Author queries:

Q1: OK as edited here? Or “annotator’s”?
Q2: Unnecessary line break?
Q3: Unnecessary line break?
Q4: Unnecessary line break?
Q5: Unnecessary line break?
Q6: Unnecessary line break?
Q7: Is the phrase “is sensible” okay? Would “is sensitive to” be preferable?
Q8: Arxiv URL okay?

SIAM J. MATH. DATA SCI.
Vol. 0, No. 0, pp. 1–26

Probabilistic Registration for Gaussian Process Three-Dimensional Shape Modelling in the Presence of Extensive Missing Data*

Filipa M. Valdeira†, Ricardo Ferreira‡, Alessandra Micheletti†, and Cláudia Soares§

Abstract. We propose a shape fitting/registration method based on a Gaussian processes formulation, suitable for shapes with extensive regions of missing data. Gaussian processes are a proven powerful tool, as they provide a unified setting for shape modelling and fitting. While the existing methods in this area prove to work well for the general case of the human head, when looking at more detailed and deformed data, with a high prevalence of missing data, such as the ears, the results are not satisfactory. In order to overcome this, we formulate the shape fitting problem as a multiannotator Gaussian process regression and establish a parallel with the standard probabilistic registration. The achieved method, the shape fitting Gaussian process (or SFGP), shows better performance when dealing with extensive areas of missing data when compared to a state-of-the-art registration method and current approaches for registration with GP. Experiments are conducted both for a two-dimensional small dataset with several transformations and a three-dimensional dataset of ears.

Key words. Gaussian processes, shape modelling, registration, variational Bayes

MSC codes. 60G15, 62F15, 62G08

DOI. 10.1137/22M1495494

1. Introduction.

Consider the problem of predicting a complex shape, like a human ear, from a dataset of similar point clouds and partially observed points of this shape. In this setting, we face both the registration and modelling of different point clouds. There are increasing areas of application for three-dimensional (3D) shape modelling, in particular when it comes to the human body, spread over medical applications (segmentation [15], prosthesis design [42], surgical planning [10]), surveillance (face recognition [34], tracking [21]), or human-machine interaction (expression/emotion detection [38], virtual humans [37]). A large number of approaches focuses on the human head, with particular incidence on the face region, and increasingly better models have been achieved for coarse-grained requirements [6, 36].

However, small and detailed areas are still not accurately represented and attempts to tackle this problem are currently emerging [35]. Our driving example is the modelling of an ear, representative of a fine-detailed region with extensive data problems. Given the challenging

* Received by the editors May 12, 2022; accepted for publication (in revised form) February 24, 2023; published electronically DATE.
https://doi.org/10.1137/22M1495494
Funding: This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Project BIGMATH, grant 812912.
† Department of Environmental Science and Policy, Universitá degli Studi di Milano, 20133 Milan, Italy (filipa.marreiros@unimi.it, Alessandra.Micheletti@unimi.it).
‡ μRoboptics, Lisbon, Portugal (ricardo.ferreira@roboptics.pt).
§ NOVA LINCS, Computer Science Department, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal (claudia.soares@fct.unl.pt).

shape of this face segment, the 3D scan procedure leads to broad regions of missing data and a high level of noise. Our approach is nonetheless generic enough to be applied to any other shape, as seen in Figure 5 of section 5 with the 2D fish data.

A standard approach to obtain a statistical shape model from a given dataset is through 3D morphable models (3DMM), first proposed in [4]. Although they include both a shape and an appearance component, here the focus is only on the former. Given a set of scans, the standard procedure to obtain a 3DMM entails two main steps: *dense correspondence* and *modelling*. During the former, the original samples of unorganized point clouds are set into correspondence, i.e., a reparametrization is found such that points with the same index have the same semantic meaning (for instance, point i of each scan represents the tip of the nose). This is a requirement for the subsequent *modelling* step where the deformations of each sample are studied, in order to express shape variability in a lower dimension. Correspondence across a dataset can be achieved by deforming a generic reference shape to each target scan, without prior information on the shape characteristics. This approach is a subclass of correspondence methods denoted as *registration* [40] and the standard choice for shape modelling pipelines [4, 34]. Given that no shape prior is taken into account, registration is by itself a broad area of study, developed separately from the modelling setting.

Therefore, traditional approaches to 3DMM directly employ a state-of-the-art registration method, followed by principal component analysis, to find a low-dimensional representation of shape variability [4, 34]. Further developments on the modelling side have included increasing the size of the training set [6] for higher variability, using part-based models [41, 7] instead of global ones or the use of different shape spaces [20, 33]. Nonetheless, and particularly for shapes with nonrigid deformations, correspondence remains an open challenge and a limiting step on the quality of the resulting model [22]. This has motivated proposals on dense correspondence specifically tailored for 3D faces in recent years [11, 13, 12].

However, regardless of the registration method in place, an underlying model restricts deformations of the reference shape, such that infeasible shapes are avoided. In fact, the process of registration can be understood, on a higher level of abstraction, as *model fitting* [32, 19], i.e., deforming the reference shape according to a preexisting model to resemble a given sample. The main difference between these two concepts is that restrictions to deformations are enforced by data (or specific shape characteristics) in model fitting and by a regularization parameter (independent of shape) in registration. Despite this close relationship, they are distinct branches of the literature since, initially, the raw data is not registered and correspondence must be achieved without access to a previous model (which can only be obtained upon registration). This originates complex pipelines, where different models and assumptions on the shape deformations are sequentially considered, for increasing refinement. A unified framework is desirable to achieve more principled approaches, as suggested in [25].

We shall now formalize the previously stated conceptual problem and introduce the unified framework, as well as the motivation for its use in our setting. We consider a *reference* $R = \{r_1, \dots, r_{N_R}\} \subset \mathbb{R}^d$ and a *target shape* $S = \{s_1, \dots, s_{N_S}\} \subset \mathbb{R}^d$, with a possibly different number of points N_R and N_S, respectively. The former is a representative example of the shape being studied, as close as possible to any other shape, without missing data, outliers, or noise, while the latter is any sample of the dataset. In previous work [42] with ear data scans, we have shown that most of the effects of translations and rotations can be previously

removed, so it is here considered that the target shape and reference were already preprocessed. Additionally, scaling is kept in the model, as it represents differences in ear size. Consequently, only nonrigid deformations between the reference and a shape are explicitly modelled.

Under these assumptions, any shape is obtained from a reference through a reparametrization and a set of nonrigid deformations θ. In the discrete setting, a reparametrization can be represented by a permutation matrix P. Thus, the problem is formulated as finding the optimal deformations between S and R

$$P^*, \theta^* \in \underset{P,\theta}{\operatorname{argmin}}\, d(\mathcal{T}(PR, \theta), S), \tag{1.1}$$

where $d(R, S)$ is a measure of dissimilarity quantifying the differences between two shapes and $\mathcal{T}(PR, \theta)$ is the transformed reference after application of the deformations θ and permutation P.

The solution of (1.1) provides a deformed reference whose points are in correspondence with the target shape. However, as stated above, this problem is generally not tackled in a unified way. On the registration side, a generic constraint is applied to the deformations, while P is retrieved, disregarding any knowledge of the particular shape. On the fitting side, correspondence is assumed (for example, by previously applying a registration method) or retrieved in a trivial manner (e.g., by taking the closest point). Both of these options entail that the correspondence is obtained with a different model $\mathcal{T}(PR, \theta)$ than the fitting. Looking at the two processes in the same setting may be beneficial as it allows both of them to benefit from additional information. In [25], the authors propose to formulate the unified framework through the use of Gaussian processes (GP). The core idea is that by designing an appropriate kernel, prior beliefs about the shape structure can be incorporated independently from the registration algorithm, allowing the shape fitting and registration to leverage the same model.

This framework has proven successful and has led to increasingly improved models of the human head [4, 36]. However, developments have been made on the modelling side by increasing the size of training datasets [6] or improving model combination [36], i.e., how to bring together models originating from different datasets. By applying the GP modelling to 3D ear point clouds, we have observed that indeed this is the most promising setting [42], but it calls for an improved outlook on the registration procedure. When models of the full head are considered, ears are a small detail that is often overlooked or disregarded [4, 13], so a straightforward correspondence method is enough to provide acceptable accuracy, explaining the limited advances on this area. It is worth noticing that a recent approach within the GP framework extends the head model to include ears [35]; however, it requires the identification of 50 manual landmarks for registration of the ear region. Besides, we approach the problem under a different perspective, where the ear scans are first reconstructed and the model is subsequently retrieved, while in [35] a preexisting head model is augmented with an additional ear model.

On the other hand, the registration field has seen considerable improvements in recent years, and state-of-the-art methods are capable of dealing with increasingly larger ratios of outliers, noise, and missing data [45, 28]. Probabilistic approaches (a subclass of registration methods) have proven to be particularly suitable to this scenario. However, being decoupled from the modelling setting, they do not allow for extensive prior information regarding the

shapes. This knowledge is usually limited to the expected overall magnitude of deformations applied to the reference, controlled by a set of algorithm parameters. Instead, modelling with kernels allows for the inclusion of additional prior beliefs in a straightforward and convenient manner. Upon the development of a fitting method with a generic kernel, the replacement with a tailored one is immediate and does not call for alterations to the method itself.

Multiannotation is a popular concept in the machine learning setting, where a given data point is labelled by different sources (annotators), i.e., there is not a unique and true label assignment, but several possible ones. By modelling registration as a GP multiannotator problem, we show that it is possible to perform probabilistic registration completely within the GP framework. This unified approach benefits both from a complex prior through the kernels and from the nice properties of probabilistic assignment, particularly when dealing with outliers and noise.

1.1. Related work. Throughout this section, we review closely related work under the two main areas of interest: registration within the GP framework and probabilistic approaches to the generic registration problem.

1.1.1. Registration within the GP framework.

Nonrigid registration with iterative closest point. A proposed approach to tackle the problem registration problem/model fitting for the GP framework is a nonrigid application of the Iterative closest point (ICP) [3], where the transformation part is obtained through Gaussian process regression (GPR). This means that to each point in the reference we attribute the closest target point, based on their Euclidean distance. These correspondences are then taken as observations and GPR is used to compute deformations for the entire shape (the mean of the posterior is the reference used in the next iteration). Our approach relates to this method, in the formulation of the problem but not in the way the correspondences are retrieved. For ear shapes, given the large regions of missing data and the highly nonrigid deformations, the closest point approach leads to undesirable results [42], particularly on the bottom region, where the missing points from the posterior section cause the deformed reference to collapse.

Registration as an optimization problem. In another proposed approach in [25], the authors formulate the registration problem/model fitting for both surfaces and images. Here, we focus on the surface formulation. First, the authors do a low-rank approximation, obtaining a parametric approximation of the original kernel. The problem is then posed as a maximum a posteriori (MAP) estimation problem, where the likelihood expresses some distance measure between the target and reference shapes, and the prior is given by the GP. The authors chose the mean squared Euclidean distance from the reference to the closest target point and solve the problem with an L-BFGS optimizer [23]. Therefore, the restriction of hard-assignment when choosing correspondences is maintained in this approach, leading to similar problems as faced in the previous one.

1.1.2. Probabilistic registration. The previous approaches imply a deterministic attribution of correspondences between points, while a soft-assignment may improve robustness to noise and outliers [45].

This leads, in the registration area, to probabilistic registration methods, of which the most used and representative is the coherent point drift (CPD) [31], which considers the alignment

of two sets as a probability density estimation problem. This approach takes R (the reference) as a set of centroids coming from a Gaussian mixture model (GMM) and S (any shape) as points generated by the centroids. An important detail is that the centroids are forced to move coherently as a group, thus preserving the topological structure of the points (motion coherence constraint over the velocity field). The goal is to estimate the centroid from which each point in X was generated, thus resulting in a correspondence output.

While considered state-of-the-art, CPD still presents difficulties in overcoming a high incidence of outliers and missing data, as well as a different number of points between the reference and target. Consequently, variants of CPD have been developed in recent years to deal with such drawbacks by assigning different membership probabilities [27] or using k-connected neighbors [1] to enforce the preservation of local structures. Other variants are only applicable to rigid registration [30, 24, 44] and, consequently, do not conform to our assumptions.

An interesting recent work [18] proposes a Bayesian formulation of CPD (BCPD). Under this setting, the authors guarantee convergence of the algorithm, introduce more interpretable parameters, and reduce sensitivity to target rotation. Besides, this formulation is amenable to kernels beyond the Gaussian, thus presenting a close relationship to our work. In fact, we shall see that it is possible to establish a parallel between BCPD and our approach, under a given set of assumptions. Interestingly, in [17] we see an improved version of BCPD, where GPR is used. However, note that the introduction of GPs has the single purpose of accelerating the algorithm. The point sets are initially subsampled, after which standard BCPD is conducted. The final step employs GPR to extend the retrieved deformations to the full shape.

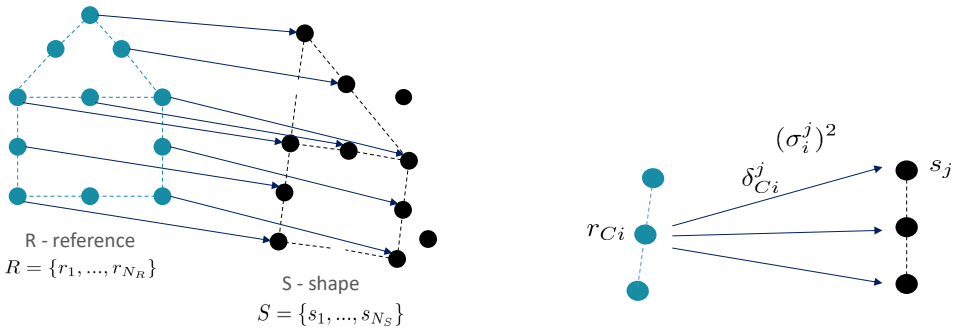
Finally, in [2] the authors propose a probabilistic registration method, using a point distribution model (PDM) as a kernel, instead of the traditional squared exponential kernel used in [31]. PDMs are the standard approach to retrieve statistical shape models from data in correspondence, through the application of principal component analysis (PCA). They also propose the use of anisotropic GMMs, oriented according to the surface normals, thus taking into account surface information. This method is closely related in the sense that it constitutes a probabilistic approach with an underlying shape modelling, thus merging soft-assignment with prior shape knowledge.

1.2. Our method.

From the previous introduction, it is possible to conclude that proposals for registration within the GP framework target hard assignment and assume a one-to-one correspondence between a shape and reference, thus motivating their extension with a soft assignment. On the other hand, probabilistic registration methods do not consider detailed prior knowledge specific to the shape. This observation motivated us to develop the shape fitting Gaussian process (SFGP), a probabilistic shape fitting/registration method within the GP framework, where one can benefit both from a complex kernel prior and a soft assignment in the correspondences. A schematic view of the main idea behind our method is presented in Figure 1.

Our main contributions are the following:

- *Shape registration/model fitting as a multiannotator GPR.* We show how the problem of registration with soft assignment can be understood within the GP framework as a multiannotator Gaussian process regression (section 2).



(a) Registration scheme. (b) Multi-annotator concept.

Figure 1. *Schematic view of our method. The registration concept is illustrated on the left, where a reference point set R is put into correspondence with a generic shape S. The main goal is to retrieve the deformation applied to the reference so that it resembles the target shape, as well as the matching between the two point sets. Note the existence of missing data, i.e., points found in R but not in S, and outliers, i.e., points found in S but not in R, introducing additional challenges to the registration procedure. On the right, the main concept behind our method is depicted. For each reference point with correspondence, r_{Ci}, several possible deformations are considered, referring to the different target shape points. Each deformation (label) has an associated variance $(\sigma_i^j)^2$ related to the level of confidence in that particular annotation.*

- *Parallel between probabilistic registration and our method, SFGP.* We provide a parallel between BCPD and our algorithm, under a few assumptions, which allows us to benefit from the probabilistic setting (section 4). We further show how their differences lead to a good performance in the presence of extensive missing data (section 5).
- *Application to a difficult registration problem—3D ears registration.* We show that our method is suitable for the registration of 3D point clouds with highly nonrigid deformations, high occurrence of missing data and outliers, by performing simulations with 3D point sets of human ears. The results show improvement with respect to state-of-the-art proposals (section 5).

2. Registration within the GP framework.

In this section we present the formulation of our problem within the GP framework. This extends the approach in [25] with the explicit modelling of missing data and the incorporation of multiple annotators to model probabilistic correspondences.

2.1. Gaussian processes.

A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. A GP $u(x)$ is fully specified by its mean $\mu(x)$ and covariance function $k(x, x')$ defined as

$$\mu(x) = \mathbb{E}[u(x)],$$
$$k(x, x') = \mathbb{E}[(u(x) - \mu(x))(u(x') - \mu(x'))]$$

and usually written as

$$u(x) \sim \mathcal{GP}(\mu(x), k(x, x')).$$

We refer to [39] for a more thorough introduction to the theory of Gaussian processes.

Outputs in higher dimensions. GPs have initially been defined for scalar outputs, but they can be extended to the vector-case under certain assumptions so that the results obtained for the scalar case remain valid [16]. In particular, a useful class of covariance functions for the vector-valued case arises from the scalar-valued covariance functions [29]. Let $A \in \mathbb{R}^{d\times d}$ be a symmetric, positive definite matrix and l a real-valued covariance function. It can be shown that the matrix-valued function $K \in \mathbb{R}^{d\times d}$ with entries k_{ij} defined by $k_{ij} = A_{ij}l(x, x')$ is a valid covariance, with A_{ij} representing the correlation between the ith and jth output component. Therefore, under the assumption that different dimensions have no correlation, it is possible to use any preexisting scalar kernel and set A as the identity matrix.

2.2. Problem formulation and notation. We consider that any shape $S = \{s_1, \dots, s_{N_S}\}$ can be obtained from a reference shape $R = \{r_1, \dots, r_{N_R}\}$, where $s_i, r_i \in \mathbb{R}^d$. In particular, the shape S is obtained by adding deformations $u(r)$ to the reference points, where $u(r)$ is modelled as a GP defined by a mean function $\mu : \mathbb{R}^d \to \mathbb{R}^d$ and a kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$, and written as $u(r) \sim GP(\mu(r), K(r, r'))$.

Note that we are working in the case of nonscalar output, but as stated above the results obtained for scalar outputs may be applied, as long as the kernel K is ensured to be valid. We further assume that we can get noisy observations of the deformations

$$\delta(r_i) = u(r_i) + \epsilon, \tag{2.1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ and σ_n^2 is the noise variance.

Under the traditional GP setting perspective, the reference points r_i can be viewed as the input data, while the respective deformations $\delta(r_i)$ correspond to the labels or output variables.

Modelling missing data and outliers. In [25] the authors assume a one-to-one correspondence between target shape and reference, so each point s_i is attributed to a reference point as $S = \{r + u(r) \mid r \in R\}$. However, in challenging scenarios such as the ears, this assumption is far from true, given the large ratio of missing data and outliers found on the target shapes. Therefore, we consider the existence of reference points without correspondence in the target and vice versa. We formulate this assumption by splitting the reference R into two subsets $\{R_C, R_M\}$, where the former set contains points with correspondence and the latter refers to missing data. In the same way, we split S into the corresponding points and outliers as $S = \{S_C, S_O\}$, such that S_C presents a one-to-one correspondence with R_C. Under these assumptions, the observed deformations can be expressed as

$$\delta = \begin{bmatrix} \delta_{C1} \\ \vdots \\ \delta_{CC} \end{bmatrix} = \begin{bmatrix} s_{r_{C1}} - r_{C1} \\ \vdots \\ s_{r_{CC}} - r_{CC} \end{bmatrix}, \tag{2.2}$$

where $R_C = \{r_{C1}, \dots, r_{CC}\}$ and $s_{r_{Ci}}$ is the shape point corresponding to the reference point r_{Ci}. A list of the notation used from this section onward can be found in Table 1.

2.2.1. Registration problem. Within this framework, and because of the Gaussianity of the distributions, for which the mode and mean coincide, the shape fitting and registration

Table 1
Notation table.

Variable	Description
$S = \{s_1, \ldots, s_{N_S}\} \subset \mathbb{R}^d$	The target shape point set, with its respective vector representation $s = (s_1^T, \ldots, s_{N_S}^T)^T \in \mathbb{R}^{N_S d}$
$R = \{r_1, \ldots, r_{N_R}\} \subset \mathbb{R}^d$	The reference point set, with its respective vector representation $r = (r_1^T, \ldots, r_{N_R}^T)^T \in \mathbb{R}^{N_R d}$
$R_C = \{r_{C1}, \ldots, r_{CC}\}$	Set of reference points with correspondence to target, where C is the number of points with correspondence
$S_C = \{s_{r_{C1}}, \ldots, s_{r_{CC}}\}$	Set of shape points with correspondence to the reference, where $s_{r_{Ci}}$ is the target point corresponding to the reference point r_{Ci}
$\Delta = \{\delta_{C1}, \ldots, \delta_{CC}\} \subset \mathbb{R}^d$	The deformations for each reference point with correspondence, with its respective vector representation $\delta = (\delta_{C1}^T, \ldots, \delta_{CC}^T)^T \in \mathbb{R}^{N_R d}$
$D_{\sigma_n^2} = \text{diag}(\sigma_{C1}^2, \ldots, \sigma_{CC}^2)$	Diagonal matrix of observation noise, where $\sigma_{C_i}^2$ is the variance of noise for observed deformation δ_{Ci}
$\mathcal{C}_i = \{j : s_j \in S, p_{i,j} > P_{MIN}\}$	Set of indices of target shape with correspondence with reference point r_i
$K = K_{RR} = [k(r_i, r_i)]_{i=1}^{N_R} \in \mathbb{R}^{N_R d \times N_R d}$	The kernel matrix of the entire reference point set, i.e., containing all the points in R
$K_{R_C R_C} = [k(r_{Ci}, r_{Ci})]_{i=1}^{C} \in \mathbb{R}^{Cd \times Cd}$	The kernel matrix of the reference points with correspondence
I_d	The identity matrix of size d
ω	Outlier probability
$D_{\varsigma^2} = \text{diag}(\varsigma_1^2, \ldots, \varsigma_{N_R}^2)$	The diagonal matrix of registration noise, where ς_i^2 is the variance associated to reference point r_i
$\tilde{A} = A \otimes I_d$	The Kronecker product of matrix A with I_d
1_N	The vector of ones with size N

problem can be formulated as a MAP problem [26], that is,

$$\max_u p(u \mid R, S).$$

In particular, the dependence is only on the observed points and deformations (not over the entire reference shape), so that

$$\max_u p(u \mid R_C, \delta).$$

However, the correspondences are not known beforehand and they depend on the deformations themselves, leading to

$$\operatorname*{argmax}_u p(u \mid R_C(u), \delta(u)). \tag{2.3}$$

In an ICP-like approach, we split our problem into two, where u is kept fixed in the inner maximization and R_C, δ are kept fixed in the outer optimization. Consequently, the final formulation of the fitting problem is written as

$$\max_u \left\{ \max_{R_C, \delta} p(u \mid R_C(u), \delta(u)) \right\}. \tag{2.4}$$

In the outer problem, the current correspondence is used to estimate the transformations on the reference. Conversely, the inner maximization computes likely correspondences, given the currently transformed reference shape.

3. Formulation with multiannotators. Given the conceptual formulation in (2.4), it is necessary to define the correspondence procedure to effectively solve the optimization. In section 3.1 we consider a hard-assignment that is then extended through multiannotators to a probabilistic one in section 3.2.

3.1. Correspondence with hard-assignment. Let us first consider the case where there is single correspondence for the reference inliers. Regardless of the method used to establish a matching, the output will be a vector of observed deformations δ with the same size as R_C, as expressed in (2.2). By (2.4) and for fixed R_C and δ, we are interested in the MAP of the GP posterior. In particular, given a set of observed inputs R_C and outputs δ, the goal is to find the most likely deformations for unobserved points R_M. This corresponds to noisy GPR, where u is the GP and R_C, δ are the training dataset. In GPR, a Bayesian approach is followed to retrieve the predictive equations for test points (in this case R_C). Applied to our setting and for the full shape input R, these equations correspond to [39]

$$u_*|R_C, \delta, R \sim \mathcal{N}(\mu_p, \Sigma_P),$$

where $\mu_p \in \mathbb{R}^{N_R d}$ and $\Sigma_p \in \mathbb{R}^{N_R d \times N_R d}$ are given as

$$\begin{aligned} \mu_p &= K_{R_C R}^T (K_{R_C R_C} + \sigma_n^2 I_{Cd})^{-1} \delta, \\ \Sigma_p &= K_{RR} - K_{R_C R}^T (K_{R_C R_C} + \sigma_n^2 I_{Cd})^{-1} K_{R_C R}, \end{aligned} \tag{3.1}$$

where $K_{R_C R_C} = [k(r_{Ci}, r_{Ci})]_{i=1}^{C} \in \mathbb{R}^{Cd \times Cd}$ is the kernel matrix of the observed points, $K_{RR_C} = [k(r_j, r_{Ci})]_{j,i=1}^{N_R, C} \in \mathbb{R}^{N_R \times C}$ is the kernel matrix of the predicted and observed points, $K_{R_C R} = K_{RR_C}^T$ its transpose, and I_{Cd} is the identity matrix of size Cd. The deformed reference is then obtained as $\bar{r} = r + \mu_p$, to be used in the inner iteration for computation of the correspondences (e.g., by taking the closest point).

Remark on the kernel matrices. Note that throughout the fitting procedure, only the points contained in R are taken into account. Thus, we lie in the discrete setting and the prediction outputs μ_p and Σ_p can be written as a vector and matrix, instead of functions. In the same way, K_{RR}, the kernel matrix of all reference points, is constant and all other kernel matrices $(K_{R_C R_C}, K_{R_C R})$ are subsets of K_{RR}. For this reason, K_{RR} is simply denoted as K. Despite this observation, we note that the formulation with GPs (instead of its discrete counterpart, i.e., the multivariate normal distribution) is relevant, as the extension to a continuous surface is desired at later steps of the pipeline.

3.2. Introducing soft-assignment. Introducing a soft-assignment equates to attributing different possible target points to a reference point, each with a different probability of correspondence. In particular, we assume that a reference point r_i has N_S possible deformations, corresponding to a match with each target point. In the traditional machine learning setting, this idea can be expressed as multiannotation. The concept refers to the multiple labellings of the same data point when there is no exact ground truth available (e.g., the attribution of a label is subjective). In this situation, a common approach is to obtain a collection of labels

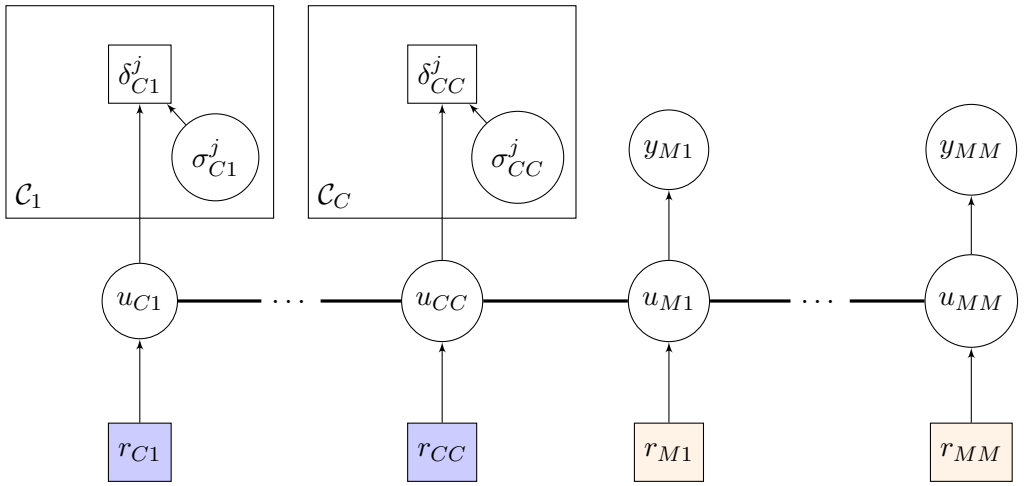


Figure 2. *Graphical model for the GP with multiple annotators in the context of shape modelling. The model follows the notation proposed in [39] for GP graphical models, where the horizontal bold line represents a set of fully connected nodes. Squared nodes represent measurements (empty squares) or constants (filled squares) and circles are latent ones. Template points with correspondence are identified with blue, while missing points are identified with orange to facilitate visual recognition. The plates represent the repetition of contained variables and, in this case, refer to the multiple annotations j of a given reference point.*

for each data point, provided by different annotators, with (possibly) different levels of confidence. The data points correspond to the reference points in R, while the labels correspond to the different possible deformations with respect to the target shape points. For a visual representation of the described model, we refer the reader to the respective graphical model in Figure 2.

Therefore, we consider δ^j_{Ci} as the deformation between the reference point r_{Ci} and the shape point s_j, with an associated variance $(\sigma^j_{Ci})^2$. The final variance and deformation of point r_i is then retrieved by weighting all of the j contributions. By appropriately defining this relationship we are able to obtain predictive equations that follow the same structure as the GPR in (3.1). Following [14], we define these two quantities as

$$\frac{1}{\sigma^2_{Ci}} = \sum_{j\in\mathcal{C}_i} \frac{1}{(\sigma^j_{Ci})^2}, \quad \hat{\delta}_{Ci} = \sigma^2_{Ci} \sum_{j\in\mathcal{C}_i} \frac{\delta^j_{Ci}}{(\sigma^j_{Ci})^2}, \tag{3.2}$$

where $\delta^j_{Ci} = s_j - \bar{r}_{Ci}$. The variance $(\sigma^j_{Ci})^2$, corresponding to the annotator's confidence in the original formulation, represents here the probability of correspondence between point C_i and s_j. Note that while the variance of annotator j is often considered the same across the data points i, here we consider that annotator j of point C_i is not necessarily the same as annotator j of point C_k, i.e., there can be at most $N_R \times N_S$ annotators.

The predictive equations can readily be obtained by assuming annotators provide the labels independently from each other, leading to the likelihood as

$$p(\delta|u) = \prod_{i=1}^{N_R} \prod_{j\in\mathcal{C}_i} \mathcal{N}(\delta^j_{Ci}|u_i, (\sigma^j_{Ci})^2).$$

Considering this expression and the prior on u, the predictive equations for an input R follow $\mathcal{N}(\hat{\mu}_p, \hat{\Sigma}_p)$, with mean and covariance [14]

$$\hat{\mu}_p = K_{R_C R}^T (K_{R_C R_C}^{-1} + \tilde{D}_{\sigma_n^2})^{-1} \tilde{D}_{\sigma_n^2} \hat{\delta},$$
$$\hat{\Sigma}_p = K_{RR} - K_{R_C R}^T (K_{R_C R_C}^{-1} + \tilde{D}_{\sigma_n^2})^{-1} K_{R_C R},$$

where $D_{\sigma_n^2} = \text{diag}(\sigma_{C1}^2, \ldots, \sigma_{CC}^2)$, $\tilde{D}_{\sigma_n^2} = D_{\sigma_n^2} \otimes I_d$ and $\hat{\delta} = [\hat{\delta}_{C1}, \ldots, \hat{\delta}_{CC}]^T$, where σ_{Ci}^2 and $\hat{\delta}_{Ci}$ are given by (3.2). The kernel matrices are defined in (3.1). We can see that these equations differ from (3.1) only on the noise covariance matrix and label vector, but follow exactly the same structure.

In order to complete our algorithm, two points need to be addressed: how to define the correspondence set R_C and how to compute $(\sigma_{Ci}^j)^2$. The computation of this value will be detailed in the next section.

4. Computation of variance. In order to obtain a theoretically sound update for elements σ_{Ci}^j in (3.2) we reformulate our problem in the standard probabilistic approach. This framework is first introduced in section 4.1 and then applied to our setting in section 4.2. In section 4.3 we derive the parameter estimation for the previous formulation. Finally, in section 4.3 we establish a parallel between the GP formulation in section 3.2 and the probabilistic approach presented in this section. We follow the Bayesian formulation proposed in [18], instead of the original in [31], as the Bayesian setting has a strong connection to the GPs. Throughout this section, we keep the notation introduced in section 2 whenever the variables refer to the exact same elements.

4.1. Probabilistic shape registration. Under the probabilistic formulation, a reference shape (R), upon an appropriate transformation, can be seen as a set of centroids of a GMM, where the target points of any shape (S) correspond to data generated by the centroids. Further, a point s_j can be an outlier with probability ω, in which case it is generated from an outlier probability distribution $p_{out}(s_j)$. If s_j is not an outlier, then it corresponds to a point r_i with probability α_i (membership probability).

In [18] the authors consider explicit similarity transformations and nonrigid ones, such that a point i of the transformed reference is given as

$$\mathcal{T}(r_i) = \beta \Gamma (r_i + v_i) + \eta, \tag{4.1}$$

where r_i is the original reference point, β is a scale factor, Γ is a rotation matrix, η is a translation vector, and v_i is a displacement vector for nonrigid transformations. We use $\rho = (\beta, \Gamma, \eta)$ to denote the set of similarity transformations. According to GMM, the generation of a target point s_j follows a multivariate normal distribution with mean $\mathcal{T}(r_i)$—the transformed reference point—and covariance matrix $\varsigma^2 I_d$, so the probability distribution to generate s_j starting from r_i is

$$\phi_{ij}(s_j; \mathcal{T}(r_i), \varsigma^2) = \frac{1}{(\varsigma\sqrt{2\pi})^d} \exp\Big(- \frac{\|s_j - \mathcal{T}(r_i)\|^2}{2\varsigma^2} \Big). \tag{4.2}$$

In order to explicitly introduce correspondences, two aditional variables are added: $c \in \{0,1\}^{N_S}$, an indicator variable that takes a value of 1 for c_j if point s_j is an outlier, and

$e \in \{1, \dots, N_R\}^{N_S}$, where $e_j = i$ if the jth target point corresponds to the ith reference point. Taking the outlier assumptions into account, we obtain the joint distribution for (s_j, e_j, c_j) as

$$p(s_j, e_j, c_j | R, \varsigma^2, v, \rho) = \{w p_{out}(s_j)\}^{1-c_j} \left\{ (1-w) \prod_{i=1}^{N_R} (\alpha_i \phi_{ij})^{\gamma_i(e_j)} \right\}^{c_j}, \tag{4.3}$$

where γ_i is an indicator function, taking a value of 1 if $e_j = i$ and 0 otherwise, and α_i is the probability that $e_j = i$, with $\sum_{i=1}^{N_R} \alpha_i = 1$. The authors take $p(\alpha)$ as a Dirichlet distribution and set a prior on the deformations as

$$p(v|R) = \phi(\delta; 0, \lambda^{-1} G \otimes I_d),$$

where $G = (g_{ii'}) \in \mathbb{R}^{N_R \times N_R}$, with $g_{ii'} = k(r_i, r_i')$ and $k(x, x')$ a kernel function; λ is a positive constant.

Finally, the full joint is given as

$$p(S, T, \theta) \propto p(\delta|R) p(\alpha) \prod_{j=1}^{N_S} p(s_j, e_j, c_j | R, v, \varsigma^2, \alpha, \rho), \tag{4.4}$$

where $\theta = (v, \varsigma^2, \alpha, \rho, c, e)$.

4.2. Formulation. We shall briefly detail the assumptions and problem formulation used in this approach, by deriving equivalents of (4.1) through (4.4) according to our assumptions. This will lead to the final expression for the joint distribution in (4.8).

Transformation model. We do not consider similarity transformations, so the transformation acting on the reference is merely given by the displacement vector v_i

$$\mathcal{T}_i = r_i + v_i. \tag{4.5}$$

Gaussian mixture model. According to the GMM we obtain a similar expression to (4.2)

$$\phi_{ij}(s_j; \mathcal{T}(r_i), \varsigma_i^2) = \frac{1}{(\varsigma_i \sqrt{2\pi})^d} \exp\Big(- \frac{\|s_j - \mathcal{T}(r_i)\|^2}{2\varsigma_i^2} \Big), \tag{4.6}$$

but we introduce an individual variance for each reference point given as ς_i^2. In section 5 it will become clear that this has a positive impact when dealing with large regions of missing data. Regarding the outliers, we follow the same assumptions as in BCPD, but for simplification, we take $p_{out}(s_j) = 1/N_S$, as was previously taken in CPD. Besides, we take equal membership probabilities $\alpha_i = 1/N_R$, meaning that a point in the target is expected to be associated with any point of the reference with equal probability. This is also the assumption in CPD and is here taken for simplification on this first formulation of the framework. Thus, we obtain the joint distribution

$$p(s_j, e_j, c_j | R, v, \varsigma^2) = \left\{ \frac{w}{N_S} \right\}^{1-c_j} \left\{ \frac{(1-w)}{N_R} \prod_{i=1}^{N_R} (\phi_{ij})^{\gamma_i(e_j)} \right\}^{c_j}. \tag{4.7}$$

Prior distributions. The prior on deformations is expressed with the previously defined kernel $k(r, r')$, except that λ is taken as 1, since it can be included within the kernel. Therefore, $\lambda^{-1}G \otimes I_d$ can simply be denoted as $K(r, r')$, as defined in 2.2 and the prior is given as $p(v|r) = \mathcal{N}(0, K)$.

4.2.1. Full joint distribution. Finally, the full joint distribution is obtained as

$$p(s, r, \theta) \propto p(v|r) \prod_{j=1}^{N_S} p(s_j, e_j, c_j | r, v, \varsigma^2), \tag{4.8}$$

where $\theta = (v, \varsigma^2, c, e)$ are the parameters to be estimated.

4.3. Solving the problem with variational Bayesian inference. In order to estimate the parameters θ in (4.8), we resort to variational Bayesian inference (VBI) [5, 43], a useful tool when dealing with challenging posterior distributions. We start by providing an overview of VBI and its formulation for out problem, followed by the update equations obtained through this approach.

4.3.1. Background. The idea behind VBI is to use a distribution $q(\theta)$ to approximate the true posterior $p(\theta|S, R)$, where the closeness between the two distributions is measured by the Kullback–Leibler (KL) divergence. Therefore, the goal is to minimize the KL divergence between q and p, i.e.,

$$\begin{aligned} q^*(\theta) &= \underset{q(\theta \in Q)}{\operatorname{argmin}} \quad KL(q(\theta) \;||\; p(\theta|S, R)) \\ &= \underset{q(\theta \in Q)}{\operatorname{argmin}} \, \mathbb{E}[\log q(\theta)] + \mathbb{E}[\log p(\theta, R, S)] + \log p(S, R), \end{aligned}$$

where Q is a predefined set of distribution families to which q belongs. However, since $\log p(S, R)$ may not be computable, the evidence lower bound (ELBO) is maximized instead

$$ELBO(q) = \mathbb{E}[\log p(\theta, R, S)] - \mathbb{E}[\log q(\theta)].$$

The ELBO is equivalent to the negative KL divergence up to a constant, and therefore maximizing the former is equivalent to minimizing the latter. The choice of a suitable form for Q is fundamental, as it should ideally lead to a sufficiently simple ELBO, while being flexible enough to provide a good approximation to the original posterior distribution.

Here, we assume that q has a strong separation form, i.e., the latent variables are mutually independent and governed by different factors. This leads to $q(\theta) = \prod_{i=1}^{M} q_i(\theta_i)$, where $q_i(\theta_i)$ is the distribution for the variable θ_i. In particular, we consider

$$q(\theta) = q_1(v) q_2(c, e) q_{31}(\varsigma_1^2) \dots q_{3i}(\varsigma_i^2) \dots q_{3N_R}(\varsigma_{N_R}^2).$$

A standard method to maximize the ELBO, and the one followed here, is the coordinate ascent variational inference. If we fix all other q_j, then we know that the optimal q_i is

$$q_i(\theta_i)^* \propto \exp\{\mathbb{E}_{-i}[\log p(\theta_i | \theta_{-i}, S, R)]\} \propto \exp\{\mathbb{E}_{-i}[\log p(\theta_i, \theta_{-i}, S, R)]\}, \tag{4.9}$$

where $\mathbb{E}_{-i}[\log p(\theta, S, R)]$ is the expectation of the joint probability with respect to the remaining $q_{j \neq i}$ and θ_{-i} corresponds to all parameters in θ except θ_i. Hence, each q_i is updated iteratively by computing $\mathbb{E}_{-i}[\log p(\theta, S, R)]$, until convergence is reached.

4.3.2. Update equations. We present the updates for each component of $q(\theta)$ in Propositions 1, 2, and 3. The proofs follow [18, 5] and can be found in the supplementary material. The final equations exhibit a similar structure to those in [18], except that ς^2, taken as a scalar in [18], is replaced by the diagonal matrix D_{ς^2}.

For ease of notation, and in preparation for the subsequent equations, we define $p_{ij} = \mathbb{E}[c_j\gamma_i(e_j)]$ as the probability of correspondence between reference point i and target point j, with the respective probability matrix $P = [p_{ij}]_{i,j=1}^{N_R,N_S} \in [0,1]^{N_R\times N_S}$. We further define $\nu_i = \sum_{j=1}^{N_S} p_{ij}$, representing the expected number of target points corresponding with r_i, as well as $\nu = P1_{N_S}$, the corresponding vector.

Proposition 1. *The deformations v follow a normal distribution $\mathcal{N}(\mu_v, \Sigma_v)$, with the update equations for mean and covariance given as*

$$\begin{aligned} \mu_v &= \Sigma_v \tilde{D}_\nu \tilde{D}_{\varsigma^2}^{-1}(\tilde{D}_\nu^{-1}\tilde{P}s - r), \\ \Sigma_v &= (K^{-1} + \tilde{D}_\nu \tilde{D}_{\varsigma^2}^{-1})^{-1}, \end{aligned} \tag{4.10}$$

where $D_{\varsigma^2} = \mathrm{diag}(\varsigma_1^2, \dots, \varsigma_{N_R}^2)$ and $D_\nu = \mathrm{diag}(\nu_1, \dots, \nu_{N_R})$.

Proposition 2. *The update for the correspondence probability is*

$$p_{ij} = \frac{(1-w)\langle\phi_{ij}\rangle}{\frac{N_R}{N_S}w + (1-w)\sum_{i'=1}^{N_R}\langle\phi_{i'j}\rangle}, \tag{4.11}$$

where $\langle\phi_{ij}\rangle = \phi_{ij}(s_j; \mathcal{T}(r_i), \varsigma_i^2)\exp\left\{-\frac{\mathrm{Tr}(\Sigma_v^i)}{2\varsigma_i^2}\right\}$, $\mathcal{T}(r_i) = r + \mu_v$, Σ_v^i is the submatrix of Σ_v related to the v_i component, and $\mathrm{Tr}(\cdot)$ is the trace of the matrix.

Proposition 3. *The update for each variance term ς_i^2 is given as*

$$\varsigma_i^2 = \frac{1}{d}\left(\frac{[\tilde{P}\mathrm{diag}(s)s]_i - 2\bar{r}_i^T[\tilde{P}s]_i}{\nu_i} + \|\bar{r}_i\|^2 + \mathrm{Tr}(\Sigma_v^i)\right), \tag{4.12}$$

where $[A]_i$ refers to the ith row of matrix A.

4.4. Parallel with GP framework. If we assume that there are no missing points, then it is possible to establish a parallel between the previous formulation and the GP framework. With this aim, we reformulate our expressions in order to obtain a similar structure to Propositions 1 through 3. Note that the update step of p_{ij} can be understood as the "getting correspondence" part, i.e., the inner optimization of problem (2.4).

Proposition 4. *Considering no missing points, i.e., $R_C = R$, and if the variance $(\sigma_{Ci}^j)^2$ in (3.2) is taken as*

$$(\sigma_{Ci}^j)^2 = \frac{\varsigma_i^2}{p_{ij}}, \tag{4.13}$$

where p_{ij} is given by (4.11), then an equivalent exists between the update equations in Propositions 1, 2, and 3 and the update equations for multiannotator GPR in (3.1).

Proof. When $R_C = R$, and since $K^T = K$, the posterior mean and covariance in (3.1) become

$$\mu_p = K(K + \tilde{D}_{\sigma_n^2})^{-1}\hat{\delta},$$
$$D_{\sigma_P^2} = K - K(K + \tilde{D}_{\sigma_n^2})^{-1}K,$$

where $D_{\sigma_n^2} = \text{diag}(\sigma_1^2, \ldots, \sigma_{N_R}^2)$, with σ_i^2 and $\hat{\delta}_i$ given by (3.2) and here restated without the notation for correspondences (as there are no missing points)

$$\frac{1}{\sigma_i^2} = \sum_{j \in N_S} \frac{1}{(\sigma_i^j)^2}, \quad \hat{\delta} i = \sigma_i^2 \sum_{j \in N_S} \frac{\delta_i^j}{(\sigma_i^j)^2}.$$

Taking the variance as in (4.13), we can write the previous equations as

$$\frac{1}{\sigma_i^2} = \frac{P_i 1_{N_s}}{\varsigma_i^2} = \frac{\nu_i}{\varsigma_i^2},$$
$$\hat{\delta}_i = \nu_i^{-1} \sum_j p_{ij}(s_j - r_i) = \nu_i^{-1}\tilde{P}_i s - \nu_i^{-1} \sum_j p_{ij} r_i = \nu_i^{-1}\tilde{P}_i s - r_i,$$

where P_i refers to the ith row of matrix P. Therefore, we have that $D_{\sigma_n^2} = D_{\varsigma^2} D_\nu^{-1}$ and $\hat{\delta} = \tilde{D}_\nu^{-1}\tilde{P}s - r$. The posterior deformations in (1) can then be written as

$$\begin{aligned}
\mu_p &= K(K + \tilde{D}_{\varsigma^2}\tilde{D}_\nu^{-1})^{-1}\delta \\
&= KK^{-1}\left[1 + \tilde{D}_{\varsigma^2}\tilde{D}_\nu^{-1}K^{-1}\right]^{-1}\delta \\
&= \left[K^{-1} + \tilde{D}_\nu \tilde{D}_{\varsigma^2}^{-1}\right]^{-1}\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}(\tilde{D}_\nu^{-1}\tilde{P}s - r)
\end{aligned}$$

and the covariance as

$$\begin{aligned}
D_{\sigma_P^2} &= K - K(K + \tilde{D}_{\varsigma^2}\tilde{D}_\nu^{-1})^{-1}K \\
&= K - \left[K^{-1} + \tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\right]^{-1}\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}K \\
&= \left[K^{-1} + D_\nu D_{\varsigma^2}^{-1}\right]^{-1}\left[K(K^{-1} + \tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}) - \tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}K\right] \\
&= \left[K^{-1} + \tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\right]^{-1},
\end{aligned}$$

thus being equivalent to the expressions for μ_v and Σ_v in (4.10). This entails that v and δ, under the assumption of no missing data, refer to the same variable and are updated with equivalent equations. Therefore, if p_{ij} and ς_i^2 are updated according to Propositions 2 and 3 (respectively), there is an equivalence between the two methods. ∎

Although this is established for the case of no missing data, we take (4.13) as a reasonable update for the annotators' variance, together with the necessary updates for p_{ij} and ς_i^2. **AQ1**

4.5. Missing data points. Given the probability matrix P, we apply a predefined threshold P_{MIN}, such that pairings with a lower value than P_{MIN} are identified as non-corresponding. So, for each point r_i, the considered correspondences to the target are $\mathcal{C}_i = \{j : s_j \in S, p_{ij} > P_{MIN}\}$. Then if a point r_i has no elements in $\mathcal{C}_i$, it is considered a missing point, meaning that $R_M = \{r_i : r_i \in R, |\mathcal{C}_i| = 0\}$ and $R_C = \{r_i : r_i \in R, |\mathcal{C}_i| > 0\}$.

Algorithm 4.1 SFGP.

Input: $r, s, K, D_{\varsigma_0^2}, \omega, P_{MIN}$

$\bar{r} = r,\ D_{\sigma_P^2} = \mathbf{0}$
while some stopping criterion is not met **do**
  $R_C, \hat{\delta}, D_{\sigma_n^2} = \text{get\_correspondences}\,(r, s, \bar{r}, D_{\sigma_P^2}, D_{\varsigma^2}, \omega, P_{MIN})$
  $\mu_p = K_{R_C R}^T (K_{R_C R_C} + \tilde{D}_{\sigma_n^2})^{-1} \hat{\delta}$
  $D_{\sigma_P^2} = K_{RR} - K_{R_C R}^T (K_{R_C R_C} + \tilde{D}_{\sigma_n^2})^{-1} K_{R_C R}$
  $\bar{r} = r + \mu_p$
  $\varsigma_i^2 = \frac{1}{d}\left(\frac{[\tilde{P}\text{diag}(s)s]_i - 2\bar{r}_i^T[\tilde{P}s]_i}{\nu_i} + \|\bar{r}_i\|^2 + \text{Tr}(D_{\sigma_P^2}) \right)$
end while

Algorithm 4.2 get˙correspondences.

Input: $r, s, \bar{r}, D_{\sigma_P^2}, D_{\varsigma^2}, \omega, P_{MIN}$

Output: $D_{\sigma_n^2} = \text{diag}(\sigma_{C1}^2, \ldots, \sigma_{CC}^2), \hat{\delta} = (\hat{\delta}_{C1}^T, \ldots, \hat{\delta}_{CC}^T)^T, R_C$

for $(i, j) \leftarrow (1, 1)$ to (N_R, N_S) **do**
  $\phi_{ij}(s_j; \bar{r}_i, \varsigma^2) = \frac{1}{(\varsigma_i \sqrt{2\pi})^d} \exp\left(- \frac{\|s_j - \bar{r}_i\|^2}{2\varsigma_i^2} \right)$
  $\langle \phi_{ij} \rangle = \phi_{ij} \exp\left\{ - \frac{1}{2\varsigma^2} Tr(\sigma_{P_i}^2 I_d) \right\}$
  $p_{ij} = \frac{(1-w)\langle \phi_{ij} \rangle}{\frac{N_R}{N_S} w + (1-w) \sum_{i'=1}^{N_R} \langle \phi_{i'j} \rangle}$
  $(\sigma_i^j)^2 = \frac{\varsigma_i^2}{p_{ij}}$
end for
$\mathcal{C}_i = \{ j : s_j \in S, p_{ij} > P_{MIN} \}$
$R_C = \{ r_i : r_i \in R, |\mathcal{C}_i| > 0 \}$
for $i \in R_C$ **do**
  $\frac{1}{\sigma_{Ci}^2} = \sum_{j \in \mathcal{C}_i} \frac{1}{(\sigma_{Ci}^j)^2}$
  $\hat{\delta}_{Ci} = \sigma_{Ci}^2 \sum_{j \in \mathcal{C}_i} \frac{\delta_{Ci}^j}{(\sigma_{Ci}^j)^2}$
end for

4.6. Algorithm for SFGP. The pseudocode for our method is found in Algorithms 4.1 and 4.2, where the former contains the main outer steps and the latter details the computation for the correspondence part.

5. Experimental results and discussion.

In this section we present the results of experiments with both 2D and 3D data, with the respective discussion. For each subsection, we first describe the datasets and settings, following with an analysis of the results.

5.1. 2D data.

5.1.1. Dataset. As 2D data, we take the *fish dataset* [8], where the reference is a 2D fish with 98 points. The target point sets are then generated by applying different kinds of

Table 2

Brief description of the different methods used in the experiments. For BCPD, the absence of normalization means that the shapes maintain their relative size, i.e., they are both normalized with respect to the target shape size, as recommended by the authors.

Name	Description
SFGP_Full	SFGP in its complete version
SFGP_bcpdReg	SFGP, where the registration variance ς^2 is taken as a scalar instead of a vector, computed according to BCPD equations
GPReg_noTresh	SFGP without the threshold for missing points P_{MIN}
GPClosestPnt	Registration with GPR, but where the correspondence part is achieved by taking the closest point, i.e., not considering multiannotators
BCPD_Standard	BCPD method with the standard parameters
BCPD_Opt_Norm	BCPD method with optimized parameters for the fish dataset, with normalization of both shapes
BCPD_Opt_noNorm	BCPD method with optimized parameters for the fish dataset, without normalization of both shapes, since this is not used in our method and could potentially benefit it in some cases

AQ2 AQ3 AQ4 AQ5 AQ6

alterations to the data. Nonrigid deformations are generated by warping the reference points with a Gaussian radial basis function. The dataset has four other variations considering outliers, missing data, rotation, and noise, all of them with a moderate level of deformations included.[1] In order to accurately replicate the ear data challenges, we further create a new dataset, based on the noise level 2 of the fish dataset. Here, we introduce structured missing data in the following way: we choose one point of the reference as the center and increasingly set the width of a squared bounding box around this point—all the points within the box are removed.

5.1.2. Setting. We consider different variations of our method, as well as different variations of BCPD, in order to show the relevance of each modification. Their description can be found in Table 2. To fairly compare our method with BCPD, we set their parameters with the same value whenever possible—consequently, we use the squared exponential kernel for our model. The remaining parameters and initial values for VBI are tuned with the deformation level 1 for both methods, by grid search. A detailed description of all settings can be found in the supplementary material.

5.1.3. Metrics. For the evaluation of results, we mainly look at the Euclidean distance error between corresponding deformed reference $\bar{r}_i$ points and the ground truth s_i^*, i.e., the complete and deformed target shape without noise, averaged over the shape, so $d(s,t) = \frac{1}{N_R}\sum_{i=1}^{N_R} \|s_i^* - \bar{r}_i\|_2^2$. This is then averaged over the entire dataset, consisting of 100 samples. However, it should be noted that BCPD will occasionally not lead to a successful registration, in which case it does not produce an output or does not produce correspondence for any point. Since this result will not be taken into account for the distance metric and often occurs in the most challenging settings, we also present the fraction of successful registration. Our

[1] A more detailed description of the dataset can be found in [8].

method does not consider a failed registration unless there are no deformations found in the first iteration.

5.1.4. Discussion. The results for all the considered methods and data variations can be found in Figures 3 and 4. Our main focus is the dataset with an increasing level of missing regions (Figure 3(a)), as this closely replicates the challenges in the ear reconstruction problem. While for the lowest level it is evident that BCPD (when optimized) performs better, as we increase the missing area, our method presents a progressive advantage. Comparing *SFGP_Full*, *SFGP_bcpdReg,* and *SFGP_noTresh* it becomes clear why those modifications are advantageous when facing extensive missing regions. It is also evident that the closest point approach has the poorest performance overall.

It is also interesting to look at the results in the presence of outliers (Figures 4(a) and 4(b)), for which we tested all methods with $\omega = 0.1$ and $\omega = 0.3$ (with the exception of *GPClosestPnt*, where this is not applicable), since this parameter reflects the expected outlier probability. While BCPD outperforms SFGP when ω is adequately adjusted to the real outlier occurrence, we note that the behavior of our method is not as dependent on this parameter. Thus, in the absence of prior knowledge, SFGP is found to be a more suitable choice.

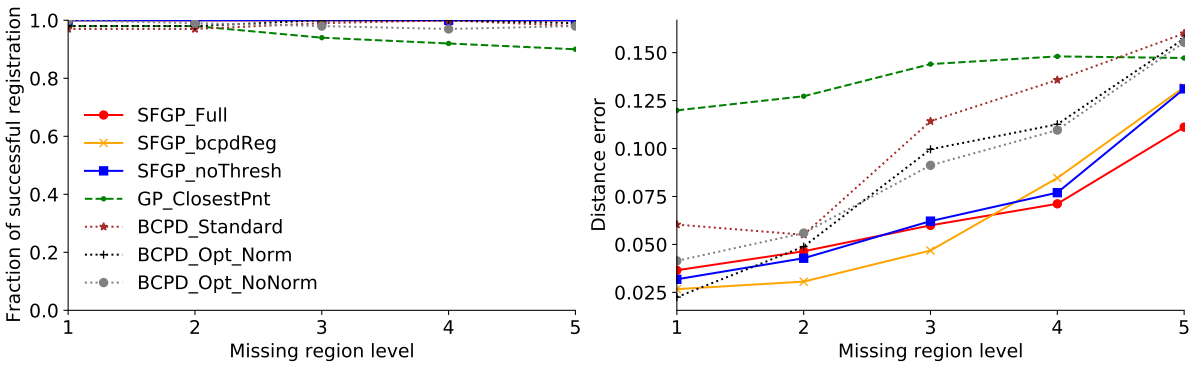
Looking at the variation of noise (Figure 3(c)) and deformations (Figure 3(b)), we see that overall an adequately fitted BCPD outperforms our method and is able to achieve lower errors, even when both parameters are previously tuned. We also note that the two variations of our method always perform better than the full proposal for these scenarios—the proposed alterations do not bring an advantage when we are not dealing with structured missing data. However, this decrease in performance is deemed acceptable given the gain it provides in Figure 3(a) and when compared with the closest point proposal always leads to lower error.

An intuition on why our method is able to cope well with extensive missing data is offered in Figure 5, where we compare the fitting results from BCPD with different levels of deformation and our proposed method. It is clear that the main challenge in achieving an adequate fitting with BCPD is that a high level of deformations leads to the collapse of the missing regions, while lower values do not provide enough flexibility to fit small details found in the nonmissing parts. With SFGP, collapsing is prevented, while allowing enough nonrigid deformation to accurately fit fine details.

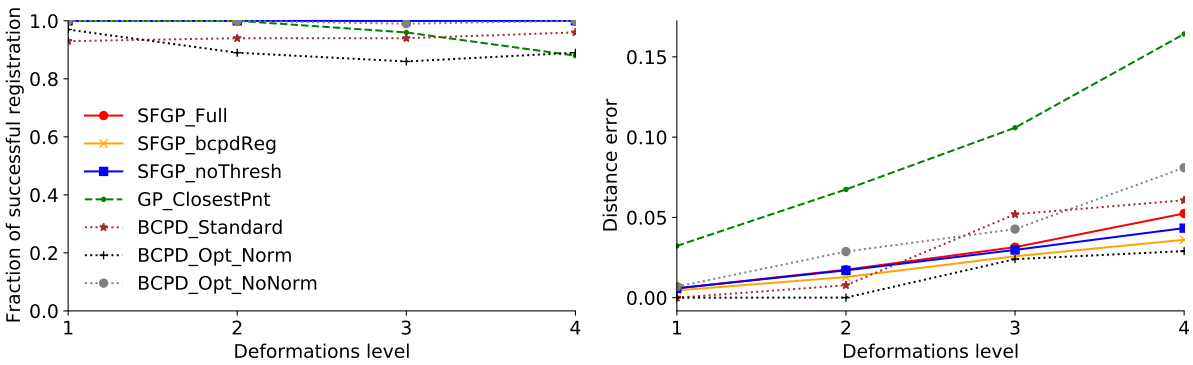
Furthermore, in Figure 6 we present additional metrics for the missing region version of this dataset. The high performance on both recall and precision presented by SFGP tells us that the lower distance error previously observed relates to an accurate identification of missing points. Additionally, we notice that the GP with closest point correspondence exhibits a very high precision, but at the cost of a low recall.

5.2. 3D Ear simulated data.

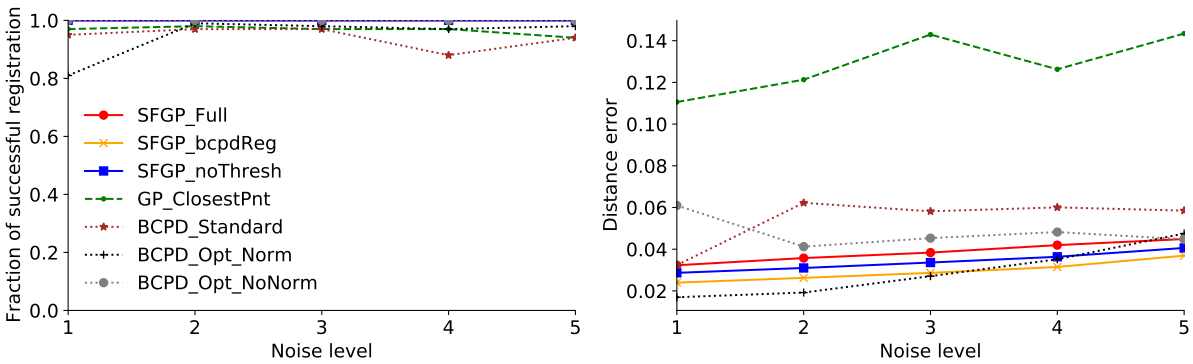
5.2.1. Dataset. In this section, we test our method with 3D ear data, the main goal of our work. The dataset is obtained from the *ear dataset* in [9], with subsequent transformations to achieve more realistic shapes, mimicking the real-life challenges observed in raw scans. We denote the transformed shapes as *simulated dataset* and an example may be found in



(a) Increasing level of missing region.

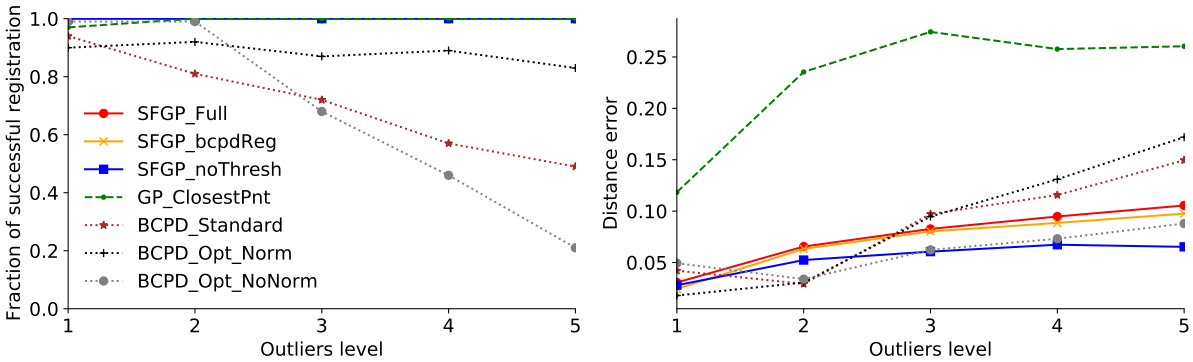


(b) Increasing level of deformations.

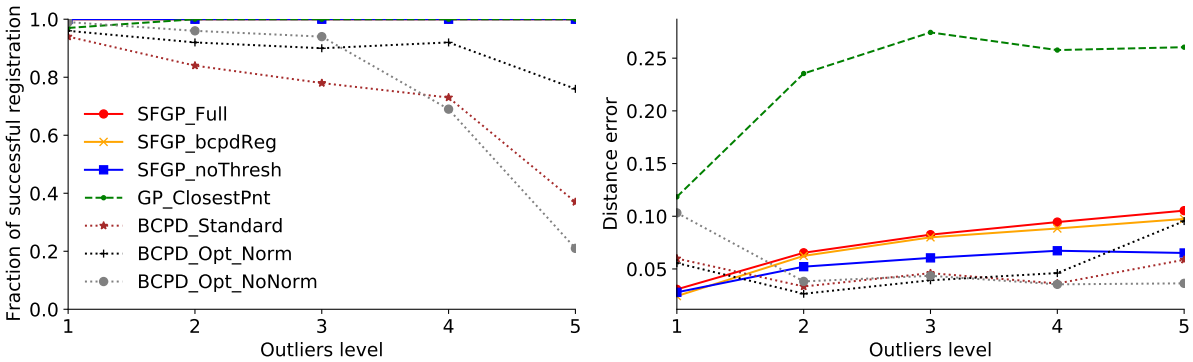


(c) Increasing level of noise.

Figure 3. *Results for the fish dataset with different types of modifications. The x-axis always depicts an increase in a given data modification, while the y-axis provides the ratio of successful items registered and the average distance error. Variants of SFGP are depicted with full lines, variants of BCPD are depicted with thin dashed lines, and GP with closest point is depicted with a thick dashed line. Increasing levels of missing region refer to increasing widths of the bounding box surrounding the selected reference points (width ranging from* 0.1 *to* 0.4*); increasing levels of deformations refer to increasing variance of the warping Gaussian radial basis function; and increasing levels of noise refer to increasing variance of the additive Gaussian noise (standard deviation ranging from* 0 *to* 0.05*).*



(a) Increasing fraction of outliers, for $\omega = 0.1$.

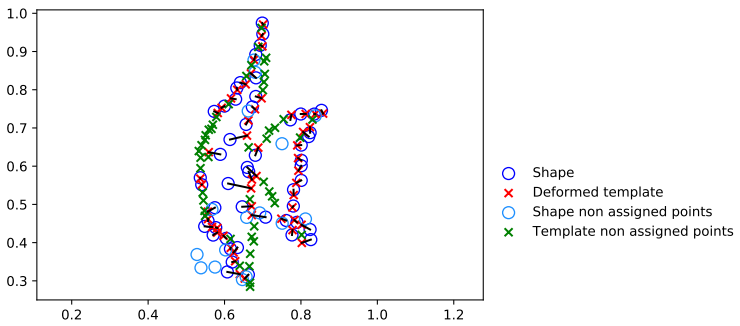


(b) Increasing fraction of outliers, for $\omega = 0.3$.

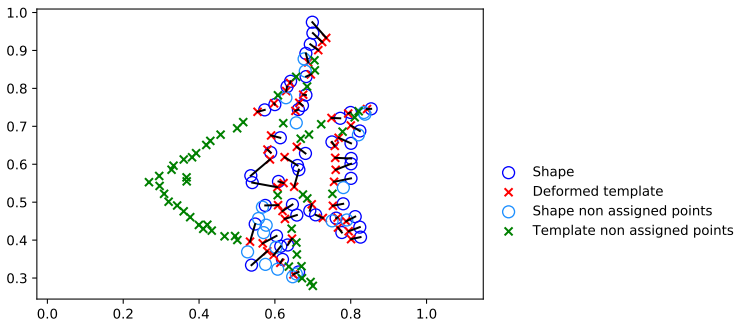
Figure 4. *Results for fish dataset with increasing fraction of outliers. The x-axis always depicts an increase in a given data modification, while the y-axis provides the ratio of successful items registered and the average distance error. Variants of SFGP are depicted with full lines, variants of BCPD are depicted with thin dashed lines, and GP with closest point is depicted with with a thick dashed line. Increasing levels of outliers refer to the increasing ratio of outlier points with respect to the reference points (ranging from* 0 *to* 2*).*

Figure 8. The transformations applied include missing data, outliers, measurement noise, and a slight rotation, translation, and scaling—a more detailed description can be found in the supplementary material (see SM2.2.1).

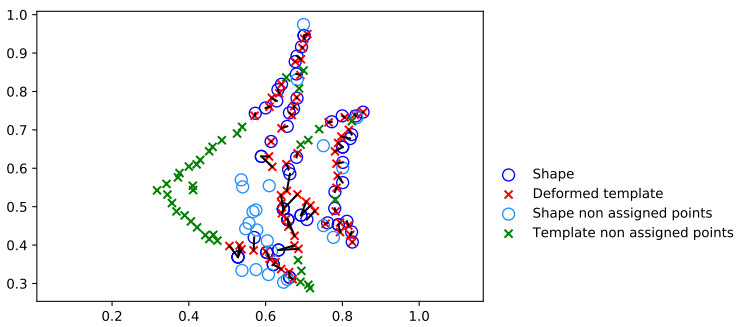
To facilitate computation, the dataset and reference were subsampled to around 3000 points. Further, we consider as possible targets only a subset of the initial 500 samples contained in the dataset, to account for the lack of variability observed between shapes (see our previous work [42] for a more detailed explanation), i.e., to ensure that the reference does not closely resemble the target shapes, we select samples presenting larger deformations with respect to the reference (measured as the average of Euclidean distance between corresponding shape points). The reference was chosen as the first shape of the dataset. Usually, one tries to achieve a reference as close as possible to all shapes in the dataset (e.g., mean shape) in order to improve shape fitting. However, since we face a problem of lack of variability in our dataset, we opt for this strategy to increase the shape difference to the targets.



(a) BCPD registration with low λ.



(b) BCPD registration with high λ.



(c) Our registration.

Figure 5. *Example for missing data performance with fish dataset. The target can be seen in blue circles, while the deformed reference after registration is represented with red and green crosses. Points with and without correspondence are identified both for the reference and target shape. The two results for BCPD are run with different values of parameter λ responsible for controlling the expected length of the deformation—small values of λ allow for more deformation and vice versa. With BCPD we can either get enough nonrigidity to fit the existing points, at the cost of collapsing the missing region, or preserve the shape of this segment at the cost of a rigid transformation that cannot appropriately fit the observed points. With SFGP, it is possible to allow a level of nonrigid deformations that fits the shape details, while correctly identifying the missing regions.*

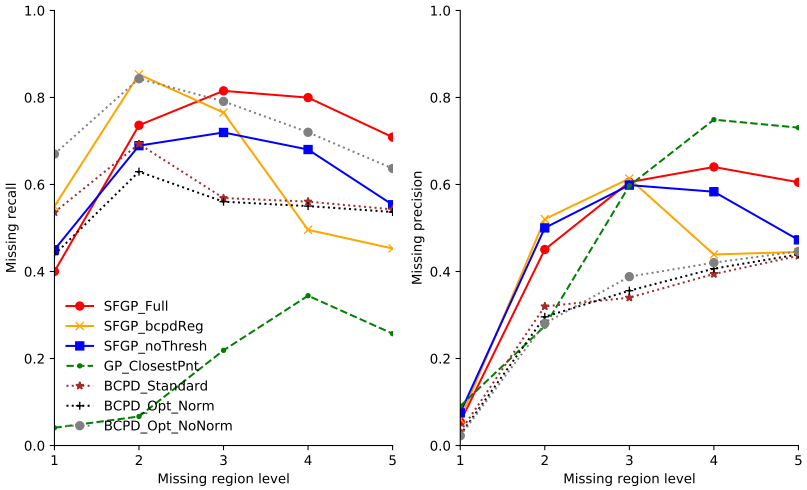


Figure 6. *Recall and precision for increasing missing region. A high recall indicates that the method is able to identify most of the missing points, while a high precision means that most of the points identified as missing are in fact missing.*

5.2.2. Discussion. Figure 7 presents the fitting results with SFGP, BCPD [18], and ANISO [2] for the simulated dataset. For fairness, we consider the SFGP with both a squared exponential kernel and a sum of a squared exponential and PCA kernel. The PCA kernel is obtained from the sample covariance of the training dataset: a subset of the original *ear dataset* not included in the possible targets of the experiment. While the squared exponential kernel introduces a similar prior to BCPD, the PCA kernel does so for ANISO. The distance error is presented separately for the missing and nonmissing regions, for additional insight into the behavior of each method.

A first comparison is due between SFGP without the PDM kernel and BCPD. While the nonmissing points present a slightly higher error for our method, this is largely compensated for by the distance error occurring in the missing regions. In practice, this entails that unreasonable shapes are less likely to be originated from the fitting with SFGP, thus proving its advantage for this particular setting of extensive missing data.

Second, the comparison between SFGP and ANISO evidences the need for a tailored method to handle outliers and missing data. The mere addition of a shape model to the probabilistic registration is not enough to overcome such obstacles (even with the anisotropic variance proposed by the authors).

As expected, we note that the addition of a shape model to SFGP improves the registration results, particularly for the nonmissing regions. Under this prior, our method presents a lower distance error not only for the missing regions but for the complete shape.

Finally, we include an example of registration with SFGP and the closest point approach in Figure 8. The limitations of the latter are clearly evidenced, as well as the capability of our method to overcome them. SFGP avoids the collapse of the posterior section when large regions of the ear are missing and adequately fits the anterior regions. Despite this, there is

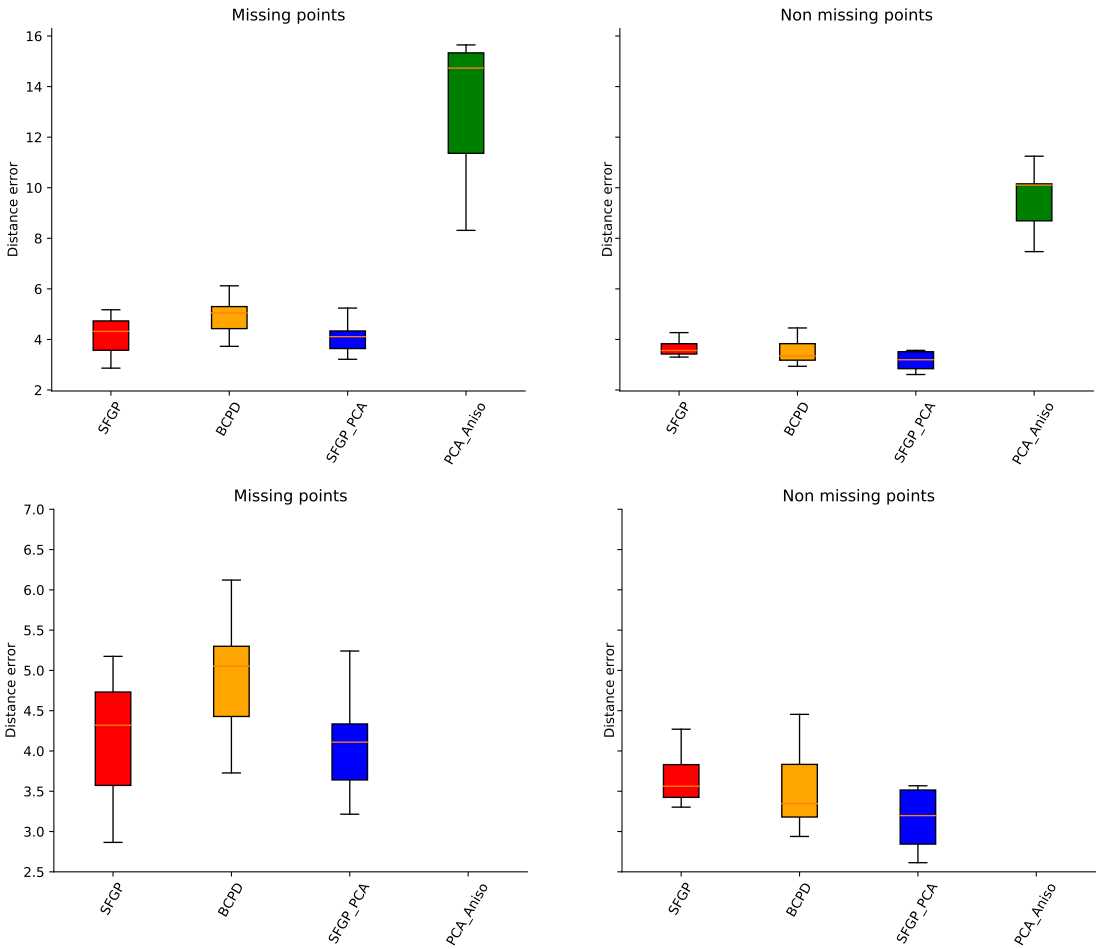
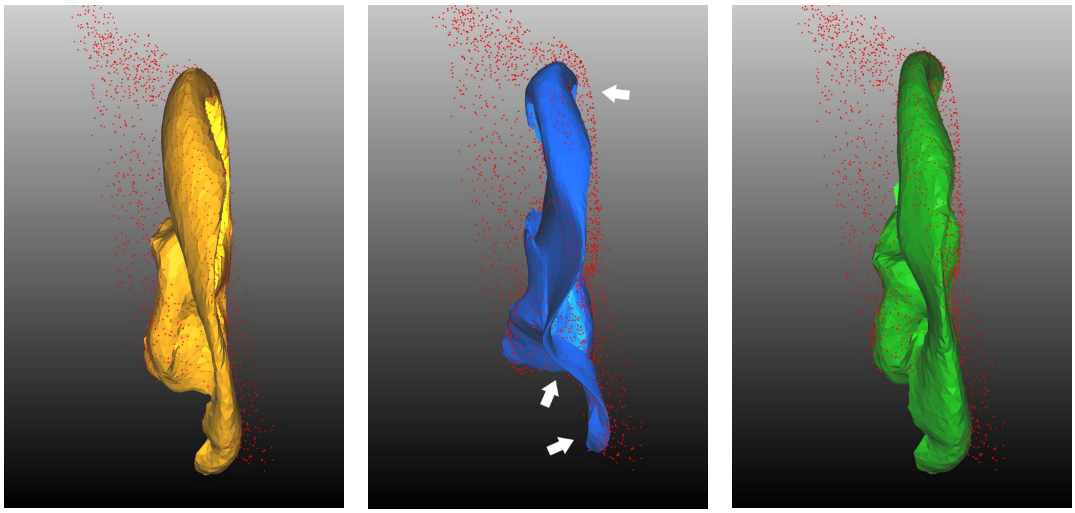


Figure 7. *Results for the fitting of a reference to the simulated dataset with a probabilistic registration method (BCPD), SFGP using a squared exponential kernel (SFGP), SFGP using a PCA kernel (SFGP_PCA), and probabilistic registration with PCA kernel and anisotropic variance (PCA_ANISO). The boxplot is obtained from the mean Euclidean distance between the true target shapes and the deformed reference, for the entire dataset. On the left, we consider only the subset of points that are missing and, on the right, the remaining ones. The bottom row contains a zoom-in of the top plots along the y-axis, for better visualization of the first three methods.*

still room for improvement as seen on the top front region, where the deformed reference does not entirely fit the target.

A remark on computational time is also due. While BCPD takes around 10 seconds per shape with acceleration and 20 minutes without, SFGP currently requires around 100 minutes. However, this is an unoptimized version and does not employ any method to deal with the high dimensionality of the data, to which GPR is sensible. On one hand, the same acceleration used to obtain P in BCPD will reduce computational time. On the other, several GPR tools for large data settings already exist and could be included to reduce time complexity. Both of these should be contemplated in future work. **AQ7**



(a) Ground truth shape. (b) Closest Point. (c) Our method, SFGP.

Figure 8. *Lateral view of shape fitting for a 3D ear shape. On the left, the ground truth mesh is represented in yellow and the simulated ear as a point cloud in red (with noise, missing data, and outliers). In the middle, the result obtained from the GP with closest point approach is depicted as a blue mesh. On the right, the result from the application of SFGP is represented in green. The middle approach presents three main limitations, indicated by white arrows. Near the top, it fails to completely fill the front part, which our method can only partly overcome. However, large improvements are seen in the two bottom problematic regions. The large missing area (middle arrow) and the presence of data only for the front part of the ear (bottom arrow) both cause the reference to collapse for the middle approach. SFGP is able to overcome these challenges, leading to an increased resemblance to the original shape.*

6. Concluding remarks and future work.

We developed a method that bridges the gap between the Gaussian process framework used in 3D morphable models and the probabilistic registration methods, by formulating the shape fitting problem in a GPR multiannotator setting. This allows us to benefit from advantages on both sides and obtain a method particularly suited for shape fitting in the presence of extensive missing data—a useful tool for challenging shapes such as the human ear.

Naturally, even if the missing points are correctly identified, the shape prediction in those regions will be as good as the prior model. Therefore, it is beneficial to have a more complex and accurate model, able to express more knowledge regarding the particular shape. As stated, the GP framework offers a very suitable setting, with kernels expressing intuitive properties of the shapes. So, defining a more appropriate kernel is the logical next step. It would also be pertinent to study how the parallel with the probabilistic registration holds when we introduce the missing point set and the threshold, to have a more theoretical insight into the properties of our method.

Acknowledgment. Computational resources were provided by the INDACO Platform, which is a project of High Performance Computing at the University of Milan http://www.unimi.it.

REFERENCES

[1] L. Bai, X. Yang, and H. Gao, *Nonrigid point set registration by preserving local connectivity*, IEEE Trans. Cybern., 48 (2018), pp. 826–835.

[2] F. Bernard, L. Salamanca, J. Thunberg, A. Tack, D. Jentsch, H. Lamecker, S. Zachow, F. Hertel, J. Goncalves, and P. Gemmar, *Shape-aware surface reconstruction from sparse 3D point-clouds*, Med. Image Anal., 38 (2017), pp. 77–89.

[3] P. J. Besl and N. D. McKay, *A method for registration of 3D shapes*, IEEE Trans. Pattern Anal. Mach. Intell., 14 (1992), pp. 239–256.

[4] V. Blanz and T. Vetter, *A morphable model for the synthesis of 3D faces*, in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 2002.

[5] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, *Variational inference: A review for statisticians*, J. Amer. Statist. Assoc., 112 (2017), pp. 859–877.

[6] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniahy, and D. Dunaway, *A 3D morphable model learnt from 10,000 faces*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 5543–5552.

[7] A. Brunton, T. Bolkart, and S. Wuhrer, *Multilinear wavelets: A statistical shape space for human faces*, in Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Comput. Sci. 8689, 2014, pp. 297–312.

[8] H. Chui and A. Rangarajan, *A new algorithm for non-rigid point matching*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2000, pp. 44–51.

[9] H. Dai, N. Pears, and W. Smith, *A data-augmented 3D morphable model of the ear*, in Proceedings of the 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), 2018, pp. 404–408.

[10] P. E. Fadero and M. Shah, *Three dimensional (3D) modelling and surgical planning in trauma and orthopaedics*, Surgeon, 12 (2014), pp. 328–333.

[11] Z. Fan, X. Hu, C. Chen, and S. Peng, *Dense semantic and topological correspondence of 3D faces without landmarks*, in Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[12] Z. Fan, X. Hu, C. Chen, and S. Peng, *Boosting local shape matching for dense 3D face correspondence*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10936–10946.

[13] Z. Fan, S. Peng, and S. Xia, *Towards Fine-Grained 3D Face Dense Registration: An Optimal Dividing and Diffusing Method*, arXiv:2109.11204, 2021.

[14] P. Groot, A. Birlutiu, and T. Heskes, *Learning from multiple annotators with Gaussian processes*, in Proceedings of the International Conference on Artificial Neural Networks, 2011, pp. 159–164.

[15] T. Heimann and H.-P. Meinzer, *Statistical shape models for 3D medical image segmentation: A review*, Med. Image Anal., 13 (2009), pp. 543–563.

[16] M. Hein and O. Bousquet, *Kernels, Associated Structures and Generalizations*, Technical report 127, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, 2004.

[17] O. Hirose, *Acceleration of non-rigid point set registration with downsampling and Gaussian process regression*, IEEE Trans. Pattern Anal. Mach. Intell., 43 (2021), pp. 2858–2865.

[18] O. Hirose, *A Bayesian formulation of coherent point drift*, IEEE Trans. Pattern Anal. Mach. Intell., 43 (2021), pp. 2269–2286.

[19] P. Huber, Z. Feng, W. Christmas, J. Kittler, and M. Raetsch, *Fitting 3D morphable models using local features*, in Proceedings of the IEEE International Conference on Image Processing (ICIP), 2015, pp. 1195–1199.

[20] D. Jiang, Y. Jin, F. Zhang, Z. Zhu, Y. Zhang, R. Tong, and M. Tang, *Sphere Face Model: A 3D Morphable Model with Hypersphere Manifold Latent Space*, https://arxiv.org/abs/2112.02238, 2021. **AQ8**

[21] M. R. Koujan and A. Roussos, *Combining dense nonrigid structure from motion and 3D morphable models for monocular 4D face reconstruction*, in Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production, CVMP '18, 2018.

[22] H. Laga, *A survey on nonrigid 3D shape analysis,* in Academic Press Library in Signal Processing: Image and Video Processing and Analysis and Computer Vision, Academic Press, New York, 2018, pp. 261–304.

[23] D. C. Liu and J. Nocedal, *On the limited memory BFGS method for large scale optimization*, Math. Program., 45 (1989), pp. 503–528.
[24] W. Liu, H. Wu, and G. S. Chirikjian, *LSG-CPD: Coherent point drift with local surface geometry for point cloud registration*, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 15273–15282.
[25] M. Lüthi, C. Jud, and T. Vetter, *A unified approach to shape model fitting and non-rigid registration*, in Proceedings of the 4th International Workshop on Machine Learning in Medical Imaging, Lecture Notes in Comput. Sci. 8184, Springer-Verlag, Berlin, 2013, pp. 66–73.
[26] M. Lüthi, T. Gerig, C. Jud, and T. Vetter, *Gaussian process morphable models*, IEEE Trans. Pattern Anal. Mach. Intell., 40 (2018), pp. 1860–1873.
[27] J. Ma, J. Zhao, and A. L. Yuille, *Non-rigid point set registration by preserving global and local structures*, IEEE Trans. Image Process., 25 (2016), pp. 53–64.
[28] B. Maiseli, Y. Gu, and H. Gao, *Recent developments and trends in point set registration methods*, J. Vis. Commun. Image Represent., 46 (2017), pp. 95–106.
[29] C. Micchelli and M. Pontil, *On learning vector-valued functions*, Neural Comput., 17 (2005), pp. 177–204.
[30] Z. Min, J. Liu, L. Liu, and M. Q.-H. Meng, *Generalized coherent point drift with multi-variate Gaussian distribution and Watson distribution*, IEEE Robot. Autom. Lett., 6 (2021), pp. 6749–6756.
[31] A. Myronenko and X. Song, *Point set registration: Coherent point drift*, IEEE Trans. Pattern Anal. Mach. Intell., 32 (2010), pp. 2262–2275.
[32] A. Patel and W. Smith, *Exploring the identity manifold: Constrained operations in face space*, in Proceedings of the European Conference on Computer Vision (ECCV), 2010, pp. 112–125.
[33] A. Patel and W. A. P. Smith, 3*D morphable face models revisited*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1327–1334.
[34] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, *A* 3*D face model for pose and illumination invariant face recognition*, in Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 296–301.
[35] S. Ploumpis, E. Ververas, E. O'Sullivan, S. Moschoglou, H. Wang, N. Pears, W. Smith, B. Gecer, and S. Zafeiriou, *Towards a complete* 3*D morphable model of the human head*, IEEE Trans. Pattern Anal. Mach. Intell., 43 (2021), pp. 4142–4160.
[36] S. Ploumpis, H. Wang, N. Pears, W. A. P. Smith, and S. Zafeiriou, *Combining* 3*D morphable models: A large scale face-and-head model*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
[37] S. Racković, C. Soares, D. Jakovetić, Z. Desnica, and R. Ljubobratović, *Clustering of the blendshape facial model*, in Proceedings of the 29th European Signal Processing Conference (EUSIPCO), 2021, pp. 1556–1560.
[38] S. Ramanathan, A. Kassim, Y. Venkatesh, and W. S. Wah, *Human facial expression recognition using a* 3*D morphable model*, in Proceedings of the International Conference on Image Processing, 2006, pp. 661–664.
[39] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, MIT Press, Cambridge, MA, 2005.
[40] Y. Sahillioğlu, *Recent advances in shape correspondence*, Vis. Comput., 36 (2020).
[41] M. Toews, D. L. Collins, and T. Arbel, *A statistical parts-based appearance model of inter-subject variability*, Med Image. Comput. Comput. Assist. Interv., 9 (2006), pp. 232–240.
[42] F. Valdeira, R. Ferreira, A. Micheletti, and C. Soares, *From noisy point clouds to complete ear shapes: Unsupervised pipeline*, IEEE Access, 9 (2021), pp. 127720–127734.
[43] M. Wainwright and M. Jordan, *Graphical models, exponential families, and variational inference*, Found. Trends Mach. Learn., 1 (2008), pp. 1–305.
[44] A. Zhang, Z. Min, J. Pan, and M. Q.-H. Meng, *Robust and accurate point set registration with generalized Bayesian coherent point drift*, in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 516–523.
[45] H. Zhu, B. Guo, K. Zou, Y. Li, K.-V. Yuen, L. Mihaylova, and H. Leung, *A review of point set registration: From pairwise registration to groupwise registration*, Sensors, 19 (2019), 1191.