A Statistical Deformation Prior for Non-Rigid Image and Shape Registration

Thomas Albrecht, Marcel Lüthi, Thomas Vetter
Computer Science Department, University of Basel
Bernoullistrasse 16, 4056 Basel, Switzerland
{Thomas.Albrecht, Marcel.Luethi, Thomas.Vetter}@unibas.ch

Abstract

Non-rigid registration is central to many problems in computer vision and medical image analysis. We propose a registration algorithm which is regularized by prior knowledge in the form of a statistical deformation model. This model is obtained from previous registrations performed on a set of noise-free training examples given by images, or shapes represented by level set functions. Contrary to similar approaches, our method does not strictly constrain the result to lie in the span of the statistical model but rather uses the model for Tikhonov regularization. Therefore, our method can be used to reduce the influence of noise and artifacts even when the model contains only a few typical examples. This automatically gives rise to a bootstrapping strategy for building statistical models from noisy data sets requiring only a limited number of high quality examples. We demonstrate the effectiveness of the approach on synthetic and medical images.

1. Introduction

Many current methods in computer vision, computer graphics and medical image analysis incorporate prior knowledge about the object under consideration in the form of statistics computed from a set of typical examples. In order to be able to extract statistical information from several objects of a class, the objects have to be brought into correspondence. That is, to every point in a reference object, one needs to find the corresponding point in all the examples. This problem, known as the registration problem, is generally formulated in a generic fashion as an optimization problem. In this paper, we present a method for incorporating prior knowledge about the object class into the registration algorithm by means of a statistical deformation model. We show that this can greatly improve the registration results in the case of noisy and incomplete data. Our particular motivation stems from two projects in medical image analysis, where a statistical model of the femur bone and the human skull have to be built from CT data. In this environment, the available data is scarce, and often very noisy (*e.g*. due to metal artifacts or pathologies). The method presented here directly gives rise to a method for bootstrapping the model building from only a few high quality examples. Once an initial statistical model has been built, it can be used to make the registration process more robust to missing data and noise, and hence to make the statistical model more expressive by incorporating data from low quality examples. As the method is suitable for both shape (or surface) registration and image registration, we can combine data from different sources and modalities. The method is independent of the dimensionality of the data. However, the memory footprint in the 3D case is typically too large for a standard PC.

The main contribution of our paper is the inclusion of a new regularization term in form of a statistical deformation model that fits naturally into a variational framework for image registration. We show how the problem can be interpreted as a maximum a posteriori solution in a Bayesian setting. In this interpretation, the space spanned by the statistical deformation model corresponds to the prior probability on the deformation fields. We regularize the covariance matrix of this prior using a shrinkage estimate and show that this penalizes large deviations from the space spanned by the deformation model.

Non-rigid registration is an extremely well researched problem and several attempts have been made to incorporate prior knowledge. For an overview of registration methods we refer to the recent survey papers by Zitova and Flusser [22] (image registration), Audette *et al*. [1] (surface registration), and in particular the book by Modersitzki [13] for a thorough discussion of variational methods for image registration, on which we base our work. In the related area of image segmentation, using statistical models to guide the segmentation has shown excellent results. See *e.g*. Cremers *et al*. [5] for a recent review. In the area of non-rigid image registration, the need for incorporating prior knowledge into the registration algorithm has also been recognized. The concept of statistical deformation models has been researched by various group. We refer to the papers of Rueck-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

ert *et al*. [15], Gee and Bajcsy [9] and the references therein. Similar to our attempt, Gee and Bajscy use a statistical deformation model to constrain the solution of the non-rigid registration problem. In contrast to our work, their solution is only sought in the space spanned by the model. In spirit closest to our work is the work of Xue *et al*. [21] and Wang and Staib [20]. However, in both approaches the statistical regularization is performed in a separate step, whereas we present an integrated formulation.

This paper is structured as follows. In Section 2 we describe the variational method we use for surface and image registration, followed by a discussion of its probabilistic interpretation. We present experimental results on simple 2D shapes and real x-ray images of femur bones in Section 3. In Section 4 we outline a bootstrapping procedure using our formulation and briefly discuss the dependence of the different parameters in our model. A discussion and outline of future work is given in Section 5.

2. Methods

In this section we present the image and shape registration algorithm we use and introduce its enhancement by the statistical deformation model. The starting point for our algorithm is the diffusion registration algorithm presented by Modersitzki in [13], which is closely related to a variational formulation of Thirion's Demons algorithm [17] and variational optical flow [10]. In order to apply this method for shape registration, we use the extension proposed by Lüthi *et al*. in [12].

2.1. Image Registration

Diffusion registration is defined as a minimization problem. Given a reference image $I_0 : \Omega \to \mathbb{R}$ and a target image $I_1 : \Omega \to \mathbb{R}$ both defined on a common image domain $\Omega \subset \mathbb{R}^d$, the task is to find a dense deformation field $u : \Omega \to \mathbb{R}^d$ such that the distance between $I_0(x + u(x))$, the reference image warped by the deformation field u, and the target image $I_1(x)$ is minimized. In order to minimize the distance between two images, it has to be quantified mathematically. If the images I_0 and I_1 are taken with the same image modality, the L^2-distance between $I_0(x+u(x))$ and $I_1(x)$ is a suitable distance measure.

Because minimizing only the L^2-distance with respect to u leads to an ill-posed problem, a regularization term is introduced into the minimization problem, according to the concept of Tikhonov regularization [18]. Therefore, diffusion registration is defined as the minimization of the following two-part functional with respect to u.

$$\mathcal{J}[u] = \mathcal{D}[u] + \alpha\mathcal{R}[u], \tag{1}$$

where

$$\mathcal{D}[u] = \tfrac{1}{2}\int_\Omega (I_0(x + u(x)) - I_1(x))^2\, dx \tag{2}$$

is the L^2 distance measure, and

$$\mathcal{R}[u] = \tfrac{1}{2}\int_\Omega |\nabla u|^2\, dx := \tfrac{1}{2}\sum_{i=1}^{d}\int_\Omega |\nabla u_i|^2\, dx \tag{3}$$

is the Tikhonov regularization term.

From the calculus of variations, it is known that any solution has to fulfill the Euler Lagrange equation:

$$(I_0(x+u(x)) - I_1(x))\nabla I_0(x+u(x)) - \alpha\Delta u(x) = 0, \tag{4}$$

for all $x \in \Omega$. Possible methods for finding a solution to this partial differential equation for image registration are, among others, the finite difference scheme proposed by Modersitzki [13], Thirion's Demons algorithm [17], or the finite element method proposed by Dedner *et al*. [7].

Although we discuss the case of diffusion registration, all methods presented in this paper can also be applied to the large class of all image registration methods that can be formulated in the form of Equation (1), with a distance measure $\mathcal{D}$ and regularizer $\mathcal{R}$.

2.2. Shape Registration

By representing shapes or surfaces by level set functions, see [14], we can use an image registration method for shape registration. We represent a shape $\Gamma \subset \mathbb{R}^d$ by its signed distance function

$$I(x) := d_\Gamma(x) = \begin{cases} \text{dist}(x,\Gamma) & x \in \text{outside}(\Gamma) \\ 0 & x \in \Gamma \\ -\text{dist}(x,\Gamma) & x \in \text{inside}(\Gamma), \end{cases} \tag{5}$$

where $\text{dist}(x,\Gamma)$ is the Euclidean distance from x to Γ. When this distance function is evaluated on a rectangular domain $\Omega \subset \mathbb{R}^d$ it can be interpreted as an image. Thus, two shapes $\Gamma_0, \Gamma_1 \subset \mathbb{R}^d$ can be registered by registering their respective distance images $I_i = d_{\Gamma_i} : \Omega \to \mathbb{R}^d$, $i = 1, 2$.

Because registering the distance images alone does not always lead to meaningful correspondence, Lüthi *et al*. [12] propose representing a surface Γ by both its distance function $I(x) = d_\Gamma(x)$ and its mean curvature map $H(x) = \text{div}\,\frac{\nabla I(x)}{|\nabla I(x)|}$, which for each point $x \in \Omega$ evaluates the mean curvature of the level set of I passing through that point.

This second feature image is included into the functional (1) by an additional distance term

$$\mathcal{C}[u] := \tfrac{1}{2}\int_\Omega (H_0(x + u(x)) - H_1(x))^2\, dx. \tag{6}$$

This is the same term as the distance term (2) with I_0, I_1 replaced by H_0, H_1.

The full functional to be minimized for this curvature-guided shape registration with an additional weighting parameter $\beta \in \mathbb{R}^+$ is:

$$\tilde{\mathcal{J}}[u] := \mathcal{D}[u] + \beta\mathcal{C}[u] + \alpha\mathcal{R}[u]. \tag{7}$$

In a similar fashion, we can add additional feature images along with their corresponding distance terms to the functional. For instance, in medical examples we can include the original x-ray or MRI image from which the surface Γ was extracted as a third feature image to gain an additional distance term $\mathcal{X}[u]$ and extend the functional to:

$$\hat{\mathcal{J}}[u] := \mathcal{D}[u] + \gamma\mathcal{X}[u] + \beta\mathcal{C}[u] + \alpha\mathcal{R}[u]. \tag{8}$$

2.3. Bayesian Interpretation

In this section we outline a probabilistic interpretation of the variational formulation of the registration problem (1), following the exposition of Wang and Staib [20]. The main idea is to interpret I_1 in Equation (1) as a Gaussian Process $\mathcal{GP}(\mu, K)$. A Gaussian process is a generalization of the multivariate normal distribution to the infinite dimensional case. More precisely, it is a collection of random variables with the property that any finite subset of it has a joint Gaussian distribution. A Gaussian process on the functions $\Omega \to \mathbb{R}$ is completely defined by its mean $\mu : \Omega \to \mathbb{R}$ and covariance function $K : \Omega \times \Omega \to \mathbb{R}$.

Consider a Gaussian process with mean $\mu = I_0(x + u(x))$ and covariance function $K(x, x') = \sigma_I^2 \delta_{xx'}$, where δ is the Dirac delta function. Under this assumption, the log probability of observing I_1 is given by

$$\ln P(I_1|u) \propto -\tfrac{1}{2\sigma_I^2} \int_\Omega (I_0(x + u(x)) - I_1(x))^2 \, dx. \tag{9}$$

Assuming independence of the different images, we can treat each of the distance terms $\mathcal{D}, \mathcal{C}, \mathcal{X}$ in Equation (8) separately:

$$\ln P(I_1, H_1, X_1|u) = \ln P(I_1|u) + \ln P(H_1|u) + \ln P(X_1|u). \tag{10}$$

Each such term is defined analogously to (9). As the following discussion is independent of how many feature images we use, we will without loss of generality assume that only I_0 and I_1 are given.

The regularization term $\mathcal{R}$ in (1) can be interpreted as defining a (Gaussian) prior distribution on all deformation fields u:

$$\ln P(u) \propto -\tfrac{1}{2\sigma_u^2} \int_\Omega |\nabla u|^2 \, dx. \tag{11}$$

Applying Bayes theorem, we get

$$\ln P(u|I_1) \propto \ln(P(u)P(I_1|u)) = \ln P(u) + \ln P(I_1|u) \tag{12}$$

and by plugging in (9) and (11) we obtain

$$\begin{aligned} \ln P(u|I_1) \propto &-\tfrac{1}{2\sigma_I^2} \int_\Omega (I_0(x + u(x)) - I_1(x))^2 \, dx \\ &-\tfrac{1}{2\sigma_u^2} \int_\Omega |\nabla u|^2 \, dx. \end{aligned} \tag{13}$$

Taking the sign into account, the minimization problem (1) can be seen as maximizing the a posteriori probability for (12), where the variances $\sigma_I^2, \sigma_u^2 \in \mathbb{R}^+$ take the role of the weighting parameter α. The prior probability on the deformation field favors smooth deformation fields but is otherwise very generic. Our goal is to learn a prior distribution from examples of the same object class and hence to be able to quantify the probability of observing a deformation field and to penalize unlikely results. In the following we assume that the objects to be registered, and hence the resulting deformation fields, follow a Gaussian distribution.

2.4. Statistical Deformation Models

Suppose now that we have already registered a set of $n \geq 1$ images $\{I_1, \ldots, I_n\}$ and possibly their curvature or other feature images to a common reference $I_0 : \Omega \to \mathbb{R}$. We now wish to build a statistical model from the resulting deformation fields $\{u_1, \ldots, u_n\}$ in order to exploit the experience gained from these previous registrations for further registration tasks.

The idea of statistical deformation models has been introduced for instance by Rueckert *et al.* in [15]. A statistical model based on principal components analysis (PCA) much like those well-known in literature, [4], [3], is built from a number of n samples. The only difference is that here the samples are not images or triangle meshes but deformation fields. The same discretization with $k \in \mathbb{N}$ degrees of freedom is chosen for all registrations on the image domain $\Omega \subset \mathbb{R}^d$. Thus, each discrete deformation field u_i, is given by k d-dimensional vectors and can therefore be identified with a (rather long) vector $u_i \in \mathbb{R}^N$ with $N = kd$.

With these vectors, a PCA model can be calculated as usual, see [2] for details. We calculate the arithmetic mean $\overline{u} = \frac{1}{n} \sum_{i=1}^n u_i \in \mathbb{R}^N$ and the sample covariance matrix $\hat{\Sigma} \in \mathbb{R}^{N \times N}$ with the k, l-th entry defined by:

$$\hat{\Sigma}_{kl} = \tfrac{1}{n-1} \sum_{i=1}^n (u_i - \overline{u})_k \, (u_i - \overline{u})_l. \tag{14}$$

This is done in order to estimate the probability distribution of the deformation fields. As in all PCA models, the samples $u_i \in \mathbb{R}^N$ are assumed to be iid. samples drawn from a multivariate normal distribution $\mathcal{N}_N(\mu, \Sigma)$ with mean $\mu \in \mathbb{R}^N$ and covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$. It is approximated by the estimated multivariate normal distribution $\mathcal{N}_N(\overline{u}, \hat{\Sigma})$.

This modeling only makes sense if our deformation fields $\{u_1, \dots, u_n\}$ are gained from registrations of accurately pre-aligned images or shapes of the same (normally distributed) object class.

Principal Components The part of PCA modeling that is actually concerned with the principal components of the model is the computation of the main modes of variation or *principal components* of the estimated multivariate normal distribution $\mathcal{N}_N(\overline{u}, \hat{\Sigma})$. Although the concepts are well-known and established, we will review them quickly to clarify notation. The principal components are given as the eigenvectors of the sample covariance matrix $\hat{\Sigma} \in \mathbb{R}^{N\times N}$. The corresponding eigenvalue to an eigenvector $v_i \in \mathbb{R}^N$ is the sample variance of the model in the direction of v_i and will be denoted by $\hat{\sigma}_i^2 \in \mathbb{R}^+$. Because $\hat{\Sigma}$ is a symmetric matrix, it does indeed have N eigenvalues, but as it is calculated according to Equation (14) from a set of only n mean free samples, there are at most $n-1$ nonzero eigenvalues. They can, along with their corresponding eigenvectors, be efficiently calculated with the help of a singular value decomposition (SVD) of an $(n\times n)$-matrix, see [8] for details. Note that n is in general much smaller than N. The SVD computes $m < n$ nonzero eigenvalues $\hat{\sigma}_i$ with corresponding eigenvectors v_i. Although we will see that this won't be necessary in practice, the remaining $N - m$ eigenvectors to the eigenvalue 0 can be calculated with the Gram-Schmidt method as the orthogonal complement in $\mathbb{R}^N$ to the span of $\{v_1, \dots, v_m\}$. The full set of N eigenvectors forms an orthonormal basis of $\mathbb{R}^N$ and, expressed in this basis, $\hat{\Sigma}$ becomes a diagonal matrix with $(\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2, 0, \dots, 0)$ as its diagonal. In other words, if we define $S \in \mathbb{R}^N$ as the orthogonal matrix with $\{v_1, \dots, v_N\}$ as its columns, we have

$$S^{\mathrm{T}} \hat{\Sigma}\, S = \hat{\Lambda} := \mathrm{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2, 0, \dots, 0), \tag{15}$$

where $\mathrm{diag}(a)$ denotes a diagonal matrix with the elements of the vector $a \in \mathbb{R}^N$ as its diagonal.

In order to evaluate the probability density function of the estimated multivariate normal distribution $\mathcal{N}_N(\overline{u}, \hat{\Sigma})$, we need to be able to invert $\hat{\Sigma}$. However, as $\hat{\Sigma}$ has a rank of only $m \leq N$, this is not possible. It would be possible to restrict our whole registration problem to the span of $\{v_1, \dots, v_m\}$, which is, up to the addition of the mean, the same as the span of the samples $\{u_1, \dots, u_n\}$. If we let $\underline{S} \in \mathbb{R}^{N\times m}$ be the matrix with $\{v_1, \dots, v_m\}$ as its columns, the projection of $\hat{\Sigma}$ to this space is given by $\underline{S}\,\mathrm{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2)\,\underline{S}^{\mathrm{T}}$. It can easily be inverted by inverting the $\hat{\sigma}_i^2$. This approach is used in a number of papers on the topic, *e.g.* [20]. However, especially if the number of samples n is rather small we may not want to strictly constrain our problem to the very limited span of these few samples.

Covariance Estimation Because we use the sample covariance matrix $\hat{\Sigma}$ defined by (14) as an estimator for an assumed covariance matrix $\Sigma \in \mathbb{R}^{N\times N}$, our PCA model assigns a variance of zero to all directions in $\mathbb{R}^N$ deviating from the span of the samples, *i.e.* the model defines the probability of observing a deformation field $u \in \mathbb{R}^N$ as zero if it is not in the span of $\{u_1, \dots, u_n\}$. In reality however, we have to assume that there are many more nonzero eigenvalues in the correct covariance matrix Σ, and we have simply not been able to recover them in our estimation because we have not yet seen enough samples.

This is why the sample covariance matrix $\hat{\Sigma}$ is in fact a very poor estimator of the covariance matrix for a small number $n < N$ of samples. This problem is of course well known and many efforts for improved covariance matrix estimation have been proposed, see [6, 16] for instance. The simplest improvement proposed in the latter, which is known as *shrinkage estimation*, is to compute a convex combination of a multiple of the N-dimensional identity matrix I_N and the sample covariance matrix $\hat{\Sigma}$. In this way, we gain a new estimator $\hat{\Sigma}_2$ of the covariance matrix:

$$\hat{\Sigma}_2 = (1-\lambda)\,\hat{\Sigma} + \lambda\,\sigma_0^2 I_N, \tag{16}$$

where $\lambda \in (0, 1]$ is known as the shrinkage intensity and determines the weight between $\hat{\Sigma}$ and $\sigma_0^2 I_N$. It can be chosen optimally according to [16]. $\sigma_0^2 \in \mathbb{R}^+$ can be interpreted as the variance we assume in all directions if we don't use $\hat{\Sigma}$ at all. So when we use $\hat{\Sigma}_2$ instead of $\hat{\Sigma}$ as an estimator for Σ, we no longer assume the variance in directions perpendicular to the span of the model to be zero, but rather to be $\lambda\,\sigma_0^2$. In the direction of an eigenvector v_i corresponding to a nonzero eigenvalue $\hat{\sigma}_i^2$ of $\hat{\Sigma}$, the variance defined by $\hat{\Sigma}_2$ is $(1-\lambda)\,\hat{\sigma}_i^2 + \lambda\,\sigma_0^2$. This can be deduced from the following calculation:

Let $S, \hat{\Lambda} \in \mathbb{R}^N$ be defined as in Equation (15). We then have:

$$\begin{aligned}
S^{\mathrm{T}} \hat{\Sigma}_2\, S &= S^{\mathrm{T}} \big((1-\lambda)\,\hat{\Sigma} + \lambda\,\sigma_0^2 I_N\big) S && (17)\\
&= (1-\lambda)\, S^{\mathrm{T}} \hat{\Sigma}\, S + \lambda\,\sigma_0^2\, S^{\mathrm{T}} I_N S && (18)\\
&= (1-\lambda)\,\hat{\Lambda} + \lambda\,\sigma_0^2 I_N && (19)\\
&= \mathrm{diag} \begin{pmatrix} (1-\lambda)\,\hat{\sigma}_1^2 + \lambda\,\sigma_0^2 \\ \vdots \\ (1-\lambda)\,\hat{\sigma}_m^2 + \lambda\,\sigma_0^2 \\ \lambda\,\sigma_0^2 \\ \vdots \\ \lambda\,\sigma_0^2 \end{pmatrix}, && (20)
\end{aligned}$$

where in Equation (19) we have used (15) and the fact that the columns of S are orthonormal and therefore $S^{\mathrm{T}} S = I_N$.

Equation (20), shows that our shrinkage estimator $\hat{\Sigma}_2$ has full rank and is in fact diagonalized by the transformation matrix S made up of the eigenvectors of $\hat{\Sigma}$. Hence, the

inversion of $\hat{\Sigma}_2$ is trivial:

$$\hat{\Sigma}_2^{-1} = S \operatorname{diag} \begin{pmatrix} ((1-\lambda)\,\hat{\sigma}_1^2 + \lambda\,\sigma_0^2)^{-1} \\ \vdots \\ ((1-\lambda)\,\hat{\sigma}_m^2 + \lambda\,\sigma_0^2)^{-1} \\ (\lambda\,\sigma_0^2)^{-1} \\ \vdots \\ (\lambda\,\sigma_0^2)^{-1} \end{pmatrix} S^{\mathrm{T}}, \tag{21}$$

which can be split up as

$$\begin{aligned} &= S \operatorname{diag} \begin{pmatrix} ((1-\lambda)\,\hat{\sigma}_1^2 + \lambda\,\sigma_0^2)^{-1} - (\lambda\,\sigma_0^2)^{-1} \\ \vdots \\ ((1-\lambda)\,\hat{\sigma}_m^2 + \lambda\,\sigma_0^2)^{-1} - (\lambda\,\sigma_0^2)^{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} S^{\mathrm{T}} \\ &\quad + S \operatorname{diag} \begin{pmatrix} (\lambda\,\sigma_0^2)^{-1} \\ \vdots \\ (\lambda\,\sigma_0^2)^{-1} \end{pmatrix} S^{\mathrm{T}}. \end{aligned} \tag{22}$$

For convenience, we define $\eta_i^2 := ((1-\lambda)\,\hat{\sigma}_i^2 + \lambda\,\sigma_0^2)^{-1} - (\lambda\,\sigma_0^2)^{-1}$ and arrive at the expression:

$$\hat{\Sigma}_2^{-1} = \underline{S} \operatorname{diag} \begin{pmatrix} \eta_1^2 \\ \vdots \\ \eta_m^2 \end{pmatrix} \underline{S}^{\mathrm{T}} + (\lambda\,\sigma_0^2)^{-1}\, I_N, \tag{23}$$

where $\underline{S} \in \mathbb{R}^{N \times m}$ denotes the matrix whose columns are the first m eigenvectors of $\hat{\Sigma}$, *i.e.* $\{v_1, \ldots, v_m\}$. Note that the final expression for $\hat{\Sigma}_2^{-1}$ no longer contains the full matrix $S \in \mathbb{R}^{N \times N}$. Therefore, as already mentioned before, in practice we never need to calculate S but rather only its much smaller sub-matrix $\underline{S}$, which can be efficiently calculated with the singular value decomposition of $\hat{\Sigma}$.

Density Function With our invertible covariance estimator $\hat{\Sigma}_2^{-1}$, we can now calculate the probability density function of the estimated multivariate normal distribution $\mathcal{N}_N(\overline{u}, \hat{\Sigma}_2)$. It is defined as $f : \mathbb{R}^N \to \mathbb{R}^+$

$$f(u) = \tfrac{1}{(2\pi)^{\frac{N}{2}} \det(\hat{\Sigma}_2)^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}(u-\overline{u})^{\mathrm{T}} \hat{\Sigma}_2^{-1} (u-\overline{u})\right). \tag{24}$$

If we let $C := (2\pi)^{\frac{N}{2}} \det(\hat{\Sigma}_2)^{\frac{1}{2}}$ and plug in Equation (23) we get:

$$\begin{aligned} f(u) &= \tfrac{1}{C} \exp\left(-\tfrac{1}{2}(u-\overline{u})^{\mathrm{T}}\, \underline{S} \operatorname{diag} \begin{pmatrix} \eta_1^2 \\ \vdots \\ \eta_m^2 \end{pmatrix} \underline{S}^{\mathrm{T}}\, (u-\overline{u})\right) \\ &\quad \cdot \exp\left(-\tfrac{1}{2(\lambda\,\sigma_0^2)}\, (u-\overline{u})^{\mathrm{T}} (u-\overline{u})\right), \end{aligned} \tag{25}$$

$$\begin{aligned} &= \tfrac{1}{C} \exp\left(-\tfrac{1}{2}\, |\operatorname{diag} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_m \end{pmatrix} \underline{S}^{\mathrm{T}}\, (u-\overline{u})|^2\right) \\ &\quad \cdot \exp\left(-\tfrac{1}{2(\lambda\,\sigma_0^2)}\, |u-\overline{u}|^2\right), \end{aligned} \tag{26}$$

where $|v|^2 = v^{\mathrm{T}} v$ is the squared Euclidean norm of a vector v. With Equation (26) , we can now evaluate the likelihood that a vector field $u \in \mathbb{R}^N$ belongs to our PCA model. Next, we would like to add a term into our registration functional (1) and its extension (7) that favors solutions that are likely to belong to our model.

Continuous Formulation The problem arises that the functionals (1) and (7) are given in a continuous variational formulation with vector fields $u : \Omega \subset \mathbb{R}^d \to \mathbb{R}^d$ and our model is given in form of discrete vector fields $u \in \mathbb{R}^N$. However, provided that we use the same reference discretization with N degrees of freedom for all registrations, Equation (26) can be used directly in the numerical implementation. Yet, we briefly introduce a continuous formulation, which, thanks to results from functional analysis, takes a similar form as Equation (26). Due to space limitations, we omit all details and state only its final form.

After replacing the discrete multivariate normal distribution $\mathcal{N}_N(\mu, \Sigma)$ with a continuous Gaussian process $\mathcal{GP}(\mu, K)$, we can perform analogous calculations as in the discrete case and end up with a density function for continuous deformation fields, in analogy to Equation (26).

$$\begin{aligned} f(u) = \tfrac{1}{C} \exp&\left(-\tfrac{1}{2} \sum_{i=1}^{m} \langle\, \eta_i v_i,\, u-\overline{u} \,\rangle_{L^2}^2\right) \\ &\cdot \exp\left(-\tfrac{1}{2(\lambda\,\sigma_0^2)}\, \|u-\overline{u}\|_{L^2}^2\right), \end{aligned} \tag{27}$$

where we have again set $\eta_i^2 := ((1-\lambda)\,\hat{\sigma}_i^2 + \lambda\,\sigma_0^2)^{-1} - (\lambda\,\sigma_0^2)^{-1}$. The eigenfunctions $\{v_1, \ldots, v_m\}$ exist in the continuous case thanks to Mercer's theorem.

2.5. Statistical Deformation Prior

Recall from Section 2.3 that the goal is to find a prior distribution which penalizes unlikely deformation fields. Equation (27) provides us with a regularized density estimation. The assumption of the statistical model is that the deformation fields for a given object class are distributed according to this density function. Thus, using the notation of Section 2.3 we obtain:

$$\begin{aligned} \mathcal{S}[u] := \ln f(u) \propto &-\tfrac{1}{2} \sum_{i=1}^{m} \langle\, \eta_i v_i,\, u-\overline{u} \,\rangle_{L^2(\Omega)}^2 \\ &- \tfrac{1}{2(\lambda\,\sigma_0^2)}\, \|u-\overline{u}\|_{L^2(\Omega)}^2. \end{aligned} \tag{28}$$

Using a convex combination of $\mathcal{S}$ and the smoothness prior $\mathcal{R}$ from Equation (3) as a regularization term, we arrive at a prior distribution that enforces both smoothness and penalizes unlikely deformations:

$$\ln P(u) \propto \xi\, \mathcal{R}[u] + (1-\xi)\, \mathcal{S}[u], \tag{29}$$

where $\xi \in [0,1]$ determines the trade-off between the two terms. Plugging this term into Equation (12) as the new prior (or Tikhonov regularization term) for u, we arrive at a statistically regularized version of the registration problem:

$$\mathcal{J}[u] = \mathcal{D}[u] + \alpha \left(\xi \mathcal{R}[u] + (1-\xi)\mathcal{S}[u]\right), \qquad (30)$$

Note that since all registrations used for building the statistical model are performed with the same reference, all deformation fields are relative to that reference, and therefore only registrations onto that reference can be constrained with this model.

Geometric Interpretation Equation (28) admits an interesting geometric interpretation. The second term on the right hand side states that the probability of a solution decreases exponentially with its distance from the mean. Hence, all the solution on a hyper-sphere around the mean are equally likely. The first term is a projection onto a hyper-plane spanned by the first m eigenfunctions (principal components) of the covariance matrix. Under the usual assumption of linearity of the object classes (see Vetter and Poggio [19]) any point on this plane represents an object of the class, and the probability of observing the object decreases exponentially with the distance from the mean. Most approaches to statistical models consider only this plane to search for solutions. In our formulation (28) this corresponds to letting the variance σ_0^2 tend to 0 and setting the shrinkage intensity $\lambda = 0$. By choosing $\sigma_0^2, \lambda \neq 0$ we search for solutions in a hyper-ellipsoid around the hyper-plane that is spanned by the principal components.

3. Experimental Results

The experiments in this section show the resilience of our method to missing and noisy data. For all the experiments, we align the shapes as a pre-processing step by rigid registration using landmark points. To avoid being stuck in a local minimum, we perform the registration using a multi-resolution discretization. As a basis for finding a minimum of the discretized version of (30), we use Thirion's Demons algorithm as implemented in the Insight Toolkit [11]. The parameters are set individually for each experiment, but are chosen to be the same for the statistically regularized method and the standard method without statistical regularization.

Hands In the first experiment we built a statistical model from the outline of 9 hands (Figure 1). With this model as prior information, we register the reference image to a new image not contained in our model, in which one of the fingers has been cut off. Figure 2 shows the result we obtain by warping the reference image with the resulting deformation field. It can be seen that with the statistical regularization the surface is accurately matched. Where there is no data due to the missing finger, it is interpolated in a statistically meaningful way. In contrast, without the statistical regularization, the registration does not find meaningful correspondences in the place where the data is missing.

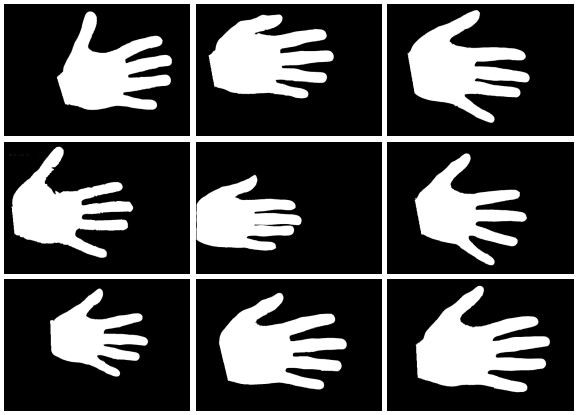

Figure 1. The shapes that are used to build the statistical model. The upper-left hand is the shape we use as the reference.

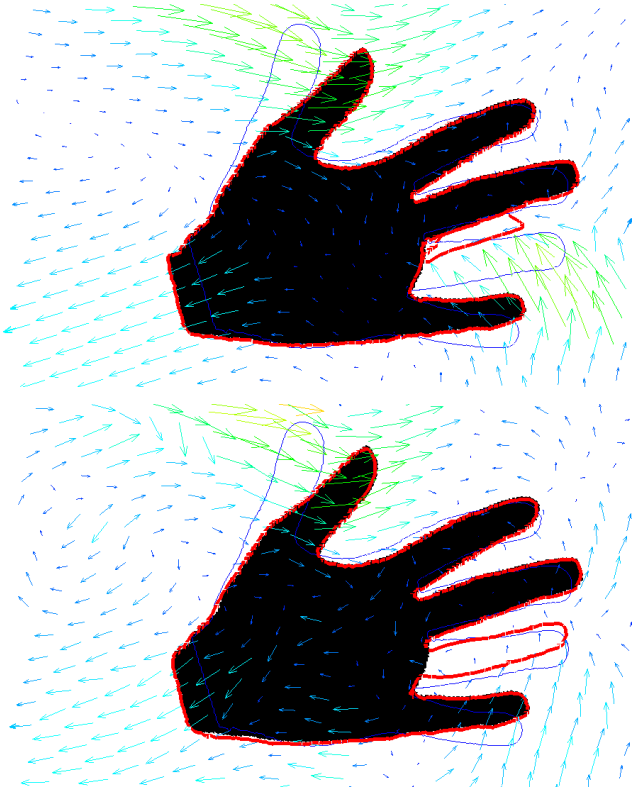

Figure 2. The reference shape (thin blue line) is registered onto a hand with the index finger missing (black shape). The red line shows the warp of the reference with the resulting deformation field, without statistical regularization (upper image), and with statistical regularization (lower image).

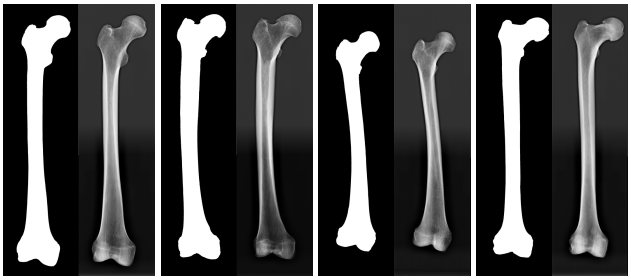

Figure 3. The first four out of 19 femur bones used to build the statistical deformation model. The first bone is the reference. Both shape and image information were used in the registrations.

Bones In the second experiment, we built a statistical deformation model from 19 femur bones. We have used the shape registration method including the x-ray image information (Figure 3) by minimizing the functional (8).

In Figure 4 we try to register the x-ray image of the reference to a noisy x-ray image of a bone not contained in the model. In this case, no shape information is used as the noisy image would not have been easy to segment. The first image in Figure 4 shows the result of attempting to register the x-ray images directly without statistical information. The outline of the reference bone is displayed in the image and gives an indication of how well the warped target image matches the reference. While the outline seems to be reasonably well matched, we have to observe that the registration distorts the image in an unnatural way and fails to establish correspondence, visualized by the vectors of the deformation field. For instance, the bone marrow canal in the shaft of the bone is completely distorted by the registration. The second registration is regularized with the statistical model and yields a more natural warp and deformation field. We observe that the contour of the registered bone does not completely coincide with the target contour, because, as always, there is a trade-off between regularization and matching accuracy.

4. Bootstrapping and Parameter Tuning

The experimental results presented in Section 3 show how shapes with severe artifacts can be accurately registered, given an expressive enough statistical model. This suggests the use of a bootstrapping strategy. Assume we have a number n of images available. We order the images according to their quality and start to register the image with the least noise. After two registrations, we can already obtain (an admittedly very rough) estimate of the deformation probability distribution. In the next registration the statistical model can already be used to slightly penalize unlikely deformations, *i.e*. those deviating strongly from the estimated mean. The more examples we include, the better the estimates become and the more weight we can put on the statistical regularizer. In this way, it is possible to include data sets with severe artifacts, once the statistical model has become reliable enough. For instance, we could use many noisy x-ray images like the one in Figure 4 for extending our statistical model.

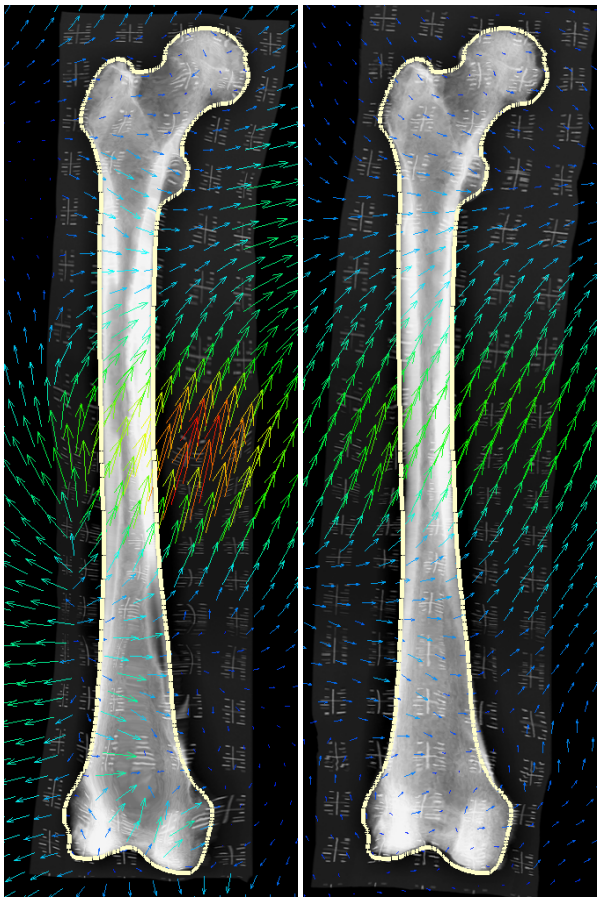

Figure 4. Registration of a noisy CT image onto the reference. In the first try, the registration matches the reference's outline (light yellow) accurately but distorts the image unnaturally, see the shaft for instance. In the second try, the statistical model is used and both the warp and the deformation field (arrows) look much more natural.

As the new regularization term is formulated as an estimator of a probability distribution, results from statistics are readily applicable. In [16], Schäfer *et al*. show how to choose the shrinkage parameter optimally. The remaining parameters should be set depending on a confidence measure of the estimator.

5. Conclusion

In this paper we have presented an approach for incorporating prior knowledge in the form of a statistical deformation model into a class of non-rigid registration algorithms. The prior knowledge is added to the registration method as

an additional regularization term, complementing the usual smoothness term. The statistical deformation model is built from a set of previous registration results, *i.e*. dense deformation fields. Our experimental results show that the new regularization greatly improves registration results in the case of noisy and incomplete data. Indeed, the prior knowledge allows us to register data-sets which are otherwise too noisy to lead to useful results.

In contrast to other approaches, our method does not restrict the solution to lie in the span of the examples of the model. It can therefore be applied to generate new deformation fields that can themselves be used to increase the expressiveness of the model. This observation leads to a bootstrapping strategy for building a statistical model from noisy or incomplete data-sets, requiring only a few high quality examples.

The use of statistical deformation models has the advantage that it can be applied to registration of surfaces represented by level-sets as well as images of any modality. For some applications, however, it might be a limitation that the deformation fields are always relative to a reference image, *i.e*. registration can only be performed from the reference image used to build the statistical model. It also requires an initial alignment of all images to the reference image, using for instance landmark-based rigid registration. For our application of building statistical models, this is a small price to pay compared to the largely increased number of data-sets we are able to register using this method.

In this work we have not addressed the question of how the newly introduced parameter, *i.e*. the trade-off between smoothness regularizer and statistical regularizer, can be chosen automatically. A detailed study of how to choose the parameter optimally, based on statistical measures of confidence, is the topic of future work. Another question we are going to address in upcoming work is finding a more memory-efficient representation of the statistical deformation fields, such that the methods becomes feasible to use with high-resolution 3D data-sets on a standard PC.

Acknowledgments This work was funded by the Swiss National Science Foundation in the scope of the NCCR CO-ME project 5005-66380 and the Hasler Foundation in scope of the HOVISSE project. We would like to thank the University Hospital of Basel for providing the x-ray images.

References

[1] M. A. Audette, F. P. Ferrie, and T. M. Peters. An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis*, 4:201–217, 2000.

[2] C. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99*, pages 187–194. 1999.

[4] T. Cootes, C. Taylor, D. Cooper, J. Graham, et al. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] 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*, 72(2):195–215, 2007.

[6] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion Snakes: Introducing Statistical Shape Knowledge into the Mumford-Shah Functional. *International Journal of Computer Vision*, 50(3):295–313, 2002.

[7] A. Dedner, M. Lüthi, T. Albrecht, and T. Vetter. Curvature guided level set registration using adaptive finite elements. In *Pattern Recognition*, pages 527–536, 2007.

[8] J. W. Demmel. *Applied Numerical Linear Algebra*, chapter 3, pages 109–119. SIAM, 1997.

[9] J. Gee and R. Bajcsy. Elastic matching: Continuum mechanical and probabilistic analysis. *Brain Warping*, 1998.

[10] B. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[11] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc., 2005.

[12] M. Lüthi, T. Albrecht, and T. Vetter. Curvature guided surface registration using level sets. In *Proceedings of CARS*, pages 126–128, 2007.

[13] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford Science Publications, 2004.

[14] N. Paragios, M. Rousson, and V. Ramesh. Non-rigid registration using distance functions. *Computer Vision and Image Understanding*, 89(2-3):142–165, 2003.

[15] D. Rueckert, A. Frangi, and J. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *Medical Imaging, IEEE Transactions on*, 22(8):1014–1025, 2003.

[16] 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*, 4(1):32, 2005.

[17] J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[18] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. John Wiley & Sons, 1977.

[19] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–742, 1997.

[20] Y. Wang and L. Staib. Physical model-based non-rigid registration incorporating statistical shape information. *Medical Image Analysis*, 4(1):7–20, 2000.

[21] Z. Xue, D. Shen, and C. Davatzikos. Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. *Medical Image Analysis*, 10(5):740–751, 2006.

[22] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.