Image and Surface Registration

Ke Chen[*], Lok Ming Lui [**], and Jan Modersitzki [***]

Abstract. Image registration is one of the most useful and practical applications of image analysis. Among its manifold tasks are the tracking of changes between data from different time points or motion correction. Moreover, superimposing complementary information across image modalities is needed, as new imaging modalities emerge in the field. This chapter presents a review of the fundamental ideas and models in the field. The goal is to reflect the current state of art of image registration (IR) to motivate the readers to refine these models, and to enable the tackling of new challenges as they arise. After discussing the background (Section 2), we review main components of a variational model: a distance measure or data fidelity term (Section 3), regularization to ensure existence of solutions and constraints to further restrict the wanted transformation (Section 4). We also discuss diffeomorphic approaches which ensure local invertibility. We present the surface registration (SR) modelling in the same framework of variational models in Section 5 where the close relationship between IR and SR is also discussed, while we briefly discuss the numerical methods for IR and SR in Section 6. Finally in Section 7, we touch upon the main ideas in deep learning based approaches for registration.

Keywords: Image registration, surface registration, variational models

https://www.sciencedirect.com/handbook/handbook-of-numerical-analysis/vol/20/suppl/C

---

[*] Department of Mathematical Sciences, University of Liverpool, United Kingdom (`http://www.liv.ac.uk/%7Ecmchenke`)

[**] Department of Mathematics, Chinese University of Hong Kong, China (`https://www.math.cuhk.edu.hk/%7Elmlui/`)

[***] Institute of Mathematics and Image Computing, University of Lübeck, Germany (`https://www.mic.uni-luebeck.de/people/jan-modersitzki.html`)

Table of Contents

1 Introduction

Registration is the process of automatically establishing correspondences between geometrical data, such as images, surfaces or point clouds. Typically, registration aligns a pair, a stack or a sequence of such data. Of particular interest is the registration of two images. Often, one of these images is considered to

be fixed, stationary or a reference and is denoted by R and the other one is consider a moving, floating or template image denoted by T. The necessity of registration can be found in diverse fields of sciences and engineering, including remote sensing, computer graphics, computer vision and medical imaging; see, e.g. [9,78,30,60,40,63,93,74] and references therein. For example, in medical imaging, finding an accurate correspondence between biomedical images is crucial for statistical shape analysis of the anatomical structures. In computer graphics, a surface registration satisfying user-defined correspondences of landmark features is needed for the constrained texture mapping. In computer vision, image registration offers a crucial tool in combining, merging or fusing multiple data sources from multi-modality images such as infrared images and satellite images. Despite its importance, the registration problem inhere many mathematical challenges: for various applications appropriate models are missing and the modelling is unclear, variational models are non-convex and solutions non-unique (ill-posedness); see [84,25,11] among others. Very often, a suitable choice of the regularization and a proper registration model is non-trivial to make. Robustness of a model remains a major challenge.

Different registration models have been recently developed. In terms of the type of information to be matched, existing approaches can mainly be divided into three categories, namely, feature-based registration, intensity-based registration and hybrid registration using both features and intensity information. Examples of such features are localized points a.k.a. landmarks, structures such as vessels or segmentations of organs.

Particularly landmark-based schemes are popular in computer sciences. One of the main advantages of landmarks-based method is the straightforward incorporation of user-interaction during the registration. This provides an intuitive and user-validated assumption to achieve a meaningful deformation approximation. From the numerical point of view, landmarks-based method is more computationally efficient when modeling large deformations, since more (initial) information about the deformation are provided through feature correspondences. The downside is that landmarks are generally difficult to determine, particularly in 3D medical data. Inaccurate, incorrect or non-corresponding landmarks may degrade the registration results considerably. Moreover, for many applications the number of required landmarks is unknown but registration results may depend heavily on this number. Intensity-based registration aims to match corresponding data without reducing the data to features such as landmarks. Registration is usually obtained by matching intensity functions, such as the image intensity for image registration or the surface curvature for surface geometric registration. The main advantage of the intensity-based registration is that more image information is taken into account and the derivation of features is not required. This is quite desirable in delivering an automatic algorithm. However, the lack of human supervision may cause inaccuracy in the registration result especially in cases where landmarks can be confidently identified with known knowledge. Hybrid registration that combines landmarks-based and intensity-based methods has gained increased attention. Hybrid approaches use

both the landmarks and intensity information to guide the registration. This type of approaches can provide excellent initial guesses; see, e.g. [28,32] for extended discussion.

A generic mathematical formulation of image registration (IR) takes a variational form. Here, a joint energy is to be minimized on a set of suitable transformations. This energy typically consists of an application conform image distance or data fidelity, a regularization, and potentially additional penalty terms. The data fidelity term measures data mismatching, such as the landmark mismatching error or intensity mismatching error. The regularization term usually aims to enhance the smoothness of the registration map and may occasionally require an additional term to impose a geometric constraint on the map invertibility. An extra penalty might be added to emphasize particular aspects such as volume preservation or point to point correspondences. This term may either be phrased as a penalty or a hard constraint. We note, that there also exists a connection to a Bayesian formulation of the registration problem. There, the probability of a transformation given the two images is maximized. However, a thorough discussion is beyond the scope of this paper.

The regularization puts a bias on the set of transformation. Therefore, for different applications, different classes of regularizers have been proposed. For example, one of the common classes of mappings is the space of diffeomorphisms. Despite its computational costs, diffeomorphic registration has become more popular in recent years, especially in biomedical image registration. For many biomedical applications, a one-to-one transformation is a reasonable. Diffeomorphic mappings provide this feature but may result in a very smooth transformation fields. While this can be beneficial for some applications, it might be too restrictive for others. An example is the registration of lung images, as the lung slides along the rib cage. Other options to ensure bijectivity is to include a penalty or constraint on the Jacobian of the transformation. Other commonly used classes of transformations include rigid or affine transformations, conformal, quasiconformal mappings as well as isometric mappings.

The ingredients of the variational model will be reviewed and discussed in this chapter, mostly for a pair of mono-modality images. However the framework can be also be used for the registration of multi-modal images. Here, the fidelity term has to be replaced appropriately. In fact, fusion of images with different modalities has been challenged the medical imaging field very rapidly due to the presence of highly accessible patients' information in recent years. For example, cross platform non-rigid registration of CT with MRI images has found a significant role in different clinical application. There are many complementary modalities, such as MR images vs CT images, ultra-sound images vs PET images, CT images vs optical images, infrared images vs digital images and so on. In some instances labelling of anatomical features by medical experts are also involved to further improve the robustness, accuracy and authenticity of the registration. Being motivated by these, various multi-modality image registration models have been proposed. From a mathematical perspective, it is the same variational framework as for mono-modal images, simply with a different

the data fidelity or distance measure. Note, however, that multi-modal image registration is by far more non-convex than mono-modal registration. In practical applications issues such as pre-alignment, choice of parameters, local minima and robustness become much more severe.

The topic of surface registration (SR) is closely connected to IR. Though each problem class has own distinct features, there are common modeling ideas. In this chapter, we discuss both the similarities, commonalities as well as the differences. Some methods from one class can lend themselves to the other class; even mixed models are possible.

Both the theoretical and numerical analysis of IR and SR are very challenging because registration problems are typically non-convex, the joint energy may has arbitrary many global minimizers. Numerical approaches can easily end up in local minimizers and a proper initial guess can play a significant role. For example, the curvature regularizer discussed in Sec. 4 has an infinite dimensional kernel of harmonic functions. Moreover, not all models have been supplied with an existence theory. As a result, discretized functionals and related linear systems may not be positive definite and registration thus provides a rich source of challenging problems to design efficient and converging algorithms for. Note that current medical CT images can result in 512^3 voxels and a registration ends up in a non-convex optimization problem with about $0.5 \cdot 10^9$ unknowns; the results preferably in real time.

Finally we touch upon the artificial intelligence aspect of IR. In the era of big data, various learning based approaches are being developed. Following the success of deep learning techniques in various imaging tasks, many recent studies have been carried out to apply deep learning techniques to improve various registration models. In particular, deep learning techniques have been applied to build prediction models of spatial transformations for achieving image registration under supervised learning framework. Landmark features can also be learnt for guiding the registration. Of course, the quality of the training data plays an important role for the accuracy of the obtained registration result. Exciting aproaches also investigate to speedup optimization either by learning gradients or transformation manifolds.

In this chapter, our aim is to give a survey of different registration models and to motivate new research works in this challenging and yet exciting field. Various variational models, statistical models and learning based models will be explored.

2   Mathematical background

The goal of image registration is to find a suitable map y that maps a template T to a reference image R, such that $T(y)$ and R are aligned, that is corresponding points are placed at the same position. We also aim to provide a mathematical framework that is capable to cover as many image registration approaches as possible.This section will introduce this framework, modelling ideas and notation but leaving details to later sections.

2.1   Continuous and discrete images

From reading this volume of the Handbook, the reader should be clear that mathematical imaging can offer powerful modeling tools (namely the variational framework) mainly because we can assume that the input images T, R are continuous. Then the full power of functional analysis can be applied to imaging and it is natural to define and discuss geometry (such as gradients, curvature, H_1 norm) of T, R. Also, a continuous interpretation of image data is much more convenient if geometrical transformations are to be applied.

In the continuous interpretation that is used in this paper, a generic image is defined as a mapping $I : \Omega \to G$. Here, $\Omega \subset \mathbb{R}^d$ denotes the image domain and d is the image dimensionality, where typically Ω is a one-dimensional interval with $d = 1$, and a square with $d = 2$ or a cube with $d = 3$. The image range is denoted by G and may indicate binary images $G = \{0, 1\}$, grey scale images $G = \{0, \dots, 255\}$, real images $G = \mathbb{R}$, or multispectral images such as color images, mass spectroscopy, or tensors. In this Chapter, we focus on real images mainly with $d = 2$. However, the discussed models extend to other images.

In applications, the image data I is typically given as a d-array of size $m_1 \times \cdots \times m_d$, which may be naturally viewed as sampled from a continuous image function I. Interpolation or approximation techniques must be used to embed the data into a function space; see, e.g., [62,16]. In this sense, the notation $I(y)$ implies that an interpolation must be implemented (unless y takes integer values only). This aspect is unique in IR and SR, different from other imaging problems. However, this Chapter emphasizes the 'existence' of a continuous image I (via interpretation) that is assumed available for discrete images.

If the continuous image I needs to be visualized, the domain Ω is partitioned into a number n of cells with cell-centres $x_j = (x_j^1, \dots, x_j^d)$, and the discrete image $[I(x_j), j = 1, \dots, n]$ is displayed. In imaging, a pixelized interpretation of image data is often used for processing; see Fig. 1 for an illustration.

Finally we remark that another popularly used notation for the deformed template $T(y)$ is $T \circ y$, where the transform y may also be written as $y = x + u(x)$ later when we apply regularization to the deformation field u.

2.2   A mathematical framework for image registration

As common for ill-posed problems, we use a variational framework that aims to minimize a joint energy functional constituting a data-fidelity term or an image distance measure and a regularizer. Details on distance measures $\mathcal{D}$ and the regularizer $\mathcal{R}$ are provided in subsequent sections.

The objective is to determine the wanted transformation $y : \mathbb{R}^d \to \mathbb{R}^d$ as a minimizer of a joint energy $\mathcal{J}$ over a feasible set of transformation $\mathcal{A}$, i.e.

$$\mathcal{J} : \mathcal{A} \to \mathbb{R}, \quad \mathcal{J}(y) := \mathcal{D}(T \circ y, R) + \mathcal{R}(y), \tag{1}$$

where the data fidelity term $\mathcal{D}$, the regularization term $\mathcal{R}$, and the feasible set $\mathcal{A}$ are discussed below.

(a) image   (b) with pixel   (c) grid

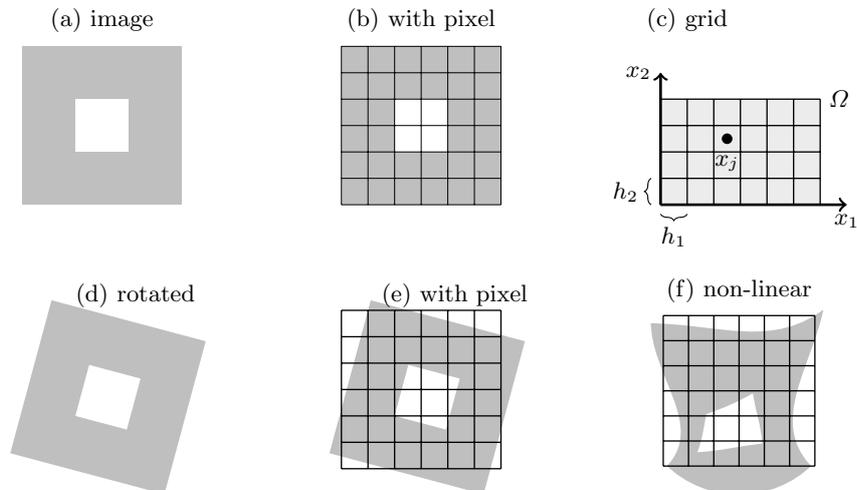(d) rotated   (e) with pixel   (f) non-linear

Fig. 1. An image (a) overlaid with a pixel grid (b), a rotated version of the image (d) overlaid with a pixel grid (e); a non-linearly deformed version of the image overlaid with a pixel grid (f). Part (c) displays a partitioned image domain Ω with 6-by-4 cells with cell-centres $x_j = (x_j^1, x_j^2)$, $j = 1, \ldots, n = 24$.

We remark that, though the approach applies to both, mono- and multi-modal images, we focus on mono-modality for ease of presentation. We also remark that is often convenient to consider the deformation u rather than the transformation y, where the deformation or displacement denotes the change, $y(x) = x + u(x)$. Note, that in contrast to the differences of transformation and deformation the notations transformed image and deformed image are used synonymously in the literature.

3   Distance measures

An important piece of the puzzle in model (1) is the similarity or dissimilarity $\mathcal{D}$ of two given images. There are essentially two approaches. The first approach is to embed the data into a function space or a space of densities and to define similarity via norms or metrics in these spaces. The second approach is based on a projection of data onto a feature space and to base the similarity on feature similarity.

3.1   Volumetric differences

Probably the most intuitive volumetric distance measure is the energy of the difference image a.k.a. the sum of squared differences (SSD),

$$\mathcal{D}^{\mathrm{SSD}}(T, R) := \|T - R\|_{L_2}^2 = \int_\Omega (T(x) - R(x))^2 \, dx$$

for a pair of mono-modality images. The underlying idea is that if image intensities match, i.e. $T(y(x)) \approx R(x)$, then the images correspond. Obviously, the difference can be replaced by any other meaningful residual function $r(T,R)$. Also, the L_2-norm $|r|^2$ may be replaced by another metric $\psi(r)$, and the integration may also include a weighted measure $d\omega$. We mention two particular options to replace the L_2-norm by an L_1-norm and a differential Huber-norm [42] respectively

$$|r|_{h,\varepsilon} = \sqrt{|r|^2 + \varepsilon^2} \quad \text{or} \quad |r|_{H,\varepsilon} = \begin{cases} 0.5|r|^2, & |r| \leq \varepsilon, \\ \varepsilon(|r| - 0.5\varepsilon), & \text{otherwise.} \end{cases} \tag{2}$$

The least-squares distance (LSD) approach of [38] relates the template to a projected version of the reference image. More precisely, using a transfer function $\gamma : \mathbb{R} \to \mathbb{R}$ similarly to a gamma-correction, the residual is phrased as $r(T, \gamma \circ R)$. Determining the optimal colour map γ might then be part of the overall optimization.

A weight function $\omega : \mathbb{R}^d \to \mathbb{R}$ might be use to either emphasize or deemphasize certain areas. For example, if $\Sigma \subset \Omega$ denotes a segmentation of a critical organ, then a weight $\omega := \chi_\Sigma$ would ignore points that are not in Σ. Here, χ_Σ denotes the characteristic function,

$$\chi_\Sigma = \begin{cases} 1, x \in \Sigma, \\ 0, x \notin \Sigma. \end{cases}$$

Note that ω is not necessarily restricted to $\{0,1\}$ and may also be interpreted as a certainty or confidence. The weight may also depend on image intensities. For example, $\omega(x; T, R) := \exp(-(T(x) - R(x))^2)$ is suggested in [24,23]. Here, $\omega \approx 1$ in areas in which $T(x) \approx R(x)$, whereas $\omega \approx 0$ where intensity differences are huge. This weight therefore also yields an implicit soft segmentation into corresponding and non corresponding structures. This concept can be beneficial if one images displays structures that are not contained in the other, e.g. tumours or air bubbles in the intestines; see also [15] for similar strategies.

The above remarks also outline limitations of SSD based distances. For multi-modal images, i.e. images that are acquired with different sensors, it may holds that $T(x) \neq R(x)$ even if the images are perfectly aligned. A nasty example is illustrated in Fig. 2, mimicking the signal "coffee" sensed by nose (smell, smooth reference, gray) and by eye (vision, characteristic function for of the cup, black). Since the intensity of the smell does not relate to the size of the cup, a minimization of $\|T \circ y - R\|_{L_2}$ is meaningless; see Fig. 2 for particular solutions. Several approaches for multi-modal distance measures have been suggested. We note that to our best knowledge, none of these approaches is able to solve the "coffee problem".

The fusion of images with different modalities is a very important application field in image registration. For example, cross platform non-rigid registration of CT with MRI images has found a significant role in different clinical applications. Fig. 3 displays a typical example where the CT image shows structural
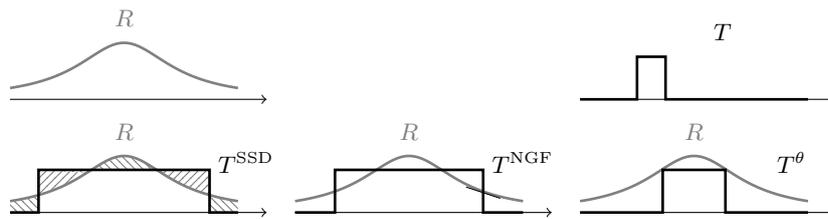
Fig. 2. "Coffee-Problem:" Multi-modal 1D registration problem aligning the smell of a cup of coffee (top row left: smooth reference R, gray) with the visual impression (top row right: characteristic function, template T, solid black). Bottom row shows solutions of L_2 minimization (left), correlation of maximum slope (center), and thresholded reference; note that the scale of smell and vision may not relate.

Fig. 3. Illustration of MR (left, template), CT data (right, reference), and the registered MR (middle, $T(y)$). The result is from the NGF model [80] with $\alpha = 1/4$ and $N = 256$.

(bone) and anatomical information and the MR image shows details on soft tissues. Registration or fusion of the two modalities thus locates complementary information in a normalized geometry.
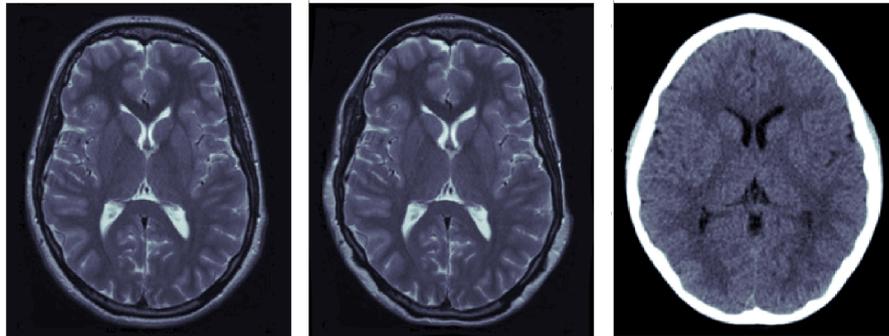
In [35] it is suggested to use the co-linearity of normalized gradient fields (NGF) of the images as a residual. More precisely, $r(x) := 1 - (g(T; x)^\top g(R, x))^2$, where $g(I; x) := \nabla I(x)/|\nabla I(x)|_{h,\varepsilon}$. The motivation is, that even if intensities of T and R do not match, intensity changes occur at corresponding positions. This assumption holds for many applications where structural changes appear at corresponding spatial positions such as the example in Fig. 3. However, it does not hold for the "coffee" example in Fig. 2.

The smoothing of the gradient fields is crucial. In the most likely advent of noise, the normalized gradient field is uninformative. The parameter ε is therefore introduced to discriminate between noise $|\nabla I(x)| \leq \varepsilon$ and signal change i.e. $|\nabla I(x)| > \varepsilon$. We note that a proper embedding of the image data into a smooth function space as outlined in Sec. 2.1 can improve the performance significantly. Variants of NGF include to use individual noise parameters ε_T and ε_R. Additionally, one can work with the L_2 inner product for embedded gradients $\nabla_\varepsilon I := (\nabla I^\top, \varepsilon)^\top \in \mathbb{R}^{d+1}$. Exploration of the different variants is a topic of current research. For example, a recent improvement over NGF can be found in [80] where a different measure of NGF is proposed.

Another commonly used similarity measure for multimodal images is based on the mutual information of the data [22,82]. Further approaches for multimodal image registration are based on so-called Wasserstein metrics [66] or normalized cross correlation [73,14].

In some applications such as histological serial sectioning, a sequence of $r > 2$ images I_j, $j = 1, \ldots, r$ needs not be registered. An obvious extension of the above framework is based on sequential concatenation of standard registration problems for two consecutive images I_{j-1} and I_j,

$$\mathcal{J}(y_1, \ldots, y_r) = \sum_{j=2}^{r} \left\{ \mathcal{D}(I_{j-1} \circ y_{j-1}, I_j \circ y_j) + \mathcal{R}(y_j) \right\}, \quad y_1(x) = x.$$

This process can be very time consuming as the transformations y_j depend on the whole image sequence. Schatten-q-norms based approaches provide a promising alternative as it naturally combines all images of the sequence. Here, a Schatten-q-(quasi) norm [83] of a data matrix A is defined as the q-norm of the vector containing the singular values of A: with a singular value decomposition $A = U\text{diag}(\sigma)V^\top$ it holds $\|A\|_{S,q} := \|\sigma\|_q$. For more details and choices of the data matrix we refer to[1,64,79,5].

3.2   Feature based differences

A simple example for feature based similarity is the location of markers, landmarks or keypoints. Markers are prominent spatial locations that have been attached to an object before imaging. Landmarks are usually outstanding points

such as tips of fingers or eye centres that can be identified after imaging. Independent of whether these points have been identified a priori or a posteriori, the idea then is to determine the image transformation by matching the features. The transformation can be computed from an interpolation or approximation approach. To be more precise, let $(t_\ell)_{\ell=1}^L$ and $(r_\ell)_{\ell=1}^L$ denote two lists of corresponding landmarks, $t_\ell, r_\ell \in \mathbb{R}^d$, for the template and reference image, respectively. The goal is to determine the transformation y such that ideally $y(r_\ell) = t_\ell$ for $\ell = 1, \ldots, L$.

In case of a parametric approach, i.e. $y \in \mathcal{A}$ is parametrizable (by parameters $\{\alpha_j\}$), a least squares approach for the parameter α_j is typically used. Here, $\mathcal{A}$ denotes the set of feasible transformation, for example rotations as in Fig. 1. More precisely, the optimal parameters are computed by minimizing the landmark distance $\mathcal{D}^{\mathrm{LM}}$ over the set $\mathcal{A}$ i.e.

$$\min_{\alpha_1,\ldots,\alpha_N} \mathcal{D}^{\mathrm{LM}}(y) := \sum \|y(r_\ell) - t_\ell\|, \quad y \in \mathcal{A}. \tag{3}$$

Obviously, there is a huge degree of freedom in tuning the residuals. Mahalanobis distances [59] can be used if the location error is not isotropically distributed, weights can be incorporated to emphasize particular correspondences or to address further uncertainties. For a linear space $\mathcal{A}$, a solution can by obtained by solving the linear system $w = Q(r)^\dagger t$, where r and t denote the collection of landmarks, $Q(r)$ is the Vandermonde-matrix and † denotes the pseudo-inverse; see [62] for details.

Other approaches are to introduce a smoothness measure or regularizer $\mathcal{R}$ and to optimize with respect to the data fidelity, the regularization or a weighted compromise. For example, the thin-plate bending energy

$$\mathcal{R}^{\mathrm{TPS}}(y) := \sum_j \int_\Omega (\partial_{1,1}y^j)^2 + 2(\partial_{1,2}y^j)^2 + (\partial_{2,2}y^j)^2 \, dx \tag{4}$$

is used in thin-plate-spline registration [71]. Variants of this techniques are based on different optimization approaches for the transformation:

1. $\mathcal{R}(y) \overset{!}{=} \min$ subject to $\mathcal{D}(y) \leq \sigma$,
2. $\mathcal{D}(y) \overset{!}{=} \min$ subject to $\mathcal{R}(y) \leq \sigma$,
3. $(1 - \sigma)\mathcal{D}(y) + \sigma\mathcal{R}(y) \overset{!}{=} \min$,

where σ denotes a noise level.

For a variety of smoothness operators $\mathcal{R}$, solutions can be computed explicitly via radial basis functions or fundamental solutions ρ. For example, for $\mathcal{R}^{\mathrm{TPS}}$ as in (4) the radial basis function and thus the transformation of the first approach are given by

$$\rho(r) = \begin{cases} r^2 \log r, & d = 2 \\ r, & d = 1 \end{cases}, \quad y^j(x) = \sum_{k=1}^L c_k^j \rho(\|x - r_j\|) + w_0^j + \sum_{k=1}^d w_k^j x_k^j,$$

where the coefficients can be computed by solving the KKT systems

$$\begin{pmatrix} A & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} c^j \\ w^j \end{pmatrix} = \begin{pmatrix} t^j \\ 0 \end{pmatrix}, \quad j = 1, \ldots, d,$$

with $A = [\rho(\|r_j - r_k\|)]_{j,k} \in \mathbb{R}^{L,L}$, $B = [1, r_k^1, \ldots, r_k^d]_k \in \mathbb{R}^{L,d+1}$ and $t^j = (t_1^j, \ldots, t_L^j)^\top \in \mathbb{R}^L$; see [62] for details and discussions.

A major advantage of these approaches is that explicit solutions are available. If the number of landmarks is not too large (say less then 10,000), the above KKT system can generally be solved with reasonable effort and accuracy. Another advantage of landmark-based method is the straightforward incorporation of expertise knowledge during the registration process. This provides an intuitive and user-validated assumption to achieve a meaningful deformation approximation.

However, this approach also has a number of disadvantages: the impact of the number of landmarks and the distribution of landmarks is difficult to control. In practice, the determination of landmarks is cumbersome, difficult and erroneous. Also, when both the number of landmarks and the deformation of landmarks are large, parametric approaches to obtain landmark-matching registration usually fail to obtain a bijective deformation and overlaps can usually be observed in the obtained registration (unless extra constraint is imposed at the cost of not matching the landmarks accurately). To overcome these issues, various non-parametric approaches based on solving optimization problems have been proposed, such as the Large Deformation Diffeomorphic Metric Matching (LDDMM) method [45,4] and the Quasi-conformal landmark registration (QCLR) method [56,48].

4 Regularization

As already outlined, image registration is an ill-posed problem and hence regularization becomes inevitable. In the registration context, the main goal of regularization is to ensure the existence of solutions. Uniqueness is a critical topic. For example, if one wants to register a square to a square, rotations of multiples of 90 degrees results in four meaningful global solutions. From an application point of view it can by questionable to declare one of these solutions to be best even if it makes the mathematical problem well-posed. The problem gets more exciting if one considers a circle rather than a square. We remark, that in some approaches the ambiguities are removed by adding boundary conditions, which may not comply the application demands. A smaller issue is related to the fact that many regularizers are based on derivatives and are thus blind with respect to translations. In practice, this is rarely a problem, since either the distance measure adds information, or alternative solutions may also be meaningful.

We remark that a regularizer is often imposed on the deformation rather than the underlying transform, expressed as a function of the increment, $\mathcal{R}(y - y_{\text{ref}})$, where the reference transformation y_{ref} may be obtained from a pre-registration,

an educated guess, or simply by setting $y_{\mathrm{ref}}(x) = x$. Using this construction, it is also common to use zero Dirichlet boundary conditions for the update $u := y - y_{\mathrm{ref}}$ (though Euler-Lagrange equations tend to impose Neumann's type boundary conditions). The boundary conditions result from the variational formulation and should not be phrased artificially.

4.1 Regularization by ansatz-spaces, parametric registration

In many applications the transformation is restricted to a certain set $\mathcal{A}$, such that the wanted transformation can be expressed as a finite combination of basis functions and parameters $w \in \mathbb{R}^p$. We call this class parametric registration schemes. Examples are the set of rigid transformations, the space of affine transformation or the space of spline transformations [71],

$$
\begin{aligned}
y^{\mathrm{rigid}}(x, w) &= \mathrm{rotation}(w) \cdot x + \mathrm{translation}(w), \\
y^{\mathrm{affine}}(x, w) &= \mathrm{affine\text{-}map}(w) \cdot x + \mathrm{translation}(w), \\
y_k^{\mathrm{spline}}(x, w) &= \sum w_j^k \cdot \mathrm{spline}_j(x), \quad k = 1, \ldots, d.
\end{aligned}
$$

The registration problem is to find the optimal $w \in \mathbb{R}^p$, where p denotes the degrees of freedom. In the 2D case for example, one may consider

$$
\begin{aligned}
y^{\mathrm{rigid}}(x, w) &= \begin{pmatrix} \cos w_1 & -\sin w_1 \\ \sin w_1 & \cos w_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} w_2 \\ w_3 \end{pmatrix}, \, w \in \mathbb{R}^3 \quad \mathrm{or} \\
y^{\mathrm{affine}}(x, w) &= \begin{pmatrix} w_1 & w_2 \\ w_4 & w_5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} w_3 \\ w_6 \end{pmatrix}, \quad\quad w \in \mathbb{R}^6.
\end{aligned}
$$

In terms of basis functions, for the latter case, we can define them precisely: $\phi_1 = x_1 e_1$, $\phi_2 = x_2 e_1$, $\phi_3 = e_1$, $\phi_4 = x_1 e_2$, $\phi_5 = x_2 e_2$, $\phi_6 = e_2$. Then we have the equivalent form

$$
y^{\mathrm{affine}}(x, w) = \sum_{j=1}^{6} w_j \phi_j(x).
$$

In many publications, regularization for the parametric case is often neglected, particularly for low degrees of freedom. Note, however, that registration is an ill-posed problem even in the parametric case and thus should be regularized; see also the above square or disk example. Particularly for the affine linear transformation $y^{\mathrm{affine}}(x, w) = (I + A)x + b$ and the diffusion regularizer leads to $\mathcal{R}^{\mathrm{diff}}(y^{\mathrm{affine}}) = \|A\|_{\mathrm{Fro}}^2$; see also [63,19]

4.2 Quadratic regularizer

A huge class of regularizers is based on quadratic forms on the displacement u,

$$
\mathcal{R}(y) = \int_{\Omega} (\mathcal{B}u)^{\top} (\mathcal{B}u) \, dx,
$$

where $y(x) = x + u(x)$, $\mathcal{B}$ may be chosen differently but the Euler-Lagrange equations will contain the operator $\mathcal{B}^* \mathcal{B}$.

Diffusion regularizer The classical optical flow regularizer [41]

$$\mathcal{R}^{\mathrm{diff}}(y) := \sum_j \int_\Omega \|\nabla u_j\|^2 \ dx, \quad \mathcal{B}^*\mathcal{B} = -\Delta,$$

is used in registration as it enables fast algorithms due to its decoupled nature; see e.g. [27,10,36,20]. However, the decoupling may also lead to unnatural and unpleasing transformations.

Elastic regularizer A physically motivated extension of the diffusion regularizer is the elastic potential

$$\mathcal{R}^{\mathrm{elas}}(y) := \int_\Omega \sum_{j,k} \frac{\mu}{4} (\partial_j u_k + \partial_k u_j)^2 + \frac{\lambda}{2} (\mathrm{div}\, u)^2 \ dx, \ \mathcal{B}^*\mathcal{B} = -(\mu\Delta + (\lambda+\mu)\nabla\, \mathrm{div}),$$

with Lamé-constants λ, μ; see [29,7,26] and [62] for details on the elastic potential. Here, image features are considered to act like elastic bodies. We note that the results are often very pleasing, if displacements u are small, $|\partial_j u_k| \ll 1$.

Curvature regularizer For some applications, smooth transformations might be desirable and a second order regularizer can thus be beneficial. The linear curvature regularizer

$$\mathcal{R}^{\mathrm{curv}}(y) := \sum_j \int_\Omega \Delta u_j \ dx, \quad \mathcal{B}^*\mathcal{B} = \Delta^2,$$

has been suggested in [28]; for implementations see also [39]. This regularizer is of particular interest for registrations where volumetric distance measures and landmark conditions are combined,

$$\min \mathcal{J}(y) \quad \text{subject to} \quad \mathcal{D}^{\mathrm{LM}}(y) = 0;$$

see [28] for details.

4.3   Non-quadratic regularizer

A disadvantage of the previous regularizers is that the regularization energy is always finite, even for unwanted transformation. Thus, an unwanted and non-regular transformation might be a minimizer if one reduce the distance measure sufficiently by adjusting the parameter. We therefore also discuss nonlinear, non-quadratic and non-convex regularizers.

Mean curvature regularizer An extension of the above curvature regularizer is to use the mean curvature as regularizer

$$\mathcal{R}^{\mathrm{mean\ curv}}(y) := \sum_j \int_\Omega \kappa(u_j)^2 \ dx, \quad \kappa(u_j) := \nabla \cdot \frac{\nabla u_j}{|\nabla u_j|}, \tag{5}$$

see [21]. The extended version may even be coupled by replacing κ with $\hat{\kappa}(u_j) = \nabla \cdot (\nabla u_j / |\nabla u|)$ in (5); see [92] for details.

Gaussian curvature regularizer Gaussian curvature was investigated in [43]. For 2D, κ in (5) is replaced by

$$\kappa_G(u_j) = \frac{u_{j;1,2}u_{j;2,1} - u_{j;1,1}u_{j;2,2}}{(u_{j;1}^2 + u_{j;2}^2 + 1)^2}.$$

It is fair to remark that both curvature models are non-trivial to solve efficiently; a related model could be built based on the elastica regulariser but again efficient solution is a topic of current research [46].

Fractional derivative based regularizer Fractional derivatives are increasingly used in imaging to take advantage of non-local behaviour. We briefly review [91], suggesting

$$\mathcal{R}^{\text{frac}}(y) := \sum_j \int_\Omega \|\nabla^\alpha u_j\|_2^2 \, dx, \quad \nabla^\alpha u_j = (\partial_1^\alpha u_j, \dots \partial_d^\alpha u_j)^\top, \tag{6}$$

$\partial_k^\alpha u_j$ denoting the fractional α-order derivative [68], $1 \le \alpha \le 2$ and the particular choice $\alpha = 1.6$. The fractional α-order derivative be based on the Riemann-Liouville, Grünwald-Letnikov or Caputo definition. For homogeneous zero boundaries condition and suitable smoothness assumption on u, all three definitions are equivalent.

We remak in passing that fractional derivatives, though non-local, generate structured matrices (Toeplitz form) so existing fast solvers can be utilized [6].

Hyperelastic regularizer We finally highlight the hyper elastic regularizer as proposed in [25,11], to overcome the limitations of a quadratic regularizer and to also allow for large deformation:

$$\mathcal{R}^{\text{hyper}}(y) := \int_\Omega \alpha_l \text{length}(y) + \alpha_s \text{surface}(y) + \alpha_v \text{volume}(y) \, dx, \tag{7}$$

where $C := \text{cof}\nabla y$ denotes the cofactor matrix and $d := \det(\nabla y)$ the determinant of the matrix $F := \nabla y$, respectively. Here,

$$\text{length}(y) = \psi_l(\nabla y), \qquad \psi_l(F) = \|F - \text{Id}\|_2^2,$$
$$\text{surface}(y) = \psi_c(\text{cof}\nabla y), \qquad \psi_c(C) = \max\{\|C\|_2^2 - 3, 0\}^2,$$
$$\text{volume}(y) = \psi_d(\det(\nabla y)); \qquad \psi_d(d) = ((d-1)^2/d)^2.$$

For details, see [11]. In particular the penalty on the determinant ensures finite energies only for transformation with $\det \nabla y > 0$ and therefore enforces one-to-one transformations. Note that the penalty $\mathcal{C}$ also introduces a bias towards the identity transformation.

4.4   Registration penalties and constraints

The set of feasible transformation $\mathcal{A}$ can be used to restrict the wanted transformation. For example, by setting $\mathcal{A} = \{y : \ \mathcal{P}(y) = 0\}$, where the penalty could be a landmark match $\mathcal{P} = \mathcal{D}^{\mathrm{LM}}(y)$ (cf. (3)), the restriction y being rigid on certain areas, or the determinant of the Jacobian to be be positive.

   In many applications, these restrictions or so-called hard constraints are replaced by so-called soft constraints or penalties. Here, the objective in (1) is augmented by a penalty term,

$$\mathcal{J} : \mathcal{A} \to \mathbb{R}, \quad \mathcal{J}(y) := \mathcal{D}(T \circ y, R) + \mathcal{R}(y) + \mathcal{P}(y). \tag{8}$$

We note, that the penalty approach comes with a number of drawbacks. Firstly, for a minimizer y^* of (8) generally $\mathcal{P}(y^*) \neq 0$. Secondly, even if $\mathcal{P}(y^*) \ll 1$, it might be big on small but potentially critical structures such as tumours or lessons. Thirdly, the penalty usually involves a weighting parameter that has to be tuned for practical applications. And fourthly, typically huge values are used for weighting, which then lead to ill-conditioned optimization problems. Despite its various conceptual disadvantages, penalty approach are very common in the registration community as they are generally easy to be implemented.

4.5   Penalties for locally invertible maps

The mapping $y = y(x) = x + u(x)$ is invertible in a neighbourhood of a point x, if the determinant of the Jacobian $d(x) := \det(\nabla y(x))$ is nonzero and the inverse function theorem applies. In order to also preserve orientation, one may constrain the transformation such that $d(x) > 0$. We remark that this constraint is mathematically challenging as the determinant may leaves L_2 and solutions of the registration problem may not be evaluated pointwise; see e.g. [34,33] and the extended discussion in [11]. In [34], the poitnwise constraint $\mathcal{C}(y) = d(x) = 1$ has been introduced and extended to $k(x) \leq d(x) \leq K(x)$ in [37]. Here, the constraint applies to all points $x \in \Sigma \subset \Omega$, where Σ could be the whole domain or a subset indication critical organs or tumours; upper and lower bounds are denoted by k and K, respectively.

   A commonly used idea in image registration is to add a penalty to the joint energy. In [70], the following penalty on the determinant of the Jacobian d is introduced

$$\mathcal{P}(y) = \int (\log(d))^2 \ dx,$$

[87] considered the penalty

$$\mathcal{P}(y) = \int |d - 1| \cdot \log |d| \ dx.$$

Both approaches favour transformations with $d \approx 1$.

We remark that the hyperelastic approach of [11] might be interpreted along the same lines as the volume term is

$$\text{volume}(y) = \int \left((d-1)^2/d\right)^2 \, dx.$$

Note, however, that [11] introduced a regularizer, i.e. the existence of a minimizing element with $d > 0$ is guaranteed, but the theory also requires the length and in particular the cofactor term; see [11] for details and extended discussion.

The 2D case is simpler. In [90], a penalty based on the Beltrami coefficient is proposed:

$$\mathcal{P}(y) = \int \frac{|\mu|^2}{(|\mu|^2 - 1)^2} \, dx, \quad |\mu|^2 = \frac{\|\nabla y\|_2^2 - 2\det(\nabla y)}{\|\nabla y\|_2^2 + 2\det(\nabla y)}. \tag{9}$$

However, the determinant is not easily accessible and it adds strong nonlinearity to a model. For $d = 3$ for example, the determinant is a polynomial of derivatives of degree 3 and thus may not even be in L_2. Furthermore, the determinate being non-negative on certain grid points does not imply that the determinant is non-negative everywhere. The algorithm from [11] resolves these issues using a computational framework with finite elements.

4.6 Diffeomorphic registration

A diffeomorphic transform y is invertible, differentiable and smooth function (that offers one-to-one mapping) and its inverse is also smooth. For applications such as brain mapping and registration of MR images, diffeomorphic registration models are particularly suitable.

The above subsection discussed one to one maps and how to achieve them. Diffeomorphic registration models offer more than locally invertible maps. Here we review two such models.

In the first diffeomorphic registration model of Large Deformation Metric Mappings (LDMM) [4,88], an additional time component is introduced. Let $K := [0, 1]$ be a normalized time interval, $\Omega_K := \Omega \times K$, $I : \Omega_K \to \mathbb{R}$ and $y : \Omega_K \to \mathbb{R}^d$. For brevity, we write $z_t := z(\cdot, t)$ for functions depending on space and time. Moreover, we introduce the velocity v such that $\partial_t y_t(x) =: v(y_t(x))$. The objective is then to find a minimizing element (y, v) of an energy

$$\mathcal{J}^{\text{LDDMM}}(y, v) := \mathcal{D}(T, R) + \mathcal{R}^{\text{LDDMM}}(v), \tag{10}$$

constrained for all $(x, t) \in \Omega_K$ by

$$
\begin{aligned}
I_0(x) &= I(x,0) = R(x), & I_1(x) &= I(x,1) &= T(x), \\
y_0(x) &= y(x,0) = x, & \partial_t y_t(x) &= v(y_t(x)).
\end{aligned}
$$

In the second diffeomorphic registration model of [48], the key idea is to look for y in the space of quasi-conformal maps, using

$$\mathcal{R}(u) = \int_\Omega |\nabla \mu|^2 dx, \quad \mathcal{C}(u) = \int_\Omega |\mu|^p dx + \mathcal{P}(y)$$

where $\mu = \mu(y)$ is the Beltrami coefficient (9), explicitly represented by $y = (y_1, y_2)$. More specifically, if one write $y = y_1 + \sqrt{-1}y_2$ in the complex form, then $\mu(y)$ is defined as $mu(y) = \frac{\partial y}{\partial \bar{z}} / \frac{\partial y}{\partial z}$, where $\frac{\partial}{\partial \bar{z}} = \frac{1}{2}(\frac{\partial}{\partial x} + \sqrt{-1}\frac{\partial}{\partial y})$ and $\frac{\partial}{\partial z} = \frac{1}{2}(\frac{\partial}{\partial x} - \sqrt{-1}\frac{\partial}{\partial y})$. $\mathcal{P}$ denotes the landmarks constraint. Note that the one-to-one mapping property can be seen from the equivalence of $\det(\nabla y) > 0$ and $|\mu| < 1$.

4.7   Registration by inverse consistent approach

This framework is used to enforce a directionally unbiased registration. Note that the optimal y for the images T and R may not relate to the optimal z for R and T, order matters. This is, among others, due to the fact that the distance as well as the regularizer are assigned to the reference image domain. A remedy suggested in [18] is to symmetrize the problem i.e. to register R to T, as well as T to R, at the same time. Introducing the transformation $z : \mathbb{R}^d \to \mathbb{R}^d$, such that ideally $z = y^{-1}$, the objective now is to minimize

$$\mathcal{J}(y,z) = \mathcal{D}(T \circ y, R) + \mathcal{D}(R \circ z, T) + \mathcal{R}(y) + \mathcal{R}(z) + \mathcal{E}(y,z),$$

with

$$\mathcal{E}(y,z) = \|y \circ z^{-1} - \mathrm{id}\| + \|z \circ y^{-1} - \mathrm{id}\|;$$

for details, see, e.g. [18,17,2,81].

5   Surface registration

In this section, we discuss state-of-the-art approaches for surface registration (SR); see, e.g., [48,85,54,67,8,75,51] and references therein.

SR is a special class of registration problem. Instead of warping between two images, SR aims to find a mapping between two curvilinear surfaces; see also Figure 4. More precisely, the goal of SR is to establish a meaningful correspondences between two Riemann surfaces which are denoted by S^k, $k = 0, 1$. The correspondence should match the surface geometry as well as the data defined on the surfaces. As outlined in Sec. 3, SR can be interpreted as a special case of feature based image registration, where surface are derived from the data by some kind of projection. However, a major difference between SR and the conventional IR is that surfaces have an intrinsic geometry. In IR, the domains of images are Euclidean spaces, whereas in SR, the domains of interest are Riemann surfaces embedded in $\mathbb{R}^d$. In this paper we focus on $d = 3$ for ease of presentation. As such, the surface geometry is usually taken into consideration in SR. This is the reason why SR is usually also called shape matching [8].

5.1   Brief introduction to surface geometry

In practice, the surfaces to be registered often have different geometry. In SR, one main goal is to match two surfaces based on their geometry. For instance,
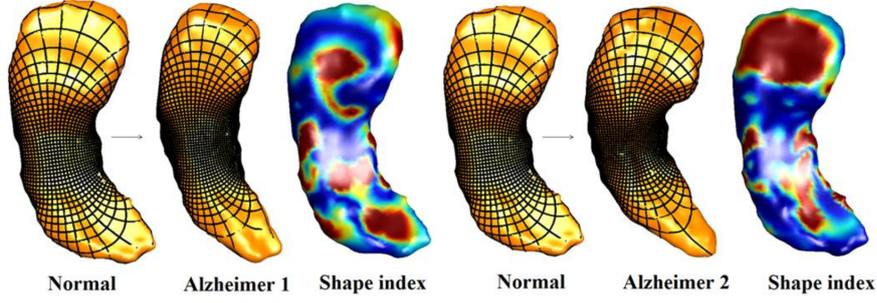
Fig. 4. Examples of surface registration between hippocampal surfaces for medical morphometry.

Fig. 5. Examples of surface conformal parameterizations of a human face (left) and a brain cortical surface onto a unit disk (right), respectively.

in brain surface registration, the sulci and gyri on each brain surfaces should match with each other. To achieve this goal, the underlying idea of SR is that the similarity of two surfaces S^k can by measured by utilizing the associate Riemanian metrics g^k, $k = 0, 1$. More specifically, geometric distortion between S^0 and S^1 can be measured by comparing the so-called pull-back metric $f^*(g^1)$ with the original metric g^0.

A Riemann surface is an arbitrarily smooth 2D manifold S equipped with an inner product g_p on the tangent space T_pS for each point $p \in S$; see e.g. [50]. The inner product g_p is also called a Riemannian metric and varies smoothly amongst points on S. More specifically, a Riemannian metric on S is a collection of inner product

$$g_p : T_pS \times T_pS \to \mathbb{R}, \quad p \in S,$$

such that for vector fields X and Y defined on S, the mapping $p \mapsto g_p(X(p), Y(p))$ is arbitrarily smooth.

The pull-back metric maps the metric of S^1 back to S^0. To this end let $f : S^0 \to S^1$ be one-to-one and differentiable. For any given point $p \in S^0$ and any given $u \in T_pS^0$, let γ be a smooth curve on S^0 with $\gamma(0) = p$ and $\gamma'(0) = u$. The differential df_p is defined by $df_p(u) := (f \circ \gamma)'(0)$ and the pull-back metric by

$$f_p^*(g^1) : T_pS^0 \times T_pS^0 \to \mathbb{R}, \quad f_p^*(g^1)(u, v) := g_p^1(df_p(u), df_p(v)).$$

Intuitively, a pull-back metric takes the inner product of two tangent vectors u and v of S^0 being pushed forward by f to two tangent vectors on S^1. The inner product differs from g^0 by the distortion introduced by f and the geometric distortion between the surfaces S^0 and S^1 can be measured by comparing the pull-back metric with the original metric g^0.

Over the last decades, different approaches for SR have been proposed. Below we will summarize some popular approaches from the literatures; see, e.g., [49,85,8,51]. Specific examples include the isometric maps [75], the conformal maps [54], and the quasi-conformal maps [48,56], which we discuss in more detail.

A map f is called isometric if $f^*(g^1) = g^0$. An isometric map captures shape deformations that preserve length, such as human poses. A map is called conformal if $f^*(g^1) = \lambda g^0$, where λ is the so-called conformal factor, a non-negative smooth function $\lambda : S^0 \to \mathbb{R}_{>0}$. A conformal map preserves the local geometry under the map, although the local area is usually distorted. It captures deformations that does not change the local shape, such as biological deformations.

A map is called quasi-conformal, if $f^*(g^1) = |\frac{df}{dz}|^2 |dz + \mu d\bar{z}|^2$, where z is the local coordinate of S^0 and μ is a complex-valued function capturing the conformality distortion. It captures more general deformations with bounded amount of local geometric distortions. Depending on various applications, SR aims to look for optimal map in a suitable class of mappings that satisfies other prescribed constraints, such as the landmark constraints and curvature-matching condition.

5.2   Parameterization-based approaches

A commonly used approach for SR is based on the global surface parameterization [54]. The basic idea is to flatten the surface to a simple domain, such as a 2D unit disk or a 2D rectangle; see also Fig. 5. For instance, a surface can be interpreted as a 2D image, whose image intensity is given by some geometric quantities, if it is parameterized onto a 2D rectangle. SR can then be considered as finding an optimal mapping between two images that matches the corresponding geometric quantities, such as the curvatures. This problem can be solved by any registration approaches discussed in the previous sections. More specifically, suppose the surfaces S^0 and S^1 to be registered are parameterized onto the domain $\Omega \subset \mathbb{R}^2$. Denote the parameterizations of S^0 and S^1 by $\psi_0 : \Omega \to S^0$ and $\psi_1 : \Omega \to S^1$ respectively. Suppose the mean curvatures H_0 and H_1 of S^0 and S^1 are to be matched. The SR problem is reduced to an IR problem to match two 2D images given by $I^0 := H_0 \circ \psi_0 : \Omega \to \mathbb{R}$ and $I^1 := H_1 \circ \psi_1 : \Omega \to \mathbb{R}$. In other words, we aim for a transformation $T : \Omega \to \mathbb{R}^2$ such that

$$I^1 \circ T(x) = I^0(x) \text{ for all } x \in \Omega.$$

Sometimes, the constraints on landmark correspondences can be enforced. Suppose p_i on S^0 should be matched to q_i on S^1 for $i = 1, 2, ..., m$. Then, the transformation T should further satisfy the condition that $T\psi_0^{-1}(p_i) = \psi_1^{-1}(q_i)$

for $i = 1, 2, ..., m$. Similar to IR, a regularization on T is usually imposed to constrain the type of transformation. For instance, a common choice is to look for an optimal conformal parameterization that matches landmarks and the surface geometry [54]. Once an optimal T is obtained, the registration map between S^0 and S^1 can be obtained by the composition map $\psi_1 \circ T \circ \psi_0^{-1} : S^0 \to S^1$. Of course, the choice of parameterization is crucial in these approaches and may introduce distortions such as conformality distortion or metric distortion. To avoid theses distortions, conformal or isometric surface parameterizations are often used; cf., e.g., [54,56].

5.3 Laplace-Beltrami eigenmap approaches

In these approaches, one takes advantage of an eigen-system of the Laplace-Beltrami operator Δ_S on the surface S [77,75,51]. Let (S, g) be a Riemann surface in $\mathbb{R}^3$, $g = (g_{i,j}) \in \mathbb{R}^{2,2}$ be the Riemann metric and $(g^{i,j})$ its inverse. For a smooth function $f : S \to \mathbb{R}$, the Laplace-Beltrami (L-B) operator Δ_S acting on f near a point p is defined by

$$\Delta_S f := (\det g)^{-\frac{1}{2}} \sum_{i,j=1}^{2} \partial_i [\det(g) \ g^{i,j} \partial_j f].$$

As the L-B operator Δ_S is self-adjoint and elliptic, it has a system of eigenvalue and corresponding eigenfunctions (λ_j, ϕ_j) with $-\Delta_S \phi_j = \lambda_j \phi_j$ and $\lambda_j \leq \lambda_{j+1}$ for all $j \in \mathbb{N}$. This L-B eigen-system can be used to define a shape signature for a surface up to isometry. Two isometric surfaces have the same L-B eigen-systems [75]. SR strategies have been investigated to match L-B eigen-systems as good as possible for surfaces undergoing isometric deformations; see, e.g. [72].

The L-B approach is especially effective when registering shapes with different poses. For instance, for every point $p \in S$, the L-B eigen-system gives rise to a m-dimensional feature vector $s_m(p) := (\lambda_j^{-1/2} \phi_j(p))_{j=1}^m$, where the number m is user supplied. In a discrete scenario, we assume that the surfaces S^k are represented by a collection of vertices $\{p_j^k\}_{j=1}^n$, $k = 0, 1$. An L-B signature of S^k is then the $n \times m$ matrix B^k, where the j-th row of B^k is given by $s_m^k(p_j^k)$. A match of the surfaces can be obtained by computing an optimal correspondence matrix C such that $||B^1 C - B^0||_{\text{Fro}}$ is minimized.

5.4 Metric approaches

In these approaches, a non-rigid shape is modelled as a metric space (X, dX), where X is a two-dimensional smooth compact connected and complete Riemannian surface (possibly with boundary) embedded into $\mathbb{R}^3$, and $dX : X \times X \to \mathbb{R}$ is a metric measuring distances between pairs of points on X. Two shapes (X, dX) and (Y, dY) are similar if the metrics between pairs of corresponding points on X and Y coincide, i.e., there exists a bijective map $\phi : X \to Y$ such that

$dY \circ (\phi \times \phi) = dX$ [8]. To simplify the computation of the shape correspondence ϕ, a low dimensional representation is often considered, by means of a minimum-distortion embedding $\phi : X \to Z$. The low dimensional representation is usually called the canonical form, which can be computed by solving the multidimensional scaling problem [8]

$$\min_{\phi:X \to Z} \max_{x,x' \in X} ||d_X(x, x') - d_Z(\phi(x), \phi(x'))||$$

where d_Z is the metric of Z. Once the embeddings $\phi(X)$ and $\psi(X)$ of X and Y are computed, an optimal correspondence can be obtained by minimizing the Hausdorff distance between $\phi(X)$ and $\psi(X)$. It is generally impossible to select a common metric space in which the geometry of any shape can be accurately represented. To solve this problem, the Gromov-Hausdorff distance, instead of the Hausdorff distance, is considered:

$$d_{GH}(X, Y) := \inf_{\phi:X \to Z, \psi:Y \to Z, Z} d_H^Z(\phi(X), \psi(Y)).$$

5.5   Functional map approaches

Another popular SR approach is based on a functional map representation [67]. For $k = 0, 1$, we consider the spaces $\mathcal{F}^k := \mathcal{F}(S^k, \mathbb{R})$ of mappings from S^k to $\mathbb{R}$. Then any one-to-one transformation $y : S^0 \to S^1$ defines the linear mapping $y_{\mathcal{F}} : \mathcal{F}^0 \to \mathcal{F}^1$, $\psi^0 \mapsto \psi^1 := \psi^0 \circ y^{-1}$.

For a given $y_{\mathcal{F}}$, the underlying transformation y can be reconstructed using point evaluation functionals δ_x [67]: for any $x^0 \in S^0$, $y_{\mathcal{F}}(\delta_{x_0}) =: \delta_{x^1}$ establishes the connection between $x^0 \in S^0$ and $x^1 \in S^1$.

Suppose now that for $k = 0, 1$, $\{b_j^k\}_{j=1}^{\infty}$ is a basis of $\mathcal{F}^k$. Thus we have the following representations, $f = \sum \zeta_\mu b_\mu^0$, $y_{\mathcal{F}}(b_\mu^0) = \sum \gamma_{\nu,\mu} b_\nu^1$, and

$$y_{\mathcal{F}}(f) = \sum \beta_\nu b_\nu^1 = y_{\mathcal{F}}(\sum \zeta_\mu b_\mu^0) = \sum \sum \gamma_{\nu,\mu} \zeta_\mu b_\nu^1.$$

With the column vectors (β_ν) and (ζ_μ) that are formed by the coefficients and the matrix representation $C = (\gamma_{\nu,\mu})$ of $y_{\mathcal{F}}$, we have $(\beta_\nu) = C(\zeta_\mu)$. With this setup, SR can be computed by different choices of basis functions as well as prescribing different conditions on $y_{\mathcal{F}}$.

For example, to obtain an isometric SR, one can choose $\{\phi_j^k\}_{j=1}^{\infty}$ to be the Laplace-Beltrami eigenfunctions on S^k and choose an optimal C that is closest to a diagonal matrix. To further enforce the landmark constraints $y_{\mathcal{F}}(\delta_{p_\ell^0}) = \delta_{p_\ell^1}$. Functional map method reduces the SR problem to optimizing a matrix C.

5.6   Relationship between SR and IR

SR and IR are closely related to each other. Many ideas from image registration can be applied to surface registration. In fact, both IR and SR can be formulated

as an optimization problem over the space of admissible registration maps as follows:

$$E_{reg}(f) = \mathcal{D}(I_1, I_2 \circ f) + \mathcal{R}(f) \tag{11}$$

where $\mathcal{D}$ is the distance measure and $\mathcal{R}$ is a suitable regularization term. Common distance measures in IR are usually chosen as the L^1 or L^2 error between image intensities. These distance measures are often used in SR. The only challenge in SR, which is different from IR, is that meaningful 'intensities' on surfaces are not readily defined. One important question in SR is to design suitable geometric quantities, such as curvatures, conformality distortions, Laplace-Beltrami eigenfunctions, to drive the registration process and obtain a geometric matching surface registration map.

In recent years, various 'meaningful' intesnities based on the surface geometry has been proposed to define the dissimilarity measure before surfaces. Another connection is the regularization terms in IR and SR. Most of the regularization terms applied in IR can be generalized to SR. Examples include the total variation (TV), harmonic energy and so on. The only difference is that on surfaces, instead of using the standard Euclidean differentiation, covariant derivatives defined according to the Riemannian metric have to be utilized. In other words, the problem of surface registration is indeed a combination of ideas from IR with Riemannian geometry.

More specifically, the connection between a surface S and an image I might be interpreted as follows. If the surface is parameterized as $S = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 = I(x_1, x_2)\}$, then the surface is characterized by the intensities of a 2D image I. For a comparison of the two surfaces, the distance measure can be chosen as the discrepancies of some geometric properties such as curvature. An example is the discrepancies of curvatures [13,55]:

$$\mathcal{D}(S^0, S^1) := \int_{S_1} \alpha(H^1 - H^0)^2 + \beta(K^1 - K^0)^2 \, ds,$$

where H^k denotes the mean curvatures and K^k the Gaussian curvatures on S^k, $k = 0, 1$. Here $\alpha, \beta > 0$ are regularization weights. Another example is to match the Laplace-Beltrami eignenfucntions defined on each surface according to their Riemannian metrics.

As for the regularization term, various choices can be chosen. For example, in order the enhance the smoothness of the surface mapping, a harmonic energy can be used:

$$\mathcal{R}_{harmonic} \int_{S_1} |\nabla_{S_1} f|^2 ds,$$

where ∇_{S_1} is the gradient on the surface in term of covariant derivatives. If one aims for a biholomorphic mapping as regularization, the Riemann mapping theorem [31] suggest to add the Riemannian metric to the curvature mismatching term $\mathcal{D}(S^0, S^1)$,

$$\mathcal{D}(y) := \mathcal{D}(S^0, S^1 \circ y) + \gamma|\mu(T)|^2.$$

Here, $\mu(y)$ is the Beltrami coefficient measuring the metric distortion under the transformation y. This gives rise to various geometric measurements. Interestingly, if $\alpha, \beta, \gamma > 0$, this gives a full geometric measurement. This is that $\mathcal{D}(y) = 0$ when two surfaces are the same up to a rigid motion. For $\beta = 0$, $\mathcal{D}(S^0, S^1) = 0$ when the two surfaces are isometric to each other [55].

6   Numerical Methods

Image registration is phrased as a variational problem and the objective is to find a minimizing element of $\mathcal{J}$ in an admissible set $\mathcal{A}$. Typically, analytic solutions are not at hand and numerical schemes are to be applied. As the data to be process can range to 3×1024^3 and time is valuable, the choice of a numerical solution strategy is non-trivial. In this paper, we only briefly outline some of the basic concepts as the topic is too big; for details see, e.g. [63].

There are mainly two directions to go. The first is called optimize-then-discretize. Here, one derives the so-called Euler-Lagrange equations (ELE), which are conditions for a minimizer of the functional $\mathcal{J}$ in (1). One then discretizes the ELE which yields a nonlinear set of equation. Finally, theses equations are solved numerically. However efficient solvers are non-trivial to design, mainly to the fidelity term and depending on the regularisers used. For example, the ELE for $\mathcal{J}$ based on the SSD distance measure and the diffusion regularizer and $y_{\mathrm{ref}} = 0$ yields

$$\mathcal{J}(y) := \mathcal{D}^{\mathrm{SSD}}(T \circ y, R) + \alpha \mathcal{R}^{\mathrm{diff}}(y) = \frac{1}{2}\|T \circ y - R\|^2_{L^2(\Omega)} = + \frac{\lambda}{2}\|\mathcal{B}y\|^2_{L^2(\Omega)}$$
$$0 = (T \circ y - R)^\top \nabla T \circ y - \Delta y,$$

see [62] for details. The transformation can be discretized using a finite differences or finite element approach; see, e.g., [62,63,48,90] and the references therein.

This approach has a number of drawbacks. First of all, the operator is derived from partial integration and the approach therefore comes with additional assumptions on z. For example, for the diffusion regularizer, the Gâteaux derivative for a suitable perturbation z, one gets

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(\mathcal{R}(y + \varepsilon z) - \mathcal{R}(y)) = \sum_j \int_\Omega (\nabla y_j)^\top (\nabla z_j) \; dx$$
$$= \sum_j \left([z_j \cdot n^\top \nabla y_j]_{\partial \Omega} - \int_\Omega \Delta y_j \cdot z_j \; dx \right).$$

With the artificial arguments that with appropriate boundary conditions the boundary term vanishes and that y is twice differentiable, the above ELE follows. Secondly, it is neither clear nor obvious that discretization of the ELE yields a symmetric Hessian system.

The so-called discretize-then-optimize approach offers a better alternative [36]. Here, the objective $\mathcal{J}$ and the wanted transformation y are discretized, and

then a sequence of finite dimensional optimization problems is to be solved. For the above example, one ends up with

$$J^h(y^h) = \frac{h}{2}\|T(y^h) - R(x^h)\|^2_{\mathbb{R}^m} + \lambda\frac{h}{2}\|B^h y^h\|^2$$

where x^h is a collection of m grid points, Sec. 2.1, $R(x^h) \in \mathbb{R}^m$ are image values on the grid, $y^h = y(x^h)$ is the representation of the transformation on the grid, $T(y^h)$ are image values of the transformed template, h is the volume element. For $d = 1$, the matrix B^h is given by

$$B^h = \frac{1}{h}\begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{m-1,m}. \qquad (12)$$

In this approach, the problem due to data fidelity takes the form of a highly non-convex term and also a non-positive definite Hessian.

Advantages of the discretize-then-optimze approach are that no additional assumptions on the solutions are used, boundary conditions are derived from the optimization problem and are automatically taken care of, and the Hessian on the discrete problem is naturally symmetric. Finally, stopping criteria and step-length control follow standard optimization rules; see, e.g., [65]. The optimization is typically based on first order or second order schemes, mainly quasi-Newton schemes. In particular in a multi-level framework, where a numerical solution of a coarse grid representation is used as a starting guess for a finer grid, one expects an excellent starting point and hence fast convergence due to a higher order scheme.

Solving the problem numerically is still challenging. Memory and time restriction prohibits the use of direct solvers for the Hessian systems. Highly efficient multigrid solvers for B^*B or $M + B^*B$ are used, where M is some approximation to the Hessian of the data-fit; see [63,20] for details.

In particular for the SSD measures and similarly to optical flow problems, some authors use a Gauß-Newton type approximation of the data-fit. With $y^{k+1} = y^k + u$, $T^k := T \circ y^k$ and $dT := \nabla T \circ y^k$,

$$\mathcal{D}(T \circ y^{k+1}, R) = \|T \circ (y^k + u) - R\|^2_\Omega \approx \|T + dT \cdot u - R\|^2_\Omega =: q(u)$$

the objective becomes a convex quadratic function.

Development of new and faster algorithms with optimal computational complexity remains an interesting topic.

7 Deep learning based registration

Machine learning techniques have advanced rapidly over the last decade, especially in image classification and segmentation. The main idea of deep-learning based algorithms is to aggregate information of various complexities in the datasets of training images to drive the registration process. In recent years,

supervised and unsupervised deep learning techniques have been successfully employed for image and surface registration [44,12,47,52,61,76,86] (in particular the latter five references on convolutional neural networks for image registration). Convolutional neural networks of various architectures have been used to learn meaningful feature descriptors [3,89,69].

Though possessing many similarities in the variation framework to other imaging models, registration is quite different in the context of deep learning. The major difference lies in the non-readily available sets of training data e.g. in medical imaging, few clinicians can accurately draw the deformation to a good accuracy. This sets the scene of the state of art in deep learning based registration which we briefly review; refer to [53,58,57].

Firstly, we discuss the supervised learning. To deal with the problem of getting enough training data, one could generate simpler transforms for various images so that training tuples of the form (T, R, y) are available, instead of the usual registration pair (T, R) being given as in §1-6. Deep learning is also employed to estimate the momentum in the framework of Large Deformation Diffeomorphic Metric Matching (LDDMM) [86] where the same idea of generating a space of priori deformations is used to design training data. Of course, once training data are given or rather generated, various deep learning architectures are applicable as with solving other imaging problems by learning. Refer to recent surveys [53,58]. Clearly the supervised model is good for problems where the types of deformations can be predicted a priori, though not robust.

Secondly, we mention the unsupervised learning where no training data are required so that more generality is offered. Then as with non-learning models, intensity-based registration models drive the registration process through matching image intensities. The key step is to replace the loss function (usually measuring the distance between the given deformation and the iterate) by an energy functional that underlines a variational registration model (see §1-6). In this idea, various deep neural networks have been proposed to transform images to match image intensities. For instance, the spatial transformer network [44] transforms an input image through a neural network such that the classification task can be simplified. Convolutional neural networks with reinforcement learning are trained to predict descent directions of the transformation towards optimal alignment [52,61]. Convolutional neural network has also been applied to predict parameters in the thin-plate spline registration model [12]. General dense displacement vector field can also be predicted by training a deep convolutional neural network, without applying a fixed parametric registration model [76]. Other ideas of merging small sets of training data and unsupervised learning may also be considered, to design new learning models.

8   Conclusions

Registration between corresponding data is a crucial process and has important applications in various fields. In this chapter, we give an overview of the registration problem. The basic mathematical background and the mathemati-

cal framework for registration problems are firstly introduced. State-of-the-art registration models are then described. In particular, we carefully examine various existing distance measures and regularization terms. Their advantages and disadvantages are also discussed. Furthermore, mathematical models for surface registration, which aims to find meaningful one-to-one correspondence between two surfaces embedded in 3D space, are also reviewed. The connection between surface registration and image registration is also studied. Numerical methods for different registration models are also briefly explained. As a popular research direction, various deep learning-based registration models invented recently are also described. In the future, we expect registration models, which combine mathematical models and machine learning techniques, will define a trend for solving more challenging registration problems.

References

1. I. Arganda-Carreras, C. O. S. Sorzano, P. Thévenaz, A. Muñoz-Barrutia, J. Kybic, R. Marabini, J. María Carazo, and C. Ortiz de Solorzano. Non-rigid consistent registration of 2d image sequences. *Physics in Medicine and Biology*, 55 20:6215–42, 2010.
2. J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38:95–113, 2007.
3. V. Balntas, E. Johns, L. Tang, and K. Mikolajczyk. Pn-net: Conjoined triple deep network for learning local image descriptors. *preprint arXiv:1601.05030*, 2016.
4. M. F. Beg, M. I. Miller, A. Trouve, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61:139–157, 2005.
5. K. Brehmer, B. Wacker, and J. Modersitzki. Simultaneous registration of image sequences – a novel singular value based images similarity measure. In *Proceedings in Applied Mathematics and Mechanics - 89th GAMM Annual Meeting*, page https://doi.org/10.1002/pamm.201800370. Wiley, 2018.
6. T. Breiten, V. Simoncini, and M. Stoll. Low-rank solvers for fractional differential equations. *Electron. Trans. Numer. Anal.*, 45:107–132, 2016.
7. C. Broit. *Optimal Registration of Deformed Images*. PhD thesis, Computer and Information Science, University of Pensylvania, USA, 1981.
8. A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro. A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *International Journal of Computer Vision*, 89:266–286, 2010.
9. L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.
10. A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. Real-time optic flow computation with variational methods. *Computer Analysis of Images and Patterns*, 2756:222–229, 2003.
11. M. Burger, J. Modersitzki, and L. Ruthotto. A hyperelastic regularization energy for image registration. *SIAM Journal on Scientific Computing*, 35(1):B132–B148, 2013.
12. X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen. Deformable image registration based on similarity–steered cnn regression. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, 28:300–308, 2017.

13. H.L. Chan, H.F. Li, and Lok Ming Lui. Quasi-conformal statistical shape analysis of hippocampal surfaces for alzheimer disease analysis. *Journal of Neurocomputing*, 175(A):177–187, 2016.

14. S. Chelbi and A. Mekhmoukh. Features based image registration using cross correlation and radon transform. *Alexandria Engineering Journal*, 57(4):2313–2318, 2018.

15. K. L. Chen, A. Derksen, S. Heldmann, M. Hallmann, and B. Berkels. Deformable image registration with automatic non-correspondence detection. In J. F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale Space and Variational Methods in Computer Vision: 5th International Conference, SSVM 2015, Lège-Cap Ferret, France, May 31 - June 4, 2015, Proceedings*, pages 360–371, Cham, 2015. Springer International Publishing.

16. L. Chen. *Digital Functions and Data Reconstruction*. Springer, 2013.

17. Y. M. Chen and X. J. Ye. *The Legacy of Alladi Ramakrishnan in the Mathematical Sciences*, chapter Inverse consistent deformable image registration, pages 419–440. Springer, New York, 2010.

18. G.E. Christensen and H.J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 20:568–582, 2001.

19. N. Chumchob and Ke Chen. A robust affine image registration method. *International Journal of Numerical Analysis and Modelling*, 6(2):311–334, 2009.

20. N. Chumchob and Ke Chen. A robust multigrid approach for variational image registration models. *Journal of Computational and Applied Mathematics*, 236(5):653–674, 2011.

21. N. Chumchob, Ke Chen, and C. Brito-Loeza. A fourth order variational image registration model and its fast multigrid algorithm. *SIAM J. Multiscale Model. Simul.*, 9:89–128, 2011.

22. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. In *IPMI*, pages 263–274, 1995.

23. D. Drobny. *Automatic Weighting of Non-corresponding Regions in Image Registration*. PhD thesis, University of Lubeck, Lubeck, Germany, 2015.

24. D. Drobny, H. Carolus, S. Kabus, and J. Modersitzki. Handling non-corresponding regions in image registration. In H. Handels, T. M. Deserno, H.-P. Meinzer, and T. Tolxdorff, editors, *Bildverarbeitung fur die Medizin 2015*, pages 107–112, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.

25. M. Droske and M. Rumpf. A variational approach to non-rigid morphological image registration. *SIAM Appl. Math.*, 64(2):668–687, 2004.

26. B. Fischer and J. Modersitzki. Fast inversion of matrices arising in image processing. *Numerical Algorithms*, 22:1–11, 1999.

27. B. Fischer and J. Modersitzki. Fast diffusion registration. *Contemporary Mathematics*, 313:117–129, 2002.

28. B. Fischer and J. Modersitzki. Combining landmark and intensity driven registrations. *PAMM*, 3:32–35, 2003.

29. M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92, 1973.

30. C. A. Glasbey and K. V. Mardia. A review of image-warping methods. *Journal of Applied Statistics*, 25(2):155–171, 1998.

31. J. Gray. On the history of the riemann mapping theorem. *Rendiconti del Circolo Matematico di Palermo, Serie II*, page 47–94, 1994.

32. E. Haber, S. Heldmann, and J. Modersitzki. A scale-space approach to land-mark constrained image registration. In *Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision (SSVM)*, pages 1–12. Springer LNCS, 2009.

33. E. Haber, R. Horesh, and J. Modersitzki. Numerical methods for constrained image registration. *Numerical Linear Algebra with Applications*, 17(2-3):343–359, 2010.

34. E. Haber and J. Modersitzki. Numerical methods for volume preserving image registration. *Inverse Problems*, 20:1621–1638, 2004.

35. E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. In C Barillot, DR Haynor, and P Hellier, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, volume 3216, pages 591–598. Springer LNCS, 2006.

36. E. Haber and J. Modersitzki. A multilevel method for image registration. *SIAM Journal of Scientific Computing*, 27:1594–1607, 2006.

37. E. Haber and J. Modersitzki. Image registration with a guaranteed displacement regularity. *International Journal of Computer Vision*, 71:361–372, 2007.

38. S. Heldmann. Multimodal registration of MR images with a novel least-squares distance measure. In *Medical Imaging 2010: Image Processing*, San Diego, 2010. SPIE.

39. S. Henn and K. Witsch. A variational image registration approach based on curvature scale space. In *LNCS 3459*, page 143–154, 2005.

40. D. Hill, P.G. Batchelor, M. Holden, and D. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001.

41. B.K.P. Horn and B.G.Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

42. P. J. Huber. Robust estimation of a location parameter. *Annals of Statistics*, 53(1):73–101, 1964.

43. M. Ibrahim, Ke Chen, and C. Brito. A novel variational model for image registration using gaussian curvature. *Journal of Geometry, Imaging and Computing*, 1:417–446, 2014.

44. M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks. *Advances in Neural Information Processing System*, 28:2017–2025, 2017.

45. S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9:1357–1370, 2000.

46. A. M. Khan, W. Chen, A. Khalil, and J. Lin. Euler's elastica and curvature based model for image restoration. *PLOS ONE*, 13, 09 2018.

47. J. Krebs, T. Mansi, H. Delingette, L. Zhang, F.C. Ghesu, S. Miao, A.K. Maier, N. Ayache, R. Liao, and A. Kamen. Robust non–rigid registration through agent–based action learning. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, 28:344–352, 2017.

48. K. C. Lam and Lok Ming Lui. Landmark- and intensity-based registration with large deformations via quasi-conformal maps. *SIAM J. Imaging Sci.*, 7:2364–2392, 2014.

49. K.C. Lam and Lok Ming Lui. Optimized conformal parameterization with controllable area distortions. *Communications in Mathematical Sciences*, 3:52–78, 2010.

50. J. Lee. *Riemannian Manifolds: an introduction to curvature*. Springer, 1997.

51. J. Liang, R. Lai, T.W. Wong, and H.K. Zhao. Geometric understanding of point clouds using laplace-beltrami operator. *Computer Vision and Pattern Recognition (CVPR)*, pages 214–221, 2012.

52. R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D.Comaniciu. An artificial agent for robust image registration. *In: Proceedings of the ThirtyFirst AAAI Conference on Artificial Intelligence*, 28, 2017.

53. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

54. Lok Ming Lui, S. Thiruvenkadam, Y. Wang, P.M. Thompson, and T.F. Chan. Optimized conformal surface registration with shape-based landmark matching. *SIAM Journal on Imaging Sciences*, 3:52–78, 2010.

55. Lok Ming Lui, T.W. Wong, X.F. Gu, P.M. Thompson, T.F. Chan, and S.T. Yau. Hippocampal shape registration using beltrami holomorphic flow. *Medical Image Computing and Computer Assisted Intervention(MICCAI), Part II, LNCS 6362*, pages 323–330, 2010.

56. Lok Ming Lui, T.W. Wong, W. Zeng, X.F. Gu, P.M. Thompson, T.F. Chan, and S.T. Yau. Optimization of surface registrations using Beltrami holomorphic flow. *Journal of Scientific Computing*, 50:557–585, 2012.

57. A. S. Lundervold and A. Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.

58. A. M., C. Syben, T. Lasser, and C. Riess. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik*, 29(2):86–101, 2019.

59. P. C. Mahalanobis. On the generalised distance in statistics. In *Proceedings of the National Institute of Science of India*, volume 2, pages 49–55, 1936.

60. J. B. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

61. S. Miao, Z.J. Wang, and R. Liao. A cnn regression approach for real–time 2d/3d registration. *IEEE Transaction on Medical Imaging*, 35:1352–1363, 2016.

62. J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, New York, 2004.

63. J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. SIAM, Philadelphia, 2009.

64. F. K. Nejadasl and R. Lindenbergh. Sequential and automatic image-sequence registration of road areas monitored from a hovering helicopter. *Sensors*, 14(9):16630–16650, 2014.

65. J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, New York, 1999.

66. M. Oskarsson. Regularizing image intensity transformations using the Wasserstein metric. In R. R. Paulsen and K. S. Pedersen, editors, *Image Analysis*, pages 275–286, Cham, 2015. Springer.

67. M. Ovsjanikov, E. Corman, M. M. Bronstein, E. Rodolà, M. Ben-Chen, L. Guibas, F. Chazal, and A. M. Bronstein. Computing and processing correspondences with functional maps. *Proc. SIGGRAPH Courses (SIGGRAPH 2016)*, 2016 (doi: 10.1145/2988458.2988494).

68. I. Podlubny. *Fractional differential equations*. Academic Press, 1998.

69. I. Rocco, , R. Arandjelović, and J. Sivic. Convolutional neural network architecture for geometric matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

70. T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs. Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, 22:730–741, 2003.

71. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions On Medical Imaging*, 18:712–721, 1999.

72. R. M. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Symposium on Geometry Processing*, pages 225–233, 2007.

73. J. N. Sarvaiya, S. Patnaik, and S. Bombaywala. Image registration by template matching using normalized cross-correlation. In *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pages 819–822, Dec 2009.

74. O. Scherzer. *Variational Methods in Imaging*. Springer, 2008.

75. Y. Shi, R. Lai, K. Kern, N. Sicotte, I. Dinov, and A. W. Toga. Harmonic surface mapping with laplace-beltrami eigenmaps. *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 147–154, 2008.

76. H. Sokooti, B. de Vos, F. Berendsen, B.P.F. Lelieveldt, I. Isgum, and M. Staring. Nonrigid image registration using multi–scale 3d convolutional neural networks. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pages 232–239, 2017.

77. E. D. Solomentsev and E. V. Shikin. Laplace–beltrami equation. In M. Hazewinkel, editor, *Encyclopedia of Mathematics*. Springer / Kluwer Academic Publishers, 1994.

78. A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions On Medical Imaging*, 32(7), 2013.

79. M. Tektonidis and K. Rohr. Diffeomorphic multi-frame non-rigid registration of cell nuclei in 2d and 3d live cell images. *IEEE Transactions on Image Processing*, 26(3):1405–1417, March 2017.

80. A. Theljani and Ke Chen. An augmented Lagrangian method for solving a new variational model based on gradients similarity measures and high order regularization for multimodality registration. *Inverse Problems and Imaging*, 13:309–335, 2019.

81. Tony Thompson and Ke Chen. An effective diffeomorphic model and its fast multigrid algorithm for registration of lung CT images. *Computational Methods in Applied Mathematics*, 2019 (doi.org/10.1515/cmam-2018-0126).

82. P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

83. J. Watrous. *Theory of Quantum Information*. Cambridge University Press, UK. (See also https://cs.uwaterloo.ca/∼watrous/LectureNotes.html), 2018.

84. J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *Int. J. Computer Vision*, 45(3):245–264, 2001.

85. T. W. Wong, Lok Ming Lui, P.M. Thompson, and T.F. Chan. Instrinic feature extraction and hippocampal surface registration using harmonic eigenmap. *SIAM Journal of Imaging Sciences*, 5:746–768, 2012.

86. X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Quicksilver: Fast predictive image registration – a deep learning approach. *NeuroImage*, 158:378–396, 2017.

87. I Yanovsky, P Thompson, S Osher, and A Leow. Large deformation unbiased diffeomorphic nonlinear image registration: Theory and implementation. *IEEE conference CVPR' 2007 (see also UCLA CAM Report 06-71)*, 71, 2007.

88. L. Younes. *Shapes and Diffeomorphisms*, volume 171. Springer-Verlag, 2010 (see also 2019 edition).

89. S. Zagoruyko and N. Komodakis. Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

90. D. P. Zhang and Ke Chen. A novel diffeomorphic model for image registration and its algorithm. *Journal of Mathematical Imaging and Vision*, 60:1261–1283, 2018.

91. J. P. Zhang and K. Chen. Variational image registration by a total fractional-order variation model. *Journal of Computational Physics*, 293:442–461, 2015.

92. Jin Zhang, Ke Chen, and Bo Yu. An efficient numerical method for mean curvature-based image registration model. *East Asian Journal on Applied Mathematics*, 7:125–142, 2017.

93. B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

Authors' homepages:

Ke Chen: (`http://www.liv.ac.uk/%7Ecmchenke`

Lok Ming Lui: `https://www.math.cuhk.edu.hk/%7Elmlui/`

Jan Modersitzki: `https://www.mic.uni-luebeck.de/people/jan-modersitzki.html`