

## Research Article

# Finite Element Surface Registration Incorporating Curvature, Volume Preservation, and Statistical Model Information

Thomas Albrecht,<sup>1</sup> Andreas Dedner,<sup>2</sup> Marcel Lüthi,<sup>1</sup> and Thomas Vetter<sup>1</sup>

<sup>1</sup> Department of Computer Science, University of Basel, Bernoullistrasse 16, 4056 Basel, Switzerland

<sup>2</sup> Mathematics Institute, University of Warwick, Zeeman Building, Coventry CV4 7AL, UK

Correspondence should be addressed to Thomas Albrecht; thomas.albrecht@unibas.ch

Received 17 June 2013; Accepted 21 August 2013

Academic Editor: Tianye Niu

Copyright © 2013 Thomas Albrecht et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a novel method for nonrigid registration of 3D surfaces and images. The method can be used to register surfaces by means of their distance images, or to register medical images directly. It is formulated as a minimization problem of a sum of several terms representing the desired properties of a registration result: smoothness, volume preservation, matching of the surface, its curvature, and possible other feature images, as well as consistency with previous registration results of similar objects, represented by a statistical deformation model. While most of these concepts are already known, we present a coherent continuous formulation of these constraints, including the statistical deformation model. This continuous formulation renders the registration method independent of its discretization. The finite element discretization we present is, while independent of the registration functional, the second main contribution of this paper. The local discontinuous Galerkin method has not previously been used in image registration, and it provides an efficient and general framework to discretize each of the terms of our functional. Computational efficiency and modest memory consumption are achieved thanks to parallelization and locally adaptive mesh refinement. This allows for the first time the use of otherwise prohibitively large 3D statistical deformation models.

## 1. Introduction

Nonrigid registration remains one of the largest challenges in medical image analysis and computer vision. The correspondence it seeks to establish between related objects is essential for a large number of applications from shape statistics to model-based segmentation and computational anatomy [1, 2].

In medical applications, the objects that need to be brought into correspondence are usually organs which were captured with a medical imaging device, such as CT and MRI. In this paper, we use bones, skulls, and hands as examples. We propose a method to bring them into correspondence by registering a number of feature images. The unique features and contributions of our method are as follows:

- (1) the representation of surfaces by a distance and a curvature image,
- (2) the inclusion of a volume preservation term into the registration,

- (3) the introduction of prior knowledge in form of a statistical deformation model,
- (4) a continuous formulation of the registration method, including the deformation model, making the registration independent of its discretization,
- (5) an efficient finite element discretization based on the local discontinuous Galerkin method.

For us, as in many other applications, the ultimate goal is the construction of statistical shape models. Therefore, we are mostly interested in the shape of the organ's surface, which we represent by the two most prominent feature images in our method: a distance and a curvature image of the surface. Together, they provide a good description of the shape of an object. Other possible feature images, which are then simultaneously registered, encode additional information about the organs like the CT or MRI data. Registering all feature images together represents our assumption that a good registration result should bring all of these feature images into correspondence.

In addition, we need to incorporate prior knowledge about the registration result in order to be able to address the inherently ill-posed problem of image registration. The result is given in the form of a vector field, and in the most basic form of our registration method, we simply enforce the smoothness of this vector field by controlling the norm of its first derivative. However, it turns out that this regularizer, which is also found in the Demons or Diffusion registration algorithms [3, 4], allows large and unnatural looking volume change. By penalizing volume change, we impose our prior knowledge that the registration result should not compress or expand the objects excessively. This is achieved by penalizing the linearized volume change caused by the vector field.

While this generic knowledge about registration and the resulting regularization applies to most registration tasks, we can go even further and include specific prior knowledge about the objects under consideration. This is done by penalizing deformations that deviate from the space of known deformations for the specific type of object. This space is modeled by a statistical deformation model derived from previous registration results of the same object class, that is, the same organ.

One of the main contributions of this paper is the formulation, discretization, and minimization of the registration problem as a continuous functional, integrating all the different terms described above into a single analytic formula for the deformation field. This formulation allows for the simple enhancement of the scheme through further terms and for a straightforward discretization. In this paper, we present a memory-efficient and flexible schemes using adaptive finite elements; this approach is presented in a general setting. The results shown in this paper are obtained using the *local discontinuous Galerkin* method. This approach leads to a very simple formulation even of complex regularization terms and is well suited for the use with distributed grids and nonconforming adaptation. The implementation is based on DUNE-FEM [5], which uses the generic grid interface from the Dune library [6, 7]. We use the ALUGrid [8] which supports nonconforming local adaptivity and the possibility of domain decomposition and dynamic load balancing for parallel computations. These advances in efficiency and memory consumption enable us to perform registrations that were previously not possible. In particular the use of large 3D statistical deformation models, which would typically require in excess of 700 MB for each mode of variation with a uniform discretization, becomes possible for the first time. Pre- and postprocessing are performed using ITK and VTK [9, 10].

**1.1. Prior Work.** Nonrigid registration is an extremely well researched problem. For an overview of registration methods we refer to the recent survey papers by Zitová and Flusser [11] (image registration), Audette et al. [12] (surface registration), and in particular the book by Modersitzki [4] for a thorough discussion of variational methods for image registration. The most basic form of our method, that is, leaving out all the optional terms, is closely related to Thirion's Demons algorithm [3] and Modersitzki's diffusion registration algorithm [4].

The idea of surface registration using a distance or level-set representation of surfaces has been introduced by Paragios et al. [13], and the inclusion of additional feature images, especially for parametrized surfaces, is used for instance in [14]. The use of curvature images has been presented in our previous paper [15].

Volume preserving image registration was introduced by Rohlfing et al. in [16] and Haber and Modersitzki in [17]. Rohlfing et al. include a term penalizing volume change in a B-spline-based registration framework, while Haber and Modersitzki enforce strict nonlinear volume preservation in a variational formulation. In our approach, we wish to allow a limited amount of volume change and therefore use a soft constraint, that is, an additive penalty term. For efficiency, we penalize only the linear part of the volume change, and we show that this is equivalent to the linear elastic regularization term first introduced by Broit [18] and Christensen et al. in [19], even though our motivation for using this regularizer does not stem from modelling the organs as elastic bodies.

The concept of statistical deformation models and their inclusion into registration algorithms has been researched by several groups [20–23]. However, these methods either constrain the registration result strictly to the span of the model or use an interlaced algorithm performing the statistical regularization in a separate step, whereas we present an integrated formulation, which builds on our previous publication [24]. Furthermore, all of the previous methods are formulated within the discretization framework preferred by the respective group, whereas our method provides a continuous formulation of the deformation prior and, independently, a finite element discretization of this prior.

The use of finite elements for image registration goes back at least as far as [25], and we published a first finite element registration algorithm in [15]. The final model derived in this paper results in an elliptic problem with a nonlinear forcing term. The finite element discretization for general elliptic problems has now been employed for decades and can be considered standard. A summary of the standard approach of conforming, continuous finite elements can be found in [26]. Since we are using nonconform locally adapted grids with distributed memory parallelization, we employ a discontinuous finite element approach. An overview of this class of schemes can be found in [27]. The method we use is based on the *local discontinuous Galerkin* scheme introduced in [28]. To compute the second order derivatives in the elliptic regularization operator, the model is rewritten as a system of first-order equations, and these terms are discretized element-wise using intercell fluxes to enforce consistency with the analytical model. This results in a small discretization stencil, which leads to a simple implementation on nonconform distributed grid structures. The flexibility of this approach leads to a very simple formulation in spite of the rather complex elastic regularization terms. Also, the freedom in the choice of the basis functions allows us to use an orthonormal basis resulting in a diagonal mass matrix which simplifies the statistical regularization term considerably.

## 2. Registration Method

In the following, we describe our registration method. At its core, it is an image registration method and as such can be used directly on images. But as our focus lies in registering the shape or surface of organs, we describe how the method can be used to register two surfaces  $\Gamma_0, \Gamma_1 \subset \mathbb{R}^d$ . The surfaces can be segmented from medical images or acquired otherwise. We assume that they are already rigidly preregistered, for instance by Procrustes alignment [29], so that our algorithm only needs to recover the nonrigid component of the registration. Moreover, the preregistration enables us to choose a common rectangular image domain  $\Omega \subset \mathbb{R}^d$  which contains both surfaces, so that we can represent each surface  $\Gamma$  by its signed distance function  $I : \Omega \rightarrow \mathbb{R}$ :

$$I(x) := \begin{cases} \text{dist}(x, \Gamma) & x \in \text{outside}(\Gamma) \\ 0 & x \in \Gamma \\ -\text{dist}(x, \Gamma) & x \in \text{inside}(\Gamma), \end{cases} \quad (1)$$

where  $\text{dist}(x, \Gamma)$  is the Euclidean distance from  $x$  to  $\Gamma$  and the inside and outside of  $\Gamma$  have to be assigned in a meaningful way. For open surfaces, for which inside and outside cannot be defined, an unsigned distance function can be used.

The aim of a registration algorithm is to find a deformation field  $u : \Omega \rightarrow \mathbb{R}^d$  such that the target surface's distance function  $I_1$  warped with this deformation field, that is,  $I_1(x + u(x))$ , is as close as possible to the distance function of the reference surface given by  $I_0$ . The registration result of the distance functions implies a registration of the surfaces.

We formulate the registration problem as a minimization problem. It is shown in [4] that virtually all registration methods can be interpreted in this way. The deformation field  $u$  is sought as the minimum of a functional which is the sum of two terms: *distance* and *regularization* terms. Thus, the registration problem consists of finding the minimum of the functional

$$\mathcal{J}[u] = \mathcal{D}[u] + \mathcal{R}[u], \quad (2)$$

with distance term  $\mathcal{D}$  and regularization term  $\mathcal{R}$ . The former measures the distance between the reference and the registration target. At its minimum, the warp of the target is as close as possible to the reference image. The regularization term on the other hand measures the smoothness or regularity of the registration result  $u$ . The smaller it is, the more regular the solution will be. By minimizing both terms simultaneously, we try to bring the reference and target as close together as possible while keeping the deformation field reasonably smooth. We believe that there is no single generic distance or regularization term that guarantees a good registration in every scenario. The notion of correspondence is application specific, and the more knowledge about the registration task at hand we can include into the method, the higher the chances will be to obtain a result that meets our requirements. For example, the most basic regularization term we propose simply penalizes the squared  $L^2$ -norm of the gradient. By including additional regularization terms, we can also penalize volume change or the deviation from a statistical

model of the object to be registered, thus improving the registration result. The same holds for the distance measure, where the simplest approach measures the  $L^2$  difference between the two surfaces. Again considerably better results can be achieved by including additional terms, for example, the curvature of the surfaces.

**2.1. Distance Term.** The basis for the distance term  $\mathcal{D}$  is the  $L^2$  difference between the warp of the signed distance images  $I_1$  and the reference  $I_0$  of the two surfaces to be registered:

$$\begin{aligned} & \|I_1(x + u(x)) - I_0(x)\|_{L^2(\Omega)}^2 \\ &:= \int_{\Omega} (I_1(x + u(x)) - I_0(x))^2 dx. \end{aligned} \quad (3)$$

The distance images of two similar surfaces have a similar range of values, especially close to the surfaces, which makes the  $L^2$  distance measure an appropriate choice for their comparison. In order to prevent undesirable effects at the boundary, where the distance function of each surface may be cut off at a different value, we bound the distance images at a certain distance  $b$  from the surface:

$$I^b(x) := \begin{cases} I(x) & \text{if } I(x) \leq b \\ b & \text{if } I(x) > b, \end{cases} \quad (4)$$

and register these bounded distance images instead of the original  $I(x)$ . The bound  $b \in \mathbb{R}$  should be chosen so that the  $b$  level set of each surface we want to register is completely contained inside  $\Omega$ .

**2.2. Robust Distance Measures.** For noisy or otherwise difficult feature images it can be advantageous to use a robust distance measure, which dampens the influence of the overly large differences between the images; see [30] for a review of different robust cost functions. We propose using a robust distance measure based on the Geman-McClure estimator [30, 31], which has been successfully used for medical image registration in [32]. It can be easily realized by weighting the distance measure (3) with a term  $Q_I(x)$ :

$$\mathcal{D}_I[u] := \frac{1}{2} \int_{\Omega} \frac{1}{Q_I(x)} (I_1(x + u(x)) - I_0(x))^2 dx. \quad (5)$$

For the Geman-McClure distance measure, we have  $Q_I(x) = C^2 + (I_1(x + u(x)) - I_0(x))^2$ , with a regularization parameter  $C \in \mathbb{R}$  which controls the robustness of the measure. A similar term is used in Thirion's Demons algorithm [3], where the norm of the gradient of the image replaces  $C$ :  $Q_I(x) = |\nabla I_1(x + u(x))|^2 + (I_1(x + u(x)) - I_0(x))^2$ . In our experiments, both weights yielded similar results, which is not surprising as for distance functions  $|\nabla I| = 1$  almost everywhere. We have found that for distance images that represent surfaces free from artifacts or excessive noise, it is not necessary to use a robust distance measure. But it proved to be of good use for the additional feature images introduced in the following sections, such as the curvature images.

**2.2.1. Curvature Guided Registration.** When registering surfaces by means of their distance images, the problem arises that by definition, and the value of the distance function is zero on the whole surface and therefore contains no information on the surface. Therefore, the distance function  $\mathcal{D}$  is minimized whenever a point on one surface is registered onto a point on the other surface even if it does not correspond functionally or semantically to the given point. In fact, when we try to minimize the registration functional with a gradient descent scheme, the corresponding point is only sought in the direction of the gradient of the distance image, that is, perpendicular to the surface. This effect is somewhat alleviated by the regularization term, but this is often not enough to obtain a sensible registration. See the example in Section 4.3 for instance.

For bone registration, we wish to establish correspondence between points that have a similar anatomical function. So similar bumps, crests, ridges, and so forth should be matched. Such features are well described by the curvature of the surface. In fact, for a large class of objects, corresponding points on two surfaces have similar curvature. Figure 1 illustrates this for the mean curvature of human skulls. Therefore, we use the mean curvature as an additional feature to be matched.

With the surfaces represented by their distance images, the curvature is easily calculated by  $H(x) = \text{div}(\nabla I / \|\nabla I\|)$ . For each  $x \in \Omega$ ,  $H(x)$  is the mean curvature of the level surface passing through  $x$ . So if  $x$  is on the zero level set of  $I$ ,  $H(x)$  is the mean curvature of the surface at that point. Since for distance images  $\|\nabla I\| = 1$  almost everywhere, the curvature image  $H$  is even more easily computed as  $H = \Delta I$ .

The curvature images are included in the registration process with a distance term analogous to that in (5) as follows:

$$\mathcal{D}_H[u] := \frac{1}{2} \int_{\Omega} \frac{1}{Q_H(x)} (H_1(x + u(x)) - H_0(x))^2 dx. \quad (6)$$

The overall distance measure is then given as  $\alpha \mathcal{D}_I[u] + \beta \mathcal{D}_H[u]$  with  $\alpha, \beta \in \mathbb{R}^+$  controlling the influence of the distance and curvature images.

**2.2.2. Additional Feature Images.** In an obvious fashion, any number of additional feature images can be added. If we denote the  $k$ th pair of feature images by  $X_0^k, X_1^k$ , the full distance term for our registration method is given as

$$\begin{aligned} \mathcal{D}[u] &:= \sum_{k=1}^n \alpha_k \mathcal{D}_{X^k}[u] \\ &= \sum_{k=1}^n \frac{\alpha_k}{2} \int_{\Omega} \frac{1}{Q_{X^k}} (X_1^k(x + u(x)) - X_0^k(x))^2 dx, \end{aligned} \quad (7)$$

with weighting parameters  $\alpha_k$ . In our experiments, we have included the original CT scans from which the bone surfaces were segmented as additional feature images where they were available. 2D projections of two such CT scans can be seen in Figure 2. Additional image modalities or manual annotations could also be included provided that they are available for

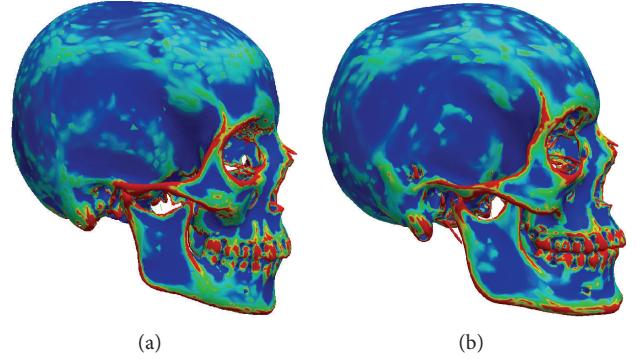


FIGURE 1: Two skulls colored according to their mean curvature. We see that corresponding points have similar mean curvature.

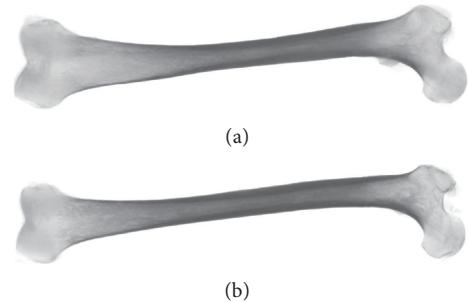


FIGURE 2: 2D projections of CT scans of two femurs.

both surfaces to be registered. Obviously for each of the surfaces, all its feature images need to be in correspondence, which is usually already the case or can be achieved with a rigid registration method. For multimodal image pairs, more sophisticated multimodal distance measures can be employed instead of the  $L^2$  distance measures used in (7) (cf. [4, 33]).

**2.3. Regularization Term.** Registration is an ill-posed problem, and any algorithm trying to minimize a distance measure without enforcing some kind of smoothness or regularity on the solution is bound to fail. We begin by introducing a very basic regularization term, which is later enhanced by adding further terms.

One of the most basic ways to control the smoothness of the deformation field  $u$  is through its first derivative  $Du$ , which we will, for simplicity, denote by  $\nabla u$ . But the reader should bear in mind that  $u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^d$ , and hence  $\nabla u \in \mathbb{R}^{d \times d}$ . We define the basic regularization term as

$$\begin{aligned} \mathcal{R}_g[u] &:= \frac{1}{2} \int_{\Omega} \|\nabla u\|^2 dx \\ &= \frac{1}{2} \sum_{l=1}^d \int_{\Omega} |\nabla u_l|^2 dx. \end{aligned} \quad (8)$$

The smaller  $\mathcal{R}_g[u]$  is, the smoother the deformation field  $u$ .

**2.3.1. Volume Preservation.** While the regularization term (8) forces the deformation field to be smooth, it still allows some quite unnatural deformations. In particular, it allows for excessive expansion or compression of the registered object by the deformation field; see Section 4.4 for an example. The compression or expansion of the warp  $x + u(x)$  can be measured by the determinant of its first derivative  $\det[D(x + u(x))]$ . A volume preserving deformation field must satisfy  $\det[D(x + u(x))] \equiv 1$ . Naturally, as we are mostly interested in registering bones of different individuals, which in general do not have the same volume, we do not wish to enforce a strict incompressibility constraint as in fluid dynamics or other registration approaches [17]. Instead, we add a soft incompressibility constraint to our existing regularization term based on the linearization of  $\det[D(x + u(x))]$ . Thus, volume change is limited but not completely prohibited.

Since  $\det[D(x + u(x))] = 1 + \operatorname{div} u + (\text{nonlinear terms})$ , it follows that the smaller the divergence of  $u$ , the closer the linearization of the determinant is to 1 and therefore the field to being volume preserving, provided the values of  $Du$  are not too large and therefore the linearization justified. Therefore we add the square of the deformation field's divergence to the functional:

$$\begin{aligned}\mathcal{R}[u] &:= \mu \mathcal{R}_g[u] + \nu \mathcal{R}_d[u] \\ &= \mu \frac{1}{2} \int_{\Omega} \|\nabla u\|^2 dx + \nu \frac{1}{2} \int_{\Omega} (\operatorname{div} u)^2 dx.\end{aligned}\quad (9)$$

In a later section, we will see that the Euler-Lagrange equations for the functional (9) correspond to those of the well-known linear elastic registration methods (see Appendix A). Although the volume preservation is only approximative, it enhances the registration result visibly; see Section 4.4.

**2.3.2. Statistical Deformation Prior.** It is well known that the regularization terms can be interpreted from two different perspectives. On the one hand, they serve as numerical stabilizers and make the numeric treatment of ill-posed problems feasible. On the other hand, the regularization term incorporates prior knowledge about the solution into the problem. In this respect, the regularization terms that we have introduced so far were generic in the sense that they only require the solution to be smooth or volume preserving. They do not take properties of the object to be registered into account. Even though registration is a prerequisite for expressing prior knowledge about a specific class of objects, it can itself benefit from such prior knowledge when the data is noisy. Therefore, we propose including prior knowledge about the specific class of objects by introducing an additional regularization term. This term is based on a probability distribution estimated from previous successful registration results and penalizes unlikely deformations. This type of regularization becomes very natural when we consider the probabilistic interpretation of the regularization approach. The regularization term  $\mathcal{R}[u]$  can be seen as defining a prior probability distribution  $p[u]$  over the function space, and the distance term  $\mathcal{D}[u]$  defines the likelihood  $p[D \mid u]$  of observing the data given deformation field  $u$ . See Wang

and Staib [21] for a detailed discussion of this probabilistic interpretation for variational image registration.

Assume that we already have a set of  $n$  deformation fields  $\{u_1, \dots, u_n\}$ , mapping from a common reference image  $I_R$  to a given set of target images  $\{I_1, \dots, I_n\}$ . We expect that a new registration result is unlikely to be very different from the ones already given, since it registers an object of the same class. Indeed, we can regard the deformation fields  $u_1, \dots, u_n$  as training data from which we can estimate a probability distribution over the possible deformations. This is the main idea behind *Statistical Shape Models* [34, 35] and the related *Statistical Deformation Models* [22], to which our problem belongs. In these approaches, the common assumption is that the objects (i.e., the shapes and the deformation fields) follow a normal distribution. Its parameters are estimated from the training data.

We introduce a continuous formulation for deformation models here, which fits naturally into our continuously defined registration framework and permits a straightforward finite element discretization. The model is characterized by the mean field  $\bar{u}$

$$\bar{u}(x) = \frac{1}{n} \sum_{i=1}^n u_i(x) \quad (10)$$

and the covariance operator  $\mathcal{C} : L^2(\Omega, \mathbb{R}^d) \rightarrow L^2(\Omega, \mathbb{R}^d)$

$$\mathcal{C} = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u}) \otimes (u_i - \bar{u}), \quad (11)$$

which acts on a deformation field  $w \in L^2(\Omega, \mathbb{R}^d)$  through

$$\begin{aligned}\mathcal{C}[w](x) &= \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(x) \\ &\quad \times \int_{\Omega} (u_i - \bar{u})(y) \cdot w(y) dy.\end{aligned}\quad (12)$$

This operator is the continuous analogon to the sample covariance matrix. It is a linear integral operator with a symmetric integral kernel given by

$$C(x, y) = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(x) (u_i - \bar{u})^T(y). \quad (13)$$

$\mathcal{C}$  is a self-adjoint (i.e., symmetric) compact operator. Thanks to the spectral theorem for compact normal operators (see [36] for instance) it admits an eigenvalue decomposition with positive eigenvalues  $\sigma_i^2$  and corresponding orthonormal eigenfunctions  $\rho_i$ . As  $\mathcal{C}$  is estimated from  $n$  examples  $u_i - \bar{u}$ , there are at most  $m \leq n$  nonzero eigenvalues;  $\sigma_i^2$  and  $\mathcal{C}$  can be represented as

$$\mathcal{C}[w](x) = \sum_{i=1}^m \sigma_i^2 \rho_i(x) \int_{\Omega} \rho_i(y) \cdot w(y) dy. \quad (14)$$

Moreover, the spectral theorem guarantees the orthogonal decomposition of  $L^2(\Omega, \mathbb{R}^d)$  into the span of the eigenfunctions  $\rho_i$  and the operator's null space  $N(\mathcal{C})$ :

$$L^2(\Omega, \mathbb{R}^d) = \operatorname{span} \{\rho_i \mid i = 1, \dots, m\} \perp N(\mathcal{C}). \quad (15)$$

The span of the eigenfunctions is the same as the span of the  $u_i - \bar{u}$ . That means that it contains all linear combinations of the training examples. On this  $m$ -dimensional space, which is frequently called the *model space* of the deformation model, the covariance operator  $\mathcal{C}$  is invertible. Thanks to the eigenvalue decomposition, the inversion of  $\mathcal{C}$  on the model space is easily achieved by inverting the eigenvalues and coincides with the formulation of the pseudoinverse  $\mathcal{C}^\dagger$  on the whole space  $L^2(\Omega)$  as follows:

$$\mathcal{C}^\dagger [w](x) = \sum_{i=1}^m \sigma_i^{-2} \rho_i(x) \int_{\Omega} \rho_i(y) \cdot w(y) dy. \quad (16)$$

We can therefore formally define a normal distribution  $\mathcal{N}(\bar{u}, \mathcal{C})$  with mean  $\bar{u}$  and covariance  $\mathcal{C}$  on  $L^2(\Omega)$ :

$$p_1(u) = \frac{1}{Z} e^{-(1/2)\langle u - \bar{u}, \mathcal{C}^\dagger[u - \bar{u}] \rangle_{L^2(\Omega, \mathbb{R}^d)}}, \quad (17)$$

where  $Z$  is a normalization parameter. Technically, while this function exists and is the natural extension of the density function of a statistical shape or deformation model to a continuous function space, it is not a well-defined density function. However, for any finite discretization  $u_h$ , the corresponding density of the multivariate normal  $\mathcal{C}[u_h]$  exists and is well defined. This distribution can be used as the basis for a statistical regularization term. However, it only represents prior knowledge about the model space. Its complement,  $N(\mathcal{C})$ , is completely ignored. A statistical regularization based on this distribution would permit deformation fields  $u \in L^2(\Omega, \mathbb{R}^d)$  that are arbitrarily far away from the model space and would regularize only their projection onto the model. An extreme measure taken by other authors is therefore to restrict the registration results strictly to the model space [20, 22]. We on the other hand wish to explicitly allow results that lie outside (but still close to) the model space, as only these results can be used as reasonable additions to the statistical model.

We define an additional distribution on  $N(\mathcal{C})$ , which imposes our prior knowledge that the registration results should not lie far from the model. In the absence of additional example data, we simply assume a normal distribution with mean  $\bar{u}$  and uncorrelated uniform variance of  $\sigma^2 \in \mathbb{R}^+$ , which controls how much the results are allowed to deviate from the model space:

$$p_2(u) = \frac{1}{Z} e^{-(1/2)\sigma^{-2}\langle u - \bar{u}, u - \bar{u} \rangle_{L^2(\Omega, \mathbb{R}^d)}}. \quad (18)$$

While this term could in principle be restricted to  $N(\mathcal{C})$ , we define it on the whole space  $L^2(\Omega, \mathbb{R}^d)$ . This corresponds to shrinkage estimation, a technique used in statistics to improve the estimation of the covariance [24, 37], which adds the additional variance of  $\sigma$  also on the model space.

To define the distribution  $p(u)$  characterizing our complete statistical model, we assume the two normal distributions to be independent and define  $p(u)$  as their product

$$p(u) := p_1(u) p_2(u) = \frac{1}{Z} e^{-(1/2)\langle u - \bar{u}, \mathcal{C}^\dagger[u - \bar{u}] \rangle} \times \frac{1}{Z} e^{-(1/2)\sigma^{-2}\langle u - \bar{u}, u - \bar{u} \rangle}. \quad (19)$$

It is then natural to introduce the functional

$$\begin{aligned} \mathcal{P}[u] &= \frac{1}{2} \gamma \int_{\Omega} (u - \bar{u})(x) \cdot \mathcal{C}^\dagger [u - \bar{u}](x) dx \\ &\quad + \frac{1}{2} \gamma \int_{\Omega} \sigma^{-2} (u - \bar{u})(x) \cdot (u - \bar{u})(x) dx \\ &\propto -\ln p[u] \end{aligned} \quad (20)$$

as an additional regularization term, which penalizes unlikely deformation fields. The constant  $\gamma$  is used as a weighting factor. We thus arrive at the following functional describing our registration model:

$$\mathcal{J}[u] = \mathcal{D}[u] + \mathcal{R}[u] + \mathcal{P}[u]. \quad (21)$$

### 3. Finite Element Discretization

In this section, we describe the discretization and minimization of the functional (21) based on a finite element method and an adaptive multiresolution pseudo-time stepping approach. We start with the description of the spatial discretization focusing on the terms from the deformation prior. This finite element discretization of the continuously defined deformation prior is one of the most important contributions of this paper, and thanks to use of a memory-efficient adaptive finite element basis makes the use of large 3D statistical deformations possible for the first time.

**3.1. Space Discretization.** To find a minimizer of the functional (21), we wish to solve the weak form of the corresponding Euler-Lagrange equation:  $J'[u, \varphi] = 0$  for all suitable test functions  $\varphi$ . The details of how the derivative  $J'[u, \varphi] = \mathcal{D}'[u, \varphi] + \mathcal{R}'[u, \varphi] + \mathcal{P}'[u, \varphi]$ , which can be interpreted as a bilinear form, is computed can be found in Appendix A. The three terms, corresponding to the distance measure, regularization term, and the deformation prior, respectively, are given by

$$\begin{aligned} \mathcal{D}'[u, \varphi] &= \sum_{k=1}^n \int_{\Omega} \frac{\alpha^k}{Q_{X^k}(x)} (X_1^k(x + u(x)) \\ &\quad - X_0^k(x)) \nabla X_1^k(x + u(x)) \varphi(x) dx, \end{aligned} \quad (22)$$

$$\mathcal{R}'[u, \varphi] = \int_{\Omega} \mu \nabla u : \nabla \varphi + \nu \operatorname{div} u \operatorname{div} \varphi dx, \quad (23)$$

$$\begin{aligned} \mathcal{P}'[u, \varphi] &= \int_{\Omega} \gamma \mathcal{C}^\dagger[u - \bar{u}](x) \cdot \varphi(x) dx \\ &\quad + \int_{\Omega} \gamma \sigma^{-2}(u - \bar{u})(x) \cdot \varphi(x) dx. \end{aligned} \quad (24)$$

The sum in (22) is over all feature images  $X_k$  used for the registration, for example, the distance maps  $I$ , the curvature  $H$ , and the CT scans.

We employ a finite element discretization to compute the deformation field. Given a function space  $V_h \subset L^2(\Omega, \mathbb{R}^3)$  of finite dimension  $N$ , we seek an approximate deformation field, which we will also denote by  $u \in V_h$ , satisfying  $\mathcal{J}'[u, \varphi] = 0$  for all test function  $\varphi \in V_h$ .

After choosing basis functions  $\varphi_1, \dots, \varphi_N$  of  $V_h$ , a function  $w \in V_h$  can be represented as  $w(x) = \sum_{k=1}^N w_k \varphi_k(x)$  with a vector of degrees of freedom (DOF) denoted by  $\mathbf{w}_h = (w_k)_{k=1}^N$ .

The derivation of a discrete version both for the regularization term  $\mathcal{R}'$  and the distance measure  $\mathcal{D}'$  is straightforward and will be briefly sketched in Section 4.1 together with details on the construction of the discrete function space  $V_h$  used in our tests. In the following, we will concentrate on the term  $\mathcal{P}'$  arising from the statistical deformation prior.

The term  $\mathcal{P}'$  defined in (24) includes the eigenfunctions of the covariance operator  $\mathcal{C}$ , defined in (12). The eigenfunctions are needed for the efficient (pseudo-) inversion of  $\mathcal{C}$  according to (16). In order to formulate a discretization and an implementation for  $\mathcal{P}'$ , we need to actually calculate these eigenfunctions. According to the ideas of functional PCA [38], the eigenfunctions can be calculated with help of a finite dimensional eigenvalue decomposition if both the argument  $w$  and the training examples  $u_i$  are given as a linear combination of a set of basis functions.

So to derive a discrete formulation of  $\mathcal{C}$ , we fix the argument  $w$  and examples  $u_i$  to be from the discrete function space  $V_h$ . They are then given as linear combinations of the basis functions  $\varphi_k$ , and we have the coefficient vectors  $\mathbf{w}_h$  and  $\mathbf{v}_{h,i} := \mathbf{u}_{h,i} - \bar{\mathbf{u}}_h$ . The covariance operator  $\mathcal{C}$  from (12) then takes the form

$$\begin{aligned} \mathcal{C}[w](x) &= \frac{1}{n} \sum_{i=1}^n \sum_k \mathbf{v}_{ik} \varphi_k(x) \int_{\Omega} \sum_j \mathbf{v}_{ij} \varphi_j \cdot \sum_l \mathbf{w}_l \varphi_l dy \\ &= \sum_k \left( \frac{1}{n} \sum_{i=1}^n \sum_{l,j} \mathbf{v}_{ik} \mathbf{v}_{ij} \int_{\Omega} \varphi_j \cdot \varphi_l dy \mathbf{w}_l \right) \varphi_k(x) \quad (25) \\ &= \sum_k (\mathbf{A} \mathbf{w}_h)_k \varphi_k(x). \end{aligned}$$

The matrix  $\mathbf{A} = (a_{kl})$  is given by

$$a_{kl} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N \mathbf{v}_{ik} \mathbf{v}_{ij} m_{jl}, \quad (26)$$

where we have used  $m_{jl}$  to denote the entries of the mass matrix  $\mathbf{M}$ ; that is,  $m_{jl} = \int_{\Omega} \varphi_j(y) \cdot \varphi_l(y) dy$ .

Equations (12) and (26) show that  $\mathcal{C}$  is restricted to  $V_h$  maps to  $V_h$  and is represented by the matrix  $\mathbf{A} = (a_{kl})$ , which, resubstituting the definition of  $\mathbf{v}_{ik}$  into (26), can be represented as

$$\mathbf{A} = \frac{1}{n} \sum_{i=1}^n (\mathbf{u}_{hi} - \bar{\mathbf{u}}_h) (\mathbf{u}_{hi} - \bar{\mathbf{u}}_h)^T \mathbf{M} =: \Sigma \mathbf{M}, \quad (27)$$

where  $\Sigma$  is the sample covariance matrix of the DOF vectors  $\mathbf{v}_{h,i} = \mathbf{u}_{hi} - \bar{\mathbf{u}}_h$ . The eigenvalue decomposition of this discrete version of  $\mathcal{C}$  can be achieved by an eigenvalue decomposition of the matrix  $\mathbf{A}$ . Compared with traditional discrete statistical models as [22, 34, 35], where only the covariance matrix  $\Sigma$  is used, in our case, the mass matrix  $\mathbf{M}$  slightly complicates matters as it makes the matrix  $\mathbf{A}$  nonsymmetric in general. In functional PCA, this problem is solved by first calculating an eigenvalue decomposition of the symmetric matrix:

$$\mathbf{M}^{1/2} \mathbf{A} \mathbf{M}^{-1/2} = \mathbf{M}^{1/2} \Sigma \mathbf{M}^{1/2}. \quad (28)$$

This is a symmetric positive semidefinite ( $N \times N$ )-matrix with rank  $m \leq n \ll N$ . Remember that  $n$  is the number of training examples  $u_i$  which is usually much smaller than  $N$ , the number of degrees of freedom of the discrete function space  $V_h$ . The positive eigenvalues  $\sigma_1^2, \dots, \sigma_m^2$  and the matrix  $\mathbf{S}$  of their corresponding orthonormal right eigenvectors  $\mathbf{r}_i$  can be computed efficiently with a singular value decomposition of an  $(m \times m)$  matrix, without the need to compute a full large  $(N \times N)$  singular value decomposition. We thus have the eigenvalue decomposition:

$$\mathbf{M}^{1/2} \mathbf{A} \mathbf{M}^{-1/2} = \mathbf{S} \operatorname{diag}(\sigma_1^2, \dots, \sigma_m^2) \mathbf{S}^T \quad (29)$$

and consequently the decomposition of  $\mathbf{A}$ :

$$\mathbf{A} = \mathbf{M}^{-1/2} \mathbf{S} \operatorname{diag}(\sigma_1^2, \dots, \sigma_m^2) \mathbf{S}^T \mathbf{M}^{1/2} \quad (30)$$

with eigenvalues  $\sigma_i^2$  and eigenvectors  $\mathbf{M}^{-1/2} \mathbf{r}_i$ . The associated functions  $\rho_i = \sum_k (\mathbf{M}^{-1/2} \mathbf{r}_i)_k \varphi_k$  are the eigenfunctions of the discretized covariance operator from (25).  $\mathcal{C}$  being a symmetric operator, we expect its eigenfunctions to be orthonormal with respect to the  $L^2(\Omega, \mathbb{R}^3)$  scalar product:

$$\int_{\Omega} \rho_i \cdot \rho_j = \mathbf{r}_i^T \mathbf{M}^{-1/2} \mathbf{M} \mathbf{M}^{-1/2} \mathbf{r}_j = \mathbf{r}_i^T \mathbf{r}_j = \delta_{ij}. \quad (31)$$

With the eigenfunctions  $\rho_i$ , we can rewrite (24) with an explicit formulation for the pseudoinverse  $\mathcal{C}^\dagger$  as follows:

$$\begin{aligned} \mathcal{P}'[u, \varphi] &= \sum_{i=1}^m \gamma \sigma_i^{-2} \int_{\Omega} (u - \bar{u})(y) \cdot \rho_i(y) dy \\ &\quad \times \int_{\Omega} \rho_i(x) \cdot \varphi(x) dx \\ &\quad + \gamma \sigma^{-2} \int_{\Omega} (u - \bar{u})(x) \cdot \varphi(x) dx, \end{aligned} \quad (32)$$

which can easily be implemented once the eigenfunctions  $\rho_i$  have been calculated.

This formulation takes the same form in the continuous, and the discrete setting and could have been calculated directly using the continuous eigenvalue decomposition Equation (16). The discretization lies only in the restriction of  $\mathcal{P}'$  to the discrete function space  $V_h$  and the calculation of the eigenfunctions via a matrix SVD as described above.

**3.2. Minimization Strategy.** Registration is achieved by minimizing the functional (21). We introduce an artificial time variable  $t$  and try to minimize the functional by computing its gradient flow, that is, we solve the time-dependent partial differential equation:

$$\partial_t u - \mathcal{J}'[u] = 0 \quad (33)$$

for a function  $u(x, t)$ , given an initial solution  $u^0 = u(x, 0)$ . Note that the term  $\partial_t u$  has to be interpreted in the weak sense, that is, in the form  $\int_{\Omega} \partial_t u \varphi$ . A time stepping scheme for (33) can be viewed as a continuous version of gradient descent; indeed, if we use an Euler scheme for the time discretization of (33), we end up with a gradient descent iteration.

Choosing an explicit Euler scheme to solve (33) leads to a severe time step restriction, coupling the time step  $\tau$  with the mesh width  $h$  via  $\tau = O(h^2)$  due to the regularization term (8). On the other hand, due to the nonlinearity of the distance term  $\mathcal{D}[u]$ , a purely implicit iteration scheme would require the solution of a large nonlinear system in each iteration. Therefore, we propose to use semi-implicit time stepping scheme. We split up the functional  $\mathcal{J}[u]$  into an explicit part  $\mathcal{J}_{\text{expl}}[u]$ , containing all terms with nonlinear derivatives and an implicit part  $\mathcal{J}_{\text{impl}}[u]$  containing all terms with linear derivatives. The iteration scheme, in its simplest form, is then defined as

$$u^{n+1} = u^n + \tau \mathcal{J}'_{\text{impl}}[u^{n+1}] + \tau \mathcal{J}'_{\text{expl}}[u^n]. \quad (34)$$

In our case, this means that  $\mathcal{J}_{\text{expl}} = \mathcal{D}$  and  $\mathcal{J}_{\text{impl}} = \mathcal{R} + \mathcal{P}$ . If additional terms are introduced, they can be added to either  $\mathcal{J}_{\text{expl}}$  or  $\mathcal{J}_{\text{impl}}$ , depending on their linearity. Each iteration step thus requires the solution of a *linear* system of equations for which we employ an iterative Krylov type solver (in this case the BiCGStab method). To obtain higher order version of this method, IMEX Runge-Kutta schemes are used [39].

For our experiments, we chose a locally adaptive multiresolution strategy; we first minimize the functional on a coarse uniform grid; that is, starting with the initial guess  $u^0 \equiv 0$ , we solve (33) up to some fixed time  $T_0$ . Then the grid is refined around the reference surface, and the coarse solution is interpolated onto the refined grid. This is then used as starting value for solving (33) on the refined grid up to some fixed time  $T_1$ . This process is repeated until a sufficiently fine resolution has been reached. The times  $T_i$  have been experimentally determined. In each step, all elements closer to the surface than a certain threshold  $\Theta$  are refined; that is,  $|I_0| < \Theta$ . This threshold is decreased in each step, so that the final solution is calculated on a grid that offers a very fine resolution close to the reference surface and becomes more and more coarse further out. An example is shown in Figure 3.

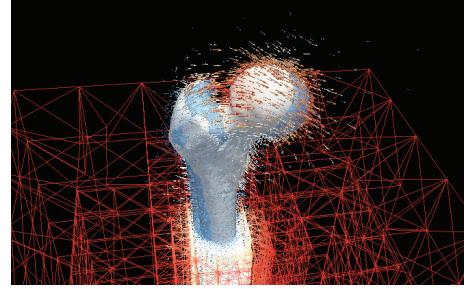


FIGURE 3: Visualization of the registration of two femurs. The deformation field that deforms one femur into the other is calculated on an adaptive grid.

## 4. Results

**4.1. Implementational Details.** The scheme is implemented in the Dune framework, a software library allowing the generic implementation of grid based numerical schemes [6, 7]. The finite element implementation is based on the DUNE-FEM module [5]. Pre- and post-processing is done using ITK and VTK [9, 10].

The spatial discretization of the deformation field  $u$  is based on a tessellation  $\mathcal{T}_h = \{T_i\}_{i \in \mathcal{I}}$  of the image domain  $\Omega$ . This tessellation must be nonoverlapping and can be uniform or adaptive and is typically made up of triangles or rectangles in 2D or tetrahedra or hexahedra in 3D. We use the ALUGrid library [8] which supports unstructured meshes in 2D and 3D with nonconforming local adaptivity and the possibility of domain decomposition and dynamic load balancing for parallel computations. Figure 3 shows a visualization of such a nonconforming locally adaptive grid.

The spatial discretization employed is based on the *local discontinuous Galerkin* (LDG) scheme. Given a tessellation  $\mathcal{T}_h$ , this scheme follows the same ideas as the standard Galerkin method [26] but employs a discontinuous ansatz space:  $V_h^q := \{v_h; v_i \in [P_q(T_i)]^d \text{ for } i \in \mathcal{I}\}$ . Here  $v_i \equiv v_{h|T_i}$  and  $P_q(T_i)$  denote the space of polynomials on the element  $T_i$  of order  $q$ . Note that there is no continuity assumption between elements, so that there is very little restriction on the tessellation. Due to the discontinuous ansatz space, the discretization of the higher order derivatives in the regularization term is slightly more involved than in the standard Galerkin method. We briefly sketch the LDG approach when applied to a partial differential equation of the form  $-\mu \Delta u - \nu \nabla \operatorname{div} u = \tilde{R}$ , where  $\tilde{R}$  combines all lower order terms. This equation corresponds to the strong form of the Euler-Lagrange equation (see Appendix A (A.9)). Rewriting the second-order terms as a first-order system for the vector valued functions  $u, w_1, \dots, w_d$ , we arrive at

$$\begin{aligned} w_k - \nabla u_k &= 0, \\ -\sum_i \partial_i (\mu w_{ki} + \nu w_{ik}) &= \tilde{R}_k \end{aligned} \quad (35)$$

with  $k = 1, \dots, d$ . This approach leads to a compact discrete form for the elastic regularization term  $\mathcal{R}'$  involving only

first-order derivatives on each element  $T \in \mathcal{T}_h$  and numerical fluxes over cell boundaries. Focusing on a single element  $T$  of the tessellation, the corresponding weak form of (35) takes the following form when we define the vector  $W_k := (\mu w_{ki} + \nu w_{ik})_i$ :

$$\begin{aligned} \int_T w_k \cdot \xi + \int_T u_k \nabla \cdot \xi - \int_{\partial T} \widehat{u_k} \xi \cdot n &= 0 \\ \int_T W_k \cdot \nabla \psi - \int_{\partial T} \widehat{W_k} \psi &= \int_T \tilde{R}_k \psi. \end{aligned} \quad (36)$$

The fluxes  $\widehat{u_k}$  and  $\widehat{W_k}$  can for example be taken as averages of the values on both sides of the boundary of  $T$  or using suitable one-sided values. For more details on the LDG method see [27, 28]. At the boundary we use homogeneous Dirichlet conditions for the deformation field  $u$ . This is a reasonable assumption if we use bounded distance images as defined in (4). In this case, the images  $I_0$  and  $I_1$  are both equal to the constant bound  $b$  near the domain boundary, and the functional only contains the regularization term  $\mathcal{R}$  in this region; that is, only the regularization operator remains for which zero boundary conditions can be prescribed.

In our implementation, we use orthogonal basis functions  $\{\varphi_l\}$ . These are easily constructed by choosing on each element  $T$  a polynomial basis  $(\phi_\alpha^T)_{\alpha=1}^r$  of  $[P^q(T)]^d$  satisfying

$$\int_T \phi_\alpha^T \cdot \phi_\beta^T = |T| \delta_{\alpha\beta}. \quad (37)$$

A basis function  $\varphi_l$  of  $V_h^q$  is then chosen to coincide with a polynomial basis function  $\phi_\alpha^T$  on one element  $T$  and vanish on all other elements. Thus the mass matrix  $M$  is a diagonal matrix of the form

$$M = \text{diag}(|T_1| I_r, \dots, |T_s| I_r), \quad (38)$$

where  $I_r$  is the identity matrix with  $r = \dim([P^q(\mathbb{R}^d)]^d)$  and  $s$  is the number of elements in the grid; that is, the total number of DOFs is  $N = sr$ . The diagonal form of  $M$  makes its storage and the calculation of the weighted sample covariance matrix  $\Sigma$  defined in (27) very efficient.

Thus, the LDG method used here not only allows us to efficiently use adaptivity and parallelization but also makes the computation of the PCA simple and efficient with respect to memory consumption and computational cost.

**4.2. 3D Surface Registration.** In this first experiment, we can observe that thanks to the level set representation of the surfaces, our method allows the accurate registration of surfaces. See Figure 3 for an example of the femur bone and Figure 4 for an example with the more complicated surface of the human skull, where our method benefits from the fact that the level set representation is independent of the topology of the surface. The topology of the skull surface is very complicated and, due to acquisition and segmentation artifacts, not necessarily the same for two different segmented skull surfaces.

In our experiments, only the reference was anatomically labeled and hand segmented with high attention to detail.

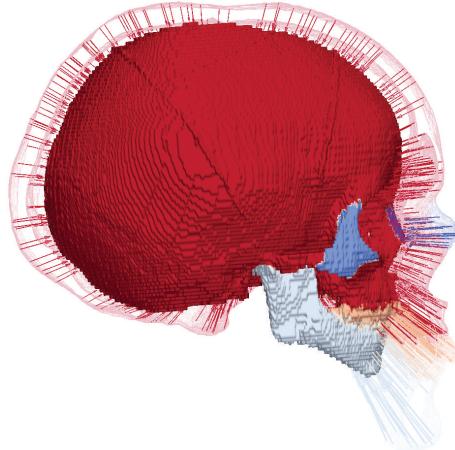


FIGURE 4: Visualization of the correspondence and the transfer of anatomical labeling between the reference (transparent outline) and a child's skull.

Figure 4 visualizes how this labelling and segmentation is transferred to another skull anatomy via the registration result. Figure 5 shows that this works over a very large range of examples from infant skulls to large adult skulls. Even though the concept of correspondence breaks down for different numbers of teeth in the reference and the target, the existing teeth are labelled correctly.

On a standard 3 GHz dual-core desktop PC with two parallel registration processes, a skull registration with 480 000 degrees of freedom (i.e., 160 000 grid points) takes about 10 minutes, a femur registration with 80 000 degrees of freedom (but more iterations) about 6 minutes. These are only indicative times to give the reader a feeling for the run time of our algorithm. Computation times can be further reduced with more parallel processes and more aggressive parameter tuning. Note that a great advantage of our method is that the adaptive discretization requires a much inferior number of degrees of freedom than a uniform discretization. A uniform discretization with a similar resolution around the surface requires about 18 million degrees of freedom, resulting in a memory consumption of over 700 MB per deformation field.

**4.3. Curvature Term.** The above experiments were performed with all of the terms introduced in Section 2 except the statistical prior. We will now report on a couple of experiments that visualize the benefit of all the additional terms, starting with the curvature term introduced in Section 2.2.1.

In Figure 6(a), a registration of a femur is performed without the curvature term; that is, the surface is only represented by the level set function and the original CT scan. At a first glance, the surfaces are well registered, and the deformation field allows us to deform the reference bone so that it coincides with the target bone. However, on closer inspection, the implied correspondence is faulty. The deformation field matches the top of the trochanter minor (an important anatomical feature of the femur) of the reference to the side of the trochanter minor on the target. Such faulty correspondence causes problems in all

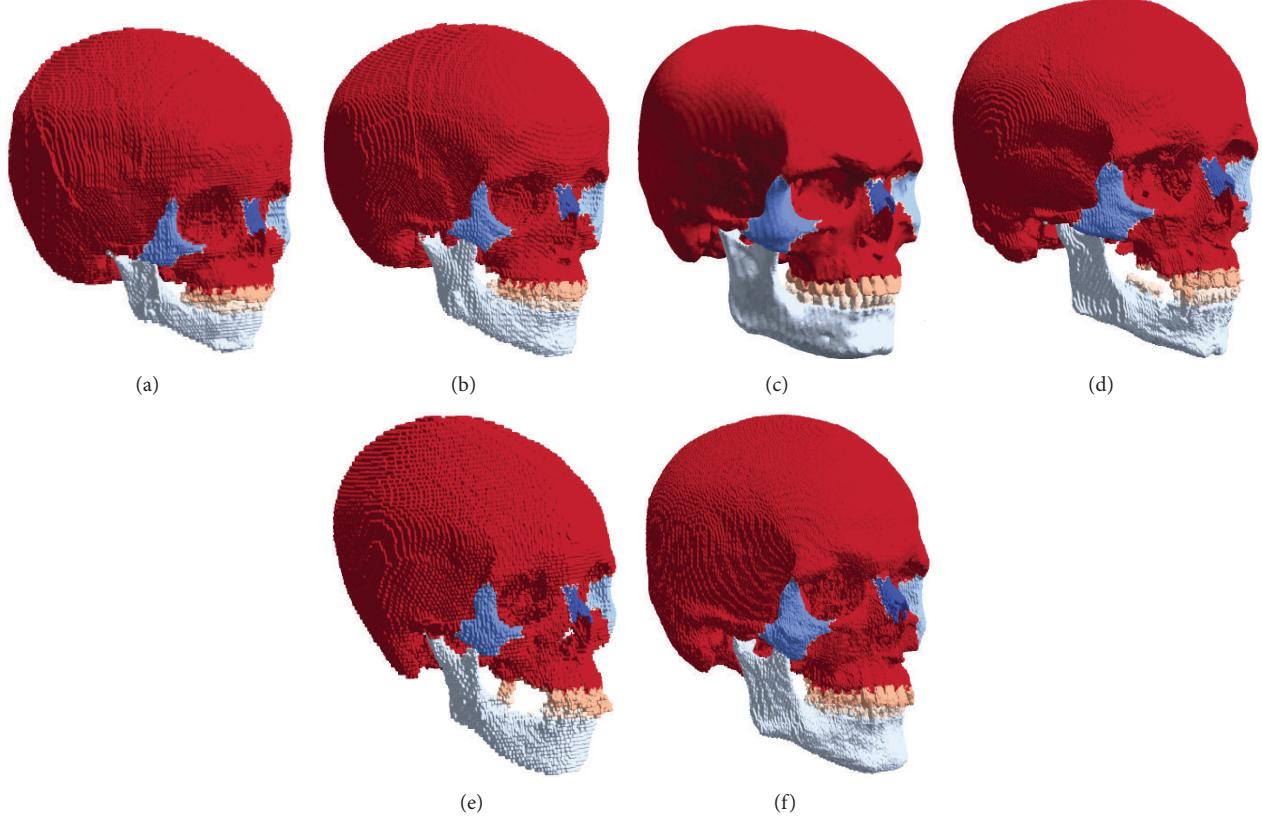


FIGURE 5: Transfer of the anatomical labelling of the reference (c) to a variety of skulls. Such a transfer requires very accurate registration results. Even a mismatch in the number of teeth is handled properly.

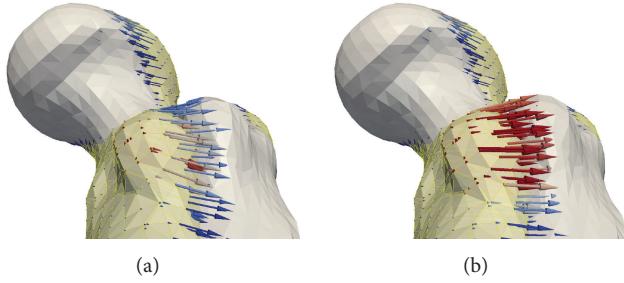


FIGURE 6: Registration of two femurs with and without curvature term. Without the curvature term, the correspondence is faulty. The corresponding features of the trochanter minor are not properly matched. The curvature term ensures matching of corresponding shape features.

subsequent applications such as building of statistical models or transferring anatomical labels.

In Figure 6(b), when the curvature term is used, the correspondence is much more sensible, and top is matched to top and sides to sides. The mean curvature proves to be a good description of anatomical features of bones. Its use in the registration method ensures superior correspondence.

**4.4. Volume Preservation.** Figure 7(a) shows the reference bone warped with a deformation field of a registration result.

To amplify the effect for visualization purposes, we have warped the reference bone with the deformation field multiplied by 2. The coloring represents the size of the triangles that make up the surface. We see that in some places the original reference grid is quite unnaturally stretched resulting in very large triangles. This is the effect of large volume expansion in the deformation field. When this volume change is penalized with the volume preservation term, the resulting mesh is much more even, while still allowing an equally good matching of the target surface.

In Figure 7(a), the weighting parameters  $\mu$  and  $\nu$  from (8) have been chosen as  $\mu = 2, \nu = 0$ , that is, no volume preservation term. In Figure 7(b), as  $\mu = \nu = 2$ , that is, equal weight of the gradient and the divergence term. Simply augmenting the weight of the gradient term, for example,  $\mu = 4, \nu = 0$  also results in less volume change, but still more than in Figure 7(b), while simultaneously decreasing the matching accuracy.

**4.5. Statistical Deformation Prior.** In these final experiments, we exhibit the use and benefit of the statistical prior. We first performed registrations of 15 intact hands for the 2D example and 50 intact femurs for the 3D example. From these, statistical models are built according to the process described in Section 3.1. They can be used as prior knowledge for registering difficult or damaged data sets. In Figure 8,

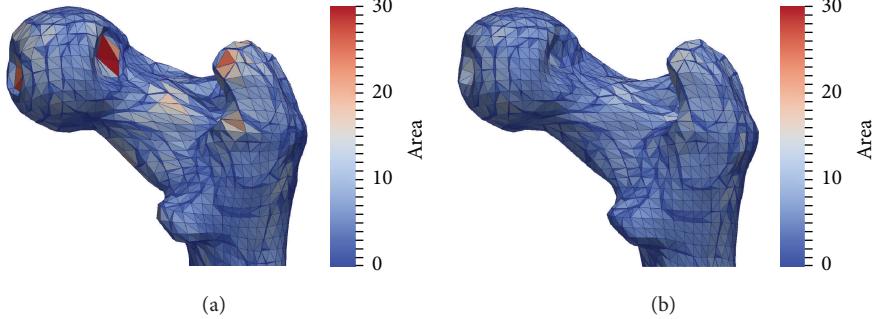


FIGURE 7: Registration of two femurs with and without volume preserving term. When the term is used, the distribution of area over the triangles of the mesh is much more even because the limited volume change prohibits strong expansion or compression of the mesh.

this prior knowledge allows the registration of a hand with a missing finger. Without this prior knowledge, the ring finger of the reference is deformed unnaturally in an attempt to match the hand with the missing finger. Note that the use of a robust distance measure and the regularization terms prevent a complete disappearance or distortion of the finger. When the prior knowledge from the statistical deformation model is used, the hand is well matched where intact. But the model enforces a registration with an intact ring finger, as all its training data sets include 5 intact fingers.

In Figure 9, a similar experiment in 3D is shown. The prior knowledge from 50 previous registrations is used in the registration of the reference bone to two damaged bones. The first bone has an artificial hip joint prosthesis and the second one a missing trochanter major. The registration method with statistical regularization registers the complete bones while complementing the missing parts from its prior knowledge. In this way, we can see how the original bone could have looked like. In these experiments, we have used 40 modes of variation (principle components). In a uniform discretization with a similar resolution around the surface, this would have required 28 GB of memory. Of course, the inclusion of the statistical prior introduces additional computational complexity. This results in an increase of computation time from about 5 minutes without to about 7 minutes with statistical prior per femur registration.

These experiments were performed with a robust distance measure (cf. Section 2.1). In the places with missing data, the regular  $L^2$  measure has a very large value and would dominate the registration process. As with all regularization techniques, a balance between over- and under-regularization has to be found. If too much weight is placed on the statistical prior, the result is pulled too close to the model.

## 5. Discussion

We have presented a registration method which allows the accurate and efficient registration of surfaces by registering their distance images. While this can be achieved with any image registration algorithm, we have shown that the most obvious choices for the distance measure and regularizer have several shortcomings, which can be relieved by adding additional distance and regularization terms to the model.

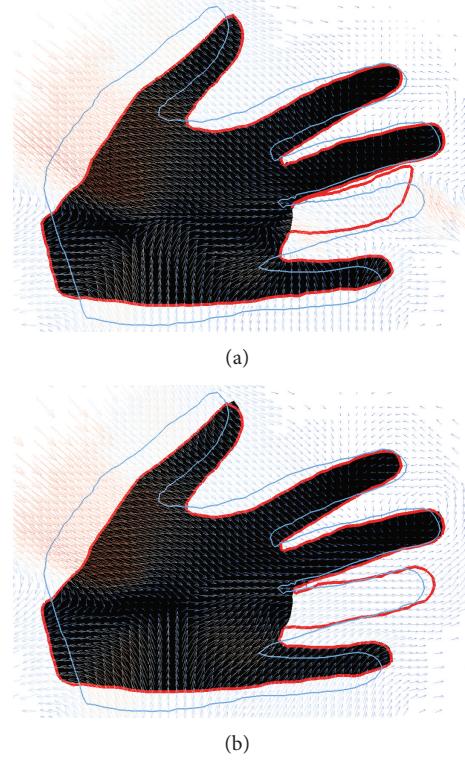


FIGURE 8: The reference shape (blue line) is registered onto a hand with a missing finger. The red line shows the warp of the reference with the resulting deformation field, without (a) and with statistical regularization (b).

We have seen that by representing the surfaces not by their distance images alone but also by a combination of feature images, most notably the curvature images, and we can obtain superior correspondence information. Concerning the regularizer, we have shown that adding a volume preserving term results in more even and naturally looking deformations and warps. Finally, we have shown that including prior knowledge in form of a statistical deformation model allows us to register considerably damaged data sets. While the statistical prior could be used in any registration task where prior registration results are available, we have found that for intact data sets

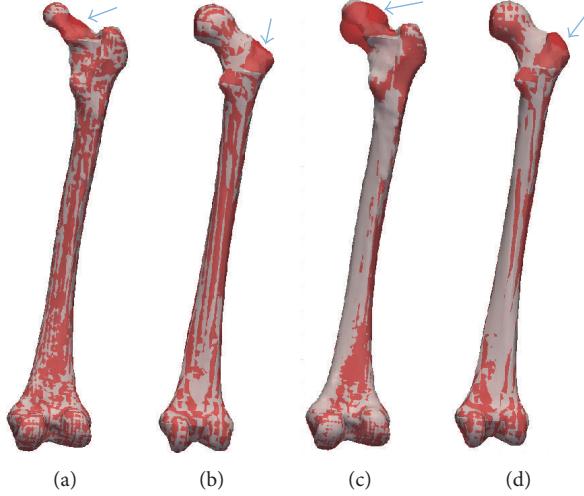


FIGURE 9: Registration of a pair of damaged bones, with and without statistical regularization. The first bone has an artificial hip joint, and in the second, the trochanter major is missing. Without statistical regularization, the damaged bones are matched exactly ((a), (b)). With statistical regularization, the method recognizes that the damaged parts do not conform with the prior knowledge and restore them automatically ((c), (d)).

the method works well enough without the prior. We use the prior in cases where the data is too corrupted to be used directly, but where we still wish to add as much as possible of its information into our existing model.

Our registration method is formulated continuously and is independent of the discretization method. We have presented a finite element discretization based on the local discontinuous Galerkin method, which allows for local grid adaption and a straightforward implementation of all the terms of our functional. The local grid adaption and parallelization limit the computational complexity and memory consumption, enabling us to perform registrations that were not possible with previous methods. In particular the inclusion of large statistical deformation models was not possible with previous methods based on uniform discretization such as the Demons Algorithm. This will become more and more important as the resolution of medical images increases in the future. The flexibility of the continuous functional and the LDG discretization allows the easy integration of other concepts and terms that are being developed in the field of image registration every day. On the other hand, the ideas and terms introduced here can be used with other discretization schemes or existing registration methods.

## Appendix

### A. Derivatives

In the following, we give details on how to derive the Euler-Lagrange equations, which characterize the minimum  $u$  of the functional  $\mathcal{J}$  given in (21). Using the standard methods from the calculus of variations, we compute the Gateaux derivatives for each of the terms that make up our registration

functional. This leads to the bilinear forms used in the finite element discretization in Section 3. We conclude this section with the derivation of the partial differential equations, that is, the strong form, which characterize the deformation field.

**A.1. Distance Measure.** As the individual terms of the combined distance measure (7) are all of the same form, it suffices to calculate the first variation of the distance measure with only one feature image  $\mathcal{D}_I[u]$  defined in (5) as follows.

$$\begin{aligned} \mathcal{D}'_I[u, \varphi] &= \frac{d}{d\epsilon} \mathcal{D}'_I[u + \epsilon\varphi] \Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon} \frac{1}{2} \int_{\Omega} \frac{1}{Q_I} (I_1(\cdot + u(\cdot) + \epsilon\varphi(\cdot)) - I_0(\cdot))^2 dx \Big|_{\epsilon=0} \\ &= \int_{\Omega} \frac{1}{Q_I} (I_1(\cdot + u(\cdot)) - I_0(\cdot)) \nabla I_1(\cdot + u(\cdot)) \varphi(\cdot) dx. \end{aligned} \quad (\text{A.1})$$

The dependency of the weighting term  $Q_I$  on  $u$  is neglected for reasons of efficiency. The other terms are derived analogously. The derivative of the combined distance measure (7) is therefore

$$\begin{aligned} \mathcal{D}'[u, \varphi] &= \sum_{k=1}^n \int_{\Omega} \frac{\alpha^k}{Q_{X^k}} (X_1^k(\cdot + u) - X_0^k) \\ &\quad \times \nabla X_1^k(\cdot + u) \varphi dx. \end{aligned} \quad (\text{A.2})$$

**A.2. Regularization Term.** We will continue by calculating the first variation of the regularization term  $\mathcal{R}_g[u]$  defined in (8).

$$\begin{aligned} \mathcal{R}'_g[u, \varphi] &= \frac{d}{d\epsilon} \mathcal{R}_d[u + \epsilon\varphi] \Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon} \frac{1}{2} \int_{\Omega} \|\nabla u(x) + \epsilon\nabla\varphi(x)\|^2 dx \Big|_{\epsilon=0} \\ &= \int_{\Omega} \nabla u : \nabla \varphi dx = \sum_{i=1}^d \int_{\Omega} \nabla u_i \cdot \nabla \varphi_i dx. \end{aligned} \quad (\text{A.3})$$

The divergence regularization term introduced in (9) is differentiated as follows:

$$\begin{aligned} \mathcal{R}'_d[u, \varphi] &= \frac{d}{d\epsilon} \frac{1}{2} \int_{\Omega} (\operatorname{div} u + \epsilon \operatorname{div} \varphi)^2 dx \Big|_{\epsilon=0} \\ &= \int_{\Omega} \operatorname{div} u \operatorname{div} \varphi dx. \end{aligned} \quad (\text{A.4})$$

Therefore, the derivative of the full regularization functional  $\mathcal{R}[u]$  from (9) is

$$\begin{aligned} \mathcal{R}'[u, \varphi] &= \mu \mathcal{R}'_g[u, \varphi] + \nu \mathcal{R}'_d[u, \varphi] \\ &= \int_{\Omega} \mu \nabla u : \nabla \varphi + \nu \operatorname{div} u \operatorname{div} \varphi dx. \end{aligned} \quad (\text{A.5})$$

*Statistical Prior.* Taking into account that the covariance operator  $\mathcal{C}$  is linear and self-adjoint and therefore this is also true for its pseudoinverse, it is easy to compute that the first variation of the statistical prior from (12) is given by

$$\begin{aligned} \mathcal{P}'[u, \varphi] &= \frac{d}{d\epsilon} \mathcal{P}(u + \epsilon\varphi) \Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon} \frac{\gamma}{2} \langle u + \epsilon\varphi - \bar{u}, \mathcal{C}^\dagger [u + \epsilon\varphi - \bar{u}] \rangle \\ &\quad + \frac{\gamma}{2} \langle u + \epsilon\varphi - \bar{u}, \sigma^{-2}(u + \epsilon\varphi - \bar{u}) \rangle \Big|_{\epsilon=0} \\ &= \frac{\gamma}{2} \langle u - \bar{u}, \mathcal{C}^\dagger \varphi \rangle + \frac{\gamma}{2} \langle \varphi, \mathcal{C}^\dagger [u - \bar{u}] \rangle \quad (\text{A.6}) \\ &\quad + \frac{\gamma}{2} \langle u - \bar{u}, \sigma^{-2} \varphi \rangle + \frac{\gamma}{2} \langle \varphi, \sigma^{-2} (u - \bar{u}) \rangle \\ &= \int_{\Omega} \gamma \mathcal{C}^\dagger [u - \bar{u}](x) \cdot \varphi(x) dx \\ &\quad + \int_{\Omega} \gamma \sigma^{-2} (u - \bar{u})(x) \cdot \varphi(x) dx, \end{aligned}$$

where  $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{L^2(\Omega, \mathbb{R}^3)}$ .

*A.3. Strong Derivative.* In order to derive the strong formulation, for the functional, we first apply integration by parts on the derivative of the regularization term (A.5). Assuming zero Neumann or Dirichlet boundary conditions, the boundary terms vanish and we get

$$\mathcal{R}'[u, \varphi] = - \int_{\Omega} \mu \Delta u \cdot \varphi + \nu \nabla \operatorname{div} u \cdot \varphi dx. \quad (\text{A.7})$$

Now we can collect all the terms making up the weak derivative as follows:

$$\mathcal{J}'[u, \varphi] = \mathcal{D}'[u, \varphi] + \mathcal{R}'[u, \varphi] + \mathcal{P}'[u, \varphi] \quad (\text{A.8})$$

from (A.2), (A.7), and (A.6) and apply the fundamental lemma of the calculus of variations to each component  $\varphi_i$  of  $\varphi$  which leads to the partial differential equation characterizing  $u$ :

$$\begin{aligned} &- \mu \Delta u - \nu \nabla \operatorname{div} u + \gamma \mathcal{C}^\dagger [u - \bar{u}] \\ &= - \sum_{k=1}^n \frac{\alpha_i}{Q_{X^k}(x)} \left( X_1^k(x + u(x)) - X_0^k(x) \right) \quad (\text{A.9}) \\ &\quad \times \nabla X_1^k(x + u(x)). \end{aligned}$$

This is the well-known equation from linear elasticity with a forcing term on the right hand side and the additional term from the statistical deformation prior.

## Acknowledgments

The authors would like to thank the University Hospital Basel and the AO Foundation for the CT data. This work was funded by the Swiss National Science Foundation in scope of the NCCR CO-ME, the Hasler Foundation in scope of the HOVISSE project, and the Landesstiftung Baden-Württemberg.

## References

- [1] D. Cremers, M. Rousson, and R. Deriche, “A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape,” *International Journal of Computer Vision*, vol. 72, no. 2, pp. 195–215, 2007.
- [2] U. Grenander and M. I. Miller, “Computational anatomy: an emerging discipline,” *Quarterly of Applied Mathematics*, vol. 56, no. 4, pp. 617–694, 1998.
- [3] J.-P. Thirion, “Image matching as a diffusion process: an analogy with Maxwell’s demons,” *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.
- [4] J. Modersitzki, *Numerical Methods For Image Registration*, Oxford Science Publications, 2004.
- [5] A. Dedner, R. Klöfkorn, M. Nolte, and M. Ohlberger, “A generic interface for parallel and adaptive discretization schemes: abstraction principles and the DUNE-FEM module,” *Computing*, vol. 90, no. 3-4, pp. 165–196, 2010.
- [6] P. Bastian, M. Blatt, A. Dedner et al., “A generic grid interface for parallel and adaptive scientific computing. Part I: abstract framework,” *Computing*, vol. 82, no. 2-3, pp. 103–119, 2008.
- [7] P. Bastian, M. Blatt, A. Dedner et al., “A generic grid interface for parallel and adaptive scientific computing. Part II: implementation and tests in DUNE,” *Computing*, vol. 82, no. 2-3, pp. 121–138, 2008.
- [8] A. Dedner, C. Rohde, B. Schupp, M. Wesenberg, and A. parallel, “load balanced mhd code on locally adapted, unstructured grids in 3d,” *Computing and Visualization in Science*, vol. 7, pp. 79–96, 2004.
- [9] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*, Kitware, 2005.
- [10] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit*, Prentice Hall, Englewood Cliffs, NJ, USA, 1996.
- [11] B. Zitová and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [12] M. A. Audette, F. P. Ferrie, and T. M. Peters, “An algorithmic overview of surface registration techniques for medical imaging,” *Medical Image Analysis*, vol. 4, no. 3, pp. 201–217, 2000.
- [13] N. Paragios, M. Rousson, and V. Ramesh, “Non-rigid registration using distance functions,” *Computer Vision and Image Understanding*, vol. 89, no. 2-3, pp. 142–165, 2003.
- [14] N. Litke, M. Droske, M. Rumpf, and P. Schröder, “An image processing approach to surface matching,” in *Proceedings of the 3rd Eurographics Symposium on Geometry Processing*, pp. 207–216, 2005.
- [15] A. Dedner, M. Lüthi, T. Albrecht, and T. Vetter, “Curvature guided level set registration using adaptive finite elements,” in *Pattern Recognition*, pp. 527–536, 2007.
- [16] T. Rohlfing, C. R. Maurer Jr., D. A. Bluemke, and M. A. Jacobs, “Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 6, pp. 730–741, 2003.
- [17] E. Haber and J. Modersitzki, “Numerical methods for volume preserving image registration,” *Inverse Problems*, vol. 20, no. 5, pp. 1621–1638, 2004.
- [18] C. Broit, *Optimal registration of deformed images [Ph.D. thesis]*, Department of Computer and Information Science, University of Pennsylvania, 1981.
- [19] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, “Deformable templates using large deformation kinematics,” *IEEE Transactions on Image Processing*, vol. 5, no. 10, pp. 1435–1447, 1996.

- [20] J. C. Gee and R. K. Bajcsy, "Elastic matching: continuum mechanical and probabilistic analysis," in *Brain Warping*, 1998.
- [21] Y. Wang and L. H. Staib, "Physical model-based non-rigid registration incorporating statistical shape information," *Medical Image Analysis*, vol. 4, no. 1, pp. 7–20, 2000.
- [22] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 1014–1025, 2003.
- [23] Z. Xue, D. Shen, and C. Davatzikos, "Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping," *Medical Image Analysis*, vol. 10, no. 5, pp. 740–751, 2006.
- [24] T. Albrecht, M. Lüthi, and T. Vetter, "A statistical deformation prior for non-rigid image and shape registration," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.
- [25] J. C. Gee, D. R. Haynor, M. Reivich, and R. Bajcsy, "Finite element approach to warping of brain images," in *Medical Imaging*, vol. 2167 of *Proceedings of SPIE*, pp. 327–337, 1994.
- [26] P. G. Ciarlet, *The Finite Element Method For Elliptic Problems. Studies in Mathematics and Its Applications*, vol. 4, North-Holland, Amsterdam, The Netherlands, 1978.
- [27] D. N. Arnold, F. Brezzi, B. Cockburn, and L. Donatella Marini, "Unified analysis of discontinuous Galerkin methods for elliptic problems," *SIAM Journal on Numerical Analysis*, vol. 39, no. 5, pp. 1749–1779, 2002.
- [28] B. Cockburn and C.-W. Shu, "The local discontinuous galerkin method for time-dependent convection-diffusion systems," *SIAM Journal on Numerical Analysis*, vol. 35, no. 6, pp. 2440–2463, 1998.
- [29] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.
- [30] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *International Journal of Computer Vision*, vol. 19, no. 1, pp. 57–91, 1996.
- [31] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bulletin of the International Statistical Institute*, vol. 52, no. 4, pp. 5–21, 1987.
- [32] C. Nikou, F. Heitz, J.-P. Armsbach, I.-J. Namer, and D. Grucker, "Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures," *NeuroImage*, vol. 8, no. 1, pp. 30–43, 1998.
- [33] M. Droske and M. Rumpf, "A variational approach to nonrigid morphological image registration," *SIAM Journal on Applied Mathematics*, vol. 64, no. 2, pp. 668–687, 2004.
- [34] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*, pp. 187–194, ACM Press, 1999.
- [35] T. F. Cootes and C. J. Taylor, "Active shape models-'smart snakes,'" in *Proceedings of the British Machine Vision Conference*, 1992.
- [36] H. W. Alt, *Lineare Funktionalanalysis*, Springer, Berlin, Germany, 1985.
- [37] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, article 32, 2005.
- [38] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2005.
- [39] L. Pareschi and G. Russo, "Implicit-explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxation," *Journal of Scientific Computing*, vol. 25, no. 1, pp. 129–155, 2005.