

Natural Gradients for Deformable Registration

Darko Zikic¹, Ali Kamen², Nassir Navab¹

¹ Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

² Siemens Corporate Research (SCR), Princeton, NJ, USA

Abstract

We apply the concept of natural gradients to deformable registration. The motivation stems from the lack of physical interpretation for gradients of image-based difference measures. The main idea is to endow the space of deformations with a distance metric which reflects the variation of the difference measure between two deformations. This is in contrast to standard approaches which assume the Euclidean frame. The modification of the distance metric is realized by treating the deformations as a Riemannian manifold. In our case, the manifold is induced by the Riemannian metric tensor based on the approximation of the Fisher Information matrix, which takes into account the information about the chosen difference measure and the input images. Thus, the resulting natural gradient defined on this manifold inherently takes into account this information. The practical advantages of the proposed approach are the improvement of registration error and faster convergence for low-gradient regions. The proposed scheme is applicable to arbitrary difference measures and can be readily integrated into standard variational deformable registration methods with practically no computational overhead.

1. Introduction

Given the target image I_T and the source image I_S , the general deformable registration problem can be stated as the minimization of the energy

$$E(\varphi) = E_D(I_T, I_S \circ \varphi) + \alpha E_R(\varphi), \quad (1)$$

consisting of the difference measure E_D and the regularization term E_R acting on the deformation φ . A popular class of methods for minimization of (1) are variational approaches, in which the unknown is the deformation function itself, and the optimization is based on the gradient of the energy.

In contrast to the gradient of the regularization term, the difference gradient ∇E_D does not have a direct physical

interpretation, since it depends in an abstract way on the input images. A major drawback of this lack of physical context is that the magnitude of the difference gradient at a certain point in the domain provides no information about the magnitude of displacement at this point. Since for variational approaches the gradient serves as the basis for the estimate of the displacement, this lack of correlation is a serious issue. A practical problem resulting from this is that the magnitude of the difference gradient is large at strong image edges, independent of the actual amount of the displacement at these points. This behaviour is present for all commonly used difference measures, and we refer to it as the *difference gradient bias*. In consequence, regions with strong edges are systematically favoured in the registration process, while the convergence and ultimately precision in regions with less strong (however still clearly present) edges are impeded. This leads to biased and sub-optimal registration results.

The main idea of this paper is to counteract the lack of physical context for difference gradients by changing the way in which the distance is measured in the space of deformations. We propose to employ a distance metric which reflects the variation of the difference measure between two deformations, instead of simply computing their Euclidean distance. This new distance metric is naturally induced by the chosen difference measure and the input images. This modification of the distance distorts the space of deformations, so that it represents the problem specific information.

We realize the proposed modification by modeling the space of deformations as a Riemannian manifold, which is induced by a corresponding Riemannian metric tensor. The metric tensor will also enable us to compute the gradient on the manifold by a modification of the Euclidean gradient.

We follow the work put forward most notably by Shun-ichi Amari in the field of *information geometry* [5], in which the Fisher Information matrix (FIM) is used as a metric tensor. The *Natural Gradient* is computed with respect to the manifold induced by the FIM as the metric tensor, and was shown to have preferable convergence properties for different applications using stochastic optimization [7]. As the

deformable registration problem is not naturally formulated as a stochastic optimization problem, in this work we discuss several possibilities to adapt the natural gradient concept to deformable registration. To this end, we develop and evaluate three different alternatives for the approximation of the FIM for deformable registration.

We demonstrate that using the natural gradient for deformable registration provides a remedy to the above issues related to the Euclidean gradient: The proposed approach improves the convergence rate and precision in low-gradient images areas. Furthermore, due to the properties of the deformable registration problem, the implementation of the resulting approaches is extremely simple, and poses very little additional computational cost. Thus, it is directly applicable to existing variational approaches.

1.1. Related and Prior Work

Our approach is inspired by the use of natural gradients for stochastic optimization for learning tasks [3]. Here, it was shown that parameter space for neural network learning tasks has a Riemannian metric structure, and advantages of the natural gradient compared to the Euclidean gradient were shown, especially in terms of convergence. The natural gradient has since been successfully applied to several different problems in the area of machine learning [4].

In the area of variational deformable registration, most methods treat the difference gradient with respect to the Euclidean frame. An exception is the work in [19, 8], where the problem is treated in Sobolev spaces. However, these works have a fundamentally different goal and employ a different manifold. In [19, 8], Sobolev spaces are employed to restrict the space of deformations to functions with certain regularity properties. On the other hand, we use a manifold based on the Fisher Information metric to incorporate the information about the chosen difference measure and the input images. This does not restrict the space of possible deformations, but changes the notion of distance in this space. The term *Natural Gradient* is employed in [8], however not strictly in its original sense of [3], but for gradients in arbitrary spaces which are employed.

2. Method

This section is organized as follows. The context of variational deformation registration is briefly presented in Sec. 2.1, and basics of Riemannian manifolds and Natural Gradients are treated in Sec. 2.2. The actual contribution is presented in Sec. 2.3 and 2.4, where we transfer the concept of natural gradients to deformable registration.

The last thing before embarking upon the method is to supplement Eq. (2) by definitions of the image domain $\Omega \subset \mathbb{R}^D$, the image functions $I_{\{S|T\}} : \Omega \rightarrow \mathbb{R}$, and the deformation $\varphi : \Omega \rightarrow \Omega$. We further require a discretized

version of the displacement function u at points in space $x_1, \dots, x_N \in \Omega$. The resulting vector $u \in \mathbb{R}^{ND}$ is constructed by $u = (u(x_1)^\top, \dots, u(x_N)^\top)^\top$, where a single subvector $u(x_i) \in \mathbb{R}^D$ is the displacement vector at x_i . The same holds for the difference gradient which is discretized by $\nabla E_D(u) = (\nabla E_D(u)(x_1)^\top, \dots, \nabla E_D(u)(x_N)^\top)^\top$. Now, we are ready to start.

2.1. Variational Deformable Registration

For variational approaches, the deformation φ is expressed in terms of the displacement u , as $\varphi = \text{Id} + u$, with the identity operator Id , so we can restate (1) as

$$E(u) = E_D(I_T, I_S \circ (\text{Id} + u)) + \alpha E_R(u). \quad (2)$$

Depending on the problem at hand, a number of difference measures and regularization terms can be employed [10, 14]. Here, the choice of more powerful statistical similarity measures such as Cross correlation (CC), correlation ratio (CR) [17], or mutual information (MI) [13, 21], is important for registration applications. Since the problem in (2) is non-linear, it is solved in an iterative manner by computing an update du to an initial displacement estimate u . In this work, we apply the compositional update rule $\varphi_{\text{new}} = \varphi_{\text{old}} \circ d\varphi$ to compute the estimates in the iterations.

Most commonly, the minimization problem in each equation is solved by computing the update du based on the gradient of the energy E with respect to the displacement, resulting in a steepest descent scheme

$$du = \frac{\partial u}{\partial t} = -\nabla E(u). \quad (3)$$

which is employed for example in [1, 14]. In order to solve the non-linear PDE in (3), a discretization of the time parameter is needed [14]. Here, the common choices are the explicit discretization leading to the update rule

$$du = -\tau(\nabla E_D(u) + \alpha \nabla E_R(u)), \quad (4)$$

and the semi-implicit discretization yielding a linear PDE

$$(\text{Id} + \tau \alpha \nabla E_R) du = -\tau(\nabla E_D(u) - \alpha \nabla E_R(u)). \quad (5)$$

The second popular group of variational methods is based on the so-called *demons* approach [18, 15, 8]. It refrains from defining an explicit regularization term in the energy function, which restricts the choice of the time-discretization scheme to (4) with $\alpha = 0$. The regularity of the deformation is ensured by convolution with a smoothing kernel, mostly a Gaussian G with variance β . This results in the fluid demons scheme

$$du = -\tau G_\beta * \nabla E_D(u). \quad (6)$$

While the demons approach does not require a gradient of the regularization term, all variational methods are based

on the gradient of the difference measure. The general form of the difference gradient at a certain point in space is

$$\nabla E_D(I_T, I_S, \varphi)(x) = W_D(I_T, I_S \circ \varphi)(x) \nabla I_S(\varphi(x)) . \quad (7)$$

Here, the scalar-valued function W_D is determined by the chosen difference measure. While W_D can depend on global statistics of the images, it is important to note that (7) depends directly on the gradient of the warped image $\nabla I_S \circ \varphi$. It is this dependence which causes the bias of the difference gradient to high-gradient regions.

In summary, all variational methods require the gradient of the difference measure, and in the most of the methods, the Euclidean difference gradient is used directly as the basis for the estimate of the displacement field update. Consequently all standard variational approaches suffer from the drawbacks associated with the difference gradient. We thus perform the modification of the distance metric for the difference term. To this end, we use the general definition of the gradient by the directional derivative $\partial_v E$ for the energy from (2)

$$\partial_v E = \langle \nabla E, v \rangle = \langle \nabla E_D, v \rangle + \alpha \langle \nabla E_R, v \rangle . \quad (8)$$

Here, we will augment the difference-related term from (8) by a Riemannian metric M as

$$\partial_v E = \langle M \nabla E_D, v \rangle + \alpha \langle \nabla E_R, v \rangle . \quad (9)$$

2.2. Riemannian Manifolds and Natural Gradients

The general definition of distance for two vectors v and $w = v + dv$ with sufficiently small dv reads

$$d_M(v, w)^2 = \langle M(v)dv, dv \rangle . \quad (10)$$

If the metric tensor M is a symmetric positive definite operator varying smoothly in the domain, it is called the Riemannian metric tensor, and it induces a Riemannian manifold. The Euclidean distance is the special case of (10), with constant $M = \text{Id}$. From the definition of the gradient in (8), one sees that the gradient in a certain space depends on the chosen metric tensor. It follows [2] that the gradient ∇_M , which is computed with respect to the Riemannian manifold induced by the metric tensor M , can be expressed in terms of the Euclidean gradient ∇ at the point v as

$$\nabla_M = M^{-1}(v) \nabla . \quad (11)$$

For stochastic optimization problems, it was shown that the Riemannian structure of the parameter space of a statistical model is defined by a metric tensor corresponding to the Fisher information matrix (FIM) [2]. For a random variable X with a probability density function P and parameters θ , the Fisher information matrix F is defined as

the covariance of the partial derivatives of the log likelihood function, that is

$$F(\theta)_{i,j} = \mathbb{E} \left[\frac{\partial \log P(X; \theta)}{\partial \theta_i} \frac{\partial \log P(X; \theta)}{\partial \theta_j} \right] , \quad (12)$$

where the expectation is evaluated with respect to $P(X; \theta)$. The gradient computed with respect to the metric (12) is called the *Natural Gradient*.

The basic difference between the stochastic optimization and the standard deterministic optimization methods which are used for deformable registration, is that the former operate on random variables and assume the existence of training sets. Since this is not naturally given for deformable registration problems, the Fisher information matrix from (12) will have to be approximated for our purposes.

2.3. Natural Gradients for Deformable Registration

In this section we transfer the concept of natural gradients to deformable registration. This will be done by following the work on natural gradients and using the Fisher information matrix as the metric which induces the Riemannian manifold with respect to which the difference gradient will be computed. To this end, we will have to approximate the expectation in the definition of the Fisher information in (12). This will lead us to three different models, which will be discussed in detail in Sec. 2.4, and evaluated in Sec. 3.

In order to define natural gradients for deformable registration, we have to cast the deformable registration problem as a statistical model. This can be done by employing the probabilistic model for difference measures [16], according to which a difference measure can be written as

$$E_D(I_S, I_T, u) = -\log P(I_T, I_S; u) . \quad (13)$$

By setting the random variable X to the input images I_S and I_T , and the parameters θ to the parameters of the deformation, that is the displacement u , we can combine (12) and (13), and write the Fisher information matrix in terms of deformable registration as

$$F(\theta)_{i,j} = \mathbb{E} \left[\frac{\partial E_D(u)}{\partial u_i} \frac{\partial E_D(u)}{\partial u_j} \right] \quad (14)$$

$$= \mathbb{E} [\nabla E_D(u) \nabla E_D(u)^\top] , \quad (15)$$

where we drop the image arguments to E_D for brevity.

In the setting of deformable registration, the evaluation of the expectation in (12) has to be approximated, since it is based on the assumption that a set of different training images is given, which agree on same parameters. This setting is not naturally given for deformable registration in (15), since we do not have a sampling set of target and source images, which are related by the same displacements. In the following we propose three different options for treating this issue.

1. We compile the sampling set by taking into account the information from sufficiently small regions around single points in the image domain. More precisely, we define a sampling set of gradients $\nabla E_D(u)_i$ by collecting subgradients $\nabla E_D(u)(y_i)$ with $y_i \in \mathcal{N}(x)$ in the vicinity of x . The resulting metric is referred to as the *FIM Region Metric* and treated in Sec. 2.4.1.
2. We assume no a priori distribution and take into account only the given input images. The sampling set thus now consists only of the input images, so that in this case, the expectation in (15) collapses to the value for the difference gradient on the given image pair. The resulting metric is referred to as the *FIM Frequentist Metric* and treated in Sec. 2.4.2.
3. As the last option, we derive the metric tensor by generalizing the *FIM Frequentist Metric*. This goes beyond a mere adaptation of the FIM and provides us with the flexibility to adapt the manifold according to the desired properties of the deformable registration problem. We refer to the result as the *Direct Metric* and give details in 2.4.3.

A general issue with (15) is that for deformable registration, the positive definiteness of F cannot be guaranteed, since the partial derivatives can equal zero, which can lead to a rank deficient and thus not invertible matrix. We counteract this in the construction of the metric by regularization

$$M = \sigma \text{Id} + F, \quad (16)$$

where σ is a noise parameter, compare e.g. [11].

2.4. The Metric Gallery

In this section we discuss the metric tensors resulting from the different assumptions considered in Sec. 2.3.

Common to all derivations in this section is the assumption of independence of subvectors $\nabla E_D(x)$ of the difference gradient at different points $x \in \Omega$. This is a common assumption for all approaches treating difference measures in a probabilistic setting, compare e.g. [16].

This assumption results in a simple structure of the Riemannian metric tensor. While in general, the discretization of the metric tensor M based on the FIM from (15) is a dense matrix of size $ND \times ND$, the independence assumption allows us to treat only a block-diagonal matrix, with blocks M_x of size $D \times D$, corresponding to single locations x in the image domain. In consequence, the general system for computation of the general gradient

$$\nabla_M E_D(u) = M^{-1}(u) \nabla E_D(u), \quad (17)$$

simplifies to

$$\nabla_M E_D(u)(x) = M_x^{-1}(u) \nabla E_D(u)(x). \quad (18)$$

The resulting block-diagonal system (18) is preferable to (17), since only N inversions of a small $D \times D$ system have to be performed, opposed to the inversion of the global $ND \times ND$ system. It is an interesting observation that properties of difference measures for deformable registration leads naturally to a simple version of the metric tensor for deformable registration, while for general large-scale learning problems, a lot of effort has to be invested in finding a good approximation to the metric tensor [11].

Now, we discuss the single realizations of the metric tensors for deformable registration, resulting from Sec. 2.3.

2.4.1 FIM Region Metric

In this approach, we assume that difference gradient subvectors in a small region $\mathcal{N}(x)$ of x constitute a sampling set for the $\nabla E_D(u)(x)$ entry of the difference gradient. This is denoted by $\nabla E_D(u)(y)$ for $y \in \mathcal{N}(x)$. The associated probability of the single gradients $\nabla E_D(u)(y)$ is based on the distance of their location y to the current point x . We model this in a standard way by a Gaussian distribution as

$$P(y) = G(y; \mu=x, \gamma). \quad (19)$$

This is comparable to the assumption made in the Lucas-Kanade optical flow approach [12].

This approach results in the metric tensor

$$M_x(u) = \sigma \text{Id} + \sum_{y \in \mathcal{N}(x)} P(y) \nabla E_D(u)(y) \nabla E_D(u)(y)^\top. \quad (20)$$

The actual computation of the natural gradient is performed in a block-wise manner by (18).

2.4.2 FIM Frequentist Metric

Under the assumption that we only take into account the difference gradient resulting from the given input images and the displacement u in every iteration, the expectation computation from (15) collapses and with (16) we arrive at the metric tensor M , composed of blocks

$$M_x(u) = \sigma \text{Id} + \nabla E_D(u)(x) \nabla E_D(u)(x)^\top. \quad (21)$$

For this case, the computation of the natural gradient is exceptionally simple, since $\nabla E_D(u)(x)$ is an eigenvector to $M_x(u)$ with the eigenvalue $\sigma + \|\nabla E_D(u)(x)\|^2$. Thus, by plugging (21) into (18), the natural gradient is computed by rescaling of the point-wise subvectors $\nabla E_D(u)(x)$ of the Euclidean gradient as

$$\nabla_M E_D(u)(x) = \frac{1}{\sigma + \|\nabla E_D(u)(x)\|^2} \nabla E_D(u)(x) \quad (22)$$

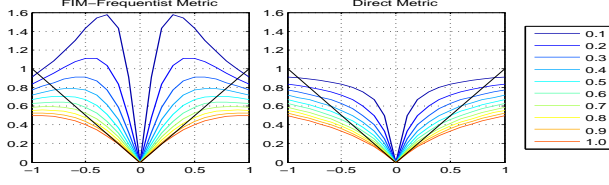


Figure 1: The application of the *FIM Frequentist Metric* and the *Direct Metric* induces a robust norm for point-wise subvectors of the difference gradient $\nabla E_D(x)$, depending on the noise level σ . x-axes depict the magnitude of the Euclidean gradient, and the y-axes the magnitude after the application of the inverse metric tensor.

2.4.3 Direct Metric

Motivated by the results in (22) and the computational efficacy of this approach, we consider a generalization for setting up a valid metric M by defining the blocks by

$$M_x(u) = \sigma \text{Id} + f(x)f(x)^\top, \quad (23)$$

with an arbitrary function $f : \Omega \rightarrow \mathbb{R}^D$.

This gives us the ability to shape the distance metric according to our requirements. We illustrate this on the following example. If we chose f as

$$f(x) = \frac{1}{\sqrt{\|\nabla E_D(u)(x)\|}} \nabla E_D(u)(x), \quad (24)$$

then $\nabla E_D(u)(x)$ is an eigenvector to M_x with an eigenvalue $\sigma + \|\nabla E_D(u)(x)\|$, and the natural gradient reads

$$\nabla_M E_D(u)(x) = \frac{1}{\sigma + \|\nabla E_D(u)(x)\|} \nabla E_D(u)(x). \quad (25)$$

In contrast to (22), now the subvector $\nabla E_D(u)(x)$ is normalized by its length, regularized by the noise level σ . When this gradient is used as estimate for the displacement, it gives approximately the same weight to all points for which a possible improvement in the difference measure is detected, if they are above a certain noise level. Please compare also Fig. 1.

Two interesting observations considering this approach and the frequentist metric from Sec. 2.4.2 can be made. Since (22) and (25) only changes the length of the subvector $\nabla E_D(u)(x)$, due to the form of the rescaling, it can be seen as an application of a robust norm to the single subvectors (however not the complete difference gradient vector). This is visualized in Fig. 1 for different choices of the noise parameter σ . It is important to note that while (22) does not change the direction of the single subvectors of the difference gradient, this is done in an independent way for each subvector, so that the complete difference gradient vector is re-oriented, and not simply re-scaled. Furthermore, please

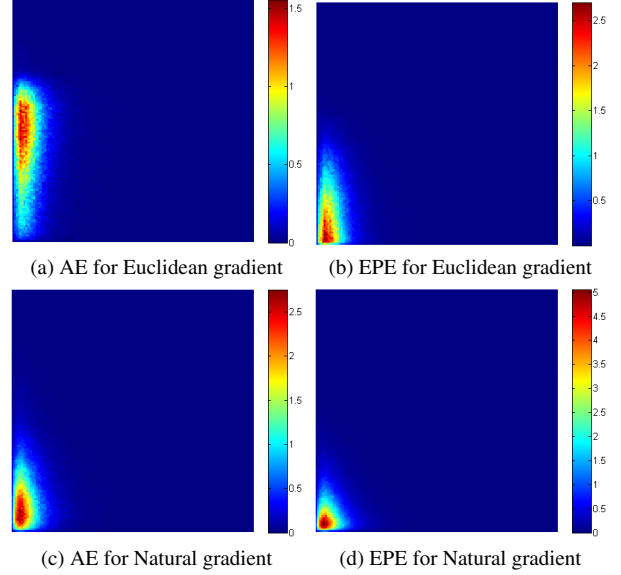


Figure 2: Distribution of the angular error (AE) and the end-point error (EPE) (y-axes) with respect to magnitude of ∇I_S (x-axis). The proposed approach clearly reduces the amount of large errors for structures with less prominent gradients.

note that the additional computational cost of this approach requires only a multiplication of the entries of the Euclidean vector by a scalar value. For registration applications, this does not present a computational overhead. Combined with the improvement of the convergence rate by the proposed approach, this reduces the overall runtime.

3. Experiments

We demonstrate the properties of the proposed method by results obtained by a standard variational approach with regularization defined in the energy term. We use the diffusion regularization component $E_R(u) = \int \sum_{k=1}^D \|\nabla u_k(x)\|^2 dx$, and employ the semi-implicit discretization from (5). The difference measures and their Euclidean gradients are implemented according to [10]. We also tested the proposed approach with a demons scheme. The results are very similar to the ones obtained by the first method, in terms of both, the final error, and the improvement of the convergence rate.

The tests are performed on several bio-medical images. Here, we do not follow a certain application, but merely want to show that the observed behaviour is not restricted to a certain class of images. In order to perform a quantitative evaluation, we generate ground truth displacements by a B-spline based FFD-grid of a 25×25 resolution. The grid control points are assigned random, uniformly distributed displacements of up to ± 5 pixels. In the experiments, we

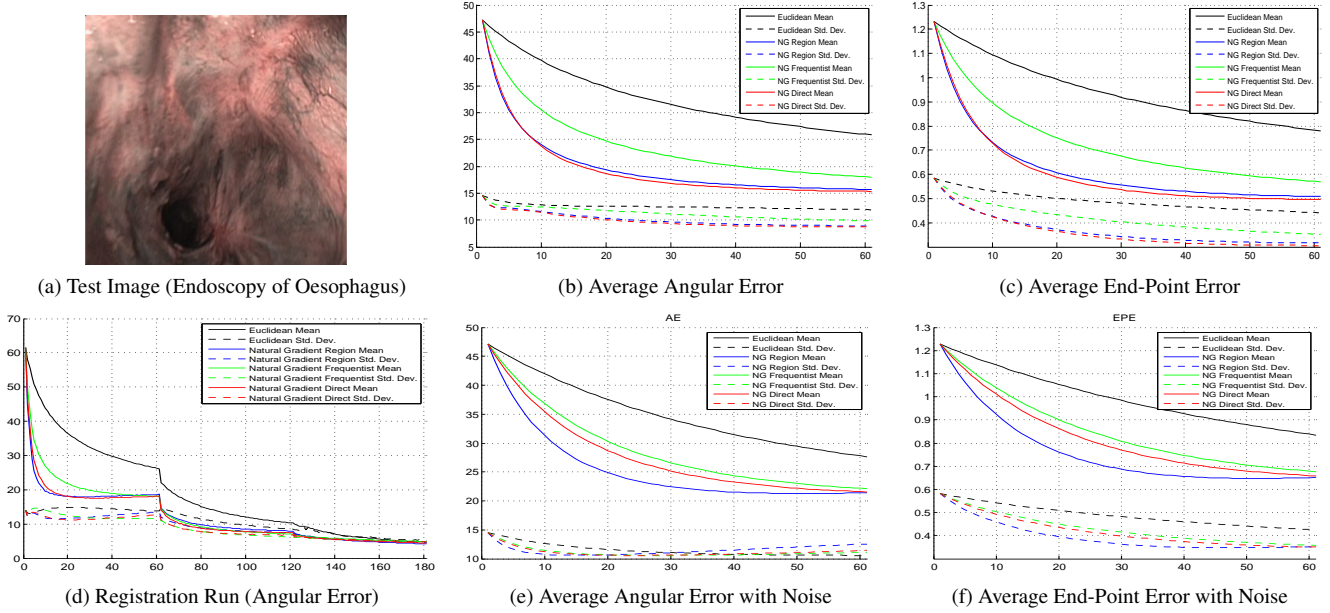


Figure 3: Averaged results of the random study with 100 runs on random displacement fields. We perform a registration with an image pyramid in (d), and the runs (e) and (f) are performed with Gaussian noise ($\sigma = 5\%$ of intensity range). The angular and end-point error exhibit comparable behaviour in all tests. We observe a clearly faster convergence of the proposed methods based on the natural gradient, compared to the standard usage of the Euclidean gradient.

monitor the mean and the standard deviation of the angular error (AE) and the end-point error (EPE) of the computed displacements over iterations, please compare [6] for definitions.

3.1. Distribution of Error

In all experiments we observe that the errors are reduced mostly in low-gradient image regions. This is quantified by comparing the error at a certain image point, to the magnitude of the image gradient at this point. We visualize this for an exemplary registration in Fig. 2. Plotting the error values versus the norm of the image gradients at single points in a joint histogram shows that the proposed approach most significantly reduces the error for low-gradient regions. Here, the example is given for the Direct metric. The other proposed approaches exhibit a similar behaviour.

3.2. Random Study

In the random study we evaluate the convergence of the standard approach employing the Euclidean gradient compared to the proposed versions of the natural gradient. To this end, we perform 100 runs on a narrow band endoscopy image of the oesophagus, please compare also Fig. 3. The used difference measure is the sum of squared differences (SSD). In each run, a new ground truth deformation u_{GT} is generated. Then, based on a given source image, the target

is created by warping the source image by u_{GT} . The registration is run with the same parameters for all different approaches. The time step τ and the regularization parameter α are held constant over iterations in order to assure comparability of the methods. In order to achieve a fair comparison, these parameters are tuned manually for the standard approach based on the Euclidean gradient. Then, the proposed methods are run with these parameters. The only parameter which is adjusted for the proposed method is the noise level parameter σ . For the noiseless cases, we use $\sigma = 0.01$ for the Frequentist and Direct metric approaches and $\sigma = 0.005$ for the approach based on the region metric. The standard deviation for the Gaussian for the region probabilities in Eq. (19) is set to $\gamma = 1.5$ in all tests. In order to demonstrate robustness to noise, we repeat this same test, but with Gaussian noise with standard deviation of 5% of the intensity range independently added to both images. Additionally to the random study on one level of the image pyramid, we also perform a qualitative example of a registration run over all pyramid levels. Here, we perform unnecessarily many iterations of the proposed method. The finding (based also on further registrations) is that also the final precision of the proposed approach tends to be better and the error in this case exhibits a smaller variance. To sum up, in all the performed random tests we detect a faster convergence of the proposed methods based on the natural gradient.

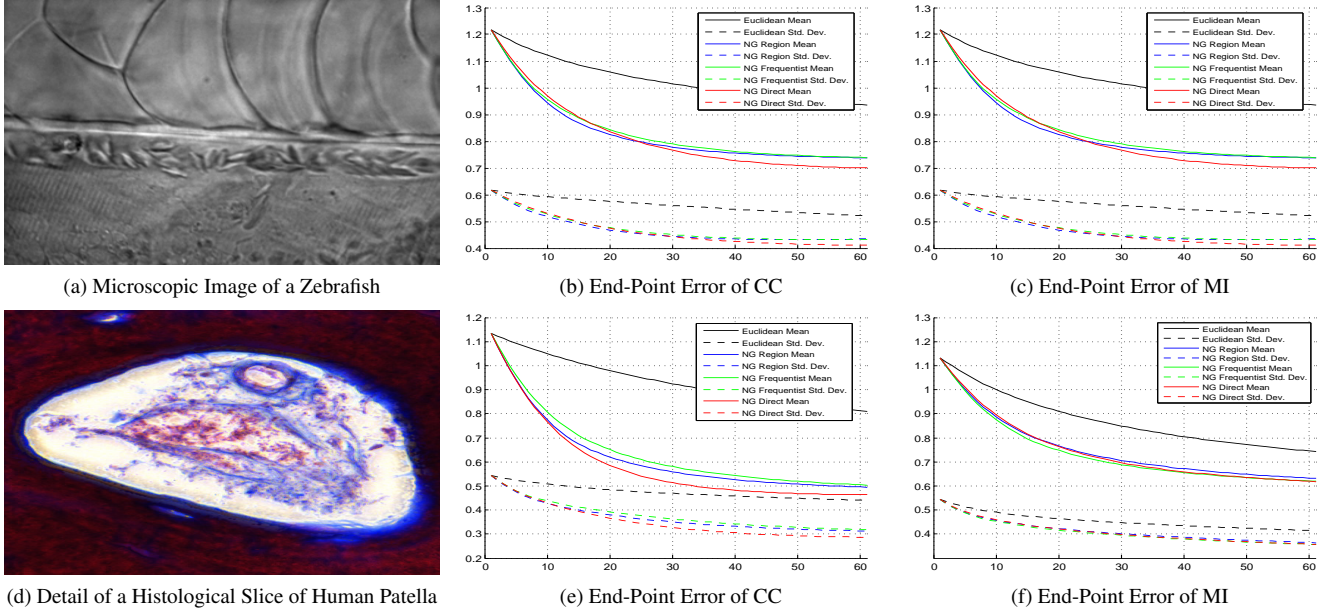


Figure 4: Results on further images with cross correlation (CC) and mutual information (MI) as difference measures. Again, we observe faster convergence of the natural gradient. Please see Sec. 3.3 for details.

3.3. Further Images, Different Measures

We also perform tests on further images and with other, more general difference measures. The other images are a bright field microscopic image of the dorsal aorta of a zebrafish larvae, and a detail of an image of a histological slice from the patella bone. The treated difference measures are the Cross Correlation (CC) and the Mutual Information (MI). For MI, the σ values for the proposed methods had to be increased by a factor of 10. The same settings were used for all images. For the CC, we simulate a multi-modal scenario multiplying the target image by a scalar value, and for the MI by inverting the target image. The results are summarized in Fig. 3, and support the findings of improved convergence for further images and difference measures.

3.4. Optical Flow Example

As the last experiment, and more as a side note, we test the performance of the proposed approach for optical flow, which is closely related to the registration problem. It is interesting that in spite of the generality of the scheme as to the ability to include any difference measure, it is also able to compute accurate flow fields. We use the Rubber Whale sequence from the Middlebury Flow data base [6] with known ground truth, and SSD as difference measure. Please note that since we are only interested in evaluating our modification of the difference gradient, here we also employ the diffusion regularization, which is known to be suboptimal for optical flow [1]. The resulting flow has the

following statistics: angular error (AE) mean of 6.181 with standard deviation of 10.860, and the end-point error (EPE) of 0.196 with standard deviation of 0.341. Please compare [6] and Fig. 5 for visuals.

4. Summary and Discussion

The main idea of this paper is to counteract the lack of physical context for gradients of difference measures in deformable registration. This is done by modifying the way in which distances between deformations are measured, by providing the metric tensor with information about the difference measure and the input images. In order to reach this goal, we adapt the concept of Natural gradients from stochastic optimization. The different choices for the required approximation of the Fisher Information matrix lead us to three different resulting approaches. Our experiments demonstrate the preferable properties of the proposed approach. A detailed comparison between the different proposed approaches is subject to further, more detailed tests, and a line of further work.

The use of natural gradients for deformable registration exhibits two advantages. It counteracts the difference gradient bias by improving the accuracy in low-gradient regions, and exhibits a clearly faster convergence compared to the standard approaches. For least-squares difference measures such as SSD, faster convergence can be achieved by using more efficient optimization schemes. Some works, e.g. [9, 20] employ least-squares methods to this end. However,

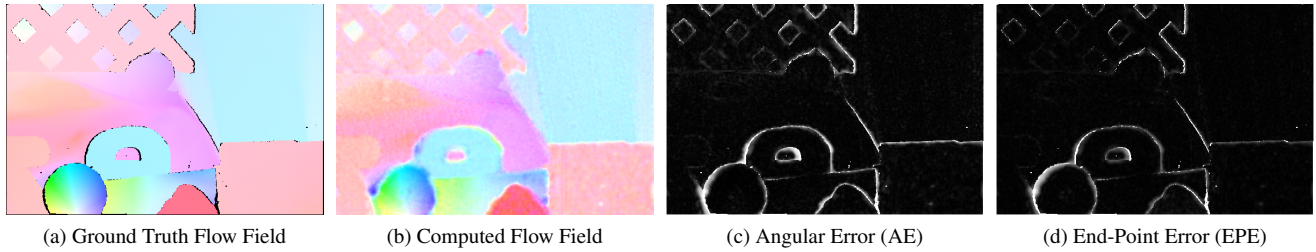


Figure 5: Results for optical flow computation on the Rubber Whale sequence. In spite of generality to application of arbitrary difference measures, the proposed approach is capable of accurate optical flow results.

least-squares methods have the drawback that they cannot be directly used for many difference measures such as CC, CR or MI, which are no least-squares measures. So a further advantage of our approach is that it improves convergence for *arbitrary* difference measures, since it is not based on the least-squares assumption.

The advantages of natural gradients come at a very cheap computational cost, and the actual implementation of the derived methods is surprisingly simple. Two of the three proposed alternatives (Frequentist and Direct metrics) require practically no computation overhead, while the Region metric is slower but also comparably efficient, since it only requires convolution of small regions, and inversion of small systems. The proposed scheme is generally applicable to arbitrary difference measures and its integration into standard variational methods requires only minimal changes in existing implementations.

Acknowledgements We would like to thank Maximilian Baust of CAMP for discussion and comments. We greatly appreciate the feedback and suggestions for improvement of the final version by the anonymous CVPR reviewers.

References

- [1] L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *Int. Journal of Computer Vision*, 2000.
- [2] S.-i. Amari. *Differential-Geometrical Methods in Statistics*. Springer, 1985.
- [3] S.-i. Amari. Neural learning in structured parameter spaces—natural riemannian gradient. *Advances in Neural Information Processing Systems*, 1996.
- [4] S.-i. Amari. Natural gradient works efficiently in learning. *Neural computation*, 1998.
- [5] S.-i. Amari and S. Douglas. Why natural gradient? *IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, 1998.
- [6] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *IEEE Int. Conference on Computer Vision*, 2007.
- [7] L. Bottou. Stochastic learning. *Advanced Lectures on Machine Learning*, 2004.
- [8] C. Chefd’hotel. *Geometric Methods in Computer Vision and Image Processing: Contributions and Applications*. PhD thesis, L’Ecole Normale Supérieure de Cachan, 2005.
- [9] E. Haber and J. Modersitzki. A multilevel method for image registration. *SIAM Journal on Scientific Computing*, 2006.
- [10] G. Hermosillo, C. Chefd’Hotel, and O. Faugeras. Variational methods for multimodal image matching. *Int. Journal of Computer Vision*, 2002.
- [11] N. Le Roux, P. Manzagol, and Y. Bengio. Topmoumoute online natural gradient algorithm. *Neural Information Processing Systems (NIPS)*, 2007.
- [12] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Int. Joint Conference on Artificial Intelligence*, 1981.
- [13] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. on Medical Imaging*, 1997.
- [14] J. Modersitzki. *Numerical methods for image registration*. Oxford University Press, 2004.
- [15] X. Pennec, P. Cachier, and N. Ayache. Understanding the demon’s algorithm: 3d non-rigid registration by gradient descent. *Medical Image Computing and Computer Assisted Intervention*, 1999.
- [16] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *Int. Journal of Imaging Systems and Technology*, 2000.
- [17] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. *Medical Image Computing and Computer Assisted Intervention*, 1998.
- [18] J. Thirion. Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis*, 1998.
- [19] A. Trounev. Diffeomorphisms groups and pattern matching in image analysis. *Int. Journal of Computer Vision*, 1998.
- [20] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 2009.
- [21] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *Int. Journal of Computer Vision*, 1997.