Using Landmarks as a Deformation Prior for Hybrid Image Registration

Marcel Lüthi, Christoph Jud and Thomas Vetter

Department of Computer Science
University of Basel, Switzerland
{marcel.luethi, christoph.jud, thomas.vetter}@unibas.ch

Abstract. Hybrid registration schemes are a powerful alternative to fully automatic registration algorithms. Current methods for hybrid registration either include the landmark information as a hard constraint, which is too rigid and leads to difficult optimization problems, or as a soft-constraint, which introduces a difficult to tune parameter for the landmark accuracy. In this paper we model the deformations as a Gaussian process and regard the landmarks as additional information on the admissible deformations. Using Gaussian process regression, we integrate the landmarks directly into the deformation prior. This leads to a new, probabilistic regularization term that penalizes deformations that do not agree with the modeled landmark uncertainty. It thus provides a middle ground between the two aforementioned approaches, without sharing their disadvantages. Our approach works for a large class of different deformation priors and leads to a known optimization problem in a Reproducing Kernel Hilbert Space.

1 Introduction

The problem of establishing point-to-point correspondence between two images is central to computer vision and medical image analysis. It is usually addressed using image registration methods, which aim to find a mapping such that each point in a target image is mapped to its corresponding point in the reference image. Formally, the registration problem can be defined as follows: Given a reference and target image $I_R, I_T : \Omega \to \mathbb{R}$ find the deformation field $u : \Omega \to \mathbb{R}^d$ from a class of admissible deformations that optimally aligns the warped reference image $I_R(x + u(x))$ with the target $I_T(x)$. This problem is usually formulated as a minimization problem of the functional

$$J[u] := \mathcal{D}_I[I_T, I_R, u] + \lambda\mathcal{R}[u] \tag{1}$$

where $\mathcal{D}_I$ is a similarity measure for the images, $\mathcal{R}$ a regularization term that penalizes non-smooth solutions and λ a regularization parameter. In cases where the images are noisy or corrupted, minimizing (1) may not lead to the desired correspondence and hybrid registration schemes have shown to be a powerful alternative. In addition to the images, these schemes let the user specify a set of landmark points $L_T, L_R \subset \Omega$ for the reference and the target image.

Various different methods for integrating these landmarks are known. They broadly fall into two classes. The first class of methods integrate the landmarks as a hard constraint [9, 5, 3, 2]. The resulting problem is to find a deformation field that minimizes (1), but subject to additional landmark constraints:

$$J'[u] = \mathcal{D}_I[I_T, I_R, u] + \lambda \mathcal{R}[u] \;\; s.t. \; \mathcal{D}_{LM}[L_R, L_T, u] \leq \varepsilon, \quad (2)$$

where $\mathcal{D}_{LM}$ measure the landmark distance. Most such methods require a perfect interpolation of the landmarks (i.e. $\varepsilon = 0$ in (2)) [5, 3, 2]. The resulting optimization problem is a difficult numerical problem. To the best of our knowledge, solutions have only been presented for the case of the thin-plate spline model. The second class of methods do not strictly enforce the landmark constraint, but add it as a penalty into (1) [14, 13, 6, 8]. This leads to the minimization problem of the functional

$$J''[u] = \mathcal{D}_I[I_T, I_R, u] + \eta \mathcal{D}_{LM}[L_T, L_R, u] + \lambda \mathcal{R}[u], \quad (3)$$

where η is an additional parameter. This approach has the advantage that the optimization problem is easier to handle and its integration into existing registration schemes is straight-forward. Moreover, we argue that it is more natural to treat the landmarks as a soft constraint as they are usually only approximately known. The main problem with this formulation is that it treats the landmark and regularization term independently, even though landmarks clearly provide a-priori information about the deformations. The regularization parameters λ and η become mutually dependent, which makes parameter tuning difficult.

In this paper we propose a formulation of hybrid registration that combines both approaches, by integrating the landmark information into the regularization term in (1). The uncertainty of the landmarks positions σ is the only additional parameter that we introduce. Its value is at least approximately known from the experimental setup. If we set $\sigma = 0$, our solution matches the landmark points perfectly, whereas for $\sigma > 0$ an approximate solution is achieved. The main idea is to model the admissible deformations as a (vector-valued) Gaussian process $u \sim \mathcal{GP}(0, k)$. Its covariance function k determines the regularization properties of the deformation field. Landmark registration becomes the problem of Gaussian process regression [10], where the displacement at the landmark points are our observations. Gaussian process regression does not only provide us with a MAP solution for the landmark registration problem, but yields the full posterior distribution $p(u|L_T, L_R)$. This distribution is again a Gaussian process $\mathcal{GP}(\mu_{LM}, k_{LM})$, whose parameters are known in closed form. Its covariance function k_{LM} spans a Reproducing Kernel Hilbert Space (RKHS), whose associated norm $\|\cdot\|_{k_{LM}}$ penalizes functions according to the posterior probability $p(u|L_T, L_R)$. Thus, it is an ideal regularization term for the registration problem (1), leading to a hybrid scheme. The resulting formulation of the registration problem has recently been addressed by Schölkopf et al. [13]. They proposed a general solution strategy that can be applied to any covariance function. In contrast to our work they do not include the landmarks information into the covariance function, but model them as a soft constraint as in (3).

Unlike most previous approaches for hybrid registration, which have been formulated for a fixed deformation model such as the thin-plate splines [5, 3, 9], elastic body splines [17] or B-splines [14, 8], our approach works for any valid (i.e. positive definite) covariance function. In particular it includes the deformation models that arise from Green's functions of regularization operators, which have been proposed for hybrid registration in [2].

We present experiments for a simple toy example and a 2D X-ray image of the hand. Although our current implementation does not allow to register large images, our results clearly illustrate the benefits of our approach. Our experiments show that different covariance function lead to different regularization properties, but the parameter σ that models the landmark accuracy keeps its intuitive meaning for the different models. Furthermore we illustrate that the ability to model the landmark inaccuracies is important to obtain good registration results.

2 Background

The core idea of our method is to use Gaussian process regression to obtain a deformation prior that implicitly incorporates the landmark term. In this section we briefly review Gaussian processes and their application to regression.

2.1 Gaussian Processes

Stochastic processes allow us to define a probability distribution over a function space. Formally, a stochastic process is a collection of random variables $f(x), x \in \Omega$ where Ω is an index set. A Gaussian process is a stochastic process with the property that for any finite number of observations, $x_1, \ldots, x_n \in \Omega$ the values $f(x_1), \ldots, f(x_n)$ are jointly normally distributed [10]. A Gaussian process is completely defined by its mean $\mu : \Omega \rightarrow \mathbb{R}$ and a covariance function $k : \Omega \times \Omega \rightarrow \mathbb{R}$. We write $\mathcal{GP}(\mu, k)$ to specify a Gaussian process. To simplify the discussion, we consider here only Gaussian processes $\mathcal{GP}(0, k)$ with zero mean, the extension to the general case is, however, straight-forward.

By specifying a covariance function $k(x, y)$, we define which functions are likely under the given process. The covariance function specifies for each pair of points x, y their covariance $E[f(x)f(y)]$. Covariance functions are also often referred to as kernels, and we will use the terms interchangeably. Many known covariance functions require that nearby values are strongly correlated, which leads to that they effectively favor smooth functions. Indeed, many covariance functions arise as Greens functions of common regularization operators [16].

Vector-valued Gaussian Processes To be able to use Gaussian process for modeling deformation fields, above concepts need to be generalized to the case in which each random variable $u(x)$ is a d dimensional random vector. The covariance function, becomes a matrix valued function $k(x, y) : \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$, with $k(x, y) = E[u(x)^T u(y)]$. It can be shown that the vector-valued case can be reduced to the scalar case. Thus all known results for real-valued Gaussian processes carry over to this more general setting. We refer to the article by Hein et al. for further details [4].

A useful class of covariance functions for the vector valued case arise from the scalar valued covariance functions. Let $A \in \mathbb{R}^{d \times d}$ be a symmetric, positive definite matrix and l a real valued covariance function. It can be shown that the matrix valued function k with entries k_{ij} defined by

$$k_{ij}(x, y) = A_{ij} l(x, y), \tag{4}$$

is a valid covariance function [7]. The entry A_{ij} determines the correlation between the i-th and j-th output component. In cases where we do not have any a-priori knowledge about their correlation, we can choose $A = \mathcal{I}_d$ as the identity. In this case each dimension is considered as independent.

2.2 Gaussian Process Regression

Assume that we are given an i.i.d. sample $S = \{(x_1, y_1,), \ldots, (x_n, y_n)\} \subset \Omega \times \mathbb{R}^d$ and let $u \in \mathcal{GP}(0, k)$ be a vector valued Gaussian process with $k : \Omega \times \Omega \to \mathbb{R}^{d \times d}$. Gaussian process regression lets us infer the distribution $p(u|S)$. We assume that $y \sim \mathcal{N}(u(x), \sigma^2 \mathcal{I}_d)$. This means, instead of observing the actual values $u(x)$ we observe noisy instances y thereof. The likelihood of the data is given as $p(S|u)) = \prod_{i=1}^{n} \mathcal{N}(u(x_i), \sigma_i)$. Under this assumption the posterior distribution $p(u|S) \propto p(u)p(S|u)$ is known in closed form [10]. It is again a Gaussian process $\mathcal{GP}(\mu_p, k_p)$ and its parameters are

$$\mu_p(x) = K_X(x)^T (K_{XX} + \sigma^2 \mathcal{I})^{-1} Y \tag{5}$$

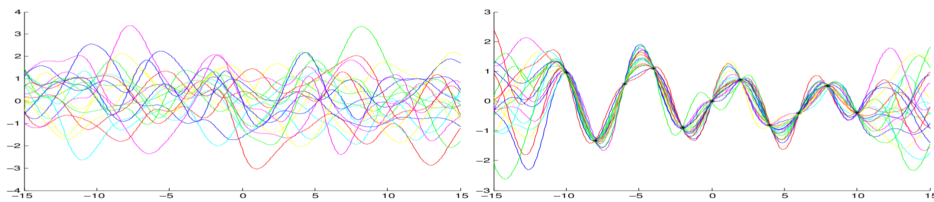
$$k_p(x, x') = k(x, x') - K_X(x)^T (K_{XX} + \sigma^2 \mathcal{I})^{-1} K_X(x'). \tag{6}$$

Here, we defined $K_X(x) = (k(x, x_i))_{i=1}^{n} \in \mathbb{R}^{nd \times d}$, $K_{XX} = (k(x_i, x_j))_{i,j=1}^{n} \in \mathbb{R}^{nd \times nd}$ and $Y = (y_1, \ldots, y_n)^T \in \mathbb{R}^{nd}$. Note that K_X and K_{XX} consist of sub-matrices of size $d \times d$.

Under this posterior distribution, only functions that agree with the given sample S are likely observations. This is illustrated in Figure 1. Figure 1a shows random samples from a one-dimensional prior distribution where we used the Gaussian covariance function, defined by $k(x, x') = \exp(-\|x - x'\|^2)$. Figure 1b shows the corresponding posterior distribution after the sample has been observed.

3 Hybrid registration using a landmark prior

In this section we show how Gaussian process regression yields a solution of the landmark registration problem, and how the resulting posterior process can be used as a regularization term. Furthermore, we outline a generic procedure how the resulting registration functional can be solved.



(a) (b)

Fig. 1: (a) Random samples from a Gaussian process. (b) Random samples from the posterior process, after a number of points have been observed. Only functions that agree with the observations are likely to be observed.

3.1 Landmark Registration

Let $u \sim \mathcal{GP}(0, k)$ be a vector valued Gaussian process with covariance function $k : \Omega \times \Omega \to \mathbb{R}^{d \times d}$ that defines our prior assumptions about the possible deformations. Usually, we choose k, such that the Gausssian process favours smooth deformation fields. Further let $L_R = \{l_R^1, \dots, l_R^n)\}$ and $L_T = \{l_T^1, \dots, l_T^n)\}$ be the given landmarks. These landmarks provide us with known deformation at the landmark points

$$L = \{(l_R^1, l_T^1 - l_R^1), \dots, (l_R^n, l_T^n - l_R^n)\} =: \{(l_R^1, y_1), \dots, (l_R^n, y_n)\},$$

and thus we have a sample set S on which we can apply Gaussian process regression. The likelihood function is given by

$$p(L|u) = \prod_{i=1}^{N} \mathcal{N}(u(x), \sigma^2 \mathcal{I}_d),$$

and corresponds to our assumption that the inaccuracies of the landmarks can be modeled as independent Gaussian noise. We know from Section 2 that the posterior $p(u|L) \propto p(u)p(L|u)$ is under these assumptions again a Gaussian process $p(u|L) \sim \mathcal{GP}(\mu_{LM}, k_{LM})$, whose parameters are known in closed form and are defined by (5). This posterior distribution thus defines a distribution over deformation fields which incorporates the landmark constraints. Its mean deformation μ_{LM} is the MAP solution to the landmark registration problem and provides an optimal trade-off between our a-priori information about the deformation field and the landmark constraints.

3.2 Combined landmark and image registration

Our starting point for combining the landmark prior with image registration is the probabilistic formulation of the registration problem by Christenson et al. [1]. They pointed out that the registration problem (1) can be interpreted as the following MAP estimation problem:

$$\arg\max_u p(u)p(I_T|I_R, u), \qquad (7)$$

where $p(u) \propto \exp(-\mathcal{R}[u])$ is a Gaussian process prior over the admissible deformation fields and $p(I_T|I_R, u) \propto \exp(\lambda^{-1}\mathcal{D}[I_R, I_T, u])$ is the likelihood. Given the displacements from the landmarks L, it is natural to reformulate (7) as

$$\arg\max_u p(u|L)p(I_T|I_R, u) \tag{8}$$

and to choose $p(u|L) \sim \mathcal{GP}(\mu_{LM}, k_{LM})$ using the closed form solution derived in the previous section.

To solve (8) we exploit the well known fact that finding the MAP solution of a Gaussian process model corresponds to solving minimization problem in the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{F}_k$ defined by the covariance function k (see e.g. Wahba [16]). In the following discussion we chose the sum of squared differences as a distance measure:

$$\mathcal{D}[I_T, I_R, u] = \int_\Omega (I_T(x + u(x)) - I_R(x))^2 \, dx. \tag{9}$$

The problem corresponding to (8) becomes

$$\arg\min_{u \in \mathcal{F}_k} \|u\|^2_{k_{LM}} + \lambda^{-1} \int_\Omega (I_T(x) - I_R(x + \mu_{LM}(x) + u(x)))^2 \, dx. \tag{10}$$

where $\|\cdot\|_{k_{LM}}$ denotes the RKHS norm. Treating the registration problem as a minimization problem in a RKHS is the starting point of a recent paper by Schölkopf al. [13]. We briefly sketch the approach and refer to the original paper for further details. The idea is to approximate the integral in (9) by uniformly sampling N points from Ω and solve the discretized problem:

$$u^* = \arg\min_{u \in \mathcal{F}_k} \lambda \|u\|^2_{k_{LM}} + \frac{1}{N} \sum_{i=1}^N [I_T(x_i) - I_R(x_i + \mu_{LM}(x) + u(x_i))]^2 \tag{11}$$

The generalized representer theorem [12] asserts that the optimal solution u^* is given as a finite linear combination of the covariance functions k_{LM}

$$u^*(x) = \sum_{i=1}^N k_{LM}(x, x_i)\alpha_i^*, \tag{12}$$

where $\alpha_i^* \in \mathbb{R}^d$ is a vector of optimal coefficients for each output dimension. The optimal coefficients $(\alpha_1^*, \ldots, \alpha_n^*)$ can be found by plugging (12) into (11) to obtain a finite dimensional minimization problem

$$\arg\min_{\alpha_1,\ldots,\alpha_N} \lambda \sum_{i,j}^N \alpha_i^T k_{LM}(x_i, x_j)\alpha_j + \frac{1}{N} \sum_{i=1}^N [I_T(x_i) - I_R(x_i + \mu_{LM}(x_i) + \sum_{j=1}^N k_{LM}(x_j, x_i)\alpha_j)]^2. \tag{13}$$

This problem can be solved using any standard optimization scheme.

3.3 A note on the implementation

The above approach provides us with a general method to find the MAP solution to (8) for the case when the prior $p(u)$ is a Gaussian process. This is a extremely general formulation and includes many different registration schemes as a special case. This generality comes at the price that each optimization nstep requires the evaluation of the covariance function in a double sum over N points. For most real images N is too large for this to be feasible. Schölkopf et al. [13] proposed to use compactly supported covariance functions. The matrix $K_{XX} = (k_{LM}(x_i, x_j))_{i,j=1}^N$ becomes sparse and needs to be computed only once. The double sums in (13) can be replaced by matrix-vector multiplications. Using a straight-forward implementation of this approach, we can register an 128×128 image in about 15 minutes on a standard PC (single core). For the registration of 3D images this approach remains infeasible. To be able to solve (8) more efficiently, some of the generality might have to be sacrificed and the solution scheme targeted to special covariance functions.

4 Results

In this section we illustrate our method on a toy example and a X-ray image. We choose the following covariance functions: The (cubic) B-spline, defined by

$$k_b(x, x') := \sum_{l \in \mathbb{Z}^2} \beta_\otimes(x - l)\beta_\otimes(x' - l) \tag{14}$$

where $\beta_\otimes$ are tensor product B-splines defined for $x \in \mathbb{R}^2$ as $\beta_\otimes(x) = \beta_3(x_1)\beta_3(x_2)$ and β_3 in turn is the cubic B-spline basis function [15]. Further, the Wu's functions defined by [11]

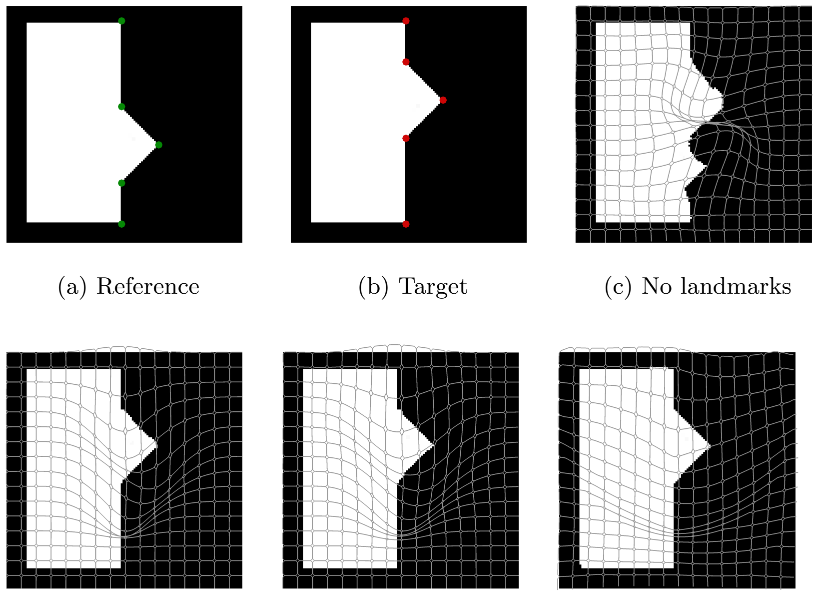
$$k_{0,0}(x, x') := k_{0,0}(r) = \max(0, (1 - r))$$

where $r = \|(x - x')\|^2/s$ and s determines the support of the kernel and finally the Wu's function defined by

$$k_{2,1}(x, x') := k_{2,1}(r) = \max(0, (1 - r))^4(4 + 16r + 12r^2 + 3r^3),$$

To obtain the corresponding matrix valued functions, we multiply the covariance function by the identity matrix $\mathcal{I}_2$ (Cf. Section 2.1). Thus, we effectively treat each output dimension as independent. We keep the regularization parameter λ, as well as the parameters of the covariance functions constant in all the examples.

Figure 2 shows the result for the toy example. A proper registration solution would match all the corners onto each other. Using only a standard smoothness prior the registration method will not be able to find the right correspondences. Figure 2c shows the result for the B-spline covariance function when the landmarks are ignored. By forcing the landmarks point to match exactly (i.e. setting $\sigma = 0$), we can enforce a correct solution (Figure 2, (d)-(f)). We see that the result looks the same in 2d-2f, but the actual deformations (as shown by the grid)



(a) Reference (b) Target (c) No landmarks

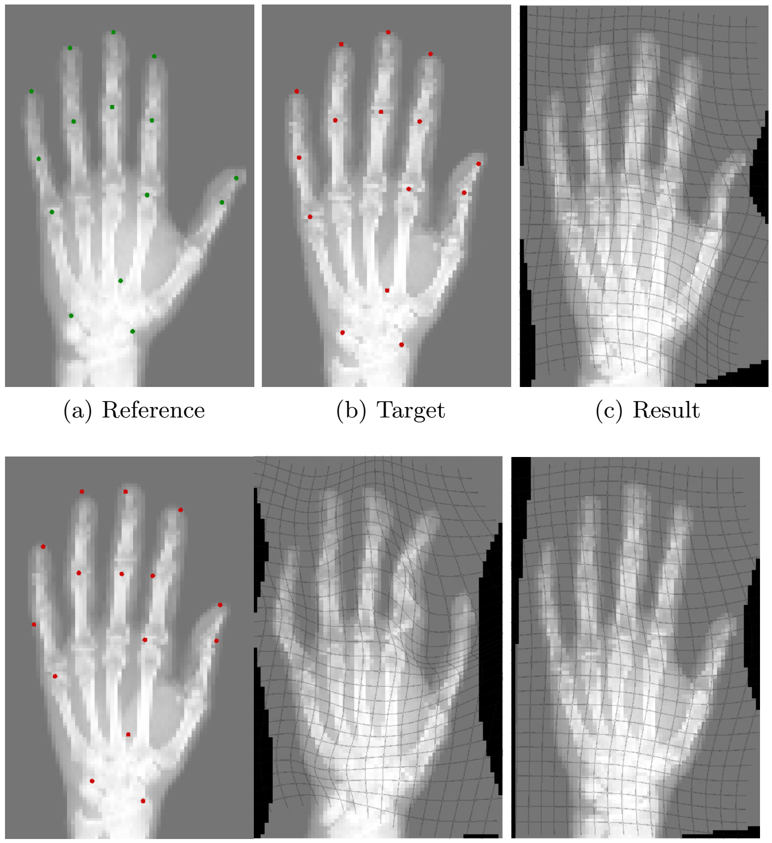
(d) B-spline kernel (e) Wu-kernel $k_{0,0}$ (f) Wu-kernel $k_{2,1}$

Fig. 2: A toy example: The goal is to transform the reference (a) onto the target (b). (c) shows a result using the B-spline kernel without landmarks. Using the landmarks we can enforce the desired matching of the corners (d)-(f). From the deformed grids we see that the deformation fields differ for different covariance functions, even though the registration result looks the same.

strongly depend on the chosen covariance functions. In our second example we illustrate the influence of the landmark accuracy on the solution. Figure 3a and 3b show the target and reference image together with the correct landmarks. Using these landmarks, we obtain the registration result depicted in Figure 3c. We then add Gaussian noise with a standard deviation of 5mm onto the landmarks position (Figure 3d). Forcing the landmarks to perfectly match, by setting $\sigma = 0$mm, leads to the result shown in Figure 3e. A much better solution is obtained if we model the landmark inaccuracy correctly by setting $\sigma = 5$mm (Figure 3f).

5 Conclusion

We have presented a novel formulation of hybrid registration. In contrast to previous approaches, we proposed to integrate the landmark information directly into the deformation prior. This avoids a new trade-off parameter, but still allows us to control how accurately the landmarks should be matched. Furthermore, the regularization term has a natural probabilistic interpretation as the posterior



(a) Reference (b) Target (c) Result

(d) Noisy Landmarks (e) $\sigma = 0$mm (f) $\sigma = 5$mm

Fig. 3: Two X-ray images of hands ((a), (b)) are registered. (c) shows the solution using the Wu kernel $k_{2,1}$. In (d) we add noise to the landmark positions. (e) shows the solution when a perfect match ($\sigma = 0$mm) is enforced. Correctly modeling the noise on the landmarks greatly improves the result (f).

$p(u|L_T, L_R)$ of the deformations u given the landmarks. Our results illustrated that we can both enforce a perfect match of the landmark points, and allow for an approximate matching. Further, we showed that by using different covariance functions, different regularization properties can be obtained, which makes this approach more versatile than previous formulations. The use of compact covariance functions makes the method practical for the registration of 2D images of moderate size. However, it is still not feasible at this point to use it for the registration of 3D images. Devising an efficient scheme for special classes of covariance functions, such that 3D hybrid registration becomes possible is an interesting challenge, which we will address in future work.

Acknowledgements

We thank our colleague, Brian Amberg, for insightful discussions regarding this work. This work has been supported by the CO-ME/NCCR research network of the Swiss National Science Foundation.

References

1. Christensen, G.E., Miller, M.I., Vannier, M.W., Grenander, U.: Individualizing neuro-anatomical atlases using a massively parallel computer. Computer 29(1), 32–38 (1996)
2. Fischer, B., Modersitzki, J.: Combination of automatic non-rigid and landmark based registration: the best of both worlds. Medical imaging pp. 1037–1048 (2003)
3. Haber, E., Heldmann, S., Modersitzki, J.: A Scale-Space approach to landmark constrained image registration. Scale Space and Variational Methods in Computer Vision pp. 612–623 (2009)
4. Hein, M., Bousquet, O.: Kernels, associated structures and generalizations. Max-Planck-Institut fuer biologische Kybernetik, Technical Report (2004)
5. Johnson, H.J., Christensen, G.E.: Consistent landmark and intensity-based image registration. Medical Imaging, IEEE Transactions on 21(5), 450–461 (2002)
6. Lu, H., Cattin, P., Reyes, M.: A hybrid multimodal non-rigid registration of MR images based on diffeomorphic demons. In: Engineering in Medicine and Biology Society, International Conference of the IEEE. pp. 5951–5954 (2010)
7. Micchelli, C.A., Pontil, M.: On learning vector-valued functions. Neural Computation 17(1), 177–204 (2005)
8. Papademetris, X., Jackowski, A.P., Schultz, R.T., Staib, L.H., Duncan, J.S.: Integrated intensity and point-feature nonrigid registration. Medical Image Computing and Computer-Assisted InterventionMICCAI 2004 pp. 763–770 (2004)
9. Papenberg, N., Olesch, J., Lange, T., Schlag, P.M., Fischer, B.: Landmark constrained non-parametric image registration with isotropic tolerances. Bildverarbeitung für die Medizin 2009 pp. 122–126 (2009)
10. Rasmussen, C.E., Williams, C.K.: Gaussian processes for machine learning. Springer (2006)
11. Schaback, R.: Creating surfaces from scattered data using radial basis functions. Mathematical methods for curves and surfaces pp. 477–496 (1995)
12. Schölkopf, B., Herbrich, R., Smola, A.: A generalized representer theorem. In: Computational learning theory. pp. 416–426 (2001)
13. Schölkopf, B., Steinke, F., Blanz, V.: Object correspondence as a machine learning problem. In: ICML '05: Proceedings of the 22nd international conference on Machine learning. pp. 776–783. ACM Press, New York, NY, USA (2005)
14. Sorzano, C.O.S., Thevenaz, P., Unser, M.: Elastic registration of biological images using vector-spline regularization. Biomedical Engineering, IEEE Transactions on 52(4), 652–663 (2005)
15. Unser, M.: Splines: A perfect fit for signal and image processing. Signal Processing Magazine, IEEE 16(6), 22–38 (1999)
16. Wahba, G.: Spline models for observational data. Society for Industrial Mathematics (1990)
17. Wörz, S., Rohr, K.: Hybrid spline-based elastic image registration using analytic solutions of the navier equation. Bildverarbeitung für die Medizin 2007 pp. 151–155 (2007)