Learning Shape Priors from Pieces

Dennis Madsen[(⊠)], Jonathan Aellen, Andreas Morel-Forster, Thomas Vetter, and Marcel Lüthi

Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland
{dennis.madsen,jonathan.aellen,andreas.forster,
thomas.vetter,marcel.luethi}@unibas.ch
https://gravis.dmi.unibas.ch

Abstract. Point Distribution Models (PDM) require a dataset in which point-to-point correspondence between the individual shapes has been established. However, in the medical domain, minimising radiation exposure and pathological deformations are reasons why healthy anatomies are often only available as partial observations. To exploit the partial shapes for learning shape models, previous methods required at least a few complete shapes and, either a robust registration method or a robust learning algorithm. Our proposed method implements the idea of multiple imputations from Bayesian statistics. We learn a PDM from a dataset consisting of *only* incomplete shapes and a single full template. For this, we first estimate the posterior distribution of point-to-point registrations for each partial observation. Then we construct the PDM from the set of registration distributions. We quantitatively evaluate our method on a 2D dataset of hands and a 3D dataset of femurs with known ground-truth. Furthermore, we showcase how to use our method on only partial clinical data to build a 3D statistical model of the human skull. The code is made open-source and the synthetic dataset publicly available.

Keywords: Statistical Shape Models · Point Distribution Models · Probabilistic registration · Multiple imputation

1 Introduction

Statistical Shape Models (SSMs) are a well-established tool for medical image analysis. They can be used to automatically quantify whether a given shape is anatomically normal, or how abnormal the shape is. This can e.g.. be used in image segmentation tasks as regularisation. Also, SSMs can be used to reconstruct the complete shape from a partial observation, which is useful for forensic investigation, reconstructive surgeries or patient-specific implant design.

Recently SSMs have been incorporated into deep learning pipelines as deformation regularisers. In [25], an SSM of the right ventricle (RV) chamber is used

---

Code available at https://github.com/unibas-gravis/shape-priors-from-pieces.

in the loss function when training a U-NET. In [27], a fully connected neural network is trained to regress the parameters of an SSM from two cardiovascular magnetic resonance (CMR) image views and patient metadata. In [28], an SSM is used to learn a 2D to 3D mapping of liver images. A direct usage of an SSM is shown in [24], where SSMs are used to reconstruct the facial surface from 2D RGB images and the skull structure from CT images. The classical SSM can also be extended to incorporate patient-specific information [2]. As an alternative to performing a Principal Component Analysis (PCA) decomposition, SSMs can also be build to incorporate the non-linear relationship between shapes [1], while still maintaining performance capabilities as a linear model.

In this work, we make use of Point Distribution Models (PDM) which is a type of SSM. PDMs provide inherent correspondence, which we consider to be an important property for a lot of automatic analysis beyond segmentation. Other models without a point correspondence assumption, are e.g.. SSM based on level sets [23] or non-parametric shape priors [11].

In all of the above-mentioned papers, the PDMs are built from complete and healthy shapes in point-to-point correspondence. However, in the medical domain, either the data is captured because there is a pathology, or it is scanned only partially to capture the essential part of the structure while minimising the radiation danger. Hence, usually, only a part of the healthy anatomy is observed. Our paper is motivated by a practical example where we need to build a skull PDM from children Computed Tomography (CT) images. To minimise the radiation exposure during the image acquisition, only partial scans are taken depending on the area of interest. Several methods address the construction of PDMs where some of the training shapes are partial. In [10], each landmark is assigned a probability of being an outlier. This is used to compute a mean of the dataset where landmarks with lower probability of being an outlier have more influence on the mean shape. In [13], the training surfaces are divided into patches and each patch is assigned a probability of being an outlier. The outlier detection is performed with PCOut [3] which identifies samples that do not fit well into the distribution. Probabilistic PCA [22] is then used to iteratively build the PDM and replace the outlier parts with healthy parts. In [7], the shape model is computed using robust PCA (RPCA) to be able to marginally improve the model through partial data. In [15], they extend this idea to have a probability of being an outlier assigned for each landmark in a shape. The authors then extend their method to a robust kernel SSM, to have a non-linear model for better compactness [16]. All of the aforementioned methods use off-the-shelf registration methods and focus on building the models robustly from the noisy registration results. Often the registrations shrink substantially where parts are missing, to which the learning algorithm then has to be robust. The majority of the methods decompose the data matrix into a low-rank matrix containing correct data and a sparse matrix with the corrupted data. These matrices are mainly found via convex optimisation and usually require non-corrupted data to be present in the dataset.

In contrast, we introduce a principled way to go from *only* partial data observations to a PDM, without the need for multiple intermediate steps. We do not need to detect if parts are outliers or to assign weights for each landmark in a shape. Our method has its foundation in how missing data is handled in Bayesian statistics. There are three ways to handle missing data [8]:

– Discard observations with any missing values.
– Rely on the learning algorithm to handle missing values in the training phase.
– Impute all missing values (complete the data-matrix before using any learning algorithm).

From [4] we know that if we choose a reasonable missing-data model, the imputed dataset is likely to provide a more accurate estimate than a strategy which discards the data with missing values. In our setting where we are only working with partial data, it is not even a possibility to discard the data. For data imputation, a range of different strategies exist. The simplest strategy is to impute the missing values with the mean or the median of the non-missing values for that feature. This approach assumes that the missing values were known in the complete-data and will bias the variance of the dataset towards zero. One would, in the worst case, end up only learning the mean shape of the dataset and nothing about the shape variance. A better option for imputation is to perform regression based on the remaining dataset. This usually overestimates correlations which are then reflected by the model, as no uncertainty is given to the missing part. Instead of single imputation, we can make use of multiple imputations. Multiple imputation was initially introduced in [20] to fill out non-responses in surveys. If the data is arbitrary missing, the Markov chain Monte Carlo (MCMC) method can be used to create multiple imputations by simulating draws from a Bayesian predictive distribution given the partial data [26].

In this paper, we propose to combine probabilistic registration and the idea of multiple imputations to build PDMs purely from partial data. For each partially observed shape, we estimate the posterior distribution of registrations using an MCMC framework. The posterior distribution, which we obtain as a set of samples, reflects not only the uncertainty in the registration of the partial data, but at the same time also contains multiple completions. These completions can be seen as multiple imputations. The PDM can then be computed using standard PCA on the complete data-matrix which contains the imputed samples for each partial data item. Quantitative experiments are performed on a 2D dataset of hands and a 3D dataset of femurs with known ground-truth. Finally, we show how a 3D skull model can be build from partial data. The main contributions of this paper are:

– To the best of our knowledge, we present the first method that learns PDMs from *only* partial data.
– We show how to extend the classical Bayesian statistical method on missing data to point-to-point registration of partial data.
– We show that multiple data imputation creates PDMs with better specificity and generalisation than if single imputation is used.
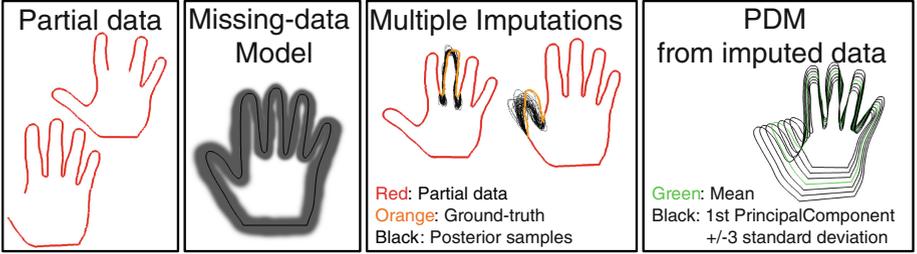
Fig. 1. Overview of our method. Given the partial data (red), the missing-data model (GPMM) is used to draw samples from the posterior distribution over registrations to impute the data. Finally, multiple imputations from each partial data item are used to compute the PDM. (Color figure online)

2   Method

A PDM is computed by performing PCA on the data matrix. As our dataset consists of partial observations, some of the entries are missing in the data matrix. Furthermore, we do not assume correspondence of the observed parts, so we do not even know the position in the matrix for the partial data. Our method implements the idea of multiple imputation from Bayesian statistics to complete the data matrix, while at the same time establishing correspondence in the dataset. For this, we independently process the M data items, each representing a partially observed shape. In contrast to previous methods, we are not only looking for the most likely imputation given a missing-data model but instead, we infer the full posterior distribution of completions

$$P(\boldsymbol{\alpha}|\boldsymbol{s}_p) = \frac{P(\boldsymbol{s}_p|\boldsymbol{\alpha})P(\boldsymbol{\alpha})}{\int P(\boldsymbol{s}_p|\boldsymbol{\alpha})P(\boldsymbol{\alpha})d\boldsymbol{\alpha}}, \tag{1}$$

where $\boldsymbol{s}_p$ denotes the partially observed shape and $\boldsymbol{\alpha}$ is the parameter vector controlling the imputation model. Using the full posterior, we account for the uncertainty of the registration *and* the reconstruction of missing areas when computing the PDM.

From the posterior distribution $P(\boldsymbol{\alpha}|\boldsymbol{s}_p)$, we randomly sample L imputations, such that the inferred data-matrix will be $X \in \mathbb{R}^{DN \times ML}$. Here N is the number of points which is multiplied with the dimensionality D of the embedding Euclidean space, $D = 2$ for the hand dataset and $D = 3$ for the femur and the skull datasets. By including multiple imputations into the PDM, we are able to take the uncertainty of the imputation under our missing-data model into account. An overview of our method is visualised in Fig. 1.

Alternatively, PCA can be performed directly on a set of distributions [6]. Instead of sampling L items from each distribution, they directly use the posterior distribution. This would help to scale the method if we are working in an even higher-dimensional space, where several thousands of imputations would

be needed for each data observation. The downside of this approach is that a Gaussian assumption is made on the noise distribution.

We use a Gaussian Process Morphable Model (GPMM) [14] as the missing-data model $P(\boldsymbol{\alpha})$. A GPMM is defined on the domain of a template mesh $\boldsymbol{s}_t$ of the object class. We define a distribution of deformation fields $u \sim GP(\mu, k)$ following a Gaussian Process GP. The distribution over deformations induces a distribution over shapes when used to warp the template mesh. The mean function μ is set to a $\boldsymbol{0}$ deformation, and the kernel function $k : \boldsymbol{s}_t \times \boldsymbol{s}_t \to \mathbb{R}^{D \times D}$ can be analytically defined, where D is the dimension of the data. As the kernel function can be analytically defined, we do not even need a large dataset to estimate our missing-data model from. A simple covariance function to use (for 3D) is:

$$k(x, x') = \begin{bmatrix} g(x, x') & 0 & 0 \\ 0 & g(x, x') & 0 \\ 0 & 0 & g(x, x') \end{bmatrix} \tag{2}$$

with $g(x, x')$ being a Gaussian kernel

$$g(x, x') = s \cdot \exp(\frac{-\|x - x'\|^2}{\sigma^2}). \tag{3}$$

In the GP framework, multiple simple kernels can be combined to provide richer priors. A statistical covariance kernel estimated from data can be used or they can be based on expert knowledge about the shape deformations of a targeted class, such as smoothness or symmetry.

Furthermore, the template mesh can either be a single full data-item from the dataset, as in our hand and femur example, or it can be handcrafted, as in our skull model example. By choosing a handcrafted template and an expert-designed kernel, we can remove the need for even a single complete example in our skull experiments.

We use a low-rank approximation of the GP, based on the truncated Karhunen-Love expansion as in [14], in order to reduce the computational complexity of the model. This gives us a linear, parametric model ruled only by a coefficient vector $\boldsymbol{\alpha} \sim \mathcal{N}(0, 1)$. The deformation model is then defined as

$$u[\boldsymbol{\alpha}](x) = \mu(x) + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i} \phi_i(x), \alpha_i \sim \mathcal{N}(0, 1), \tag{4}$$

where r is the number of retained basis functions used in the approximation and λ_i, ϕ_i are the i-th eigenvalue and eigenfunction of the covariance operator associated with the kernel function k.

The posterior distribution as described in Eq. (1) cannot be obtained analytically. But, we can evaluate the unnormalized posterior value for any shape described by the model coefficients $\boldsymbol{\alpha}$:

$$P(\boldsymbol{\alpha}|\boldsymbol{s}_p) \propto P(\boldsymbol{s}_p|\boldsymbol{\alpha})P(\boldsymbol{\alpha}). \tag{5}$$
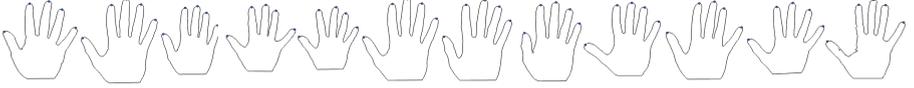
Fig. 2. The dataset of hands with marked fingertips.

Using the Metropolis-Hastings (MH) algorithm [9], we can draw samples from the posterior, based only on the point-wise evaluation of the posterior.

For the MH algorithm, we only need to specify a likelihood function and a proposal generator. The prior $P(\boldsymbol{\alpha})$ is a standard multivariate normal distribution induced by the low-rank approximation. As likelihood model $P(\boldsymbol{s}_p|\boldsymbol{\alpha})$, we use an independent point evaluator, as also used in [19]:

$$P(\boldsymbol{s}_p|\boldsymbol{\alpha}) = \prod_{i=1}^{n} \mathcal{N}(d_{l^2}(\boldsymbol{s}_p^i, \boldsymbol{\alpha}^i); 0, \sigma_{l^2}^2). \qquad (6)$$
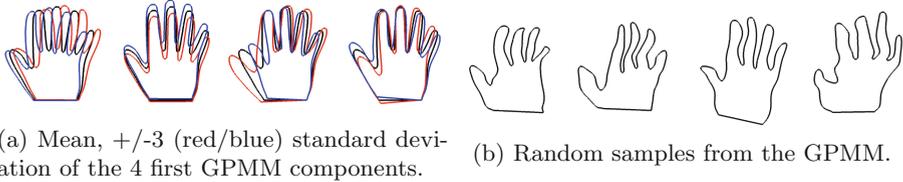
Here, d_{l^2} is the L2 distance between the i-th point from the target $\boldsymbol{s}_p^i$ and its closest point on the model instance denoted by $\boldsymbol{\alpha}^i$. We expect the observed deviations to follow a normal distribution with 0 mean and $\sigma_{l^2}^2$ variance.

The most common proposal distribution is the standard random-walk sampler. However, this is known to have slow convergence and long lag times (time between independent samples) in high dimensional spaces, as is our model space. To overcome this, we take advantage of the recent development in geometry-aware proposal strategies [17,18], which integrates correspondence estimation into the proposal step. For the exact details of the proposal distribution we refer to the provided implementation. Furthermore, we also make use of manually defined correspondences (clicked landmarks) to stabilise convergence.

For the partial data in this work, we assume that the data is a subset of the model, i.e. no overgrown regions from e.g.. metal implants or cancerous growth. Those artefacts would need to be manually removed and instead be completed by our methods.

3 Experiments

In the following experiments, we first demonstrate our method on a synthetic dataset of 2D hands and a 3D clinical dataset of femurs with artificially removed parts. As we have complete hands and femurs, we will compare our method with the ground-truth PDMs. PDMs (M) are usually evaluated based on three measurements: specificity S(M) (evaluate if the model only generates instances that are similar to those in the training set), generalisation G(M) (the ability to describe instances outside of the training set) and compactness C(M) (a model's ability to use a minimal set of parameters to represent the data variability) as described in [21]. From these measures, specificity and generalisation are the most important measures when using the model as prior information in other

(a) Mean, $+/-3$ (red/blue) standard deviation of the 4 first GPMM components.

(b) Random samples from the GPMM.

Fig. 3. Visualisation of the analytically defined 2D hand-GPMM. (Color figure online)

learning algorithms. We want the model to stay within the shape space, but also to be able to explain new data from the same shape space. For most of the experiments we do not plot the compactness as this is of less importance and very similar for all of the models. Finally, we apply our method to build a skull PDM from purely partial data.

The target meshes in all the experiments have been initially landmark aligned to a template mesh. For the synthetic experiments where parts are cut from the target meshes, we still keep the global alignment to avoid factoring pose difference into the model comparison.

All experiments are implemented in the open-source library Scalismo[1] and made publicly available. The only exception to the publicly available data and code is the partial skull dataset.

3.1   2D Synthetic Hand Experiment

For the 2D hand experiment, we make use of 12 synthetic 2D hand meshes as the targets, visualised in Fig. 2. An additional hand mesh is used as the reference mesh to construct a hand-GPMM as shown in Fig. 3. The computation of the generalisation measure for the hand models are computed by a "leave-one-out" approach. For the specificity measure, we use 1000 random samples.

The kernel used for creating the hand-GPMM is a mixture of smooth Gaussian kernels and an expert-designed kernel which is used to separately move the fingers from side to side. With the *finger* kernel we showcase how the incorporation of domain information can outperform standard kernels.

A Gaussian kernel favours smooth deformations with strong correlations between nearby points. However, this introduces a strong correlation between nearly touching sides of two neighbouring fingers, while at the same time both sides of one finger move more independently. The finger kernel aims at allowing side movements of a whole finger while preserving the overall finger shapes. An illustration of the finger kernel is visualised in Fig. 4.

We define the finger kernel for each finger individually. For this, we mark the start of each finger on both sides as A and B as well as the fingertip as C. We then define the kernel as:

$$k(x, x') = s(x, x') \Sigma e^{-\frac{1}{2\sigma^2} d_{ABC}(x, x')^2} \ . \tag{7}$$

---

[1] https://scalismo.org.
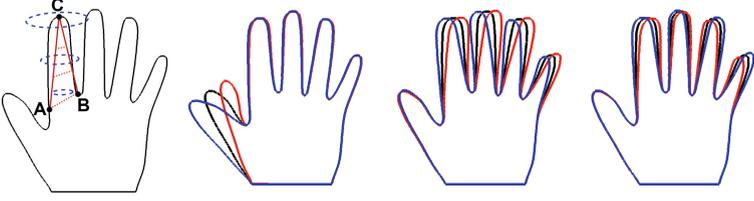
Fig. 4. Illustration of the expert-designed finger kernel. The left image shows the helper points to construct this kernel for the index finger: helper lines (red), lines connecting x and x' where $d_{ABC}(x, x') = 0$ (orange) and illustrations of the scaled covariance matrices $s(x, x')\Sigma_i$ (blue) as overlays. The three visualisations on the right show how the first three principal components of a finger kernel only model encode side-ways moving fingers. (Color figure online)

(a) 5   (b) 10   (c) 15   (d) 20   (e) 30   (f) 40   (g) 50   (h) 60   (i) 70   (j) 80
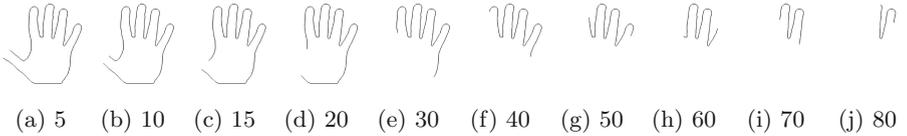
Fig. 5. Hand example of partial data. The caption lists how much of each shape has been removed (5–80%) starting from the fingertip of the thumb.

Given the closest points of x and x' on the set of lines AC and BC, $d_{ABC}(x, x')$ denotes the difference of the first entry in their barycentric coordinates. The covariance Σ is constructed by first defining a 2×2 diagonal matrix, in our example using the values 0.1γ and γ respectively. We then rotate the diagonal matrix by the angle of the x-axis and the direction MC, with M as the mid-point of AB. The scaling function $s(x, x')$ is the barycentric interpolated value between 0 and 1 using the product of the barycentric coordinates of the closest points, or 0 if x and x' do not map to the same finger. The complete expert-designed kernel is then the sum over all five finger kernels specified on the template.

Our choices of γ and σ can be seen in the published code.

Missing Finger Experiments. We clip the hands' dataset starting from a landmark on the top of each finger and cut away increasing amount (5–90%), as visualised in Fig. 5. We perform 2 different experiments with this setup. In the first experiment, a random finger is increasingly cut away from each of the hands. In the second experiment, all 12 hands are increasingly missing the thumb. In Fig. 6 we show the model measures of the missing thumb experiment. Due to space constraints, we only show the measures from this experiment, as the results for both of these experiments are very similar. We see that the models computed from multiple imputations are able to generalise much better than when using the Maximum a posteriori estimations (MAP) or mean solutions. Note, that the curve flattens for the MAP and the mean experiments as only 12
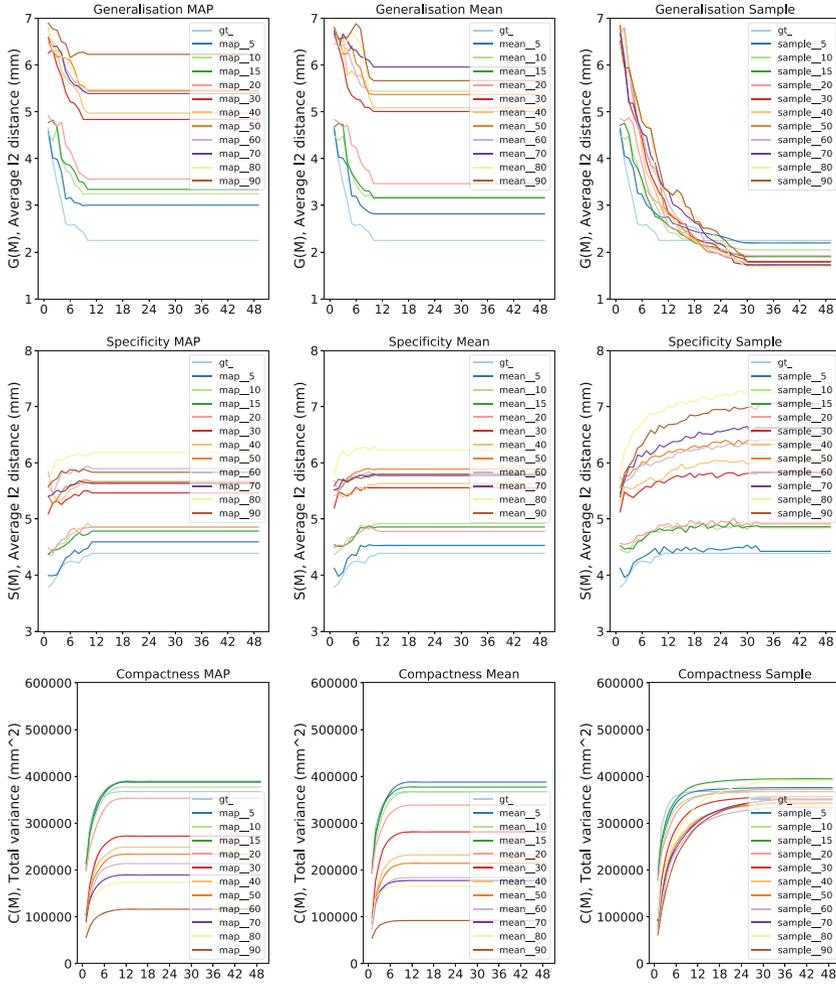
Fig. 6. Hand model measures. The left column shows the results for models computed with the Maximum a posteriori estimation (MAP) from each posterior distribution. The middle column is computed using the mean mesh from each posterior distribution. The right column shows the results using 100 imputations from each of the 12 posterior distribution. The last-mentioned model is clearly superior.

meshes are used to compute these models, whereas we use 100 imputations for each of the 12 target meshes in the multiple imputation experiment. If less than 30% of the meshes are missing, then we also obtain a better specificity measure compared to the MAP and mean solutions. Finally, we see that the multiple imputations are able to maintain much more of the total variance within the model when large parts of the targets are missing.

(a) Models including finger kernels.   (b) Models only from Gaussian kernels.
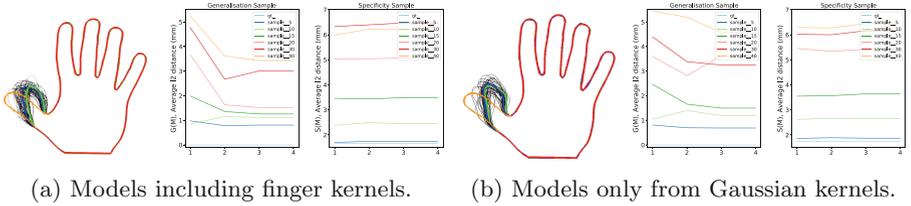
Fig. 7. Finger kernel ablation study. Red: Partial shape, Orange: Ground-truth shape, Green: MAP, Blue: Mean, Black: Random samples. The model using the finger kernel performs better than the purely Gaussian kernel model. (Color figure online)

Finger Kernel Ablation Study. In the ablation study, we demonstrate that the inclusion of expert knowledge into the kernel design improves performance slightly. We use a dataset of 5 identical meshes where only a small scaling difference is applied. When computing the generalisation measure from a leave-one-out approach, we therefore, know that the ground-truth model will be able to perfectly describe the mesh which is not in the model. From Fig. 7 we see that even though the posterior samples from the models look similar, the model with the finger kernels overall generalises better. Especially when a large part of the target is missing (>15%). The specificity measures are also marginally better by using finger kernels. The main takeaway from the ablation study is (not surprisingly), that the more expert knowledge we put into the design of a missing-data model, the better imputations we get.

3.2   3D Femur Experiment with Ground-Truth

For the initial femur-GPMM, we use a smooth Gaussian kernel as in [17]. In this experiment, we assume no additional domain knowledge. The template mesh and random samples from the femur-GPMM are shown in Fig. 8a. We use the publicly available dataset of 50 complete femur meshes that were extracted from computed tomography images[2]. We choose ten complete femur meshes as our training data, from where we clip a varying amount (5–25%) at different landmark locations. For eight of the bones, this clipping was done in a single location. For the remaining two, we spread the removal over four locations. The partial dataset is visualised in Fig. 8b. Similar to before we build PDMs from multiple imputations, the MAP solution and the mean and compare them additionally to the ground-truth PDM. For the specificity computation, the ground-truth registrations are used (these are also used to build the ground-truth PDM). For the generalisation, we use 40 registered meshes which are not included the dataset of the PDMs.

In Fig. 10 we see an example of a target shape (red) with multiple different imputations. We see how both the mean and the MAP solutions fail to perfectly predict the true shape. From the random samples we see that there is a broad

---

[2] Available at the SICAS Medical Image Repository [12].
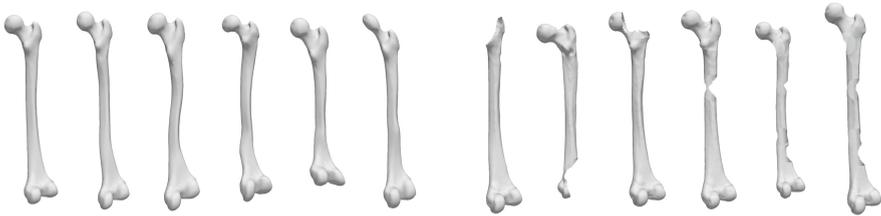
(a) Template (left) and random samples.   (b) Examples of the partial femurs.

Fig. 8. Illustration of random femur-GPMM samples and the used partial data.
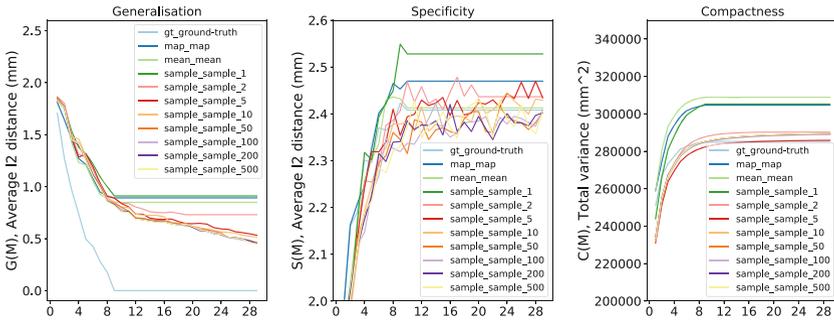
Fig. 9. Femur model measures. Multiple imputations lead to a better generalisation, while also keeping a better specificity than models built from the MAP or mean imputations.
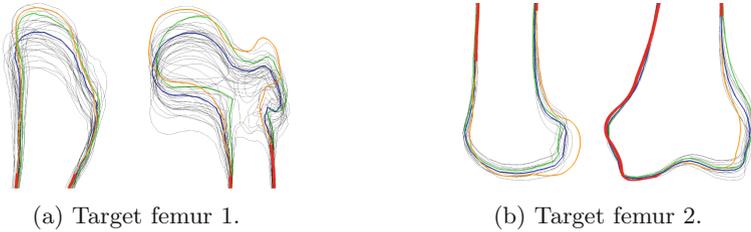

(a) Target femur 1.   (b) Target femur 2.

Fig. 10. Posterior distribution visualisations of the two first partial femurs from Fig. 8b. Red: Partial shape, Orange: Ground-truth shape, Green: MAP, Blue: Mean, Black: Posterior samples. It can be seen that the posterior samples better cover the ground-truth than both, the MAP and mean shape. (Color figure online)

distribution of possible imputations for the missing part. In Fig. 9 we compare PDMs created from the MAP, mean and multiple imputations. In the case of multiple imputations, we compare models created with a different number of samples. From the given example, we see that by using 5 or more imputations, we are able to create PDMs that generalise better, than by using the MAP or the mean solutions, while also maintaining better specificity.

3.3   3D Skull Model from Real Partial Data

In this experiment, we build a skull model from 16 partial skull pieces. For the skull-GPMM, we use a mixture of smooth Gaussian kernels and a symmetrical kernel around the sagittal plane [5]. As we have no ground-truth data, we can only qualitatively evaluate the model. In Fig. 11, examples of the partial data is shown as well as the hand-crafted skull template from which we create a skull-GPMM. We check the model deformations by varying the individual principal components. We see that the first principal component captures the size of the skull, which would also be expected.
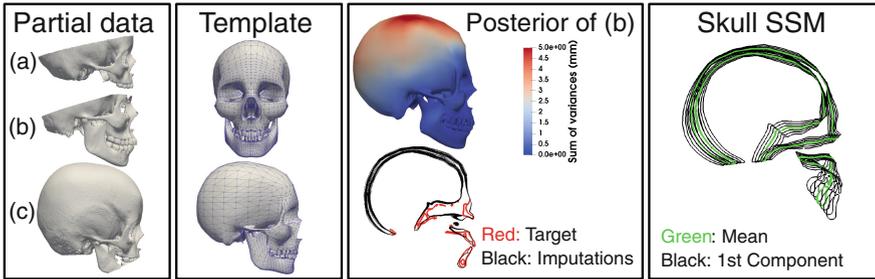
Fig. 11. Skull model experiment. The colour mesh is the point-wise sum of variances. As expected, a larger uncertainty (red) is inferred for the missing part. The visualisation of samples from the posterior in the 2D slice, show the remaining variability (black). To the right, the variation of the 1st principal component from -3 to 3 standard deviation of the resulting PDM is shown. (Color figure online)

4   Conclusion

In this paper, we introduced a principled method to create PDMs from *only* partial data observations. Our method implements the idea of multiple imputation from Bayesian statistics for point-to-point registration of partial data. With this, we can create multiple imputations of a single partial data observation to span the data-matrix. To the best of our knowledge, we are the first to create PDMs from multiple imputations of partial data. We show how this improves model specificity and generalisation. We can influence the imputations by choosing the missing-data model in the MCMC framework. Using Gaussian Process Morphable Models, we can not only design the model in an analytical way to encode smoothness or symmetry but also gradually include more and more data if complete shapes become available. Finally, we showed that our method is not only theoretically nice but that it can also be used in practice to create a PDM from only skull pieces.

References

1. Ambellan, F., Zachow, S., von Tycowicz, C.: A surface-theoretic approach for statistical shape modeling. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 21–29. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_3

2. Egger, B., Schirmer, M.D., Dubost, F., Nardin, M.J., Rost, N.S., Golland, P.: Patient-specific conditional joint models of shape, image features and clinical indicators. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 93–101. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_11

3. Filzmoser, P., Maronna, R., Werner, M.: Outlier identification in high dimensions. Comput. Stat. Data Anal. **52**(3), 1694–1711 (2008)

4. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis. CRC Press (2013)

5. Gerig, T., et al.: Morphable face models-an open framework. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 75–82. IEEE (2018)

6. Görtler, J., Spinner, T., Streeb, D., Weiskopf, D., Deussen, O.: Uncertainty-aware principal component analysis. IEEE Trans. Visual Comput. Graphics **26**(1), 822–831 (2019)

7. Gutierrez, B., Mateus, D., Shiban, E., Meyer, B., Lehmberg, J., Navab, N.: A sparse approach to build shape models with routine clinical data. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 258–261. IEEE (2014)

8. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York (2009)

9. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications (1970)

10. Jingting, M., Lentzen, K., Honsdorf, J., Feng, L., Erdt, M.: Statistical shape modeling from Gaussian distributed incomplete data for image segmentation. In: Oyarzun Laura, C., et al. (eds.) CLIP 2015. LNCS, vol. 9401, pp. 113–121. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-31808-0_14

11. Kim, J., Çetin, M., Willsky, A.S.: Nonparametric shape priors for active contour-based image segmentation. Sig. Process. **87**(12), 3021–3044 (2007)

12. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. J. Med. Internet Res. **15**(11), e245 (2013)

13. Lüthi, M., Albrecht, T., Vetter, T.: Building shape models from lousy data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 1–8. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04271-3_1

14. Lüthi, M., Gerig, T., Jud, C., Vetter, T.: Gaussian process morphable models. IEEE Trans. Pattern Anal. Mach. Intell. **40**(8), 1860–1873 (2017)

15. Ma, J., Lin, F., Honsdorf, J., Lentzen, K., Wesarg, S., Erdt, M.: Weighted robust PCA for statistical shape modeling. In: Zheng, G., Liao, H., Jannin, P., Cattin, P., Lee, S.-L. (eds.) MIAR 2016. LNCS, vol. 9805, pp. 343–353. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-43775-0_31

16. Ma, J., Wang, A., Lin, F., Wesarg, S., Erdt, M.: A novel robust kernel principal component analysis for nonlinear statistical shape modeling from erroneous data. Comput. Med. Imaging Graph. **77**, 101638 (2019)

17. Madsen, D., Morel-Forster, A., Kahr, P., Rahbani, D., Vetter, T., Lüthi, M.: A closest point proposal for MCMC-based probabilistic surface registration. In: Proceedings of the European Conference on Computer Vision (ECCV), August 2020

18. Madsen, D., Vetter, T., Lüthi, M.: Probabilistic Surface Reconstruction with Unknown Correspondence. In: Greenspan, H., et al. (eds.) CLIP/UNSURE -2019. LNCS, vol. 11840, pp. 3–11. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32689-0_1

19. Morel-Forster, A., Gerig, T., Lüthi, M., Vetter, T.: Probabilistic fitting of active shape models. In: Reuter, M., Wachinger, C., Lombaert, H., Paniagua, B., Lüthi, M., Egger, B. (eds.) ShapeMI 2018. LNCS, vol. 11167, pp. 137–146. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04747-4_13

20. Rubin, D.B.: Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In: Proceedings of the Survey Research Methods Section of the American Statistical Association, vol. 1, pp. 20–34. American Statistical Association (1978)

21. Styner, M.A., et al.: Evaluation of 3D correspondence methods for model building. In: Taylor, C., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 63–75. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45087-0_6

22. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **61**(3), 611–622 (1999)

23. Tsai, A., et al.: A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imaging **22**(2), 137–154 (2003)

24. Xiao, D., et al.: Estimating reference bony shape model for personalized surgical reconstruction of posttraumatic facial defects. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11768, pp. 327–335. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32254-0_37

25. Yang, H., Liu, Z., Yang, X.: Right ventricle segmentation in short-axis MRI using a shape constrained dense connected U-Net. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 532–540. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_59

26. Yuan, Y.C.: Multiple imputation for missing data: concepts and new development. In: Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference, vol. 267 (2000)

27. Yue, Q., Luo, X., Ye, Q., Xu, L., Zhuang, X.: Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 559–567. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_62

28. Zhou, X.-Y., Wang, Z.-Y., Li, P., Zheng, J.-Q., Yang, G.-Z.: One-stage shape instantiation from a single 2D image to 3D point cloud. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 30–38. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_4