Generation of variable anatomical models for surgical training simulators

R. Sierra a,*, G. Zsemlye a, G. Székely a, M. Bajka b

a Department of Information Technology and Electrical Engineering, Computer Vision Laboratory, Sternwartstrasse 7, 8092 Zurich, ETH Zürich, Switzerland
b Clinic of Gynecology, Department of OB/GYN, University Hospital of Zürich, Switzerland

Received 24 March 2005; accepted 9 November 2005
Available online 4 January 2006

Abstract

The generation of variable surgical scenes is a key element for effective training with surgery simulators. Our current research aims at a high fidelity hysteroscopy simulator which challenges the trainee with a new surgical scene in every training session. We previously reported on methods able to generate a broad range of pathologies within an existing healthy organ model. This paper presents the methods necessary to produce variable models of the healthy organ. In order to build a database of uteri, a volunteer study was conducted. The segmentation was carried out interactively, also covering the establishment of an anatomically meaningful correspondence between the individual organs. The variability of the shape parameters has been characterized by principal component analysis. A new method has been developed and tested, allowing the derivation of realistic new instances based on the stochastic model and complying with non-linear shape constraints which are defined and interactively controlled by medical experts.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Statistical shape analysis; Surgical simulator; Anatomy; Uterus

1. Introduction

The increasing popularity of minimal invasive procedures can be seen as the major motivation for the very significant efforts during the past decade to develop surgical training simulators (Tendick et al., 2000; Basdogan et al., 2001). These procedures have fundamentally changed the interaction paradigm between the surgeon and the operation site, as compared to classical, open surgery. The new interfaces, screens for visual feedback and specialized tools for surgical actions, created an environment which is particularly well suited for virtual reality (VR) based simulations. While the interface between the surgeon and the operation site remains unchanged during simulation, the patient can be replaced by a computer system. Such a simulation setup allows for safe training of future surgeons leading to reduced risks for the patients and improved quality of the performed interventions.

In contrast to flight simulators, which are nowadays well established in aviation, the introduction of surgical simulators in the clinical practice is still a long way ahead. There is consensus about the possible benefits of simulation based training and education (Székely et al., 2000; Montgomery et al., 2001). However, important questions about quantitative assessments of these effects remain open. Most importantly, the level of realism necessary for effective training has still to be identified for the whole spectrum of possible scenarios ranging from the training of individual visual and manipulative skills to the execution of complete surgical procedures. Effectiveness has certainly to be measured in a broad sense, i.e., by incorporating improvement of surgical skills on different levels, costs and risks.

We currently aim at a high fidelity hysteroscopy simulator, which should provide the technical environment to answer the aforementioned questions. Hysteroscopy is the distension of the cavum uteri performed through the
natural opening of the organ. It allows for a minimally invasive inspection and treatment of the uterus under organ saving conditions, which is especially important for women in their childbearing years. Hysteroscopy is the second most often performed procedure after laparoscopy in gynecology. Frequent indications are bleeding disorders, dysmenorrhea (pain during menstruation) and infertility. Most interventions are performed on women between 30 and 60 years of age.

By evaluating the training effects at different levels of realism up to highest fidelity, the optimal trade-off between realism (implying costs) and surgical performance becomes measurable. A key component of such a broader evaluation is the ability to provide different surgical scenes in every training session. Only in doing so, is it possible to exclude adaptation effects to the simulator system, which will prevent to compromise any assessment and thus the clinical value of the system. It is interesting to note in this context that this aspect has received little if no academic attention, while commercially available training simulators mostly include more than one surgical scene (Simbionix, 2004; Immersion Corporation, 2004).

The creation of variable surgical scenes can be divided into two different tasks, namely the generation of variable models of the healthy organ and of different pathologies. Numerous approaches to produce realistic models of pathologies have been reported previously (Sierra et al., 2002, 2003a,b). In this paper, the methods necessary to provide variable models of the healthy organ in the simulator environment are presented. In contrast to pathologies, a model of the healthy organ can well be derived based on the inspection of individual anatomical instances on a sufficiently large population.

In the following, the state-of-the-art for the generation of variable anatomical models is reviewed. Next, the main characteristics of the uterus are summarized and the details of the clinical study are presented. The methods for the segmentation, statistical analysis, and organ shape prediction are then described. Finally, the results for the generation of variable models are presented.

2. Characterization of anatomical variability

Anatomical structures of different individuals show a potentially large, but still strongly limited natural variability which can be represented in a statistical framework. In order to generate variable models of the anatomy, the distribution of organ shapes within an investigated population (the so-called training set) has to be described. In general, individual examples are segmented from volumetric datasets and surface meshes are generated which delineate the organ’s boundaries. A suitable, parametric description of the surfaces is necessary to capture the variability and characterize it in a possibly compact way. The concept has been successfully used to represent prior anatomical knowledge in applications such as the automatic (Cootes et al., 1995) and interactive (Hug et al., 2000) segmentation of medical imagery.

Several methods have been proposed for parametric shape descriptions, e.g., deformable superquadrics augmented with local deformation modeling (Terzopoulos and Metaxas, 1991; Vemuri and Radasavljevic, 1994), series expansions (Staib and Duncan, 1992; Brechbühler et al., 1995), or simply the coordinates of the surface vertices (Cootes et al., 1992). The shape’s statistics is then modeled by the a priori probability \( p(p) \) of a parameter vector \( p \) which is estimated from the training set. The estimation of the probability \( p(p) \) requires that the samples from the given population are coherently described, i.e., that the corresponding anatomical features of the different instances are consistently mapped throughout the whole training set. Thus, the correspondences between the single instances have to be defined. Such a point-wise inter-individual mapping between instances of the training set can be established in an implicit manner (Brechbühler et al., 1995) or explicitly if the shapes are described by point distribution models. In the latter case, both interactive methods using generalized landmarks on the surface (Zsemlye and Székely, 2003) and fully automatic approaches based on the characteristics of the point distributions (Kotcheff and Taylor, 1998; Davies et al., 2002) have been proposed.

Possible variations of the anatomy can be characterized precisely by the joint distribution function of the shape parameters \( p_i \). As the single components of the vector \( p \) are usually highly correlated, a substantial reduction of parameters is possible. Principle component analysis (PCA) is a well known approach for their decorrelation (Cootes et al., 1995). Formally, if there are \( N \) meshes in the study population, each with \( n \) vertices, and \( p_1, \ldots, p_N \) are their parametric representations (vectors of length \( 3n \) containing \( x, y \) and \( z \) coordinates of all vertices), then PCA yields the following descriptors:

\[
\bar{p} = \frac{p_1 + \cdots + p_N}{N}
\]

which is the mean vector, and the covariance matrix

\[
\Sigma = \frac{1}{N-1} \Delta p \Delta p^T,
\]

where the matrix \( \Delta p = (p_1 - \bar{p}, \ldots, p_N - \bar{p}) \) represents the distance of every instance from the mean. A new instance \( \bar{p} \) can be constructed by the proper selection of the control vector \( b \)

\[
\hat{p} = \bar{p} + Ub = \bar{p} + \sum_{j=1}^{k} b_j U_j,
\]

where the matrix \( U \) contains the eigenvectors \( U_j \) of the covariance matrix \( \Sigma \). It has been shown that in general the first \( k \) (\( k \ll N \)) eigenvectors corresponding to the largest eigenvalues capture most of the variability even for complex shapes. Therefore, the linear combination of a few eigenmodes results in a compact description of the population’s variability.
A new instance is fully specified by the parameters \( b_1, \ldots, b_k \). Their control is relatively easy if the distribution of the original shape parameters can be well approximated by a multi-variate Gaussian distribution, which can fully be characterized by its first and second order moments. The number of parameters \( (k) \) allows for a trade off between tractability and accuracy of the model description. This fact is commonly exploited for the segmentation of new datasets, where generic optimization algorithms will iteratively drive the selection of the vector \( b \) leading to an anatomically meaningful adaptation of the mean shape to the underlying dataset. Unfortunately, the shape changes associated with the individual \( b_j \) parameters are rarely sufficiently intuitive to enable the interactive control of the new object’s appearance by a medical expert.

A method to overcome this limitation has been proposed by Hug et al. (2000). Instead of modifying the shape by the coefficients of the eigenvectors, it is controlled by selected boundary points, which can be directly identified by the user in the given datasets. The positions of these landmarks introduce restrictions for the possible shape of the organ. In order to fulfill the imposed constraints, the most probable shape \( \hat{p} \) that matches a selected landmark can be computed by the minimization of the Mahalanobis distance (Morrison, 1978), leading to

\[
\hat{p} = \hat{p} + U \hat{A} U^T (U_j \hat{A} U^T)^{-1} \Delta p_j,
\]

where \( \Delta p_j \) denotes the distance of the actual \( (p_j) \) and mean \( (\hat{p}_j) \) landmark position. \( U_j \) is the sub-matrix of \( U \) containing the rows \( 3j - 2, 3j - 1, \) and \( 3j \) of \( U \) and \( \Delta \) is the matrix containing the eigenvalues of the covariance matrix \( \Sigma \) in the diagonal. The segmentation can be refined iteratively by defining additional landmarks and re-evaluating Eq. (2). The method was successfully used for the semi-automated segmentation of the corpus callosum in two dimensions (Hug, 2001).

### 3. Model generation

The uterus is a hollow, muscular organ, shaped somewhat pear like. Its cavity has three natural orifices: the cervical ostium and the two tubal ostiae. The uterus of an adult woman can be separated in a lower third part, the cervix, and the upper two-thirds, denoted as body of the uterus. This ratio is different during childhood, where body and cervix are about equal in size. The body has a dome-shaped top, called fundus, and is joined by the uterine tubes at its widest part. The cervix has a more tubular structure and extends downward into the upper portion of the vagina. The uterine cavity reaches from the cervical orifice to the internal isthmus, building the cervical channel, and from there to the tubular ostiae, forming a triangle.

The size of the uterus varies largely during pregnancy, and will not return to its previous size after childbirth. These changes have been extensively studied for the design of intra-uterine contraceptive devices (Kurz et al., 1984; Hasson, 1982). For the determination of biometric values, the parity and age, i.e., pre- or postmenopausal, are considered. In gynecology, the total length of the uterus is defined as the sum of the length of the cervix (cervical ostium-isthmus) and the body (isthmus-outer border of the fundus), measured separately during ultrasound-screening. For a regular uterus, a filled urinary bladder will push the body of the uterus in dorsal and cranial direction. The filling of the rectum though will elevate the cervix and not the fundus of the uterus in cranial direction (Drenckhahn and Zenker, 1994). The angle enclosed by the cervical ostium, the isthmus and the fundus is therefore dependent on the status of the neighboring organs.

In our study, magnetic resonance imaging (MRI) images were taken from 26 volunteers. Motivated by the targeted application, the volunteers were recruited to cover the range in age and pregnancy status which can be expected for patients treated by hysteroscopy, i.e., women after puberty with different numbers of deliveries. The scanning protocol included one sagittal, one or possibly two transversal, and one oblique MRI scans on a GE™ 1.5 T machine. To reduce the noise originating from the natural, constant pelvic activity, Buscopan® was injected. Explanation of the study protocol along with its potential risks and benefits was performed by either residents or attending physicians. Informed consent was obtained from all participating volunteers before initial administration of Buscopan®. Each MRI scan included 16–25 slices with a resolution of 0.46 × 0.46 × 3 mm³. By limiting the number of acquired slices, the scanning time of each volume could be kept below 10 min, reducing the noise in the volume introduced by the filling of the bladder during the scans. To extract the small anatomical landmarks of the cervical ostium as well as the tubal ostiae, it was necessary to take several orthogonal scans and combine the information of those scans. Due to the bladder filling, the inter-scan, intra-subject deformation is in principle non-rigid, leading to slight differences in uterine shapes for consecutive scans.

To quantify the uterus deformation caused by the bladder filling during the acquisition process, the tip of the fundus has been identified in the different scans. The position of this landmark changed in the order of 3–6 mm between subsequently scanned volumes, which is well comparable with the slice distance used. Accordingly, related effects of the segmentation were negligible at the selected resolution level. Rigid registration was therefore sufficient to provide additional cues for the determination of the organ boundaries and small anatomical structures such as the tubal ostiae, where the sagittal volume failed to provide enough information during the segmentation process. We used the method proposed by Maes (1998) for the registration based on the maximization of mutual information.

For a statistical analysis of the unstructured surface meshes, an appropriate point distribution model has to
be provided. Zsemlye and Székely (2003) proposed to initialize a new surface mesh based on manually selected landmarks on the segmented surface, which can be also higher dimensional like anatomically meaningful curves. These landmarks are discretized and triangulated in a standardized way to generate a sparsely sampled point distribution. The triangular mesh is subsequently refined along lines approximating the connecting geodesic paths by a quaternary subdivision enforcing an implicit correspondence between well identifiable landmarks based on surface distance. A comparative study (Styner et al., 2003) has demonstrated, that the resulting correspondence is of reasonable quality. Global optimization (Davies et al., 2002) can further improve the statistical properties of the parameters' distribution while resulting only in minor shifts in the coordinates of the vertices of the surface mesh.

We followed an identical strategy for the generation of the uterus model, handling both the segmentation and correspondence definition in an integrated manner. The current implementation is based on the subdivision of one predefined surface mesh which is subdivided and deformed manually according to the underlying volumes. Conceptually, the correspondences are imposed by the initial mesh and new points are introduced by geodesic interpolation. The mesh topology is carried along with the segmentation, so that all meshes share the same triangulation. Thus, the surface mesh can be reconstructed unambiguously from any new point distribution generated by the statistical framework. The actual segmentation process consists of three steps that are iteratively applied: homogeneous subdivision, grid-based deformation and Gaussian or curvature based smoothing. In order to prevent an enlargement of the orifices, the corresponding vertices have to be fixed prior to the application of the smoothing operators. Fig. 1 shows a resulting model embedded into a sagittal MRI cross-section.

Prior to the statistical analysis, the meshes are translated and rotated into a common organ based coordinate system, defined by the three natural orifices. The right tubal ostium of a uterus in normal position in a patient based coordinate system is transformed into the origin of a right-handed coordinate system. The cervical ostium is placed on the positive x-axis and the left tubal ostium on the xy-half-plane with negative y-values. No further spatial normalization was applied.

4. Shape prediction

In the context of surgical training simulators, it is desirable that the physicians can rely on a well defined set of parameters which can be modified independently, thus providing a tool for the intuitive instantiation of new models. In contrast to a segmentation framework, the specification of absolute coordinates of selected landmarks for the definition of a new case is therefore hardly an appropriate choice. The measurements performed on routine gynecologist examinations, i.e., the values measured with 2D ultrasound devices in the clinical practice, offer a much more natural alternative. Here, the length, width, and height of both the fundus and the cervix of the uterus are captured. In this case, no information about the absolute position of related landmarks is given, just distance measurements between them. The relationship between the predictors (distances between the landmarks) and dependent variables (coordinates of the vertices of the surface mesh) is not linear due to the Euclidean distance function.

The methods described in Section 2 rely, however, on a linear relationship between the predictors and the model as will be shown. In the following derivation, x₁ represents the first, and x₂ the second principal landmark. Without loss of generality, it can be assumed that they correspond to the first and second vertex of the mesh, respectively. In the first step of Hug’s method, all study instances are transformed by Eq. (2) in order to fulfill the conditions imposed by x₁:

$$p'_i = p + U(b_i - R_i(p_{i(1)} - x_1)) = p_i - UR_ip_{i(1)} + UR_ix_1,$$

where $$p_{i(1)} = (p_{3i-2}, p_{3i-1}, p_{3i})^T$$, and $$R_i = AU_i^T(U_iA_i^T)^{-1}$$. From Eq. (2).

The new average vector is given by

$$\bar{p}' = UR_ix_1 + p - UR_ip_{i(1)}$$

and the displacement from the mean is

$$\Delta p'_i = p'_i - \bar{p}' = \Delta p_i - UR_i\Delta p_i(1).$$

Note that $$\Delta p' = (\Delta p'_1, \ldots, \Delta p'_n)$$ is not dependent on the predictor $$x_1$$, neither are the covariance matrix, the eigen-

![Fig. 1. Typical model resulting from the described segmentation approach. The uterine cavity is visible in the right image, where the model is cut by a clipping plane.](image-url)
vectors, and eigenvalues of the new instances \( \mathbf{p}'_i \). Thus, the application of the prediction scheme with the new eigenvectors \( (U') \) and eigenvalues \( (\Lambda') \) results in a linear relationship with respect to both \( x_1 \) and \( x_2 \):

\[
\hat{\mathbf{p}} = \mathbf{p}' + U'R'_2(x_2 - \mathbf{p}'_2) = U'R_1x_1 + U'R'_2x_2 + \mathbf{p} - U'R_1\mathbf{p}'_1 - U'R'_2\mathbf{p}'_2,
\]

where \( R'_2 = \Lambda'U'_2 (U'_2\Lambda'U'_2)^{-1} \).

Therefore, the problem has to be reformulated to allow for a non-linear prediction scheme. The goal is to find the most probable shape that fulfills the constraints, i.e., distance measurements, imposed by the user. Let \( r \) be the number of distance measurements between the \((2i-1)\)th and \((2i)\)th vertex of the mesh and \( d_i^2 \) the squared distance between these vertices:

\[
d_i^2 : \mathbf{R}^{3n} \rightarrow \mathbf{R}, \quad i = 1, \ldots, r,
\]

\[
d_i^2(\mathbf{p}) = \sum_{j=0}^{2} (p^{d_{i,j}} - p^{d_{i,j-1}})^2.
\]

For a given set of predictors \( \{\xi_1, \ldots, \xi_r\} \), the mesh representation \( \hat{\mathbf{p}} \) has to be found, which fulfills the criterion for the distances between its landmarks

\[
d_i^2(\hat{\mathbf{p}}) = \xi_i; \quad i = 1, \ldots, r
\]

and is the most probable one, i.e., has the smallest Mahalanobis distance from the mean:

\[
\hat{\mathbf{p}} = \arg \left( \min_{\mathbf{p} \in \mathbb{R}^{3n}} \left( (\pi - \mathbf{p})^T \Sigma^{-1} (\pi - \mathbf{p}) \right) \right). \tag{3}
\]

Eq. (3) is a constrained minimization problem, which can be solved using the method of Lagrange multipliers (Bronstein et al., 1997). However, the prediction method cannot be applied directly if the number of training examples is limited. First, numerical methods are required to solve the resulting problem, which tend to become slow and unstable if the dimensionality of the variables is high, which is the case for surface meshes consisting of several thousands of vertices. Second, if the number of variables (3n) is smaller than the number of instances in the study population (N), the covariance matrix \( \Sigma \) is rank deficient and cannot be inverted.

In order to overcome these limitations the number of variables is reduced by performing PCA and projecting the parameter vectors into the eigenspace (see Eq. (1)). In other words, the study instances are expressed in the coordinate system of a linear subspace which is defined by the mean instance and a limited number of eigenvectors. The statistical analysis is performed in this subspace and the result is transformed back to the object space as a final step. The shape models expressed in this coordinate system are

\[
\mathbf{b}_i = U^T(\mathbf{p}_i - \mathbf{p}), \quad i = 1, \ldots, N \quad \text{and} \quad \mathbf{B} = U^T \Lambda U.
\]

The Mahalanobis distance also has to be defined in this subspace to solve the minimization problem. The covariance matrix of the new instances is of course diagonal and the values on the diagonal are the same as the eigenvalues of the covariance matrix of the original variables. Formally, if \( \Sigma = \frac{1}{N} \Lambda \mathbf{P} \mathbf{A} \mathbf{P}^T \) is the covariance matrix of the original variables that is decomposed into \( \Sigma = U \Sigma U^T \) after PCA, then the covariance matrix \( \Sigma_b = \frac{1}{N} \Lambda_b \mathbf{P} \mathbf{A} \mathbf{P}^T \) of the new variables is nothing else than \( \Lambda_b \).

After this transformation, the goal function of the minimization problem becomes

\[
f(\mathbf{b}) = (\mathbf{b} - \overline{\mathbf{b}})^T \Lambda^{-1} (\mathbf{b} - \overline{\mathbf{b}}) = \mathbf{b}^T \Lambda^{-1} \mathbf{b}
\]

with the constraints

\[
d_i^2(\hat{\mathbf{p}} + \mathbf{Ub}) = \xi_i, \quad i = 1, \ldots, r.
\]

This function has to be solved for the variable vector \( \mathbf{b} \). The resulting parameter vector \( \mathbf{b} \) for a given set of predictors \( \{\xi_1, \ldots, \xi_r\} \) is transformed back to the object space, providing the point distribution model describing the uterus’ shape to be used for the training scene:

\[
\hat{\mathbf{p}} = \mathbf{p} + \mathbf{Ub}.
\]

As stated, the constrained minimization problem can be solved using the method of Lagrange multipliers. The Lagrange function in this case has the form

\[
F(\lambda_1, \ldots, \lambda_r, b_1, \ldots, b_k) = f(b_1, \ldots, b_k) - \sum_{j=1}^{r} \lambda_j c_j(b_1, \ldots, b_k),
\]

where \( c = (c_1, \ldots, c_r)^T = (d_1^2 - \xi_1, \ldots, d_r^2 - \xi_r)^T \). The optimization can be performed by solving the following system of equations:

\[
\frac{\partial F}{\partial \lambda_1} = \cdots = \frac{\partial F}{\partial \lambda_r} = \frac{\partial F}{\partial b_1} = \cdots = \frac{\partial F}{\partial b_k} = 0 \tag{4}
\]

which can be performed numerically by using the Newton–Raphson method (Press et al., 2002). Appendix A presents the details of the derivation of an iterative optimization of the parameters \( \mathbf{b}_{n+1} \) as a function of \( \mathbf{b}_n \). The iteration is started from the mean shape, i.e., \( \mathbf{b}_0 = 0 \) and \( \lambda_0 = 0 \), so that \( \mathbf{p}_0 = \hat{\mathbf{p}} + \mathbf{Ub}_0 \), and is stopped, when the constraints are satisfied to a given precision, i.e., \( c^T \mathbf{e} < \epsilon \). The precision threshold can be chosen as small as \( \epsilon = 10^{-5} \).

The selected predictors \( \{\xi_1, \ldots, \xi_t\} \) are the squares of the biometric measurements used in gynecology (length, width, depth of fundus and cervix), plus the overall square length of the uterus measured from the cervical ostium to the fundal tip. This additional value, together with the fundus and cervix lengths, allows to control the flexion of the uterus, i.e., the angle between the cervix and the fundus, but requires that the three retroflected uteri are excluded from the training database. The selected set of measures can be directly extracted from the segmented models. As the segmentation inherently provides the correspondence between the different instances, the evaluation of the distances can be automatically performed after the specification of the related landmark vertices in a single model.
5. Results and discussion

Based on the method described above, an interactive tool has been developed for the intuitive derivation of new instances of the uterus. Fig. 2 depicts the GUI of the tool based on the statistical analysis of the database, presenting the mean shape and the widget with seven sliders. The user can now manually adjust each slider independently to modify the value of the corresponding predictor. Along with each slider two values are presented, first the specified and second the measured length of the predictor. These values will only diverge if the algorithm is not able to generate a uterus based on the given predictors, in which case the last generated shape is kept. The range of possible values can be set independently for each predictor through a configuration file. The Newton–Raphson optimization is a second order scheme and thus converges relatively fast, i.e., the number of correct digits is doubled in every iteration. In general, for the present implementation less than six iterations are required. Hence, new shapes can be evaluated in real time as the user moves the sliders. Even the underlying statistical analysis including the transformation of the shapes into a common coordinate system, which is performed offline as a preprocessing step, can be computed in a few seconds on a standard 3 GHz PC.

The variation of the mean shape induced by the ‘fundus depth’ predictor is illustrated in the upper row in Fig. 3. The values for the presented images range from 26 to 49 mm, covering about 1.5 times the standard deviation found in the present database. The measured values for all other predictors, i.e., the distances between the other landmarks, did not change after the deformation, as requested. The lower row of Fig. 3 depicts uteri with varying the overall length ranging from 57 to 75 mm. Obviously, the flexion of the uterus cannot be extended further. The uterus length would exceed the sum of fundus and cervix length, an impossible constraint that prevents the convergence of the Newton–Raphson algorithm.

As the shapes were not normalized to a standard size, all datasets can be integrated into one common database to directly predict new instances based on the provided parameters. While size normalization could in principle improve the prediction of the shapes, we observed no qualitative difference for the given uteri. The ‘cervix width’ and ‘fundus width’ predictors are most sensitive to perturbations. Larger changes of these values would make the adjustment of the remaining predictors also necessary to avoid unnatural shapes. These findings may indicate that the width is correlated to the remaining predictors. However, it has to be noted that the segmentation of the lateral walls of the uterus is relatively difficult and error prone, as the arteria uterina follows exactly this landmark and strongly blurs the MRI scans. It is therefore not trivial to assign the different behavior of the predictors to a single source. To assess the results of the presented framework, a systematic analysis of the prediction performance has been carried out.

In the following, three properties of the statistical framework are discussed (Davies, 2002). First, the model should be able to represent unseen instances of the object class, i.e., it should generalize the main characteristics of the shapes based on the given set of objects. Second, the compactness of the model can be assessed, i.e., its ability to represent an object by as few parameters as possible. Third,
the statistical framework should produce only valid instances of the class. In the following experiments, the predicted shapes \( P_A \) are registered by a translation and a rotation to the original dataset \( P_0 \) by minimization of the sum of the squared corresponding-point distances. In doing so, any dependency on the specific, organ based coordinate system can be removed prior to the analysis of the results.

The generalization of the model is evaluated by a leave-one-out experiment. The statistical analysis is performed on all but one dataset and the parameters of this instance are provided for the prediction of a new case. The result should then approximate the original example. Fig. 4 illustrates the best (upper row) and worst (lower row) prediction on the right, and the corresponding original mesh on the left. The error metric selected is the maximal corresponding point distance. The point-wise distances from the predicted to the original case are color encoded. The minimal point distance is represented in red whereas the maximal distance is colored in purple and evaluates to 12.44 mm for the worst example. The maximal \( d_{\text{max}} \) and the mean \( d_{\text{mean}} \) corresponding-point distances between the predicted and original shape are compiled in Table 1 for the complete set of shapes.

The leave-one-out experiment has been extended in the current context and a more in depth analysis of possible solutions has been performed. In general, the projection of the left-out object onto the PCA-space of the reduced database will not meet the imposed distance constraints. Thus, the results of the leave-one-out experiment are suboptimal in terms of corresponding point distances. Fig. 5(a) indicates the various shapes that have been computed additionally. \( k \) is the number of shapes in the database minus one. \( C \) is the non-linear subspace representing the shape constraints imposed. \( P_0 \) is the missing object, \( P_R \) its projection into the shape space \( \Sigma \) of the reduced database, \( P_A \) the estimated shape using the presented algorithm and \( P_R \) the shape closest to \( P_R \) which fulfills the constraints of left out object. \( P_R \) has been determined by the presented optimization algorithm, but minimizing the Mahalanobis distance of the resulting shape \( P_R \) to \( P_R \) instead of to the mean shape. The computed mean distances across all instances evaluate to \( \|P_0 P_R\| = 1.5 \text{ mm}, \|P_0 P_R\| = 1.8 \text{ mm}, \|P_0 P_R\| = 1.0 \text{ mm}, \) and \( \|P_0 P_R\| = 2.2 \text{ mm} \). As illustrated in Fig. 5(a), \( P_A \) and \( P_R \) are not necessarily identical. As expected, the selection of the shape with minimal Mahalanobis distance to the mean is not optimal for the prediction of the left-out object, as \( \|P_0 P_A\| > \|P_0 P_R\| \). However, the difference in the measured mean distances is relatively small.

### Table 1

<table>
<thead>
<tr>
<th>Volunteer</th>
<th>( d_{\text{max}} )</th>
<th>( d_{\text{mean}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.81</td>
<td>1.12</td>
</tr>
<tr>
<td>2</td>
<td>3.72</td>
<td>1.32</td>
</tr>
<tr>
<td>3</td>
<td>4.05</td>
<td>1.20</td>
</tr>
<tr>
<td>4</td>
<td>4.07</td>
<td>1.32</td>
</tr>
<tr>
<td>5</td>
<td>4.16</td>
<td>1.22</td>
</tr>
<tr>
<td>6</td>
<td>4.33</td>
<td>1.55</td>
</tr>
<tr>
<td>7</td>
<td>4.49</td>
<td>1.94</td>
</tr>
<tr>
<td>8</td>
<td>4.78</td>
<td>1.55</td>
</tr>
<tr>
<td>9</td>
<td>5.58</td>
<td>2.00</td>
</tr>
<tr>
<td>10</td>
<td>6.00</td>
<td>1.42</td>
</tr>
<tr>
<td>11</td>
<td>6.12</td>
<td>1.80</td>
</tr>
<tr>
<td>12</td>
<td>6.79</td>
<td>2.13</td>
</tr>
<tr>
<td>13</td>
<td>6.82</td>
<td>2.21</td>
</tr>
<tr>
<td>14</td>
<td>6.92</td>
<td>2.62</td>
</tr>
<tr>
<td>15</td>
<td>6.94</td>
<td>2.65</td>
</tr>
<tr>
<td>16</td>
<td>7.94</td>
<td>2.77</td>
</tr>
<tr>
<td>17</td>
<td>8.20</td>
<td>2.61</td>
</tr>
<tr>
<td>18</td>
<td>8.29</td>
<td>3.32</td>
</tr>
<tr>
<td>19</td>
<td>8.33</td>
<td>2.94</td>
</tr>
<tr>
<td>20</td>
<td>10.69</td>
<td>4.44</td>
</tr>
<tr>
<td>21</td>
<td>10.88</td>
<td>3.22</td>
</tr>
<tr>
<td>22</td>
<td>12.44</td>
<td>3.81</td>
</tr>
</tbody>
</table>

Overall mean \( \|P_0 P_R\| \) = 6.56 \( \text{ mm} \), \( d_{\text{mean}} \) = 2.23 \( \text{ mm} \).

All values are provided in millimeters.
and both values are in the order of magnitude of the distance from the original to the projected shape. These findings indicate that it can be safely assumed that the optimization scheme employed converges in the subspace \( \mathcal{R} \) into the direction of the projected object \( P_P \), and finds a solution in its proximity that fulfills the input parameters. In order to investigate the compactness of the scheme, efforts have been made to assess the relative importance of the selected set of predictors. The specific choice presented is clearly motivated by clinical and diagnostic constraints and the parameters are likely to be correlated. It might be possible to find a subset that is nearly as accurate for the prediction of new instances. In order to estimate the influence of a predictor, in the following all shapes have been included into the database, as opposed to the leave-one-out experiment before. Every shape is then predicted with varying numbers of predictors and its distance to the original one is evaluated. The result can also be interpreted as the remaining variability for one dataset after the specification of a number of shape constraints. Fig. 6 illustrates the results of the experiment. For each case the maximal, minimal, mean, and variation of the corresponding point distances, averaged over the database, are provided. The first graph shows the behavior of the distances for increasing numbers of predictors in an arbitrary order. It can be observed that the overall length, i.e., the last predictor specified, has a large impact on the accuracy of the shape. The order of the predictors has thus been rearranged empirically to first apply the most dominant parameter. The results are shown in the second graph. As can be seen, it is well possible to sort the chosen parameters by their relative importance. In order to quantitatively investigate this issue, the eigen-decomposition \( \sigma = \frac{1}{n} \Delta \delta \delta^T = \mathbf{u}_i \mathbf{u}_i^T \) of the predictors \( \zeta \) has been computed. Fig. 5(b) illustrates the cumulative variation for different numbers of eigenvectors \( \mathbf{u}_i \). The first eigenvector \( \mathbf{u}_1 \) already captures about 80% of the variation, confirming our original hypothesis regarding the correlation between the selected predictors. However, none of the eigenvectors is in significant alignment with any of the predictors. Therefore, using an eigenvector for the prediction of new shapes conflicts with our initial requirement of using intuitive parameters. In conclusion, it is not recommended to only use a subset of the predictors for the derivation of new instances. If the usage of a smaller number of predictors is of paramount importance, extensive tests of the different combinations of parameters and the resulting accuracy have to be carried out.

The third property, i.e., the model specificity, can be assessed by generating new instances and comparing those cases with their nearest sample in the database. As an exhaustive exploration of the parameter space cannot be reasonably evaluated, the parameters have been randomly selected using a uniform distribution to avoid any bias in the evaluation process. In order to respect the constraints
imposed by the parameter distribution observed in the training set, the selected random samples were uniformly distributed in the range $\pm 6\sigma_i$. For this experiment, approximately 300 shapes have been classified into valid and invalid uteri by a clinical expert. Fig. 7 illustrates the percentage of false shapes as a function of the Mahalanobis distance $D_{M_0}$ in the eigenspace of the predictors. The data have been accumulated according to $D_{M_0}$ into 12 bins which are shown as a continuous curve using linear interpolation. For $D_{M_0} < 16$ all shapes have been identified as valid, i.e., the model is completely specific. There is a large step at $D_{M_0} \approx 16$, after which most shapes are labeled as invalid. Only for $D_{M_0} > 55$ all uteri are marked as invalid. An automatic classification of the resulting shapes for a given set of predictors into realistic and invalid uteri would be desirable in the targeted application. The Mahalanobis distance of the predictors from the mean. This information can be accumulated according to $D_{M_0}$ into 12 bins which can handle complex topologies such as the uterus and its cavity. Future research will address this problem and its impact on the presented algorithms.

A second interesting aspect for future research is the derivation of new instances solely based on the patient’s history. Biometric studies, which were based on ultrasound (US) imaging and conducted on several hundreds of women, have shown that the selected predictors vary significantly depending on parity as well as on the menopausal state (Merz et al., 1996). Hence, age and parity could serve as input parameters to define a new set of predictors and estimate a corresponding shape. However, these studies have all been performed with US devices. Based on 2D and 3D US data acquired during our study in parallel with the MR images, first attempts have been made to establish a relationship between the routine clinical US measurements and the precise geometric description of the uterus extracted from the segmented (i.e., MR based) models. Unfortunately, the fusion of all related data has been more difficult than originally expected. Further research will be needed, also covering issues in standardization of routine investigations, before the 2D based US measurements can be used to predict the overall 3D shape of the uterus. Such a model would offer a much larger training set containing thousands of observations. This would allow to deal with eventually significant deviations of the underlying distribution from the ideal Gaussian shape. The resulting framework would provide a solid basis to approach our ultimate target, namely to build a complete prediction chain from the patient’s anamnesis to a new uterus model.

6. Summary and outlook

We have presented a framework for the generation of variable models of the uterus anatomy based on a statistical description. The definition of point correspondences necessary for the statistical analysis is implicitly incorporated into the segmentation process. The derivation of new instances is based on a predefined set of parameters that are in agreement with common measurements during routine gynecological organ inspections. The method has been specifically designed to handle these predictors, i.e., to deal with the non-linear relations between the predictors and the dependent variables. Based on the underlying statistics it is also possible to estimate the quality of the resulting meshes by measuring the deviation of the selected predictors from the mean. This information can be provided to the clinician to discard unrealistic instances.

The correspondences imposed by the segmentation process may not be optimal on the final meshes. However, the landmark errors evaluated for different organ models and correspondence methods indicate mean errors in the order of 3–4.5 mm (Styner et al., 2003). These values are larger than the errors expected from the current approach. The sampling over subdivided areas is very homogeneous so that the geodesic distances can be assumed to provide good correspondences. Nevertheless, it cannot be estimated to what extent even small changes in the correspondence definition influence the validity and performance of the statistical framework. To the best of our knowledge there is currently no software for the global optimization of point correspondences available that can handle complex topologies such as the uterus and its cavity. Future research will address this problem and its impact on the presented algorithms.

Acknowledgments

This work has been performed within the frame of the Swiss National Center of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR CO-ME) supported by the Swiss National Science Foundation. We thank Dr. Simon Wildermuth and Teodorovic Ninoslav from the Radiology Department of the University Hospital in Zürich for their support in acquiring the MRI-data of the uteri.
Appendix A. Iterative optimization of shape parameters

In the following, the derivation of the equations for the iterative computation of the vectors \( \mathbf{b}_i \) is presented. In order to optimize Eq. (4), the root of the function

\[
\mathbf{G}(\lambda, \mathbf{b}) = (G_1(\lambda, \mathbf{b}), \ldots, G_{d+1}(\lambda, \mathbf{b}))^T
\]

where \( \mathbf{G} \) is the gradient of \( \mathbf{G} \). The Jacobi matrix is then the function

\[
\mathbf{J}(\lambda) = \left( \begin{array}{c} \frac{\partial f}{\partial b_1} \\ \vdots \\ \frac{\partial f}{\partial b_d} \end{array} \right)
\]

has to be found. If \( (\lambda_0^T, \mathbf{b}_0^T)^T \) is an initial approximation of the root of the function \( \mathbf{G} \), then an iteration

\[
\vec{\lambda} = \mathbf{G}(\lambda, \mathbf{b}) + \mathbf{J}(\vec{\lambda}) (\lambda_{n+1} - \lambda_n)
\]

yields the root of \( \mathbf{G} \). Here \( \mathbf{J} \) denotes the Jacobi matrix of \( \mathbf{G} \). Let

\[
\lambda_i = \frac{\partial f}{\partial b_j}
\]

be the the gradient of \( f \),

\[
A_{ij} = \frac{\partial^2 f}{\partial b_i \partial b_j}
\]

the Hessian of \( f \),

\[
B_{ij} = \frac{\partial c_{ij}}{\partial b_j}
\]

the Jacobi Matrix of \( c \), and

\[
C_{ij} = \frac{\partial^2 c_{ij}}{\partial b_i \partial b_j}
\]

the Hessian of \( c_{ij} \). Using this notation, the Jacobi matrix \( \mathbf{J} \) can be expressed as a block matrix

\[
\mathbf{J} = \left( \begin{array}{cc} 0 & -\mathbf{B} \\ -\mathbf{B}^T & \mathbf{A} - \sum_{i=1}^{r} \lambda_i \mathbf{C}_{ii} \end{array} \right) = \left( \begin{array}{cc} 0 & -\mathbf{B} \\ -\mathbf{B}^T & \mathbf{A}' \end{array} \right),
\]

where \( \mathbf{S} = -\mathbf{BA}^{-1}\mathbf{B}^T \) is the Schur complement of \( \mathbf{A}^{-1} \) in the matrix \( \mathbf{J} \). Finally, the iterative formula can be rewritten as

\[
\left( \begin{array}{cc} \lambda_{n+1} \\ \mathbf{b}_{n+1} \end{array} \right) = \mathbf{J}^{-1} \left( \begin{array}{cc} \mathbf{c} \\ -\mathbf{B}\lambda_n \end{array} \right) = \left( \begin{array}{cc} \mathbf{S}^{-1} (\mathbf{c} - \mathbf{BA}^{-1}\lambda) - \lambda_n \\ \mathbf{A}^{-1}\mathbf{B}^T \mathbf{S}^{-1} (\mathbf{c} - \mathbf{BA}^{-1}\lambda) - \mathbf{A}^{-1}\mathbf{B}^T \end{array} \right).
\]

The iteration can be expressed in the following explicit form:

\[
\lambda_{n+1} = \mathbf{S}^{-1} (\mathbf{c} - \mathbf{BA}^{-1}\lambda_n) = (\mathbf{BA}^{2n+1})^{-1} (\mathbf{c} - \mathbf{BA}^{-1}\lambda_n)
\]

and

\[
\mathbf{b}_{n+1} = \mathbf{b}_n + (\mathbf{A}^{-1}\mathbf{B}^T) \mathbf{S}^{-1} (\mathbf{c} - \mathbf{BA}^{-1}\lambda_n) - \mathbf{A}^{-1}\lambda_n
\]

Note that the matrices \( \mathbf{c} \), \( \mathbf{BA} \) and \( \mathbf{B} \) are dependent on the variables \( \lambda \) and/or \( \mathbf{b} \), and shall be updated in each step of the iteration.

References


