Confidence regions for statistical model based shape prediction from sparse observations

Rémi Blanc and Gábor Székely

Abstract—Shape prediction from sparse observation is of increasing interest in minimally invasive surgery, in particular when the target is not directly visible on images. This can be caused by a limited field of view of the imaging device, missing contrast or an insufficient signal-to-noise ratio. In such situations, a statistical shape model can be employed to estimate the location of unseen parts of the organ of interest from the observation and identification of the visible parts. However, the quantification of the reliability of such a prediction can be crucial for patient safety. We present here a framework for the estimation of complete shapes and of the associated uncertainties. This paper formalizes and extends previous work in the area by taking into account and incorporating the major sources of uncertainties, in particular the estimation of pose together with shape parameters, as well as the identification of correspondences between the sparse observation and the model. We evaluate our methodology on a large database of 171 human femurs and synthetic experiments based on a liver model. The experiments show that informative and reliable confidence regions can be estimated by the proposed approach.

Index Terms—Statistical shape models, Shape prediction, Uncertainty estimation

I. INTRODUCTION

In the seminal paper of Cootes et al. [1], statistical shape models have been widely used for segmentation purposes, see e.g. [2] and references therein. For about a decade, these models are increasingly used for shape prediction from sparse observations [3,4,5,6,7] or for the prediction of the shape of an organ from that of a neighboring structure [8,9]. Shape prediction is also appealing in the context of intra-operative navigation with imaging devices having a limited field of view such as ultrasound [10]. In other cases, accurate morphological knowledge can only be obtained through histological processing, for example in functional neurosurgery for which even ex-vivo imaging is up to now unable to provide sufficient contrast [11,12]. The predictive capabilities of a statistical shape model are therefore of very high interest for interpolating the entire shape, or at least the parts of interest, from structures that are identifiable in the available images. Nevertheless, and in particular for medical applications where the health of the patient is at stake, the uncertainty related to such a predicted shape can be as important as the prediction itself and should be taken into account for evaluating the associated risks.

A number of papers have investigated the shape variability remaining in a statistical shape model after conditioning it on sparse information [13,14,15], usually assuming a multivariate Gaussian distribution and relying on regularization techniques to overcome the ill-posed nature of the inversion problem. A few recent contributions [16,17,18] have explicitly addressed the question of the estimation and the evaluation of confidence regions around the predicted shape, with the aim to provide quantitative and localized indications about the likelihood of presence of the shape contours around the prediction. A systematic analysis of the sources of uncertainty in model-based shape prediction has been proposed in [18]. However, in these papers the authors rely on the assumption that correspondences between the sparse observations and the statistical model are known together with the pose, which can lead to severe under-estimations of the prediction uncertainty due to the sparsity of the available data.

In this paper, we alleviate such assumptions, and explicitly estimate and incorporate the uncertainties related to pose and correspondence estimation in the pipeline. We only consider that observation data are available in the form of points, contours or surface patches which are known to belong to the organ of interest. In section II, we propose a brief overview of statistical shape models and their use for shape prediction, and present the notation used throughout the paper. The various sources of uncertainty in statistical model based shape prediction, as well as previous work related to their estimation are described in section III, where we also highlight the open issues and explain the contribution of the current paper. We present our approach for the estimation of prediction-related uncertainties in section IV. We first show in section IV.A how existing approaches can be used to incorporate uncertainties related to pose estimation, while section IV.B extends the uncertainty estimation to the case of unknown correspondences. In section V, we review the computation of confidence regions from the predictive distribution and the evaluation of their performance using a set of test cases either from a global, or a case-specific point of view. We also

This work was supported in part by the Swiss NSF, NCCR CoMe. G. Székely is with the Computer Vision Laboratory, ETH Zürich, Switzerland. R. Blanc was with the Computer Vision Laboratory when most of the present work was performed. He is now with the IMS laboratory, University of Bordeaux, France. e-mail: rblanc33@gmail.com.
propose a new case-specific correction, which exploits the relationship between the quality metric of the confidence regions and the matching metric. Experimental results of the proposed methodology are presented in section VI on a database of 171 human femurs, where the quality of the estimated confidence regions is also assessed. A synthetic experiment is proposed in Section VII, which further investigates the properties of the estimated confidence regions and the limits of the method with respect to limited numbers of training samples. Section VIII concludes the paper with a discussion of various aspects of the proposed method.

II. STATISTICAL MODELS AND SHAPE PREDICTION

Statistical Shape Modeling

Let us denote \( z_i, i \in \{1, \ldots, n\} \) the set of \( d \)-dimensional training shapes (usually \( d \in [2,3] \)) represented by \( n \) column vectors. These shape descriptors are, by construction, in anatomical correspondence across the training set, meaning that any point \( k \) from shape \( z_i \) is anatomically corresponding to the same point \( k \) of shape \( z_j \). Methods to obtain such correspondences include minimization of the description length of the model through re-parameterization of curves or surfaces \([19,20]\) or group-wise registration \([21]\). The mean shape \( m \) and the eigenvectors \( U \) related to the non-zero eigenvalues \( \Lambda \) of the covariance matrix are then estimated. A dimensionality reduction step is usually conducted by retaining eigenvectors related to the largest eigenvalues. The sample distribution of the shape parameters \( \theta_0 = U^T(z_i - m) \) is obtained by projecting each training sample \( z_i \) onto the shape space, i.e. the subspace defined by \( U \).

The shape distribution \( P(\theta) \) derived from the training set is usually assumed to follow a multivariate normal (MVN) probability law, with zero mean and covariance matrix \( \Lambda \). Other models can also be employed, such as a non-parametric kernel density \([22]\). For more details on statistical shape modelling, the reader can refer to e.g. \([23]\). By drawing parameters \( \theta \) from the distribution \( P(\theta) \), new plausible shapes can be generated:

\[
\text{z}(\theta) = m + U\theta
\]  

(1)

In practical applications of shape matching, such a model needs to be positioned with respect to the scene. Depending on the applications, different types of pose-related transformations can be considered, typically rigid or affine transformations. We denote the corresponding pose parameters by \( \pi \), which we restrict here to translations and rotations. Indeed, the resolution of 3D medical images is usually known, so that provided the model of the shape is learned whilst preserving the natural dimensions of the organ of interest, no scaling correction is necessary. A complete parametric representation of a shape from the model is therefore written as:

\[
\text{z}(\pi, \theta) = R_\pi (m + U\theta) + T_\pi
\]  

(2)

where \( R_\pi \) and \( T_\pi \) are transformations which apply the same rotation and translation to each point of the shape. The error between a shape representation \( z(\pi, \theta) \) and a gold-standard shape \( z_0 \) is denoted \( \eta = z_0 - z(\pi, \theta) \).

Shape Prediction from Partial Observation

Shape prediction considers the estimation of the complete shape, i.e. the estimation of the parameters \( (\pi, \theta) \), from partial observations denoted \( obs \). Since we consider the case of sparse observation, we strictly enforce solutions which belong to the shape space in order to avoid unlikely results.

In order to guide the prediction, some sort of correspondence has to be established between the observations and the shape model. While approaches relying on implicit correspondences have been proposed, e.g. using level set representations \([24,25]\), we will restrict ourselves to explicit correspondences in this paper. We represent these correspondences through a function \( K(z(\pi, \theta), obs) \), or simply \( K \) in the following, which realizes a mapping between the observation data and a subset of the elements in the shape descriptor \( z(\pi, \theta) \). In practice, the function \( K \) identifies a set of \( n_K \) point indices \( z_K(\pi, \theta) \) (or \( z_K \)) from the shape descriptor \( z(\pi, \theta) \) and associates them to a set of \( n_K \) points denoted \( x_K \) extracted from the observation data \( obs \). If these correspondences are correctly estimated, for any \( i \in [1; \ldots; n_K] \), the i-th point of \( z_K \) anatomically corresponds to the i-th point in \( x_K \). We denote \( m_K \) and \( U_K \) the corresponding subsets of the mean shape and the deformation modes.

Three main strongly interrelated families of approaches have been followed for shape prediction from spatial information: (1) regression-based methods, mainly relying on multi-linear models \([3,7,8]\) such as Principal Component Regression, Partial Least Square regression or Canonical Correlation Analysis; (2) estimation of a conditional distribution, generally a Gaussian model with additional regularization \([14,17]\) or optimization of the number of modes \([18]\), but also using a kernel density model \([22]\); and (3) optimization of the model parameters through the minimization of a metric quantifying the distance between the observation and the model \([4,5,16,25]\).

In the following, we optimize \( (K, \pi, \theta) \) through the minimization of a metric \( D_{obs} \) related to the discrepancy between \( x_K \) and \( z_K \). Furthermore, we are interested in estimating the distribution of the prediction error \( P(\theta|obs) \). When both \( (K, \pi) \) are known, the metric \( D_{obs} = \|x_K - z_K\|_2 \) is linear with respect to the shape parameters:

\[
D_{obs} = \|m_{\pi, K} + U_{\pi, K} \theta - x_K\|_2
\]  

with \( m_{\pi, K} = R_\pi m + T_\pi \) and \( U_{\pi, K} = R_\pi U_K \).
III. PROBLEM SPECIFICATION AND OPEN ISSUES

In [18], three major sources of uncertainties have been considered: (1) the limited representativeness of the statistical model, e.g. due to the small size of the training set or a bias in the selection of the training samples, (2) the limited statistical dependences between the predictor and the shape to predict, and (3) the uncertainties related to the observed predictors, typically observation noise. However, two sources of errors have been ignored when analyzing prediction power which may have similar influence on the quality of the results, namely the uncertainty in the identification of correspondences between the observation and the model, and in the estimation of the pose and shape parameters. In order to clearly identify the present contribution, we first review how these problems have been addressed in the literature, before concentrating on extensions related to pose and correspondence establishment in Section IV.

Model Related Uncertainties

Even if using the optimal parameters \((K_*, \pi_*, \theta_*)\), statistical models such as described above can represent a new shape \(z_0\) only up to a residual error \(\varepsilon = z_0 - z(\pi_*, \theta_*)\). This property, also called the generalization ability of the model [23 p.78], is related to the quality of the shape model itself, in particular how far the training samples represent the population to be described, but also to some extent design choices such as the dimensionality reduction applied to generate a compact model [26]. Considering a shape \(z_0\) and its projection onto the subspace of the shape model, the probability density of the projection error is denoted \(P_k\). As demonstrated in [18], this distribution can be approximated through resampling the available training data, by repetitively training models using subsets of the examples available and computing the projection errors for the left-out shapes. We write the corresponding density:

\[
P(\eta|\text{obs}, K_*, \pi_*, \theta*) \sim P_k
\]

(4)

Limited predictive properties

Limited correlations between the predictors and the shape to predict are an intrinsic problem to shape prediction, for which no solution exists besides using more, or better, predictors. Such issues imply that no single solution exists for a given prediction problem, but rather a probability density of plausible solutions.

Considering the simple case of a relatively low dimensional multivariate distribution \(P(x) = P(x|y)^T\) and known predictors \(x_0\), the conditional distribution \(P(y|x = x_0)\) can be estimated either analytically for simple distribution models, or numerically for more complex models [27]. The conditional expectation typically represents the predicted shape, while the conditional distribution represents the uncertainty.

In the context of shape modelling, with distributions of high dimensionality, the estimation of the conditional distribution of the shape parameters \(\theta\) has been performed using a regularized conditional Gaussian model, as in [15], or a kernel-density model [22]. However in these works, the pose parameters \(\pi_\ast\) as well as the exact correspondences \(K_\ast\) between the observation and the model were assumed to be known, meaning that the conditional distribution calculated in these papers corresponds to \(P(\theta|\text{obs}, K_\ast, \pi_\ast)\). Translating this probability density of the parameters in terms of shape-related density, this corresponds to:

\[
P(\eta|\text{obs}, K_\ast, \pi_\ast)
\]

(5)

In [16], this distribution was estimated in a single step, i.e. without separately estimating \(P_k\), through the following bootstrap procedure. First, the shape prediction is performed (using any method from the literature) on the actual data, \(z_K\) is identified, and a shape is estimated. For each bootstrap sample, a set of \(n\) shapes is drawn with replacement for the training database, and used for learning a statistical model. For each sample not used for learning the bootstrap model, the same prediction algorithm is employed and the predicted shape is compared to the ground truth to obtain the prediction error \(\eta\). The full bootstrap experiment provides a set of prediction errors, from which the density (5) can be estimated.

In order to better take into account the problems related to pose estimation, it is proposed in [17] to re-align the training shapes with respect to the observed landmarks, so that the parametric shape model intrinsically shows a smaller variability in the regions close to the observations, and larger when moving farther away. For example, a statistical model of a human silhouette trained with all samples aligned with the feet as reference will display a large variability in the head area due to inter-individual height variability. Assuming we observe only the head while knowing the pose precisely, we would immediately know the height of the individual because the point coordinates are expressed in the reference space, i.e. with the feet as origin. Re-aligning the model with respect to the eyes reformulates the problem in a more intuitive way, shifting the variability represented in the model toward the feet. However, in our notation this can only be interpreted as a modification of the term \(P_k\) rather than as a proper marginalization over an uncertain pose.

Observation related uncertainties

In both [17] and [18], uncertainties related to the observation noise are also incorporated, relying on approximate knowledge of the distribution of the observation noise \(P(\text{obs})\). Both papers suggest that the observation noise partially accounts for errors in the correspondence establishment. Therefore, the authors propose to model observation errors that are mostly tangential to the shape surface, as the position in the normal direction is typically constrained by the presence of edges in the images.

The noise model is exploited through marginalization over \(P(\text{obs})\) of the conditional distribution of the shape given the
observation of a subset of the landmarks of the models, resulting in:

\[ P(\eta|K_*,\pi_*) = \int P(\eta|obs,K_*,\pi_*)P(obs)\ d\eta \]  

(6)

Finally, the magnitude of the observation related uncertainties still needs to be modelled, as well as the dependences between different shape locations, an issue which has been largely ignored up to now. Consequently, it remains unclear how \( P(\text{obs}) \) should be realistically modelled, and how such choices will influence the final estimation.

**Open problems addressed in this paper**

As already mentioned, no fully convincing solution has been proposed up to now to tackle uncertainties related to the estimation of correspondences between the model and the observation data, nor to the estimation of the pose. This can be particularly problematic in the case of very sparse observations which can lead to considerable uncertainty on the rotations. Especially for elongated shapes, with observations on one end of the shape as in Fig. 2(A), a small uncertainty on a rotation angle can result in large errors at the other extremity of the object.

Likewise, the uncertainty related to the establishment of correspondences has only been approached indirectly, through the modelling of an additive noise on the position of the points of the model that are assumed to be observed. This issue raises particular problems, as a change in the estimated correspondences leads to a change of the goal function \( D_{obs} \).

The objective of this paper is to estimate the prediction uncertainty \( P(\eta|\text{obs}) \) in a data-driven fashion, with the only assumption that the observation is pre-processed and available as points, lines or surface patches lying on the surface of the modelled object. Our contribution compared to previous work is to take into account all uncertainties related to the estimation of correspondences between these observations and the model, and of the pose and shape parameters:

\[ P(\eta|\text{obs}) = \int P(\eta|\text{obs},K,\pi,\theta)P(K,\pi,\theta|\text{obs})\ dK \ d\pi \ d\theta \]  

(7)

Because it is difficult to estimate or to model a realistic distribution of the observation uncertainty \( P(\text{obs}) \) in practice, we will restrict our investigations to the case of reliable observations, and use a simple ridge regularization term (diagonal matrix) with a low weight. Nevertheless, when a more accurate model is available, the solutions presented in [17] or [18] may be employed within the proposed framework.

**IV. UNCERTAINTIES IN STATISTICAL MODEL BASED SHAPE PREDICTION**

We propose an incremental description of the workflow, incorporating first the pose-related uncertainty in section A and the correspondence-related ones in section B. The complete estimation procedure is summarized in Fig. 1.

**A. Unknown pose and known correspondences**

Let us consider first the question of joint pose and shape estimation, and still assume perfectly known correspondences \( K_* \). Contrary to (3), the metric is now non-linear with respect to the rotation parameters. As the partial derivatives of the shape model with respect to both pose and shape parameters can be easily expressed analytically, a first approach is to rely on numerical optimization strategy, such as a Levenberg-Marquardt (LM) optimization, as used e.g. in [10].

Another possibility is to use a linear approximation of the associated transformation, as proposed in [5], and to concatenate the corresponding 'pseudo-eigenvectors' to the original modes of variation: \( \tau = [\pi,\theta] \). Under this approximation, the shape can be expressed as a linear combination of all parameters: \( z(\pi,\theta) = U\tau + m \), where \( \tau = [\pi,\theta] \) is the concatenation of the pose and shape parameters. This allows for an analytical estimation of the optimal parameters:

\[ \tau = \left( U_K^T U_K + \lambda \Lambda' \right)^{-1} U_K^T (s_K - m_K) \]  

(8)

where the matrix \( \Lambda' \), and the scalar \( \lambda \), are regularization terms. Typically, the parameter \( \lambda \) is chosen proportional to the amount of noise in the observation, and \( \Lambda' \) is a diagonal matrix which contains the variance of the shape model parameters as seen on the training set for the \( \theta \)-related terms. In [5], it is proposed that the value corresponding to the largest variance is also used for the pose related terms. More complex regularization terms can be used if specific assumptions on the observation noise are available, as e.g. in [17,18]. Though the linearization of the rotation is valid only for small angles, larger rotations can be coped with by using an iterative scheme in which the pose information is incorporated into the mean shape \( m_{\pi,K} \) and eigenvectors \( U_{\tau,K} \) at each iteration.

With respect to the estimation of uncertainties associated with the shape prediction, the resampling-based approach proposed in [16] and summarized in section III to estimate the density (5), originally in the context of shape parameter estimation alone, accommodates for the incorporation of pose estimation as well. This scheme provides a non parametric, data-driven estimation of the distribution:

\[ P(\eta|\text{obs},K_*) \]  

(9)
between the observation set, we start by estimating the closest vertex on the current model \( z(\pi, \theta) \). From this initial set of corresponding pairs, we reject those which bring multiple observation points to correspond to a single model point, keeping only the closest pair. This correspondence establishment procedure is embedded in a variant of the Iterative Closest Point algorithm [29,30], alternating the estimation of correspondences for specific parameter values, and the simultaneous optimization of pose and shape parameters given the current correspondences, as described in section IV.A. The process is iterated until no significant changes are observed on the metric. Though often employed, such a scheme is known to be sensitive to initialization, and is not guaranteed to converge to the global optimum. Thus, the estimation of correspondences suffers from uncertainties which can influence the precision of the predicted shape, and need to be taken into account. In theory, the resampling-based approach could again be followed for the estimation of unknown correspondences:

\[
P(\eta|obs) = \int P(\eta|obs,K)P(K|obs)dK
\]  

However, this would necessitate to sample the space of possible correspondences using e.g. a Markov Chain Monte Carlo approach [31], and to repeat a cross-validation study as in section IV.A. to estimate \( P(\eta|obs,K) \) for every new set of proposed correspondences \( K \), which would make the estimation computationally intractable.

Nevertheless, as the position and orientation of the patient with respect to the imaging device are generally roughly known, we assume that \( P(\eta|obs,K) \) remains constant as long as the correspondences \( K \) are sufficiently likely. In order to estimate the variability related to the correspondence establishment procedure, i.e. to evaluate \( P(K|obs) \), we repeat the optimization procedure for several random seed parameters \((\pi, \theta)\). A tabu-search heuristic [32] is employed to speed-up the process, by stopping the optimization of the current seed when reaching areas of the search space that are too close to an already visited location. Initial positions are randomly chosen either around a previous local optimum or on a broader range in order to get a good coverage of the search space.

From the \( n \) optimizations performed, we obtain a set of parameters and a set of corresponding metrics \((\pi^{(i)}, \theta^{(i)}, K^{(i)}, D_{obs}^{(i)})_i \). The best result in the set is identified with the optimum \((\pi^{*}, \theta^{*}, K^{*}, D_{obs}^{*})\), and the corresponding shape \( z(\pi^{*}, \theta^{*}) \) is considered as the prediction. Using these optimal correspondences, a single cross-validation prediction experiment is performed in order to estimate \( P(\eta|obs,K^{*}) \), as in section IV.A. On the other hand, the distribution \( P(K|obs) \) is estimated from the set of samples \((\pi^{(i)}, \theta^{(i)}, K^{(i)}, D_{obs}^{(i)})_i \). We selected samples with an associated metric less than 25% higher than the found optimum, and computed a weighted MVN distribution, using weights decreasing with the associated metric. We will discuss this procedure in detail in section VIII. Finally, we approximate the uncertainty related to the prediction through the convolution of both densities:

\[
P(\eta|obs) = \int P(\eta|obs,K)P(K|obs)dK
\]  

Assuming that both distributions \( P(\eta|obs,K^{*}) \) and \( P(K|obs) \) are MVN with zero mean and covariance \( \Sigma_{CV} \) and \( \Sigma_{K} \) respectively, the resulting distribution \( P(\eta|obs) \) will have covariance \( \Sigma_{\eta} = \Sigma_{CV} + \Sigma_{K} \).

V. CONFIDENCE REGION ESTIMATION AND EVALUATION

For the sake of completeness, we briefly summarize below the methodology proposed in [16] to estimate confidence regions from the predictive distribution and to evaluate their accuracy. Additionally, we propose a case-specific correction of the confidence region size, based on previous reports indicating substantial correlations between the quality of the matching of observed parts and the average quality of the confidence regions over the complete shape.

Confidence region estimation

As in [16], we compute a confidence ellipsoid for each landmark \( i \) of the shape, by marginalizing the predictive
distribution at each shape point. Each confidence ellipsoid $C_{ij}$, at significance level $\alpha$ is derived from the corresponding marginal covariance in $\Sigma_j$. Its Mahalanobis ‘radius’ is $D_{ij}^2 = F^{-1}(1 - \alpha, d)$, where $F$ is the cumulative chi-square distribution with $d$ degrees of freedom.

**Evaluation of the quality of confidence regions**

In order to quantify how informative the estimated confidence regions are, their effective coverage probabilities are evaluated. This is performed through the comparison of the nominal and effective frequencies with which an estimated confidence region really contains the corresponding shape point by means of probability versus probability plots (P-P plots) as proposed in [16]. We focus in particular on the shape-wise quality measure:

$$\phi_{ij}^{(\alpha)} = \frac{1}{n_p} \sum_{i=1}^{n_p} I \left( A_{ij}^{(\alpha)} \right)$$

(12)

where $n_p$ is the number of shape points, and $I(A_{ij}^{(\alpha)})$ is the indicator function of the random event taking value 1 when landmark $i$ of shape $j$ is inside the estimated $\alpha$-confidence region, and 0 otherwise.

**Case-specific region correction**

In [18], it was shown that the variance of $\phi_{ij}^{(\alpha)}$ depends on the correlations between the different landmarks of the shape:

$$\text{Var}[\phi_{ij}^{(\alpha)}] = \sigma^2 (1 - \alpha) \frac{2}{n_p} \sum_{i,j} \text{Cov} \left( I(A_{ij}^{(\alpha)}), I(A_{ik}^{(\alpha)}) \right)$$

(13)

For shapes containing large numbers of points, the first term should be negligible. However, the second term, related to the correlation of the prediction error between pairs of points is influenced by parameter mis-estimations (getting stuck in a local minimum during the prediction), or if the test shape is an outlier with respect to the available training set. This is problematic because even if the confidence regions can be reliable ‘on average’, a large variance implies that for a given subject the confidence regions may be too narrow.

Nevertheless, it was reported in [16] that significant correlations exist between the optimal matching metric $D_{obs}$ and the shape-wise quality of the confidence regions $\phi_{ij}^{(\alpha)}$. The fundamental idea behind statistical model based approaches is that test and training samples emerge from the same population, which has to be reasonably approximated by the training data. In this case, if a significant correlation is observed between $D_{obs}$ and $\phi_{ij}^{(\alpha)}$ during the cross-validation learning phase, it should also exist for new test shapes. Therefore, we propose to inflate/deflate the size of the estimated confidence regions through a multiplicative coefficient determined by the matching metric $D_{obs}$, leading to case-specific correction aimed at reducing the variance of $\phi_{ij}^{(\alpha)}$ while maintaining its mathematical expectation. The function relating this correction coefficient to the metric $D_{obs}$ is learned from the training data in the following heuristic way.

From the cross-validation experiment we obtain samples of the joint distribution $(\phi_{ij}^{(\alpha)}|D_{obs}(\alpha)), D_{obs}(\alpha))$ and the Pearson correlation $R = R(\phi_{ij}^{(\alpha)}|D_{obs}, D_{obs}(\alpha))$. We then select a set of representative points along the quantiles of the distribution of the matching metric $D_{obs}$. For each of these points $k$, a scaling factor $c(D_{obs}(k))$ is optimized so that the confidence regions derived from the covariance $\Sigma_{ij} = c(D_{obs}(k)) \Sigma_{CV} + \Sigma_K$ match the nominal value $\phi_{ij}^{(\alpha)} = 0.5$. The function $c(D_{obs})$ is then linearly interpolated so that the correction factor can be applied on any test shape. The selection of the representative points is performed by first clustering the training elements according to the quantiles of the distribution of their matching metric $D_{obs}$. Then, for each cluster, a single representative element $k$ is selected so that it is the closest element of the cluster to the point $(\bar{x}_d, \bar{y}_d)$, where $\bar{x}_d$ is the average $D_{obs}$ value within the cluster, and $\bar{y}_d$ with ordinate the $(1 - R^2/2)$ percentile of the distribution of $\phi_{ij}^{(\alpha)}$ values in the cluster. This selection results in a more conservative correction when the correlation $R(\phi_{ij}^{(\alpha)}|D_{obs}, D_{obs}(\alpha))$ is low, while preserving the original size of the confidence regions when the correlation is high. This is exemplified in Fig. 3(a-b).

**VI. EXPERIMENTAL RESULTS**

We evaluate the above methodology for confidence region estimation on a large database of 171 human femurs. They were segmented from CT images using supervised tools from the Amira software (Mercury systems). Correspondences between the shapes were obtained using a diffeomorphic demons registration [33] with poly-affine regularization [34],
enabling the generation of a Point Distribution Model of the femur surface.

The dataset has been randomly split into a training set of 121 samples, which is used for learning the statistical shape model and performing cross-validation experiments, while the remaining 50 samples were kept exclusively for testing and evaluation purposes. In all experiments, including the cross-validations, statistical shape models are learned keeping 99% of the variance. For cross-validation, we compared Leave-One-Out (LOO) and bootstrap resampling. For each of the 50 bootstrap samples used, 121 shapes are drawn with replacement in the original pool of 121 training samples.

Observation data are obtained by simulating the manual contouring of the femur on a number of slices, as illustrated on Fig. 2 (A,B). Corresponding contours are automatically extracted from the test shapes, and used to estimate the predictive distribution (11). The optimal match \( \mathbf{z}(\mathbf{x}^*, \theta^*) \) is considered as the predicted shape, and confidence regions are derived from \( P(\mathbf{z}_{\text{obs}} \mid \eta^*) \) through confidence ellipsoids around each shape landmark as described in section V.

The estimation of correspondences between the observation data and the shape model, i.e. the function \( f \) from section II, is performed using a two-step variant of the closest point algorithm. First, for each point of the observation data (observed lines are represented as connected points, used to generate a triangulated surface), the closest vertex in the shape model is selected, forming the set \( z_K \). Then, for each point in \( z_K \), the closest point on the observation data (i.e. possibly on a line or a triangle) is selected as the corresponding point, forming the target vector \( x_K \).

With respect to the shape prediction algorithm (section IV.A), we opted for the linear approximation approach using equation (8), due to its faster convergence. However, this scheme proved to be occasionally unstable when the observation data do not impose strong constraints on the rotations, as in Fig. 2(A) with observation restricted at one end of an elongated shape. In such cases, the algorithm was replaced by the slower but more robust LM-based optimization. Actually, when both optimization methods converge, we observed no statistically significant differences in terms of prediction accuracy, however none are immune against getting stuck in local minima.

Concerning the tabu-search heuristic (section IV.B), we performed 100 such optimizations using random initializations. In order to reach a good coverage of the search-space, we alternated between selecting (1) completely new seeds by drawing shape parameters from \( P(\theta) \), translation parameters uniformly in the range \( \pm 2 \text{mm} \) and rotations parameters uniformly in the range \( \pm 2.5^\circ \) around the true values; and (2) seeds around previous local minima, drawing shape parameters within the range \( \pm 0.05 \) (normalized values), translation in the range \( \pm 0.1 \text{mm} \) and rotations in the range \( \pm 0.1^\circ \) around the previous values. Around each visited set of parameters \((\pi,\theta)\), a small box centered on \((\pi,\theta)\) with a side length equal to 0.002% of the parameter value is marked as already visited. In subsequent optimizations, any new point falling within such a region is immediately considered to end on the corresponding local minimum.

The optimal prediction obtained through this procedure was considered as the predicted shape, and the corresponding correspondences were used to estimate \( P(\mathbf{z}_{\text{obs}} \mid \eta^*, K^*) \) using the bootstrap procedure described in section IV.A.

The set of local minima resulting from the optimization was then used to compute a distribution of acceptable prediction results \( P(K \mid \mathbf{z}_{\text{obs}}) \) with respect to the estimation of correspondences, and convolved with \( P(\eta_{\text{obs}} \mid K^*) \) to obtain the final uncertainty estimate \( P(\eta_{\text{obs}}) \), assuming multivariate Gaussian distributions.

**Generalization ability**

The generalization ability of the model constitutes a quantitative baseline about the uncertainty regions. It corresponds to the best possible scenario in which the correspondences, the pose, but also the optimal shape parameters are known. Quantitative values for the average prediction error and the average dimensions of the 95% confidence ellipsoids are given in Tab. I, where a comparison of confidence regions obtained using either a LOO or a bootstrap-based resampling approach is shown.

The results indicate that bootstrap-based confidence regions tend to be slightly larger than LOO-based regions. This is due

### Table I

**Properties of the Estimated 95% Confidence Regions**

<table>
<thead>
<tr>
<th></th>
<th>Average projection error (mm)</th>
<th>Shape-wise effective coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOO</td>
<td>BOOTSTRAP</td>
</tr>
<tr>
<td>Average projection error (mm)</td>
<td>1.21 / 0.28</td>
<td>0.91 / 0.096</td>
</tr>
<tr>
<td>Shape-wise effective coverage</td>
<td>0.95 / 0.062</td>
<td>0.49 / 0.25</td>
</tr>
<tr>
<td>Equivalent radius (mm)</td>
<td>0.69 / 0.35</td>
<td></td>
</tr>
</tbody>
</table>
to the lower number of different training elements used during each cross-validation experiment resulting in slightly larger generalization errors. At the 95% level, bootstrap-based regions appear to be fairly precise and accurate in terms of coverage probability, while the LOO-based estimates are somewhat overconfident. The conclusion is that, not so surprisingly, the distribution of the generalization error has in practice slightly longer tails than those estimated with a Multivariate Gaussian model on the training set. In the following we focus on bootstrap confidence regions.

**Known pose and correspondences**

In this experiment, the pose and correspondences are perfectly known, and do not need to be estimated. This corresponds to the assumptions previously made in e.g. [16]. The two observation settings illustrated in Fig. 2 are investigated. The corresponding 95% confidence regions are also displayed in Fig. 2(a,b), as well as the corresponding P-P plots (Fig. 2(c,d)). Regions that are far away from any observations, and/or for which the natural variability is higher appear to be less reliably estimated, as in observation (A). Nevertheless, quantitative results in Tab. II indicate that the estimated confidence regions are still reliable, and even slightly larger than necessary on average. However, a larger variance is observed in the shape-wise quality of the regions, indicating that when very sparse observations are available, the approach is more sensitive to outliers.

The suggested case-based correction enables to reduce this variability by taking advantage of the correlations between matching metric and the shape-wise quality of the confidence regions learned in the training phase. These correlations and the estimated correction function are illustrated for both observation settings on Fig. 3(a-b). Applying this correction function to the test samples, we obtain new confidence regions. The corresponding P-P plots are given in Fig. 3(c-d).

![Cumulative Variance](image)

**Fig. 5.** (a) sample liver shape, in green: observed areas, in red: areas set as possibly observed (used to constrain the correspondence establishment process). (b) Cumulative distribution of the eigenvalues of the synthetic model.

![Cumulative Distribution](image)

**TABLE II**

| CONFIDENCE REGIONS WITH KNOWN POSE AND CORRESPONDENCES |
|----------------|----------------|----------------|
| CORRECTION | OBSERVATION (A) | OBSERVATION (B) |
| Average prediction error (mm) | 2.71 / 1.84 | 1.27 / 0.30 |
| Shape-wise effective coverage | 0.94 / 0.96 | 0.96 / 0.96 |
| Equivalent radius (mm) | 3.55 / 2.72 | 0.58 / 0.65 |
| (mm) | 4.33 / 8.85 | 0.50 / 0.33 |

**TABLE III**

| CONFIDENCE REGIONS WITH UNKNOWN POSE AND CORRESPONDENCES |
|----------------|----------------|----------------|
| CORRECTION | OBSERVATION A | OBSERVATION B |
| Average prediction error (mm) | 5.51 / 4.71 | 1.25 / 0.33 |
| Shape-wise effective coverage | 0.96 / 0.97 | 0.94 / 0.93 |
| Equivalent radius (mm) | 25.71 / 27.29 | 0.75 / 1.13 |
| (mm) | 30.36 / 34.87 | 0.33 / 0.50 |

Average prediction errors and properties of the associated 95% confidence regions, without and with case-specific corrections, with (Tab. III) or without (Tab. II) incorporating the estimation of pose and correspondences. Equivalent radius stands for the radius of the sphere with a volume equal to that of the confidence ellipsoid. The values in the cells correspond to the average / standard deviation of the quantity indicated in the row title.

**Unknown pose and correspondences**

Finally, we examine the more realistic situations in which no prior information about pose or correspondences between the observation and the model are available. Compared to the previous experiment, the results in Tab. III indicate that both the average prediction error and the size of the confidence regions are much larger in the case of observation (A). This confirms that the estimation of pose and correspondence have a very high impact on the accuracy of the prediction when the pose parameters are not much constrained by the observations. However, in case (B), the observation data are sufficient to enable a reliable estimation of the pose and correspondences, and the prediction performance remains similar to those displayed in Tab. II. In both cases, the method maintains good properties in terms of coverage probability of the confidence regions and reliably quantifies the prediction uncertainties.

**VII. SYNTHETIC EXPERIMENTS**

In order to further investigate the properties of the proposed approach, we consider another set of experiments on the anatomy of the liver shape. The shape of 12 human livers, at maximum exhale position, has been extracted from MRI...
acquisitions, and meshed for statistical shape modelling as described in [35]. Partial observation data are extracted as displayed on Fig. 5 (a).

Since we only have limited data for performing statistical analysis, we artificially enlarged the available training set, using the non-rigid scaling and movement techniques described in [36] and references therein, which generate smooth, localized non rigid deformations of the shapes. A radius of 100mm was chosen, a base scaling factor of 0.5, and the length of the random movement vector was chosen uniform in [0,15mm]. For each type of deformation, and each original shape, we selected 7 random locations, resulting in a total of 180 shapes. Applying PCA, we obtained a richer model, considered as the population shape distribution in the subsequent synthetic experiments. The distribution of eigenvalues of this model is displayed in Fig. 5 (b).

From this known distribution we sampled training sets containing respectively 50 and 20 samples, and 100 independent test shapes. Confidence regions were estimated using the proposed methodology and LOO cross-validation.

Using 50 training samples, the estimated confidence regions appear to be reliably estimated, as shown on Fig. 6. However, with a smaller training set, while the confidence regions get larger, they also tend to become less reliable (Fig. 7). Similar results were observed when the correspondences were set as perfectly known (with an overall reduction of the size of the regions, but similar degradation of the effective coverage probabilities). This can be attributed to a less accurate representation of the true shape distribution, and especially an inability to represent more extreme samples and smaller details only described by the low-energy modes of deformation.

Detailing the contributions of \( P(\eta_{obs}, K) \) and \( P(K|obs) \), both terms contribute to the increase of the overall uncertainty when using fewer training samples. The increased uncertainty related to \( P(K|obs) \) indicates that the global optimum is more difficult to find due to a noisier goal function. Though it was not the main cause for the loss of accuracy in the confidence region estimation in the present experiment, the approximation (11) may also become too coarse in such situations.

VIII. CONCLUSION AND DISCUSSION

We presented a methodological and algorithmic framework for the quantification of uncertainty regions in statistical model-based shape prediction. The framework applies for the estimation of confidence regions when predicting a complete shape from a sparse observation, relying on a set of exemplary shapes. Compared to previous work in the area, we incorporate important additional sources of uncertainty, in particular those related to the estimation of pose parameters, and to the identification of correspondences between the sparse observation and the shape model. Our approach is based on a global optimization framework that enables taking into account the typical sensitivity to initialization of model-based shape matching techniques, and relies on resampling experiments to quantify simultaneously the (limited) amount of correlations between the predictors and the shape to predict, and the generalization ability of the statistical shape model.

We proposed a practical implementation of the approach, and evaluated our framework on a large database of human femurs involving 121 training shapes and 50 test shapes, in a set of experiments displaying different sparse observations. Another set of experiments has been proposed, using a synthetic model based on the liver shape to evaluate the robustness of the proposed approach with respect to the training set size, and the actual shape distribution. In most cases, the estimated confidence regions are shown to be reliable, and even slightly larger than necessary. The proposed experiments simulate a supervised statistical model-based segmentation framework, where the user is requested to delineate parts of the shape. Synthetic experiments using few training shapes showed that the method slowly looses accuracy in the estimation of uncertainties when the population variability is only coarsely represented by the training data. Nevertheless, it still provides approximate indications about the reliability of the prediction.

**Distribution assumption**

In our implementation, we relied on assumptions that all probability distribution are Gaussian, in particular \( P(\eta_{obs}, K) \) and \( P(K|obs) \). However, both distributions are actually estimated from a set of samples: \( P(\eta_{obs}, K) \) is obtained through a non-parametric bootstrap, while \( P(K|obs) \) is calculated from the multi-seed, global optimization. Therefore, more sophisticated distribution models may theoretically be employed. A potential benefit could be expected in terms of quality of the confidence regions, at the
Concerning the estimation of \( P(\mathbf{K}|\mathbf{obs}) \), the proposed formulation remains a heuristic method for enlarging the confidence regions when large uncertainties arise from the matching procedure; the resulting distribution \( P(\mathbf{\eta}|\mathbf{obs}, \mathbf{K}) \) is therefore only an approximation. However, if the matching uncertainty is not taken into account and the estimated correspondences are wrong, then the distribution \( P(\mathbf{\eta}|\mathbf{obs}, \mathbf{K}) \) would largely underestimate the effective uncertainty. We advocate that when the estimated correspondences are suspicious (i.e. in the presence of local minima with significantly different parameters but similar goal function values), the prediction experiment should be reconsidered with more observation data acquired, or manual supervision is used to constrain the prediction process.

**Case-specific correction**

Significant correlations between the matching metric and the quality of the confidence regions were observed, especially for the femur model. We proposed to exploit these correlations observed on the training data by learning a correction factor which enabled us to improve the accuracy of the estimated confidence regions and lower the variance of their quality between individual cases. While mostly heuristic, the approach indicates that useful information remains in the residuals of the shape matching algorithm. The effectiveness of this approach is rooted in the assumption that the training set is representative of the population, or equivalently that there won’t be severe outliers in the test sets. The limits of this assumption are observed in the synthetic experiment with few training shapes.

**Observation uncertainties**

We did not perform experiments involving observational uncertainties, essentially because of the difficulty to model them accurately in practice. However, these can be integrated in the estimation pipeline through a regularization-like term in the prediction stage, by replacing the \( \lambda \) term by the expected observation noise in (8), as proposed in [17] or [18], and in the LM scheme by replacing the point-to-point distances by the corresponding likelihood of observation.

**Run-time considerations**

The current implementation, essentially self-written in C++, is still largely sub-optimal in run-time efficiency, as more emphasis was put on stability. A single prediction, given known correspondences, takes about 1s for the femur model (~20000 vertices), and 0.2s for the liver model (~3900 vertices) on a recent laptop computer. The complete shape prediction and confidence region estimate took about 10 minutes for the liver model. Nevertheless, many directions for optimization can be pointed out. Besides the direct optimization of the current code, the global optimization method can be efficiently parallelized. The parameters of the
tabu-search may also be optimized to speed-up convergence. In our experiments, the global optimum was typically found relatively early. However, care should be taken that enough seeds are used to keep reasonable chances of finding the global optimum. Reducing the number of vertices in the model by optimizing its resolution shall also greatly decrease the computational time. Finally, for a given observation task with reasonable assumptions about which areas of the model will be observed, one may both constrain the identification of correspondences, and pre-compute the cross-validation experiment \( P(\eta_{obs}, K) \).

**Perspectives**

The framework is expected to be of high interest in interventions for which the morphology of the organ of interest is assumed to be unaffected by pathology, and for which the surgical target is not directly visible on pre-operative images, either due to a low contrast, or to a limited field of view such as ultrasound. Potential applications of the method can be indirect targeting, e.g. in functional neurosurgery, when a surgical target is not directly visible on pre-operative images, is assumed to be unaffected by pathology, and for which the seeds are used to keep reasonable chances of finding the global optimum.

**REFERENCES**


