

Brownian Warps for Non-Rigid Registration

Mads Nielsen · Peter Johansen · Andrew D. Jackson ·
Benny Lautrup · Søren Hauberg

Published online: 15 May 2008
© Springer Science+Business Media, LLC 2008

Abstract A Brownian motion model in the group of diffeomorphisms has been introduced as inducing a least committed prior on warps. This prior is source-destination symmetric, fulfills a natural semi-group property for warps, and with probability 1 creates invertible warps. Using this as a least committed prior, we formulate a Partial Differential Equation for obtaining the maximally likely warp given matching constraints derived from the images. We solve for the free boundary conditions, and the bias toward smaller areas in the finite domain setting. Furthermore, we demonstrate the technique on 2D images, and show that the obtained warps are also in practice source-destination symmetric and in an example on X-ray spine registration provides extrapolations from landmark point superior to those of spline solutions.

Keywords Non-rigid registration · Brownian motion · Central limit theorem · Invariance

M. Nielsen · P. Johansen · S. Hauberg (✉)
Department of Computer Science, University of Copenhagen,
Universitetsparken 1, 2100 Copenhagen, Denmark
e-mail: hauberg@diku.dk

P. Johansen
e-mail: peterjo@diku.dk

M. Nielsen
Nordic Bioscience A/S, Herlev Hovedgade 207, 2730 Herlev,
Denmark
e-mail: madsn@diku.dk

A.D. Jackson · B. Lautrup
Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17,
2100 Copenhagen, Denmark

A.D. Jackson
e-mail: jackson@nbi.dk

B. Lautrup
e-mail: lautrup@nbi.dk

1 Introduction

In any non-rigid registration algorithm, one must weigh the data confidence against the complexity of the warp field mapping the source image geometrically into the destination image. This is typically done through spring terms in elastic registration [3, 9, 10], through the viscosity term in fluid registration [8], by controlling the number of spline parameters in spline-based non-rigid registration [1, 23] or by finding the “smallest” warp as a geodesic flow according to a specific norm on warps [4, 17].

The regularizer ensuring a simple (or “small” to use the terminology of Joshi et al. [15]) has profound influence on which solution is obtained, and the properties of the solution. When the data support is weak or the warp is large, the influence of the regularizer increases. Furthermore, if the regularizer in turn defines distances between admissible warps, it may be used as the foundation of making statistics of warps.

The diffeomorphic approaches [4, 9, 15, 17] have nicer theoretical properties in terms of having well defined inverse, allowing for composition, etc. However, all these approaches define the size of a warp as a length of path through the space of the admissible warps from the identity warp to the fiducial warp (or set of warps exhibiting fiducial properties, eg. specific point matches) and subsequently find the shortest path. Hence, it compares to integrating some “work” along a specific path of warps.

We wish to define a regularizer in a Bayesian setting as a prior on warps. This should be a prior on the group of diffeomorphisms and be “natural” in the sense that it behaves well under the group action of composition of warps. The simplest approach is then to define it as a Brownian walk in the group of diffeomorphisms. This paper constitutes the first coherent collection of material from a number of con-

ference papers on the topic [14, 18–20], clarifies a number of misconceptions in earlier manuscripts and give new theoretical properties on invariance and details for implementation.

Before going into the specifics of Brownian warps, their properties and implementations, we make the following observations.

2 Definitions

From a mathematical point of view we wish the warp W to fulfill the following principles

Smoothness W is continuous and sufficiently differentiable. If this is not the case, the warp would be able to “tear holes” in the image.

Invertible For all admissible W , the inverse W^{-1} exists. If the inverse did not exist, the warp would be able to fold.

Inverse smoothness The inverse is also smooth. If we are to work with the inverse of a warp, then obviously it should have the same properties as the warp itself.

Path-connected All warps are path-connected to the identity within the space of all admissible warps. Otherwise an underlying continuous process would not exist. In practice, reflections are neglected by this principle.

This in short means that the warp is a diffeomorphism of positive Jacobi determinant, we denote this positive diffeomorphisms.

We also wish to define a regularizer that allows us to compute warps. Intuitively this regularizer should act similar to a norm on warps, but we relax a bit on the formal requirements and formulate the following criteria for the regularizer $R(W)$ to fulfill:

Positive definite $R(W) \geq 0$ and $R(W) = \arg \min_W R(W) \Rightarrow W = I$. If $R(W)$ was a proper norm we would need $R(I) = 0$ to be the minimum, but to avoid normalization issues, we simply define $R(I)$ to be the minimum of R .

Triangle inequality $R(W_1) + R(W_2) \geq R(W_2 \circ W_1)$, where \circ denotes warp composition. If R is to behave similar to a norm, this criteria is obvious.

Source-destination symmetry $R(W) = R(W^{-1})$. Again, if R is to behave similar to a norm, it is obvious that the warp that moves a point from A to B should have the same complexity as the warp that moves a point from B to A .

Smoothness The regularizer is continuous in W and its variations are well defined. Otherwise optimization would not necessarily be tractable, and gradient-based methods ill-defined.

These properties are fulfilled by the previously mentioned diffeomorphic approaches [4, 9, 15, 17], but not by approaches that are linear in warp coordinates like linear spline-based methods [5, 7, 23].

We are now ready to define a warp: A non-rigid registration may be modeled by a warp field $W : \mathbb{R}^D \mapsto \mathbb{R}^D$

mapping points in one D -dimensional image into another D -dimensional image. We give the definition:

Definition 1 (Warp Field) A warp field $W(x) : \mathbb{R}^D \mapsto \mathbb{R}^D$ maps all points in the source image $I_S(x) : \mathbb{R}^D \mapsto \mathbb{R}$ into points of the destination image $I_D(x) : \mathbb{R}^D \mapsto \mathbb{R}$ such that $I_S(W(X))$ is the registered source image. W is invertible and differentiable (i.e., a diffeomorphism) and has everywhere a positive Jacobian $\det(\partial_{x_j} W) > 0$.

The identification of a warp field on the basis of images is a matter of inference. Below we will apply the Bayes inference machine [6], but a similar formulation should appear when using information theoretic approaches such as the minimum description length principle [21].

We wish to determine the warp field W that maximizes the posterior

$$p(W|I_S, I_D) = \frac{1}{Z} p(I_S, I_D|W)p(W),$$

where Z is a normalizing constant (sometimes denoted the partition function), $p(I_S, I_D|W)$ is the likelihood term, and $p(W)$ is the warp prior. The likelihood term is based on the similarity of the warped source and destination image and may, in this formulation, be based on landmark matches [7], feature matches [16, 22], object matches [2], image correlation [16], or mutual information [24]. The major topic of this paper is the prior $p(W)$ that expresses our belief in the regularity of the warp field prior to identifying the images.

3 Brownian Warps

We seek that distribution of warps which is the analogue of Brownian motion. We wish this distribution to be independent of warps performed earlier (i.e., invariant with respect to warps). This property is of fundamental importance particularly when determining the statistics of empirical warps, creating mean warps etc. We also want the distribution to be a simple function of the regularizer, and choose the maximum entropy solution, which is normally denoted the Gibbs distribution

$$p(W) = \frac{1}{Z} \exp[-R(W)].$$

The assumption of independence of previously performed warps then gives

$$p(W = W_2 \circ W_1) = \int p(W_2 = W \circ W_1^{-1})p(W_1)dW_1.$$

This corresponds to the semi-group property of Brownian motion: The distribution of positions after two moves corresponds to two independent moves and, through the central limit theorem, leads to a Gaussian distribution of positions.

Since this also holds for a concatenation of many warps, we can construct a warp as

$$W_B = \lim_{N \rightarrow \infty} \prod_{i=0}^N \circ W_i,$$

where the W_i are independent infinitesimal warps. This corresponds exactly to the definition of a Brownian motion on the real axis if the concatenation product is replaced by an ordinary sum.

In order to find this limiting distribution when all W_i are independent, we investigate motion in the neighborhood of a single point following along all the warps and make the following lemma:

Lemma 1 (Local structure) *Let $J_{W_i} = \partial_{x_j} W_i$ be the local Jacobian of W_i . Then, the Jacobian of a Brownian warp is*

$$J_{W_B} = \lim_{N \rightarrow \infty} \prod_{i=0}^N J_{W_i}.$$

Proof This is obviously true due to the chain rule of differentiation. \square

Assume that an infinitesimal warp acts as the infinitesimal independent warp round all points. We assume independence along the warp to constitute a Brownian motion. Furthermore, we assume spatial independence among points in their first order structure. This is the simplest possible assumption. Notice however that this does not imply that points move independently. It acts as a first order regularization on the warp including the spatial diffusion under gradient descend implied by the second order terms in the warp originating from the variation with respect to first order terms. Hence, this independence still assume spatial correlation in warp. Higher order correlation should be constructed in a warp invariant fashion which is far from trivial and left to later research.

In the case of spatial independence and independence of the infinitesimal warps along the warp, all entries in the local Jacobian are independent and identically distributed round the identity. Hence, we may now model

$$J_{W_B} = \lim_{N \rightarrow \infty} \prod_{i=0}^N I + \sigma \frac{1}{\sqrt{N}} H_i, \tag{1}$$

where H_i is a $D \times D$ matrix of independent identically distributed entries of unit spread. The denominator \sqrt{N} is introduced to make the concatenation product finite, and σ is the spread or the “size” of the infinitesimal warps. The details does not matter as they disappear in the limit, but one may think of σ as the standard deviation of the local motion.

To summarize, the limiting distribution of (1) is the distribution of the Jacobian of a Brownian Warp. In turn, this defines the Brownian distribution on warps, as we have no reason to assume other structure in the distribution.

Unfortunately, the solution to (1) is not given in the literature on random matrices. Gill and Johansen [12] solve the problem for matrices with positive entries and Högnäs and Mukherjea [13] solve, among other cases, the situation when the matrices are symmetric. However, Jackson et al. have solved the case for only two dimensions [14] and we are presently considering the solution for three. Here, we present only the result.

Theorem 1 (2D Brownian Jacobian) *The limiting distribution of (1) where H_i have independent entries of unit spread and $W : \mathbb{R}^2 \mapsto \mathbb{R}^2$, is given as*

$$p(J_{W_B}) = G(S/\sigma) \sum_{n=0}^{\infty} g_n(F/\sigma) \cos(n\theta), \tag{2}$$

where G is the unit spread Gaussian, g_n are related to the Jacobi functions, and the parameters are given as follows:

Scaling $S = \log(\det(J_{W_B}))$

Skewness $F = \frac{1}{2 \det(J_{W_B})} \|J_{W_B}\|_2^2$

Rotation $\theta = \arctan\left(\frac{j_{12} - j_{21}}{j_{11} + j_{22}}\right),$

where j_{ij} are the individual entries of J_{W_B} .

It is shown in [14] that the limiting distribution does not depend on features of the infinitesimal distribution other than its spread, σ . The parameter σ may be viewed as a measure of rigidity. The effects of the parameters are shown in Fig. 1.

It has been proved, that this distribution creates invertible warps (with probability 1), is invariant under inversion of the warp, and is Euclidean invariant [19]. Here we prove that the distribution is invariant under simultaneous and identical warping of source and destination.

Theorem 2 (Local diffeomorphic invariance) *The distribution of warps given as spatially independent Jacobians each*

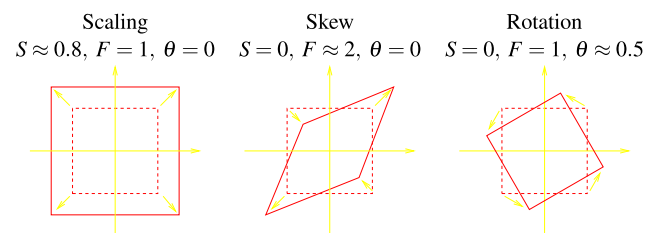


Fig. 1 The independent action of the parameters on a unit square

distributed according to (2) is invariant with respect to a diffeomorphism simultaneously acting on source and destination.

Proof A source and destination are related by a local Jacobian J such that $n_2 = Jn_1$, where n_1, n_2 are local frames in the source and destination image respectively. An arbitrary diffeomorphism acts locally on the frames with its Jacobian h . Acting on source and target simultaneously makes $n_2h = J'n_1h$. As all h, n_1, n_2 are invertible, obviously $J = J'$. \square

This theorem only hold as a local property, but is in general valid for a whole warp if an invariant measure is used for integration over the full warp field. Construction of such a measure is, however, not trivial in the general case. We will do so for the pairwise image matching problem below.

For computational purposes it may be convenient to approximate the above distribution by a distribution which is also independent in F and θ . This can be done in many ways without loosing the symmetry and diffeomorphic invariance. However, the convolution property of concatenation of warps will no longer hold exactly. We suggest the following approximation.

$$p(J_{W_B}) \approx G_\sigma(S)G_{\sigma/\sqrt{2}}(\theta)e^{-(F/\sigma)}, \tag{3}$$

where G_σ is a Gaussian of spread σ . This approximation has a relative error at less than 3% for all reasonable values of S, θ, F when $\sigma > 0.4$.

Taken from local points to a global distribution of a full warp, we may assume spatial independence of the local Jacobian of the warp. This does not correspond to assuming local independent motion of points, but that the local spatial differences in motion are distributed independently, just like independent increments (gradient) of neighboring points of a function in turn leads to Tikhonov regularization for functions. Taking this Markov Random Field approach, we may say that we formulate a first order MRF on the point motion function. The above distribution may then be viewed as Gibbs distributions, and the energy or minus-log-likelihood of a full field then reads

$$E'_s(W) = -\log p(W) + c = \int_{\Omega} S^2 + 2\theta^2 + 2\sigma F dx,$$

where c is an arbitrary irrelevant constant. However, the integration variable is not invariant under the warp, and the functional will not lead to warp invariance. This may be obtained by using a warp invariant integration measure $d\tilde{x}$:

$$E_s(W) = -\log p(W) + c = \int_{\Omega} S^2 + 2\theta^2 + 2\sigma F d\tilde{x}, \tag{4}$$

where $\tilde{x} = x\sqrt{\det(J)}$ are integration variables invariant under the warp chosen to ensure global as well as local warp

invariance. It may at first glance seem *ad hoc* to introduce this invariant measure. However it also follow directly from the probabilistic theory if one takes into account that after some (of the infinitely many) warps, it is more probable to see the areas that have increased in size. This is handled elegantly in the theory by Markussen as an effect of transforming the Itô integral of the spatio-temporal warp [17] into a Stratonovich formulation.

4 A PDE Solution

In general the warp energy (4) is augmented by an image or landmark matching term, so that the full functional to minimize for a given warp inference task reads

$$E(W) = E_s(W) + \lambda E_I(W),$$

where E_I is an image matching functional such as cross-correlation, mutual information, or landmark distance. Unfortunately the energy functional (4) is non-linear in the coordinate functions, and simple tricks such as eigenfunction expansions and derived linear splines are not possible. Therefore we will optimize this functional using a PDE approach gradient descend. We only concentrate on E_s as E_I is thoroughly treated elsewhere [11].

We treat the energy minimization problem using a gradient descend scheme:

$$\partial_t W = -\frac{\delta E}{\delta W} = -\frac{\delta E_s}{\delta W} - \lambda \frac{\delta E_I}{\delta W}.$$

We introduce the notation $Q(D, \theta, \|J\|_2^2) = S^2 + 2\theta^2 + 2\sigma F$ such that

$$E_s \equiv \int_{\Omega} Q(D, \theta, \|J\|_2^2) d\tilde{x}$$

and notice that Q does only depend on W in first order so that

$$\frac{\delta E_s}{\delta W} = -\left(\begin{matrix} \partial_x \\ \partial_y \end{matrix}\right)^T \frac{\partial Q}{\partial J},$$

where $\frac{\partial}{\partial J}$ denotes symbolic differentiation with respect to J , and ∂_x denotes spatial partial differentiation with respect to x (and similar for y). Since J is a 2×2 matrix varying in x , $\frac{\partial Q}{\partial J}$ becomes a 2×2 matrix of functions in x . That is,

$$\frac{\partial Q}{\partial J} : x \in \mathbb{R}^2 \mapsto \mathbb{R}^{2 \times 2}$$

and thereby $\frac{\delta E_s}{\delta W} : \mathbb{R}^2 \mapsto \mathbb{R}^2$.

Here we first concentrate on E'_s (not using the invariant integration variable \tilde{x} but plainly dx):

$$\frac{\partial Q'}{\partial J} = \frac{2 \log D - 2\sigma F}{D} \frac{\partial D}{\partial J} + \frac{\sigma}{D} \frac{\partial \|J\|_2^2}{\partial J} + 4\theta \frac{\partial \theta}{\partial J},$$

where J is the Jacobian matrix of W and $D = \det(J)$. Using the invariant coordinates (substituting $dx \mapsto \sqrt{D}dx$) this yields

$$\frac{\partial Q}{\partial J} = \frac{Q/2 + 2 \log D - 2\sigma F}{\sqrt{D}} \frac{\partial D}{\partial J} + \frac{\sigma}{\sqrt{D}} \frac{\partial \|J\|_2^2}{\partial J} + 4\sqrt{D}\theta \frac{\partial \theta}{\partial J}.$$

On an infinite domain the symbolic differentiation in these equations are very simple as all terms are co-linear or quadratic in the entries of J . Numerical issues do however arise on a finite domain since $\frac{\partial D}{\partial J}$ is non-zero only at the boundary.

Using E'_s directly serves the problem that the solution is no longer source-target symmetric as emphasis in the energy varies from point to point with respect to the local scaling. Using E_s in its full form using the invariant integration variable solves this problem.

On a bounded domain, this will lead to a simultaneous minimization on the size of the domain, to minimize the functional, and hence a bias toward shrinking warps. It will no longer give meaningful warps directly. This may be solved by fixing the size of the invariant domain directly using a Lagrange multiplier in the optimization problem:

$$E_{s\text{-bounded}} = \int_{\Omega} Q(D, \theta, \|J\|_2^2) d\tilde{x} + \lambda \int_{\Omega} d\tilde{x}. \tag{5}$$

We directly solve for λ using the fact that the time evolution of $\int_{\Omega} d\tilde{x}$ vanishes if

$$\lambda = -\frac{E'_s}{\int_{\Omega} d\tilde{x}}.$$

By simple calculus of variation we obtain:

$$\frac{\partial}{\partial J} Q_{s\text{-bounded}} = \frac{\lambda\sqrt{D} + Q/2 + 2 \log D - 2\sigma F}{\sqrt{D}} \frac{\partial D}{\partial J} + \frac{\sigma}{\sqrt{D}} \frac{\partial \|J\|_2^2}{\partial J} + 4\sqrt{D}\theta \frac{\partial \theta}{\partial J}, \tag{6}$$

where λ must be updated along the evolution. As λ is an integral measure, this actually is not a PDE but a partial integral-differential equation. So far, we have no proofs of stability of uniqueness of the solution. However, it works in the practical solution. It does not fall within the class for which uniqueness has been proved [11]. It also works on a totally different function space, since in previous work [11] the warps have been living in component-wise Sobolev spaces which has a non-empty intersection with the space of diffeomorphisms. However some diffeomorphisms are not

in the Sobolev space, and some members of the component-wise Sobolev space does fold and are obviously not diffeomorphisms.

This algorithm guarantees that the resulting warp is a diffeomorphism. It corresponds to some degree to the large deformation diffeomorphisms by Joshi and Miller [15] in the sense that their formulation also seek a solution composed over many time steps. However, we have succeeded in integrating out the time, and found the closed form solution for the resulting functional. Hence, we find the solution directly by optimizing the warp, and not by optimizing the warp, and all the intermediate steps, from source to destination. An interesting theoretical link between the two approaches is found in Markussen [17], where a warp-time discretization is performed, but where a Brownian motion formulation is used.

5 Implementation

We now turn to discretization of the above partial differential equation. First we discretize the energy in (5) as

$$E_{s\text{-bounded}} \approx \sum_{(x,y) \in \Omega} \sqrt{D} \cdot Q(D, \theta, \|J\|_2^2) + \lambda \sum_{(x,y) \in \Omega} \sqrt{D},$$

where D , θ , and $\|J\|_2^2$ are all functions of the local Jacobian in (x, y) . We can then compute the derivative of the energy with respect to the warp in the point (x_0, y_0) as

$$\begin{aligned} & \frac{\partial}{\partial W(x_0, y_0)} E_{s\text{-bounded}} \\ & \approx \sum_{(x,y) \in \Omega} \left[\frac{\lambda\sqrt{D} + Q/2 + 2 \log D - 2\sigma F}{\sqrt{D}} \frac{\partial D}{\partial W(x_0, y_0)} \right. \\ & \quad \left. + \frac{\sigma}{\sqrt{D}} \frac{\partial \|J\|_2^2}{\partial W(x_0, y_0)} + 4\sqrt{D}\theta \frac{\partial \theta}{\partial W(x_0, y_0)} \right]. \end{aligned}$$

To compute this gradient we need to evaluate the derivative of D , $\|J\|_2^2$, and θ with respect to $W(x_0, y_0)$. To do this, we first discretize the Jacobian. As a first approximation we compute the elements of the Jacobian at each grid point as backwards differences

$$\begin{aligned} J(x, y) & \approx \begin{pmatrix} u_x^-(x, y); & u_y^-(x, y) \\ v_x^-(x, y); & v_y^-(x, y) \end{pmatrix} \\ & = \begin{pmatrix} u(x, y) - u(x - 1, y); & u(x, y) - u(x, y - 1) \\ v(x, y) - v(x - 1, y); & v(x, y) - v(x, y - 1) \end{pmatrix}, \end{aligned}$$

where $W = (u, v)^T$. Using this approximation we can now compute the needed derivatives as

$$\frac{\partial D}{\partial u} \approx v_y^- - v_x^-$$

$$\frac{\partial \|J\|_2^2}{\partial u} \approx 2u_x^- + 2u_y^-$$

$$\frac{\partial \theta}{\partial u} \approx \frac{(u_x^- + v_y^-) - (u_y^- - v_x^-)}{(u_x^- + v_y^-)^2 + (u_y^- - v_x^-)^2},$$

and similar for the derivatives with respect to v . These derivatives are defined in terms of the discretization. To make the discretization symmetric in x and y we choose to use all four combinations of forward and backward differences in both coordinates. The final approximation of the energy is then computed as the average of these four approximations. When using this approximation we see that each grid point (x, y) that is not on the border appears 12 times in the energy. This is illustrated in Fig. 2. When computing the derivative of the energy we see that 12 frames will contribute to $\frac{\partial D}{\partial u(x,y)}$, $\frac{\partial \|J\|_2^2}{\partial u(x,y)}$, and $\frac{\partial \theta}{\partial u(x,y)}$ (and similar for $v(x, y)$).

At the boundary, the contributions from the discrete Jacobian leaving the domain are neglected, as the free boundary conditions are implemented in this way.

For time discretization of the PDE we use a simple explicit scheme. That is, we repeat the following step until convergence

$$W_{t+1} = W_t - \gamma \frac{\delta E}{\delta W}.$$

In this paper we assume that exact landmark matches are available. This basically means that the warp is known in

certain grid points. For this reason, the warp is only iteratively updated in grid points where the warp is unknown.

6 Results

We see from the energy formulation that the rigidity parameter σ determines the relative weight of the skewness term to the scaling and rotation terms. For illustration of the independent terms, see Fig. 3. For large deformations, the difference to spline-based methods, becomes obvious as for example thin plate splines can introduce folds in the warping (see Fig. 4).

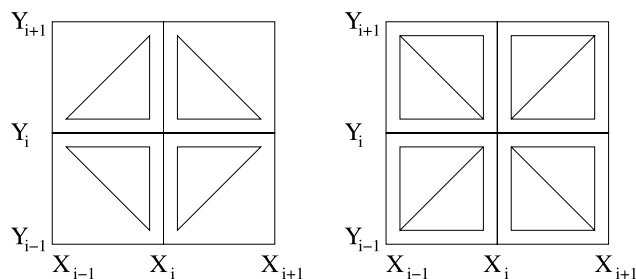


Fig. 2 In every point (x_i, y_i) , the local Jacobian is estimated from the 12 local discrete frames including the point. To the *left*, the four frames, where the point contributes centrally, are illustrated, whereas the 8 frames where the point contributes to in extremal position are illustrated to the *right*

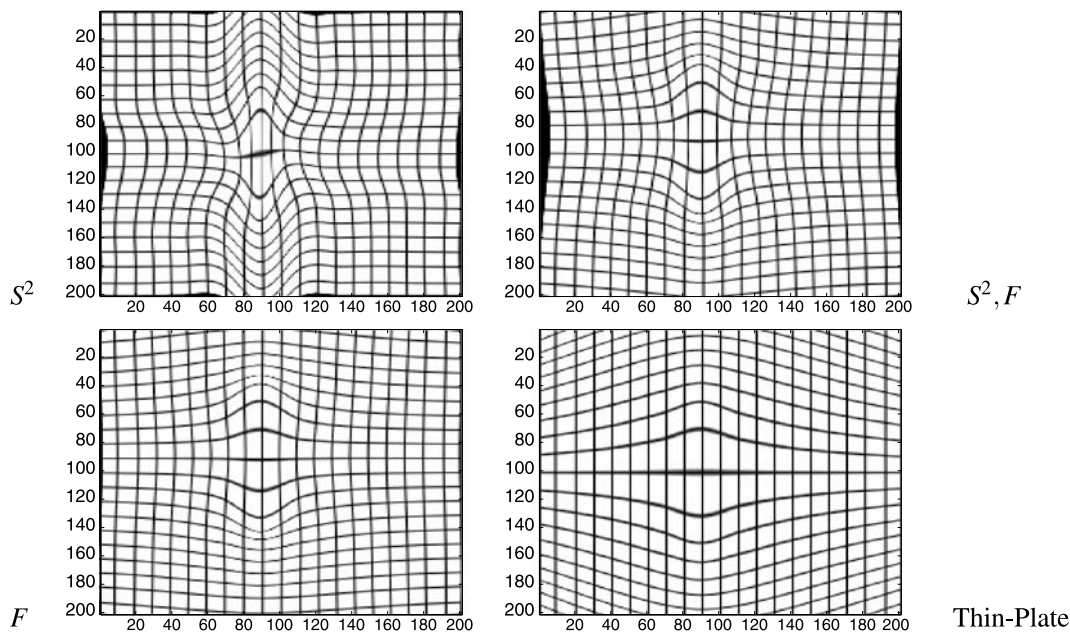


Fig. 3 Illustration of deformation of a regular grid. Two points in the center have been moved up and down respectively, while the corners are kept fixed. We see that the scaling term (*top left*) aims at keeping the area constant. The skewness term (*bottom left*) aims at keeping the

stretch equally large in all directions. *Top right* is a combination of scaling and skewness ($\sigma = 1$). *Bottom right* is a thin plate spline for comparison

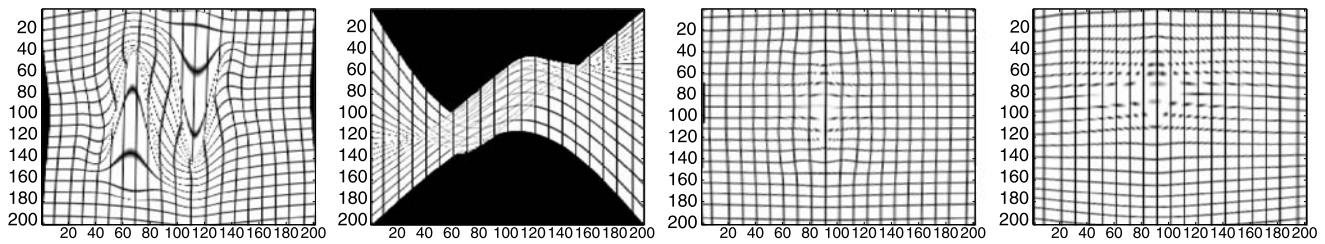


Fig. 4 Leftmost are two images of large deformations: *Left* is the maximum likelihood Brownian warp, *right* is a thin plate spline. Rightmost two images are two consecutive warps where landmark motions are

inverse: *Left* is Brownian warps, *right* is thin plate spline. Brownian warps do not give the exact inverse due to numerical imprecision, but closer than the thin plate spline

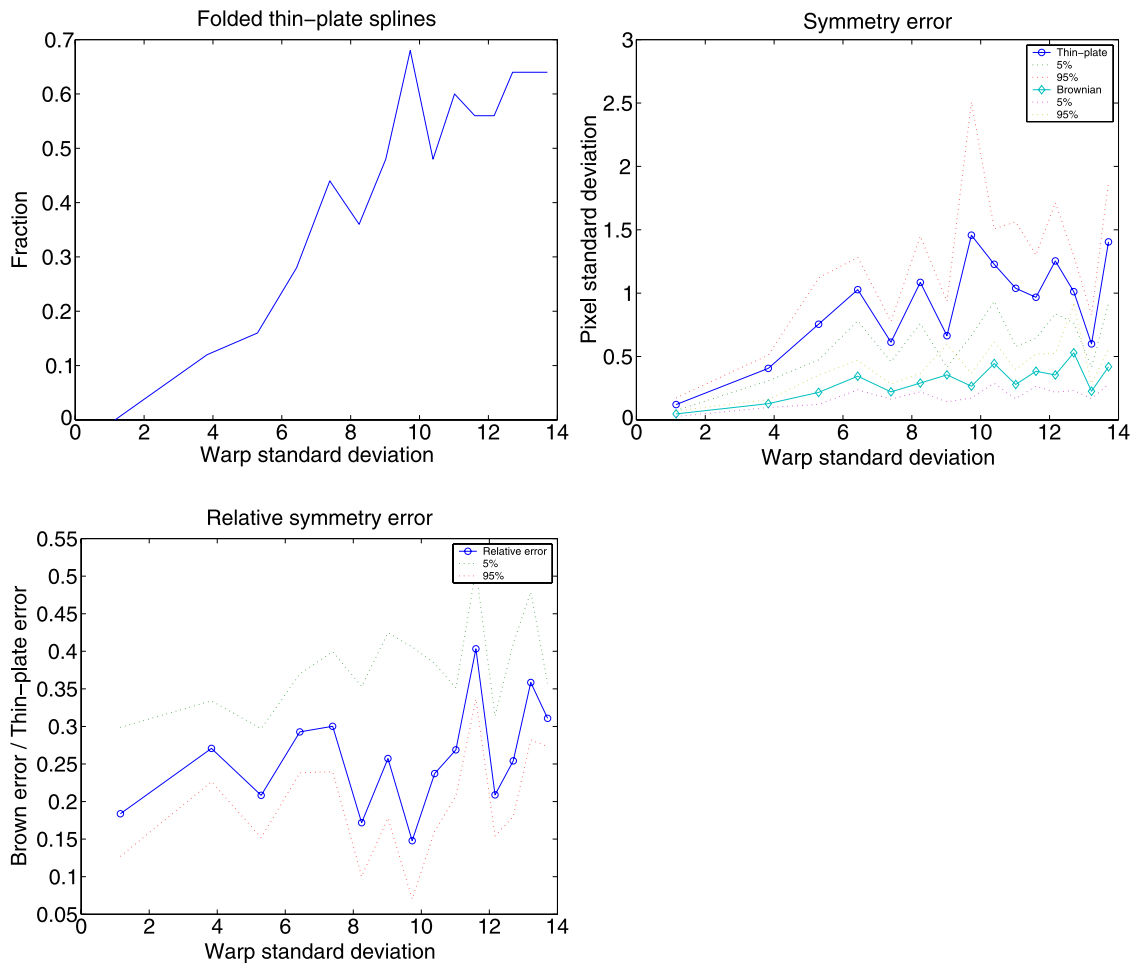


Fig. 5 Top-left is the fraction of thin-plate warps that contains a fold (is not invertible) as function of the spread of the random motion of the two interior point. Top-right is the absolute error (as difference to the identity warp) in pixel position when warping forward and concatenat-

ing with the backward warp. Below is the same for the relative error of the Brownian warps and the thin-plate warps. 25 runs for each standard deviation on a 50×50 grid was performed. All error bounds are bootstrapped 90% confidence intervals

For testing the source-target symmetry we conducted the following experiment. We kept the boundary fixed and moved two random points in the interior with a Brownian motion to new random positions (see Fig. 5).

The figures clearly show that the warp generated by the above algorithm is statistically significant more symmet-

ric than thin-plate spline warps. The motion of points after warping forward and back are less than a third than in the case of thin plate splines. Hence, not only is the theory symmetric, implementations show significant improvements. However, the warps are not totally symmetric, which in our opinion is due to the spatial discretization, as the er-

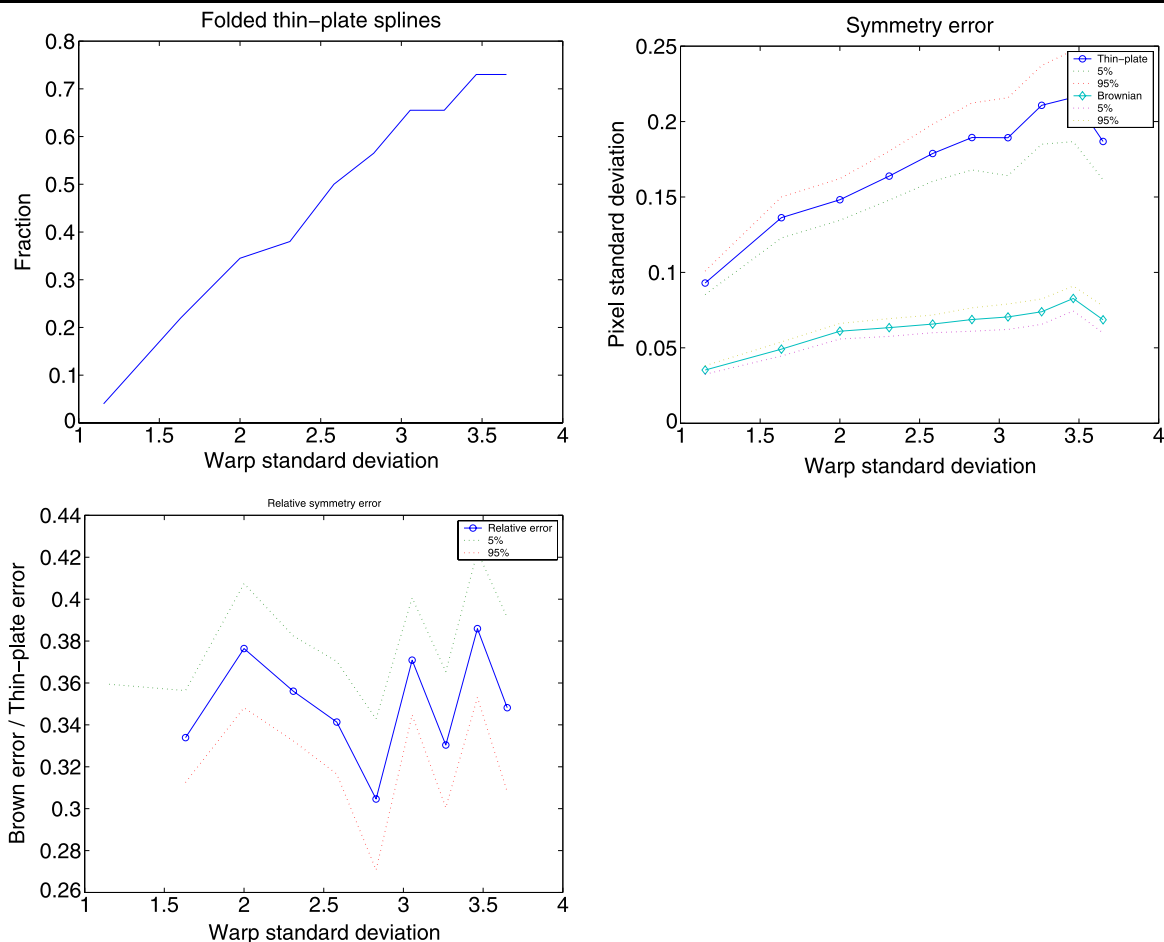


Fig. 6 As previous figure, but with 200 runs for each standard deviation on a 10×10 grid

ror is smaller on a 50×50 grid than on a 10×10 grid (see Fig. 6).

We evaluate the Brownian warps and thin-plate-spline warps for doing extrapolation of registrations of biological shapes. From a few point matches, we wish to generate a registration of the full outline. This is of course not accurate and we evaluate the quality of the registration based on the distance of the curves.

10 fractured vertebrae (the individual bones in the spinal column) have been annotated by 6 points used for fracture scoring in traditional clinical practice (see Fig. 7). These 6 points are used as landmarks for a thin-plate-spline non-rigid registration and a Brownian warp registration of the fractured vertebrae to a normal template vertebra (see Fig. 8). All vertebrae have also been annotated by a full contour. The registration can be used for transporting the full contour to the frame of the normal template vertebra. The hypothesis is now that since the diffeomorphic registration is theoretically more appropriate, it may also generalise the registration of the 6 points to the full contour better. Hence we wish to measure the distance between the template con-

tour and the contour of the fractured vertebra after registration based on the 6 height points.

The result of the registration is that two curves C_1, C_2 in the plane are given. Many different distances between these may be measured. Assuming one curve is “correct” (the template curve C_1) we wish to estimate the distance from this to an approximated curve C_2 of the registered fractured vertebra. Since no correspondence is given between points on the curves, we measure, for each point on the correct curve, the distance to the closest point on the approximate curve. This is, in mathematical terms, the Hausdorff distance from the template boundary C_1 to the warped fractured vertebra boundary C_2 :

$$d_H(C_1, C_2) = \int_c \min_{s_2} d(C_1(s_1), C_2(s_2)) ds_1$$

where C_1, C_2 are the two curves, s_1, s_2 are their respective natural parameters, and $d(\cdot, \cdot)$ is the Euclidean distance between two points.

In order to not let the arbitrary endpoint of the curves influence the measure, we have chosen to cut the template

Fig. 7 The template vertebra and the 6 annotated points normally used for fracture scoring, but here as landmarks for registration

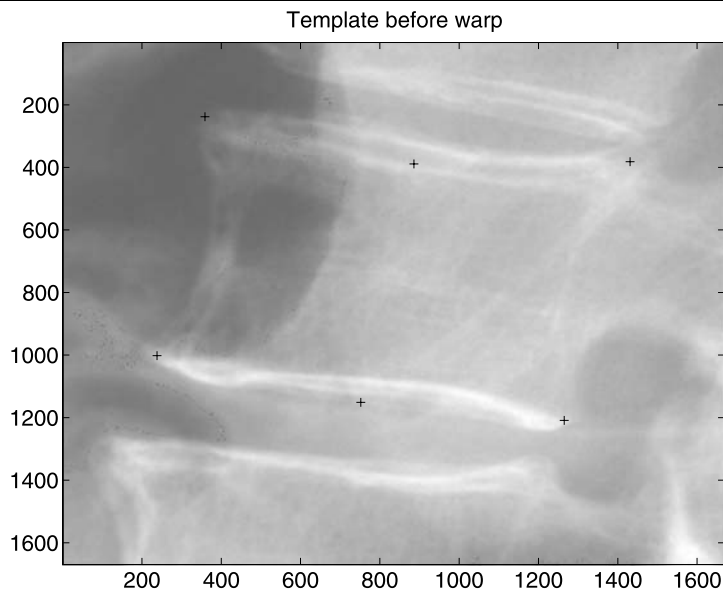
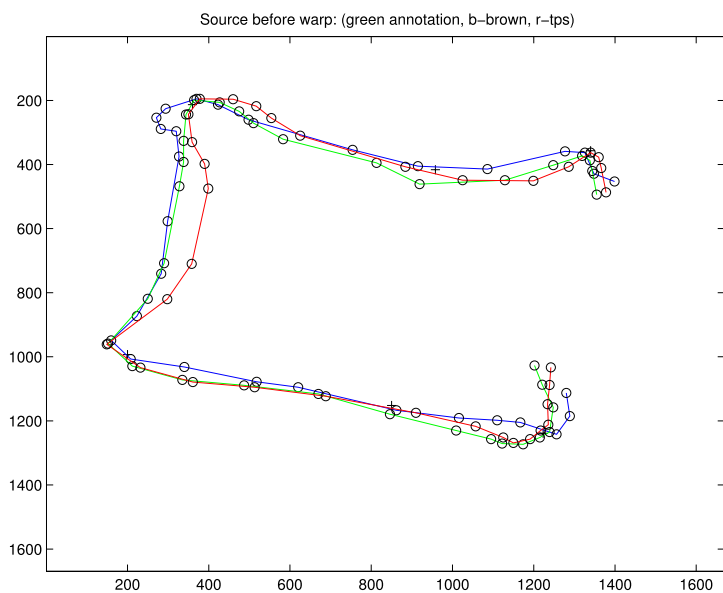


Fig. 8 (Color online) An example of a fractured vertebra and its boundary in green. In blue is the Brownian warped template boundary. In red is the thin-plate spline warped template boundary



in each end, so that it covers less of the vertebra than any of the test vertebrae.

The results are as follows:

Warping method	Average d_H	Standard dev.
Thin-plate-spline	4.83 mm	1.41 mm
Brownian warps	3.41 mm	1.27 mm

The average Brownian warp Hausdorff distance is significantly smaller than the thin-plate warped Hausdorff distances with a $p < 0.05$ using a two-sided unpaired heteroscedastic Student's t-test.

The results here are that the theoretically well-founded diffeomorphic Brownian warps actually also produce a significantly better warp for the above practical purposes.

Matching fractured vertebrae using only 6 points is a difficult task, but the Brownian warps removed more than 30% of the warping error compared to the standard linear approaches. The hope is that this will generalize to other anatomical registration tasks. If this is indeed the case, diffeomorphic warping may lead to improved registration in many other cases, and subsequently to better atlas-based segmentations, more compact shape models, and more compact appearance models.

7 Conclusion

We have exploited a prior for warps based on a simple invariance principle under warping. This distribution is the

warp analogue of Brownian motion for additive actions. An estimation based on this prior guarantees an invertible, source–destination symmetric, and warp-invariant warp. When computational time is of concern, approximations can be made which violate the basic semi-group property while maintaining the invariances. For fast implementations, we recommend an approximation including only the skewness term, as this has nice regularizing properties.

We have developed a PDE scheme for implementing an algorithm computing the maximum-likelihood warp. We have tested this in the case of exact landmark matching, and shown that it does not fold (as theory predicts) as linear approaches will do, and shown that also in discrete approximation, the scheme yields solutions very close to being source-target symmetric. The approach has shown promising results for medical image registration.

Future work includes a joint optimization scheme with other image matching terms as used earlier [11].

References

- Amini, A.A., Curwen, R.W., Gore, J.C.: Snakes and splines for tracking non-rigid heart motion. In: ECCV96, pp. II:251–261 (1996)
- Andresen, P.R., Nielsen, M.: Non-rigid registration by geometry-constrained diffusion. *Med. Image Anal.* **6**, 81–88 (2000)
- Bajcsy, R., Kovacic, S.: Multiresolution elastic matching. *CVGIP* **46**, 1–21 (1989)
- Faisal Beg, M., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* **61**(2), 139–157 (2005)
- Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992)
- Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Berlin (2006)
- Bookstein, F.L.: *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge (1991)
- Bro-Nielsen, M., Gramkow, C.: Fast fluid registration of medical images. In: *Proc. Visualization in Biomedical Imaging (VBC'96)*, pp. 267–276 (1996)
- Christensen, G.E., He, J.: Consistent nonlinear elastic image registration. In: *MMBIA01*, p. 37 (2001)
- Christensen, G.E., Miller, M.I., Vannier, M.: A 3d deformable magnetic resonance textbook based on elasticity. In: *AAAI Spring Symposium Series*, pp. 153–156. Standford University Press, Standford (1994)
- Faugeras, O., Hermsillo, G.: Well-posedness of eight problems of multi-modal statistical image-matching. Technical Report, INRIA, August 2001. Research Report 4235
- Gill, R.D., Johansen, S.: A survey of product-integration with a view toward application un survival analysis. *Ann. Stat.* **18**(4), 1501–1555 (1990)
- Högnäs, G., Mukherjea, A.: *Probability Measures on Semigroups*. Plenum, New York (1995)
- Jackson, A.D., Lautrup, B., Johansen, P., Nielsen, M.: Products of random matrices. *Phys. Rev. E* **66**(6), 5 (2002). Technical Report, article 66124
- Joshi, S., Miller, M.M.: Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Image Process.* **9**(8), 1357–1370 (2000)
- Maintz, J., Viergever, M.: A survey of medical image registration. *Med. Image Anal.* **2**(1), 1–36 (1998)
- Markussen, B.: Large deformation diffeomorphisms with application to optic flow. *Comput. Vis. Image Underst.* **106**, 97–105 (2007)
- Nielsen, M.: Evaluation of Brownian warps for shape alignment. In: *Proc. of SPIE Medical Imaging 2007* (2007)
- Nielsen, M., Johansen, P., Jackson, A.D., Lautrup, B.: Brownian warps: A least committed prior for non-rigid registration. In: *MICCAI* (2), pp. 557–564 (2002)
- Nielsen, M., Markussen, B.: From Bayes to PDEs in image warping. In: Paragios, N., Chen, Y., Faugeras, O. (eds.) *Mathematical Models in Computer Vision: The Handbook*. Springer, Berlin (2005)
- Rissanen, J.: *Stochastic Complexity in Statistical Enquiry*. World Scientific, Singapore (1989)
- Rohr, K.: *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer Academic, Dordrecht (2001)
- Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. *Med. Imaging* **18**(8), 712–721 (1999)
- Viola, P.A.: Alignment by maximization of mutual information. Technical Report AITR-1548 (1995)



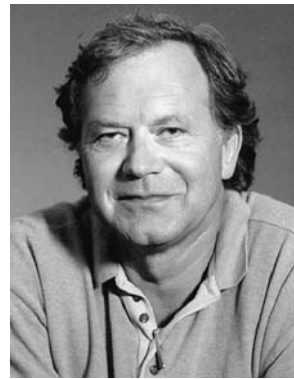
Mads Nielsen received a MSc in 1992 and a PhD in 1995 both in computer science from DIKU, Department of Computer Science, University of Copenhagen, Denmark. During his PhD studies he spent one year (1993–94) at the Robotvis lab at INRIA, Sophia-Antipolis, France. In the second half of 1995 he was post-doc at the Image Sciences Institute of Utrecht University, The Netherlands. In 1996 he was joint post-doc at DIKU and 3D-Lab, School of Dentistry, University of Copenhagen, where he served as assistant professor 1997–99. In 1998–99 he served as external associate professor at Institute of Mathematical Modelling, Technical University of Denmark. April 1999 he became the first associate professor at the new IT University of Copenhagen. Since June 2002 he was professor the same place heading the Image Analysis Group. From 2007 the group moved to DIKU, where he is heading the merged image groups and simultaneously heading Nordic Bioscience Imaging A/S. He is serving on the editorial board of JMIV, IJCV and Medical Image Analysis. He was general chair of Scale-Space 1999, MICCAI 2006 and co-chair of MMBIA 2007, MICCAI 2005 and ECCV 2002.



Peter Johansen (b. 1938) is mag. scient. 1962 in mathematics, and dr. phil. 1973 from University of Copenhagen, Denmark. He was assistant professor at the Technical University 1964–1967, and post. doc at MIT, USA 1967–1969. He was professor at University of Copenhagen 1973–2008. He has worked with automata theory, with Lindenmayer systems, and since 1979 with computational image analysis. His areas of interest is the interaction between pattern recognition and data compression, and image analysis, in particular image scale space and image warping.



Andrew D. Jackson (b. 1941) received his Ph.D. in physics from Princeton University in 1967. He was professor of physics at the State University of New York, Stony Brook from 1968 to 1996 and is currently professor at the Niels Bohr Institute, Copenhagen. Jackson's research activities cover a variety of areas in theoretical physics with particular emphasis on subatomic physics, many-body physics and biophysics.



Benny Lautrup (b. 1939) is a professor in theoretical physics at the Niels Bohr Institute at the University of Copenhagen. Is known for his part in the Nakanishi-Lautrup formalism, a concept in quantum field theory. Has during his career worked at Nordic Institute of Theoretical Physics (Denmark), Brookhaven National Laboratory (USA), CERN (Switzerland), and Institut des hautes Etudes Scientifiques (France). He has published the books *Neural Networks—Computers with Intuition* with Søren

Brunak (original in Danish and also translated into German), and *Physics of Continuous Matter: Exotic and Everyday Phenomena in the macroscopic World* in 2005. He also writes popular articles about physics and participates in the public debate in Denmark.



Søren Hauberg (b. 1980) received has MSc degree in Computer Science at the Department of Computer Science, University of Copenhagen in 2007. He mainly works in Image Processing and Pattern Recognition. He is presently considering what to do with the rest of his life.