linear. In Figure 3-6(b) we see that the main pattern of variation is captured, but the curvature of the paths is not captured at all.

$\text{SPEC} = [1:64, 1] + [1:2, 2]$. This model is called special because the nonzero parameter values do not form a rectangular block. It captures the curvature in the paths through a term which is second order in time and linear in space. Figure 3-6(c) shows that this model yields quite a good fit to the data.

LIN-FULL = [1:2, 1:4]. This model allows variation in the two linear directions in space, but at arbitrary rates in time. The paths are curved but do not match the data very well.

NULL = [1:64, 1] + [1:2, 2:4] FULLLIN + LINFULL. This is the most general model that has 0 roughness penalty and fits the data surprisingly well. Visually the fit is similar to the special model.

- 2 *Reduced Rank*. Start with a rectangular block of coefficients in A , take a singular value decomposition, and retain the dominant components. The simplest such model which fits the data well can be written as FULL.2 = [1:64, 1:4, 2]; i.e. take the dominant 2 components of the singular value decomposition of the whole matrix A . This fitted model is very closely related to the exploratory analysis based on two principal components in Le and Kume (2000).

Figure 3 shows results for some of the fitted models for the disgust expression. We notice that this expression is mainly characterized by mouth and eyebrow movements. Specifically, we observe that the eyebrows are lowered and drawn together while the lower eyelids cheeks and the upper lip are raised.

Figure 4 shows results for some of the fitted models for the happiness expression. Here, we observe that the deformation is mainly characterized by the movements of the mouth since the eyes move only slightly. Specifically, we observe a slight narrowing of the eyelids and a raising of the lip corners describing an upward curving of mouth and expansion on vertical and horizontal direction.

Figure 5 shows results for some of the fitted models for the sadness expression. In this case, as expected, we notice that the features for this expression are represented by narrowed eyes, eyebrows brought together and a down-turned mouth.

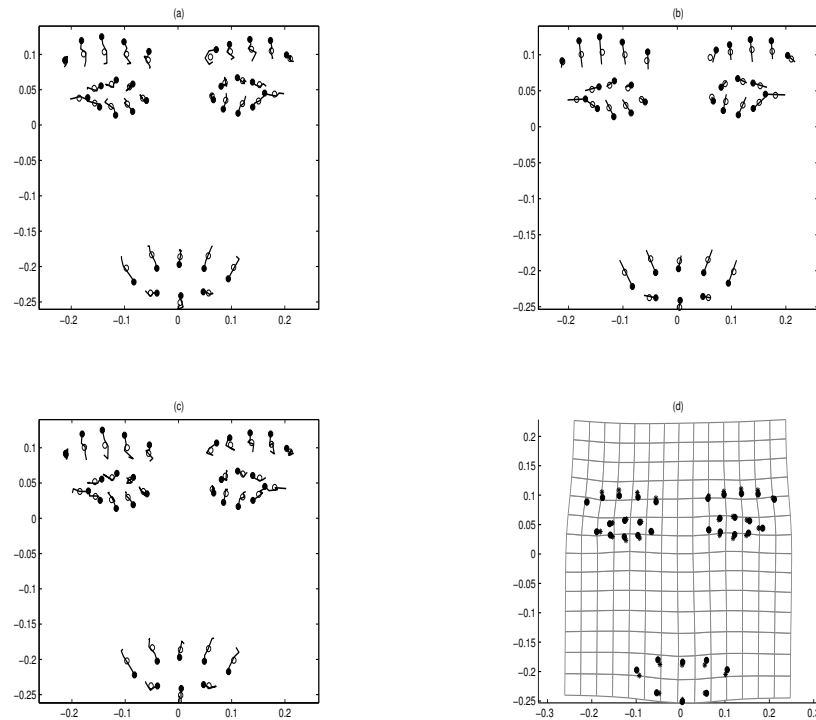
Finally, Figure 6 shows results for some of the fitted models for the surprise expression. In this case, we observe that this expression appears with a vertical expansion of the mouth, widened eyes and slightly raised eyelids and eyebrows. However, an asymmetry is surprisingly observed between the movement of the upper eyelids for the two eyes.

5. Discussion

A glance at the matrix A shows that the largest values fall in the first column and the first two entries in the second column. This observation suggests that the "special" parametric model will provide a good-fitting parsimonious model, as verified by a small RSS value. Further, the reduced rank model "full.2" also captures these effects and yields a similar fit. Overall the most useful fitted model seems to be the "special" one. The variation can be decomposed into two components: a linear (in time) variation in a general spatial direction, together with a quadratic component (in time) which is restricted to a linear transformation in space. An illustration of the deformation involved between the starting and final times under this model is plotted (with no expansion factor) in Figure 3-6(d).

Finally, notice that for simplicity of presentation, we have only focused on two specific examples of roughness penalties, though other choices are also possible. Then, a non-parametric spline-like model can be defined by minimizing a combination of a goodness of fit term plus λ times the penalty term, where the minimization is over the finite di-

Figure 3: Fitted models for the FGNET data (Disgust): (a) full, (b) full-lin, (c) spec, with the growth patterns blown up by a factor of 3 for clarity. Part (d) shows the grid deformation, without an expansion factor, between the initial and final times for the spec model.

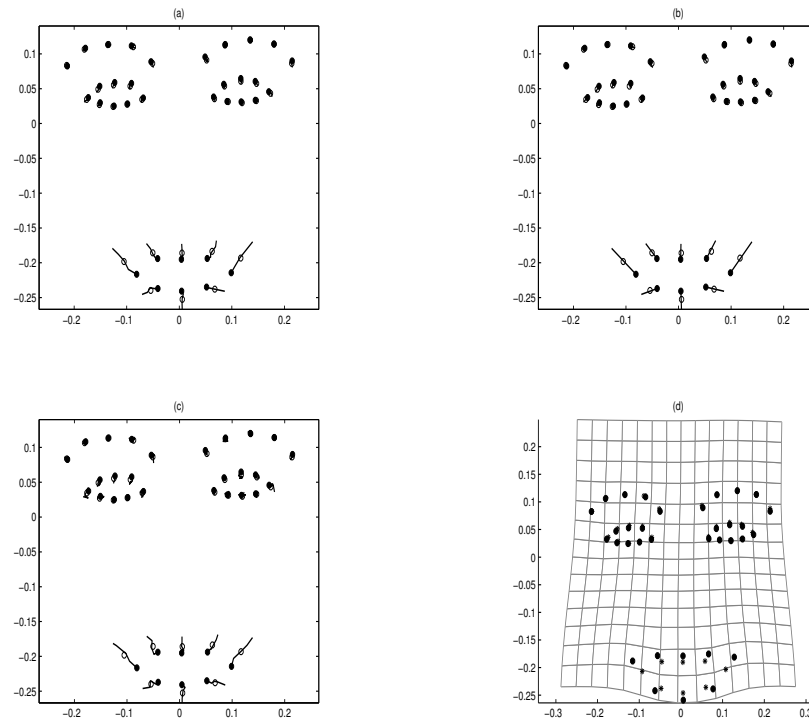


mensional vector space generated by products of the principal warp functions (including linear functions) for space and time, respectively (Kent et al., 2001).

References

- Bookstein, F.L. (1989). Principle Warps: thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 460-468.
- Bookstein, F.L. (1991) *Morphometric Tools for Landmark Data*. Cambridge University Press.
- Bookstein, F.L. (1995) A standard formula for the uniform shape component in landmark data. *NATO Proceedings*.
- Dryden, I.L. and Mardia, K.V. (1998) *Statistical Shape Analysis*. Wiley, Chichester.
- Ekman P. and Friesen W. (1971) Constants across cultures in the face and emotions, *Journal of Personality and Social Psychology*, 17, 124-129.
- Kent J.T., Mardia K.V., Morris R.J. and Aykroyd R.G. (2001) Functional models of

Figure 4: Fitted models for the FGNET data (Happiness): (a) full, (b) full-lin, (c) spec, with the growth patterns blown up by a factor of 5 for clarity. Part (d) shows the grid deformation, without an expansion factor, between the initial and final times for the spec model.



growth for landmark data. In *Proceedings in Functional and Spatial Data Analysis*. K.V. Mardia, J.T. Kent and R.G. Aykroyd (Eds). Leeds University Press. ISBN 0-85316-220-4.

Le, H. and Kume, A. (2000) Detection of shape changes in biological features. *J. Microscopy*, 200, 140147.

Mardia, K.V. (1995) Shape advances and future perspectives. In *Proceedings in Current Issues in Statistical Shape Analysis*, 57-75. K. V.Mardia and C. A. Gill (Eds). Leeds University Press.

Mardia K.V., Kent, J. T., Goodall C.R., Little, J.A. (1996). Kriging and spline with derivative information. *Biometrika*, Vol. **83**, No.1, 217-285.

Mardia K.V., Goodall C., Redfern E.J., and Alonso F.J. (1998) The Kriged Kalman filter (with discussion). *Test*, **7**, 217-252.

Figure 5: *Fitted models for the FGNET data (Sadness): (a) full, (b) full-lin, (c) spec, with the growth patterns blown up by a factor of 6 for clarity. Part (d) shows the grid deformation, without an expansion factor, between the initial and final times for the spec model.*

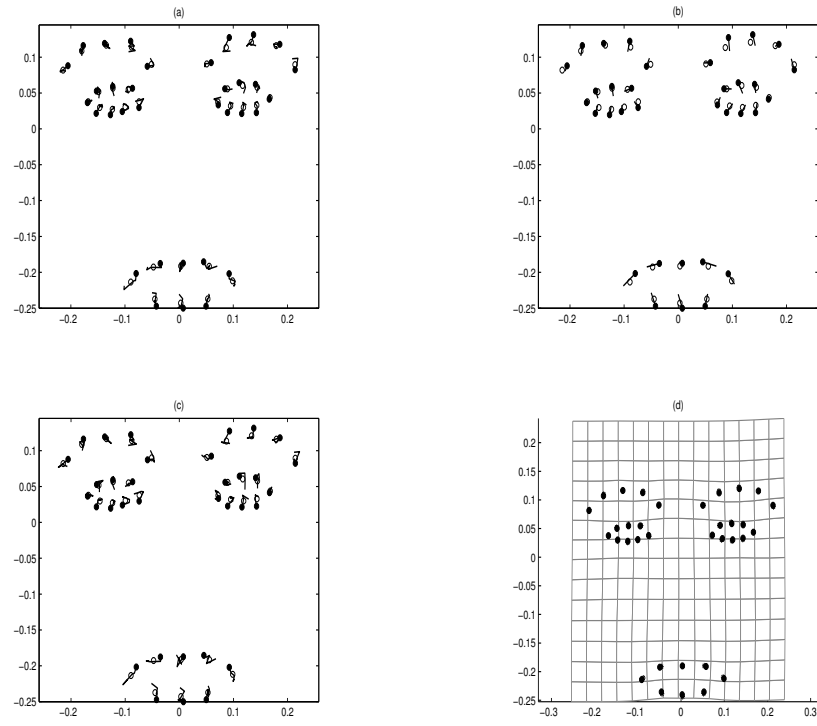


Figure 6: *Fitted models for the FGNET data (Surprise): (a) full, (b) full-lin, (c) spec, with the growth patterns blown up by a factor of 3 for clarity. Part (d) shows the grid deformation, without an expansion factor, between the initial and final times for the spec model.*

