Image Stacks as Parametric Surfaces: Application to Image Registration

Birmingham Hang Guan and Anand Rangarajan, *Member, IEEE,*

***Abstract*—We introduce a framework in which a stack of images is considered to be a 2-dimensional parametric surface embedded in a higher dimensional space. This is a simple yet powerful idea, known in the literature but not exploited to its fullest. We discuss the properties of image stacks as parametric surfaces (ISPS), apply this framework to image registration by presenting the image stack surface relative area (ISSRA) registration measure. We show the power of ISSRA as an effective objective function for image registration. Essentially, it shows good performance across a variety of different categories of registration problems: pairwise, groupwise, affine, and non-rigid. Mutual information (MI)—a classical and effective approach for registration—is widely considered to be a good choice for multimodal, pairwise registration while being difficult to extend to the groupwise setting. We discuss the deficiency of MI in the groupwise case from a theoretical point of view, present its connection to ISSRA in the pairwise case, and then show the ready extensibility of ISSRA to the groupwise setting. Experiments and comparisons are performed on different categories of image registration to showcase ISSRA's wide range of applicability to registration problems in practice.**

***Index Terms*—image stacks, parametric surfaces, relative area, area element, image registration, mutual information, congealing**

I. Introduction

IMAGE registration is a classical yet difficult problem in computer vision and medical imaging. Numerous approaches have been introduced over the past 40 years with different problem settings explored. Registration taxonomies carve up the problem space using many different criteria ([65]): pairwise or groupwise, monomodal or multimodal, affine or non-rigid, feature-based or intensity-based, etc. Most approaches have been designed to solve one specific category, and few methods work well across different settings.

Mutual information (MI) ([105], [109], [56]) is one of the most popular approaches mainly designed for pairwise multimodal registration. It is well-known that extending MI to the groupwise setting is non-trivial. Previous work in [92], [9] and [122] have all stated that estimating higher dimensional densities or joint histograms is "difficult," or "computationally impractical." Hence, in the groupwise setting, it is common to see MI approaches convert groupwise registration into a set of pairwise problems ([9], [41], [8]). In fact, the reason for this "deficiency" of MI is straightforward. As we show in Section III, the higher dimensional sample space of multiple images has zero Lebesgue measure, which leads to the higher dimensional joint density not being defined. Hence, the extension of MI to the groupwise case by estimating higher dimensional differential entropies is incorrect. Based on this observation, we present a new and novel approach: the image stack surface relative area (ISSRA) registration measure based on modeling image stacks as parametric surfaces (ISPS). As we show in Section IV, this new objective function works well for a variety of categories of registration problems—including

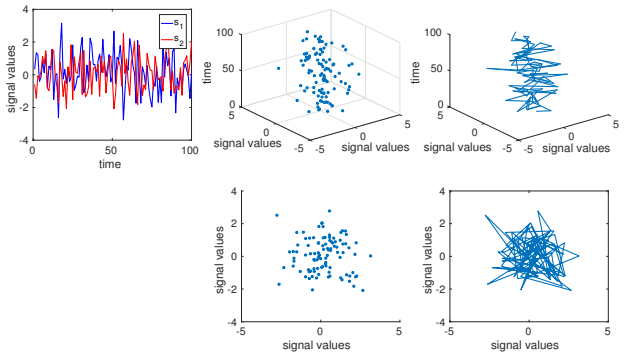


Figure 1: A pair of white noise signals, whose samples sit on a 1D curve. **Upper left**: The two signals $s_1(t), s_2(t)$; **Upper mid**: The 3D scatter plot of the signal pair: for a set of time samples $T = \{t_1, \ldots, t_n\}$, each point in the plot represents a point $(t_i, s_1(t_i), s_2(t_i)) \in \mathbb{R}^3$; **Upper right**: As the time samples become more dense and eventually fill the domain D, the scatter plot becomes a 1D parametric curve embedded in 3D space; **Lower mid**: The scatter plot of the alternative mapping: $t \mapsto (s_1(t), s_2(t))$: each point in the plot represent a point $(s_1(t), s_2(t))$; **Lower right**: Similarly, as the time samples fill the domain, the scatter plot becomes a 1D parametric curve embedded in 2D space.

The authors are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA. E-mail: bkwan,anandr@ufl.edu.

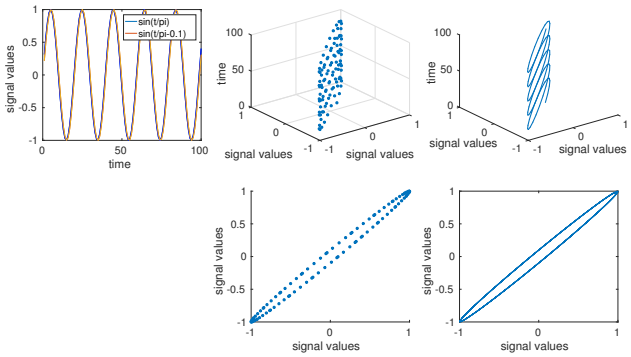


Figure 2: A pair of sine waves: $s_1(t) = \sin(t/\pi), s_2(t) = \sin(t/\pi - 0.1)$. **Upper left**: The two signals $s_1(t), s_2(t)$; **Upper mid**: The 3D scatter plot; **Upper right**: the parametric curve embedded in 3D space; **Lower mid**: The 2D scatter plot of the alternative mapping (disposing the dimension of time); **Lower right**: the 1D parametric curve embedded in 2D space. Compared to Fig. 1, this pair of more similar signals has a joint curve with less parametric arc length than the two white noise signals.

both the pairwise and groupwise setting.

Intuitively, the image stack as a parametric surface (ISPS) model regards a set of images defined on the same image domain as a 2D surface embedded in a higher dimensional Euclidean space. Hence the area of the surface (and related measures) may indicate the similarity between the images. Since human beings are not able to easily visualize higher dimensional spaces, we begin with a simpler problem: the similarity of a pair of 1D signals. (As we show in Section III, even for a pair of images—2D "signals"—the surface is embedded in $\mathbb{R}^4$. But, for a pair of 1D signals, the image of the mapping becomes a curve embedded in $\mathbb{R}^3$. The embedding can therefore be visualized.)

Consider two signals s_1, s_2 (1D functions) mapping time to signal values:

$$s_i : D \subset \mathbb{R}^+ \to R \subset \mathbb{R},$$
$$\text{with } s_i : t \mapsto x_i$$

for $i = 1, 2$, where $t \in D$ is time, and $x_i \in R$ are signal values. (This can also be written as $x_i = s_i(t)$.) This "stack" of signals $\{s_1, s_2\}$ can be considered as a mapping $\mathbf{s}$ from the time domain to 3D:

$$\mathbf{s} : D \subset \mathbb{R}^+ \to R^3 \subset \mathbb{R}^3,$$
$$\text{with } \mathbf{s} : t \mapsto (t, s_1(t), s_2(t)).$$

For each time sample $t \in D$, we have a point $(t, s_1(t), s_2(t))$ in $\mathbb{R}^3$. When we seek to visualize this set of points, we obtain a 3D scatter plot, as shown in Fig. 1. Alternatively, discarding the first dimension (the time values), we have another mapping:

$$\tilde{\mathbf{s}} : D \subset \mathbb{R}^+ \to \mathbb{R}^2,$$
$$\text{with } \tilde{\mathbf{s}} : t \mapsto (s_1(t), s_2(t))$$

which results in a 2D scatter plot (see Fig. 1). Typically, these scatter plots are regarded as sets of samples from 2D or 3D joint distributions, and are used to estimate joint densities and even differential entropies or MI. However, note that s_1, s_2 are functions of t, which is a 1D variable. Assuming that s_1, s_2 are differentiable almost everywhere, no matter how densely we sample in the time domain, these samples will never fill the 2D or 3D space (see Fig. 1). In fact, since the above mappings $t \mapsto (t, x_1, x_2)$ or $t \mapsto (x_1, x_2)$ are actually parametric curves, the sampled points will finally become the mapped curves embedded in 2D or 3D space. And hence, it is incorrect to estimate their joint densities, since the set containing all samples has zero measure.

Fortunately, since the samples lie on a curve, although joint densities cannot be estimated (since they don't exist), there are other available geometric quantities. The arc length is one reasonable choice. Assuming that s_1 and s_2 are close (as in the sine waves case above), the curve looks very smooth (see Fig. 2). When s_1 and s_2 are totally different (as in the different white noise case), the "joint curve" looks very entangled (see Fig. 1). The arc length is higher in the latter than in the former. This observation implies that the arc length of the joint curve may serve as a measure of similarity between signals. The principle can be extended to 2D images. In Section III, we explore these principles more rigorously through a theoretical perspective, and show that higher dimensional joint densities cannot be estimated from image intensities. Here, the surface area of ISPS, i.e. the "arc length" for two-dimensional signals, can serve as a measure of image similarity.

We preliminarily designed ISSA (image stack surface area) in our previous work ([31]), and applied it to affine monomodal registration. However, ISSA has its disadvantage, and is hard to be applied to non-rigid and multimodal registration. After re-examining [71] and analyzing the properties of the underlying model ISPS, we were able to design a better registration measure—ISSRA—in this work. We also discovered the strong connection of ISPS to the model of MI, which supports ISSRA being a good registration objective function, and indicates further potential in ISPS.

This paper is organized as follows: Section II gives a brief introduction to previous approaches in each category of registration, and specifies the scope of this paper. Section III sets up the definitions, works out important properties of ISPS and ISSRA, and discusses the connection between ISSRA and MI. Section IV showcases em-

pirical results for several image registration categories, comparing ISSRA with MI and/or congealing (CG). Section V concludes with a discussion and speculates on the potential of the ISPS model. Supplemental materials are included beginning with Section VI.

II. Previous Work

Over the past few decades, an enormous amount of work has contributed to image registration and from a variety of perspectives, including novel metrics for difficult settings, new deformation models, optimization strategies, feature selection, etc. Since this paper focuses on a new intensity-based metric, recent developments focusing on other aspects of registration are not discussed in detail. This includes feature-based registration ([86], [118]), new deformation models ([90]), and deep learning approaches. With the resurgence of interest in deep learning at the present time in the worlds of computer vision and image processing, many researchers have begun applying deep neural networks (DNN) to image registration. For example, the works in [38], [111] learn new features using DNNs for feature-based registration. Also, the approaches in [116], [25] employ DNNs to learn deformation fields. However, since these approaches are not related to (and are in fact complementary to) metrics for *intensity-based registration*, they are not the focus of this paper. (More discussion on DNN-based approaches is available in the Supplemental Material.)

A. Pairwise registration

Prior to mutual information [105], [109], [56], most intensity-based approaches used the sum of squared differences (SSD) of intensities (or the correlation coefficient) as the registration measure. The main drawback here is the assumption that the intensities of registered images are the same (except for noise) which is inappropriate in the multimodal setting. MI computed the joint entropy of two images thereby circumventing the SSD measure. This was novel and powerful for both monomodal and multimodal problems and had good noise performance. MI is considered to be an intensity-based image registration approach since the focus was on a new registration measure and not feature extraction.

After MI was put forth for registration, many articles followed. These extended MI in various ways, applied MI to different related problems, or discussed the nature of MI [99]. The work in [94] introduced a normalized version (NMI) (which turned out to be related to an information metric [121]), while [12] extended it to a modified entropy by introducing densities related to overlap regions in order to improve overlap invariance. Different density estimators were attempted in [120], [71], [72]. Validation and evaluation in the medical imaging domain became quite important [61]. There are also approaches using entropies other than the Shannon entropy [52]. For example, the Jensen-Shannon divergence was used in [52], the Renyi entropy in [34] and the cumulative residual entropy based on the cumulative joint distribution functions in [75]. Although introduced two decades ago, MI and its variants are still the state of the art in the pairwise, multimodal setting. Recent software suites ([46]) still employ MI as their principal registration measure for pairwise, multimodal registration.

In addition to MI, other approaches using region-based methods to perform registration and segmentation [117] and Kullback-Leibler distances requiring human annotation [15], [32] have followed. Approaches going beyond the basic paradigm of information-theoretic intensity-based image registration include the use of intensity gradients for multimodal registration [33], inner products of densities as the similarity metric [23], [24] and the introduction of neural networks (multi-layer perceptrons and not the current deep learning stacks) into the realm of image registration [55], [84].

B. Groupwise registration

Congealing (CG) is another well-known registration approach [49], [63], mainly designed for groupwise image registration. CG attempts to minimize the sums of entropies of pixel stacks of a set of images. The authors originally used CG on binary images (MNIST) and claimed that CG did not work well on gray-scale images and was inappropriate in the multimodality setting. Furthermore, due to the limitation of using stack entropies, it only works well when the number of images is large. The work in [39] extended CG to natural images by applying it to feature-based registration: SIFT features of each image and at every pixel location are divided into different clusters with the cluster number used to compute the stack entropy. A following paper [38] extends this idea using deep neural networks to learn the features instead of simply adopting SIFT features. This allowed CG to move in the direction of feature-based image registration belying its intensity-based origins.

However, there exist other works claiming that CG can be directly used on grayscale intensities [5], [123], [93]. The work in [5] claims that CG can be applied in the intensity-based, non-rigid groupwise setting where the number of images are still small. They use the so-called "Congealing" objective function (henceforth referred to as BGW-CG)—essentially, the aggregation of differences between Gaussian kernels on image intensities—clearly not a good choice for multimodal problems.

There are other groupwise approaches besides CG. The approach in [9] tried to apply NMI to groupwise

registration. However, because of the complexity of computing high-dimensional histograms for joint density estimation (of a set of images), they resorted to the summation of pairwise joint densities between each image and a reference image as the objective function. In contrast, high-dimensional entropies are efficiently computed for groupwise registration in [122] but the clear improvement over using pairwise entropies has not been subsequently validated [107]. In fact, other than Congealing, most other techniques convert the groupwise problem into an aggregation of pairwise problems [9], [41], [8] with the simplest approach being the SSD between each pair of intensity images [102], [101]. Recently, more papers on groupwise registration have appeared. However, most of these still use the summation of squares of differences (or distances) between images as the metric [60], [95], [2], [107], or use the summation of variants of MI [69].

C. Non-rigid registration

One of the earliest works on non-rigid registration adopted B-splines as the deformation model [96]. Subsequent work focused on improving quality and speed [80], [78]. In addition to B-splines, thin-plate splines and other radial basis functions are also popular deformation models [79], [74], [28]. Spatial frequency basis functions [3] and Fourier series [13], etc. have seen application. The work in [29] introduced novel regularization terms that preserve global and local topological structures.

Most older work [3], [79], [40] continued to use SSD (or related) as the similarity metric. MI and Congealing have also been used for pairwise [1], [47] and groupwise [5] non-rigid registration respectively.

D. Other related work

We are not the only researchers treating a stack of images as a parametric surface. The essential idea has been introduced long ago but mostly forgotten [89], [45], [44]. However, this previous work is mainly focused on image denoising and enhancement and not image registration. As far as we know, there does not exist any previous work discussing the relationship between the parametric surface model and MI.

III. Regularized Area of Image Stack Parametric Surfaces

A. An Image Stack as a Parametric Surface

Let $\Omega \in \mathbb{R}^2$ be the domain of an image I. I can then be represented as a mapping $I : \Omega \to \mathcal{I} \in \mathbb{R}$, where $\mathcal{I}$ is the set of image intensities. For example, typically, Ω is set to $[-1, 1] \times [-1, 1]$ and $\mathcal{I}$ is set to $[0, 1]$. It is natural to consider an image I as a *differentiable* function

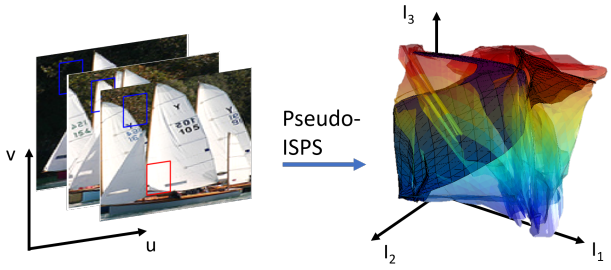


Figure 3: An example of a pseudo-ISPS indicating the intuition behind ISPS: A stack of three images defined on the (u, v) plane is mapped to $\mathbb{R}^3$ by the mapping $(x_1, x_2, x_3) = (I_1(u, v), I_2(u, v), I_3(u, v))$. The mapped graph is a 2D surface, but not a differentiable manifold. It contains self-intersections due to the non-bijectivity of the pseudo-ISPS. The image of an ISPS is differentiable, but is embedded in a higher dimensional space, which is difficult to visualize.

from Ω to $\mathcal{I}$. In this way, the graph of the image I can be considered as a two dimensional surface embedded in $\mathbb{R}^3$: to represent it parametrically, we have $S : \Omega \to \mathbb{R}^3$ so that $S(u, v) = (x(u, v), y(u, v), z(u, v))$ where $x(u, v) = u$, $y(u, v) = v$, and $z(u, v) = I(u, v)$. The above model of an image is a standard device in image processing and analysis.

Similar to the works in [49] and [89], it is also natural to generalize the above basic image model to an image stack: consider a set of images of the same size $\{I_1, I_2, \ldots, I_N\}$. The images can be defined on the same domain Ω, and we can stack them so that at each pixel location $(u, v) \in \Omega$, we get a vector $(I_1(u, v), I_2(u, v), \ldots, I_N(u, v)) \in \mathbb{R}^N$. Adding the dimensions of $x(u, v) = u$ and $y(u, v) = v$, we can model this image stack as a mapping $S : \Omega \to \mathbb{R}^{N+2}$. Representing it as a parametric surface, we have

$$S(u, v) = (x_1(u, v), x_2(u, v), x_3(u, v), \ldots, x_{N+2}(u, v))$$

where

$$\begin{cases} x_1(u, v) = u \\ x_2(u, v) = v \\ x_3(u, v) = I_1(u, v) \\ x_4(u, v) = I_2(u, v) \\ \vdots \\ x_{N+2}(u, v) = I_N(u, v) \end{cases} . \tag{1}$$

Eq. (1) corresponds to an image stack parametric surface (ISPS) of a set of images $\{I_1, I_2, \ldots, I_N\}$.

Fig. 3 provides geometric intuition. Please note that even for a pair of images, ISPS maps them into $\mathbb{R}^4$, which cannot be easily visualized. We think the pseudo-

ISPS (as seen in Fig. 3) is a good choice for presenting the basic intuition. A pseudo-ISPS is superficially similar to an ISPS, except that the first two dimensions $x_1(u,v) = u$ and $x_2(u,v) = v$ are dropped, With this simplification in place, we are able to show the mapping of three images in $\mathbb{R}^3$. To avoid confusion, we emphasize that a pseudo-ISPS is not a bijection, and cannot lead to the definition of ISSA (and subsequently ISSRA).

The motivation for introducing ISPS was laid out in the Introduction. Since ISPS leverages the notion of the image stack as a 2D surface embedded in a higher dimensional Euclidean space, its surface properties can be of use in image registration. Specifically, the underlying intuition is that the surface can be expected to be more smooth when the images are registered and rougher otherwise. Properties like surface area and curvature can therefore be computed and utilized as image similarity measures.

Without loss of generality, we extend S from Ω to $\mathbb{R}^2$ by extending each I_i smoothly (by interpolation) to $\mathbb{R}^2$. We set $I_i(u,v) = 0$ for $(u,v) \in \mathbb{R}^2 \backslash (\Omega \cup \Omega_\epsilon)$, for each $i \in \{1,\dots,N\}$, where Ω_ϵ is a set outside Ω where I_i's values are interpolated so that I_i is smooth on $\mathbb{R}^2$. By defining this extension, we are able to rigorously prove the following proposition and corollary (with all proofs relegated to the Supplemental Material).

Proposition 1. *Assuming that I_i $(i = 1,2,\dots,N)$ is smooth, S is a diffeomorphism.*

Corollary 2. *$S(\mathbb{R}^2)$ is a 2-dimensional submanifold of $\mathbb{R}^{N+2}$.*

Note that since we are only interested in $S(\Omega)$, in the following we only discuss the subset

$$\mathcal{S} = S(\Omega)$$

instead of the whole submanifold $S(\mathbb{R}^2)$ and use the word "manifold" to represent it, since this does not cause any confusion.

B. Image Stack Surface Area

Assuming that each image I_i for $i = 1,2,\dots N$ is smooth, $\mathcal{S}$ is a 2D smooth manifold with the atlas $\{(\mathcal{S},\phi)\}$ containing the single global chart $(\mathcal{S},\phi)$. Since the embedding of $\mathcal{S}$ in $\mathbb{R}^{N+2}$ is available by the mapping S, i.e. for each point $p = S(u,v) \in \mathcal{S}$, its coordinate in $\mathbb{R}^{N+2}$ is available as $(u,v,I_1(u,v),\dots,I_N(u,v))$, ϕ is naturally determined as a trivial bijection between p's local coordinate (x^1,x^2) and $(u,v) \in \mathbb{R}^2$.

We can introduce the Riemannian metric tensor $g = \sum_{i,j} g_{ij} \mathrm{d}x^i \otimes \mathrm{d}x^j$ where

$$g_{ij} = \langle \big(\frac{\partial}{\partial x^i}\big)_p, \big(\frac{\partial}{\partial x^j}\big)_p \rangle$$

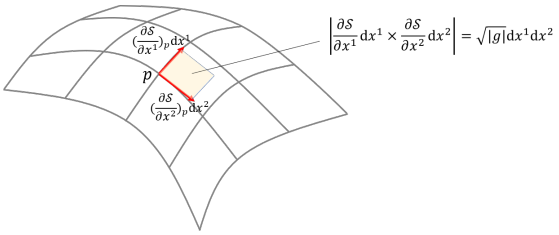


Figure 4: **Area computation of a 2D Riemannian manifold:** Given the Riemannian metric tensor, the area element (AE) at each point p of a 2D Riemannian manifold $\mathcal{S}$ can be computed with respect to the local coordinates (x^1,x^2). In the case of ISPS, the bijection of (x^1,x^2) and (u,v) is available. Thus, ISSA can be computed with respect to (u,v).

for $p \in \mathcal{S}$, $i,j = 1,2$, so as to define $\mathcal{S}$ as a Riemannian manifold. And by the bijection of local coordinates (x^1,x^2) and (u,v), we can easily represent g as a 2×2 matrix as follows:

$$g_p = \begin{pmatrix} \big(\frac{\partial}{\partial u}\big)^2 & \frac{\partial}{\partial u}\frac{\partial}{\partial v} \\ \frac{\partial}{\partial u}\frac{\partial}{\partial v} & \big(\frac{\partial}{\partial v}\big)^2 \end{pmatrix}_p .$$

Representing p as $(u,v,I_1,I_2,\dots,I_N)$, we have

$$g_{(u,v)} = \begin{pmatrix} 1 + \sum_{i=1}^N (\frac{\partial I_i}{\partial u})^2 & \sum \frac{\partial I_i}{\partial u}\frac{\partial I_i}{\partial v} \\ \sum \frac{\partial I_i}{\partial u}\frac{\partial I_i}{\partial v} & 1 + \sum_{i=1}^N (\frac{\partial I_i}{\partial v})^2 \end{pmatrix} \quad (2)$$

which can be directly computed from the image intensities.

With the definition of the Riemannian metric in place, the area of $\mathcal{S}$ can be computed (please see Fig. 4). We define the image stack surface area (ISSA) as

$$\mathrm{ISSA}(I_1,\dots I_N) = \int_\Omega \sqrt{|g_{(u,v)}|}\mathrm{d}u\mathrm{d}v$$

where $\sqrt{|g_{(u,v)}|}\mathrm{d}u\mathrm{d}v$ is called the area element (AE) at point (u,v). And numerically, ISSA can be computed as

$$\sum_{u,v} \left(\sqrt{\left| 1 + \sum_{i=1}^N (\frac{\partial I_i}{\partial u})^2 + \sum_{i=1}^N (\frac{\partial I_i}{\partial v})^2 + \mathrm{CT} \right|} \right)_{(u,v)} \quad (3)$$

with the cross term (CT)

$$\mathrm{CT} = \sum_{i=1}^N \sum_{j=1}^N (\frac{\partial I_i}{\partial u})^2 (\frac{\partial I_j}{\partial v})^2 - \sum_{i=1}^N \sum_{j=1}^N \frac{\partial I_i}{\partial u}\frac{\partial I_i}{\partial v}\frac{\partial I_j}{\partial u}\frac{\partial I_j}{\partial v}.$$

Proposition 3. *$\mathrm{CT} \geq 0$, and the "=" holds iff $\frac{\partial I_i}{\partial u}\frac{\partial I_j}{\partial v} = \frac{\partial I_j}{\partial u}\frac{\partial I_i}{\partial v}$ for each $i,j = 1,\dots,N$.*

Since CT is non-negative, we have proven that $g_{(u,v)} \geq 0$, and henceforth we get rid of the absolute

value symbol. Assuming the gradient components of both images are non-zero, $\mathrm{CT} = 0$ implies that the gradient orientations of every pair of images are parallel (or anti-parallel).

C. Image Stack Surface Relative Area

We now introduce a regularized version of ISSA, wherein the surface areas of the individual images are included to "calibrate" the overall surface area. Later, we show that this regularized version leads to improved registration accuracy and has a closer relationship to MI than ISSA. For the purposes of regularization, we define the "stronger" version of ISSA, i.e. ISSRA, by including the individual image surface areas:

$$\mathrm{ISSRA}(I_1, \ldots, I_N) = \int_\Omega \frac{\sqrt{g_{(u,v)}}}{\sqrt{N}\prod_{i=1}^{N} \sqrt[N]{\mathrm{MAE}_i(u,v)}} \mathrm{d}u\mathrm{d}v \tag{4}$$

where $\mathrm{MAE_i}$ indicates the modified area element of the ith image, and is defined as

$$\mathrm{MAE}_i = \sqrt{\frac{1}{N} + (\frac{\partial I_i}{\partial u})^2 + (\frac{\partial I_i}{\partial v})^2} \tag{5}$$

For convenience, the integrand is referred to as the RAE (relative area element).

As we know, each single image I has its surface area element $\mathrm{AE} = \sqrt{1 + (\frac{\partial I}{\partial u})^2 + (\frac{\partial I}{\partial v})^2}$. By placing the geometric average of the AE of each single image in the denominator, the surface area is regularized so that it will not decrease when each single layer of the images has less area (goes flatter). The factor of $\sqrt{N}$ and the modification to AE by replacing 1 by $\frac{1}{N}$ are aimed at regularizing the surface area so that the absolute value will not be affected by the dimension (the number of images).

Proposition 4. *RAE reaches its global minimum at a point (u, v) when $I_1(u, v) = I_2(u, v) = \cdots = I_N(u, v)$. And*

$$\min_{I_1,\ldots,I_N} \mathrm{RAE}(u, v) = 1.$$

From Proposition 4, we know that RAE reaches its global minimum when all images are identical, and in this case, ISSRA is the area of the image domain, which is the case when the surface become a subset of the plane of $I_1 = \cdots = I_N$ while each dimension shrinks by $\sqrt{N}$ except the first two. Proposition 4 also shows that ISSRA can serve as a measure of image similarity (since we can subtract one from it or use logarithms to get an image similarity measure).

D. Applying ISSRA to Image Registration

In this subsection, we mention some practical issues when applying ISSRA as an objective function to image registration and briefly describe the computational details of the optimization procedure.

Generally speaking, an image registration problem is an optimization problem:

$$\min_{T_1,T_2,\ldots,T_n} \mathrm{Obj}(I_1(T_1(u, v)), \ldots, I_n(T_n(u, v)))$$

where $I_1, \ldots, I_n$ are images to be registered while $T_1, \ldots, T_n$ are transformations applied to each image. Note that in pairwise registration where $n = 2$, I_1 is usually the fixed image and $T_1(x, y) = (x, y)$ is the identity transformation.

In our approach, the optimization problem becomes

$$\min_{T_1,T_2,\ldots,T_N} \mathrm{ISSRA}(I_1(T_1(u, v)), \ldots, I_N(T_N(u, v)))$$

i.e.the objective function becomes the relative area of the stack of transformed images. For affine registration, the transformations are affine transformation matrices, i.e.

$$T(u, v) = \begin{pmatrix} au + cv + e \\ bu + dv + f \end{pmatrix}.$$

For non-rigid registration, we adopted the cubic B-spline model as the transformations, i.e.

$$T(u, v) = (u, v) + \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(s)B_l(t)\phi_{(i+k),(j+l)}$$

where $i = \lfloor u \rfloor - 1, j = \lfloor v \rfloor - 1, s = u - \lfloor u \rfloor, t = v - \lfloor v \rfloor$. B_k and B_l are cubic B-spline basis functions, and $\phi_{i,j}$ are the control points. The B-spline term is called the displacement vector at (u, v). We also include the regularization terms alongside with the ISSRA metric for the non-rigid case, to restrict the deformation to be closer to the identity map during optimization. The regularization term is

$$\sum_N \lambda \int_\Omega \left[(\frac{\partial^2 \mathbf{Du}}{\partial u^2})^2 + 2(\frac{\partial^2 \mathbf{Du}}{\partial u \partial v})^2 + (\frac{\partial^2 \mathbf{Du}}{\partial v^2})^2 \right] \mathrm{d}u\mathrm{d}v$$

where $\mathbf{Du}$ is the displacement vector field over Ω, and λ is a parameter to adjust the relative values between ISSRA and the regularization terms for the optimization. We typically use gradient descent for optimization (and in the affine case this is augmented by a "quick and dirty" brute force search of the parameters). To obtain the results in this paper, we use the popular KNITRO optimization package ([11]) with numerical gradients.

As a registration metric, the most noticeable advantage of ISSRA is its broad suitability: it works for both

pairwise and groupwise cases, and both monomodal and multimodal settings. In Section IV, we provide different categories of registration results to show the practical effectiveness of ISSRA. Unlike MI (see Section III-E1), ISSRA is well-defined for any number of images N. Hence, it can be applied to both pairwise and groupwise registration. For its suitability for multimodal registration, there are mainly three reasons:

(1) It does not depend on the differences between elementary function values of intensities. Most metrics have the form of $\|f(I_1) - f(I_2)\|^2$, where f is an elementary function. Clearly these kinds of metrics directly depend on the intensities of images and thus do not work well in the multimodal case. ISSRA does not share such a structure.

(2) Algebraically, ISSRA is closely related to MI (see Section III-E). In each "monotonic" region, ISSRA and MI only differ by a logarithm and constants (and a reciprocal since we minimize ISSRA while maximizing MI). Furthermore, the MI model can be considered as a "compression" of the ISPS model where the monotonic regions are possibly mapped to overlapping regions by taking away the u, v dimensions. Thus, the density (the "counting" of pixels in each region) when estimating MI can be considered as the summation of areas of each monotonic region mapped to the region where the density is estimated.

(3) Geometrically, ISSRA matches the image level sets. ISSRA minimizes the aggregation of image derivatives, as well as the CT [see Eq. (3)]. CT can be transformed to

$$\begin{aligned} \mathrm{CT} &= \sum_{i,j} \frac{\partial I_i}{\partial u}\frac{\partial I_j}{\partial v}\left(\frac{\partial I_i}{\partial u}\frac{\partial J_i}{\partial v} - \frac{\partial I_j}{\partial u}\frac{\partial I_i}{\partial v}\right) \\ &= \sum_{i,j} \frac{\partial I_i}{\partial u}\frac{\partial I_j}{\partial v} \nabla I_i \wedge \nabla I_j \end{aligned}$$

which contains the aggregation of wedge products of each pair of image gradients. By minimizing the wedge products of gradients, the angles between level sets are minimized. Level set matching contributes to the matching of image structures, instead of directly matching intensities or gradient moduli, and hence, is a popular strategy for multimodal registration ([103], [71]).

The computational complexity of ISSRA is $O(mnN)$, where $m \times n$ is the size of each image. This is the same as BGW-CG ([5]), and is therefore comparable to other groupwise approaches [9], [41], [69]. It is also worth mentioning that ISSRA is computed locally. In non-rigid registration, the numerical gradients of ISSRA with respect to the B-spline parameters can be estimated locally to accelerate the computation: the perturbation of a B-spline parameter only affects its local region, thus, only the perturbation of ISSRA for this local region is needed, which is not true for global measures like MI. As the resolution of the B-spline control points increases, the difference between the practical computation time of ISSRA and MI will become even more significant.

Practically, as an approach which is based only on image gradients, Gaussian filters should be applied before gradient computation. As we know, most image registration problems lead to the optimization of heavily non-convex objective functions. Therefore, an important consideration of a registration objective is the shape of the objective with respect to transformations. In Fig. 5, 6, we show different objective functions w.r.t. a single rotation parameter with the transformation applied to one moving image. In this comparison, a variant of ISSRA is used, defined as

$$\mathrm{ISSRA}_{\mathrm{variant}}(I_1, I_2) = \int_\Omega \frac{\sqrt{|g_{(u,v)}|}}{\sqrt{2}\prod_{i=1}^{2}\mathrm{MAE}_i(u,v)}\mathrm{d}u\mathrm{d}v$$

where the nth square roots of each MAE in the denominator are removed. And a gradient matching objective function is used, defined as

$$\mathrm{GM}(I_1, I_2) = \sum_{u,v}\sqrt{(\frac{\partial I_1}{\partial u} - \frac{\partial I_2}{\partial u})^2 + (\frac{\partial I_1}{\partial v} - \frac{\partial I_2}{\partial v})^2}.$$

It is clear that NMI has the best shape. Although ISSRA does not look as good as we expected for a wide range of rotation angles, it is very smooth around the global minimum. Intuitively, ISSRA has a more concave shape because it contains square roots in its expression. And we can see that the variant of ISSRA has a much better function shape, and is closer to MI. However, although the variant of ISSRA has the above good properties and heuristically speaking, a better function shape, we still recommend ISSRA instead of this variant, for ISSRA has a more solid geometric interpretation, and better empirical performance. In the case where large affine transformations need to be applied, we set initial values using a one-iteration global search covering the whole parameter space before the application of gradient descent. In our experiments, after this global search, the initial values of parameters mostly fall into the interval containing the global minimum, where ISSRA has a good enough function shape.

E. ISSRA versus MI

1) MI is inappropriate in the groupwise setting: Consider a pair of images I_1, I_2 defined on the same domain $\Omega \subset \mathbb{R}^2$. As in the ISPS model above, we have a mapping $S' : \Omega \to \mathcal{I} \subset \mathbb{R}^2$ that gives us a pair of pixel values $(I_1(u,v), I_2(u,v))$ at each pixel location $(u,v) \in \Omega$. Considering this pair of pixel values as sampled from a joint density $p_{I_1,I_2}(i_1, i_2)$—that is, the

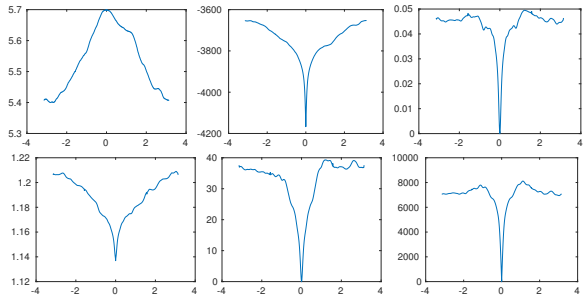


Figure 5: A comparison of different registration objectives. Each figure shows an objective function for a pair of images, where a rotation from $-\pi$ to π is applied to the moving image. **Upper left**: MI; **upper mid**: reciprocal of NMI; **upper right**: ISSRA; **lower left**: variant of ISSRA; **lower mid**: BGW-CG ([5]); **lower right**: gradient matching.

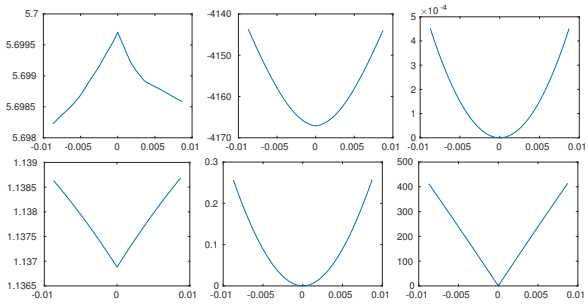


Figure 6: A comparison of different registration objectives, where a rotation from $-0.5\pi/180$ to $0.5\pi/180$ is applied to the moving image. **Upper left**: MI; **upper mid**: reciprocal NMI; **upper right**: ISSRA; **lower left**: variant of ISSRA; **lower mid**: BGW-CG; **lower right**: gradient matching.

density of the random vector (I_1, I_2)—we can estimate the density by computing the 2D histogram or through Parzen windows. And then we are able to compute the joint entropy

$$H(I_1, I_2) = -\int_{\mathcal{I}} p_{I_1,I_2}(i_1, i_2) \log p_{I_1,I_2}(i_1, i_2) \mathrm{d}i_1 \mathrm{d}i_2.$$

Similarly, we can estimate the densities of random variable I_1 (written as $p_{I_1}(i_1)$) and I_2 (written as $p_{I_2}(i_2)$), and their entropies

$$H(I_1) = -\int_{\mathcal{I}_1} p_{I_1}(i_1) \mathrm{d}i_1 \text{ and}$$

$$H(I_2) = -\int_{\mathcal{I}_2} p_{I_2}(i_2) \mathrm{d}i_2,$$

where $\mathcal{I} = \mathcal{I}_1 \times \mathcal{I}_2$. And the mutual information of I_1 and I_2 is defined as

$$\mathrm{MI}(I_1, I_2) \equiv H(I_1) + H(I_2) - H(I_1, I_2).$$

In the notation above, $\mathcal{I}$ denotes the 2D sample space of all (I_1, I_2) (modeled as real vectors so that $\mathcal{I}$ is a compact subset of $\mathbb{R}^2$).

The definitions above do not encounter any difficulty in the pairwise case. However, things dramatically change when MI is extended to 3D, because the joint density of three images *does not exist*. Consider three images I_1, I_2, I_3 defined on Ω. At each pixel location (u, v), the pixel vector is $(I_1(u, v), I_2(u, v), I_3(u, v))$. As shown in [92] and [122], in order to estimate the joint entropy of (I_1, I_2, I_3), we use these pixel vectors as samples to estimate the 3D joint histogram and density. However, the scatter plot of three continuous and differentiable images is a 2D surface embedded in $\mathbb{R}^3$ which is very sparse. When we consider the intensity vectors of the images, we are considering the mapping S' from the image domain Ω to the sample space $\mathcal{I} \subset \mathbb{R}^3$ so that $(u, v) \mapsto (I_1(u, v), I_2(u, v), I_3(u, v))$, i.e. the pseudo-ISPS.

Proposition 5. *The Lebesgue measure of $S'(\Omega)$ is zero, i.e. $m(S'(\Omega)) = 0$.*

By this proposition, it is clear that all the sample points of the random vector (I_1, I_2, I_3) are sitting on a zero-measure subset of $\mathbb{R}^3$, and therefore, the joint differential entropy and mutual information are not defined. This is the direct reason why MI is inappropriate and does not work well for more than two images.

2) The connection between joint entropy and ISSA: First, we pay attention to the two different mappings in ISPS and the mutual information model discussed above for the pairwise case: $S : \Omega \to \mathbb{R}^4$ so that $(u, v) \mapsto (u, v, I_1(u, v), I_2(u, v))$, and $S' : \Omega \to \mathbb{R}^2$ so that $(u, v) \mapsto (I(u, v), I_2(u, v))$. The only difference between the two models arises from the first two dimensions. S contains $x_1(u, v) = u$ and $x_2(u, v) = v$ as the first two parametric equations, which makes it a bijection and hence gives it good topological properties. In contrast, S' is not necessarily a bijection. In fact, the range of S' can be considered as the projection from the range of S to $\mathbb{R}^2$. In this way, intuitively we can imagine that the image of S' embedded in $\mathbb{R}^2$ is the "compression" of the image of S (the two-dimensional surface embedded in $\mathbb{R}^4$). During this "compression," if S is monotonic on the whole of Ω, i.e. S' is injective, the projection is also injective; and if S is not monotonic, each of its monotonic sub-regions are injective, but these projected regions may intersect.

Now, consider only one "monotonic region." Given a pair of images I_1, I_2 on Ω, suppose that on a subset $D \subset \Omega$, both I_1, I_2 are differentiable bijections. Then the mapping $S' : D \to \mathcal{I}$ such that $(u, v) \mapsto (I_1(u, v), I_2(u, v))$ is also a differentiable bijection, and therefore $\mathcal{I}$ is homeomorphic to D, and also a subspace

of a two-dimensional manifold embedded in $\mathbb{R}^2$ (which is just $\mathbb{R}^2$). Hence, under this mapping, we are also able to define the area of $\mathcal{I}$:

$$\mathrm{A}(\mathcal{I}) = \int_D \sqrt{|\det g|}\mathrm{d}u\mathrm{d}v \tag{6}$$

where

$$g = \begin{pmatrix} (\frac{\partial I_1}{\partial u})^2 + (\frac{\partial I_1}{\partial u})^2 & \frac{\partial I_1}{\partial u}\frac{\partial I_1}{\partial v} + \frac{\partial I_2}{\partial u}\frac{\partial I_2}{\partial v} \\ \frac{\partial I_1}{\partial u}\frac{\partial I_1}{\partial v} + \frac{\partial I_2}{\partial u}\frac{\partial I_2}{\partial v} & (\frac{\partial I_1}{\partial v})^2 + (\frac{\partial I_2}{\partial v})^2 \end{pmatrix}.$$

Note that under this mapping, there are no "1"s in the diagonal entries of g. Simplifying Eq. (6), we get

$$\mathrm{A}(\mathcal{I}) \quad = \quad \int_D |\det J|\mathrm{d}u\mathrm{d}v$$

where J is the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial I_1}{\partial u} & \frac{\partial I_1}{\partial v} \\ \frac{\partial I_2}{\partial u} & \frac{\partial I_2}{\partial v} \end{pmatrix}.$$

From the above expression, we see that the area of the mapped image $\mathcal{I}$ can be expressed using the derivatives of I_1 and I_2, in a similar manner as ISSA. We now approach this fact from a different perspective. Let (U, V) be a random vector on D with a uniform distribution:

$$p_{U,V}(u, v) = \frac{1}{A(D)} \cdot 1(u, v),$$

where $A(D)$ is the area of the domain D. Then $I_1 = I_1(U, V)$ and $I_2 = I_2(U, V)$ implies that (I_1, I_2) a random vector under a well-defined random variable transformation. And we can compute its density as follows: given that the mapping $S'(U, V) = (I_1(U, V), I_2(U, V))$ is a bijection,

$$p_{I_1,I_2}(i_1, i_2) = p_{U,V}(h_1(i_1, i_2), h_2(i_1, i_2))|\det J^{-1}|$$

where (h_1, h_2) is the inverse function of (I_1, I_2). By simplifying this expression, we have $p_{I_1,I_2}(i_1, i_2) = \frac{1}{A(D)|\det J|}$. Then we can compute the joint entropy as

$$\begin{aligned} H(I_1, I_2) &= -\int_{\mathcal{I}} p_{I_1,I_2}(i_1, i_2) \log\, p_{I_1,I_2}(i_1, i_2)\mathrm{d}I_1\mathrm{d}I_2 \\ &= \log(A(D)) + \frac{1}{A(D)} \int_{\mathcal{I}} \frac{\log|\det J|}{|\det J|}\mathrm{d}I_1\mathrm{d}I_2. \end{aligned}$$

Through a routine change of variables in the integral, we have

$$H(I_1, I_2) \quad = \quad \log(A(D)) + \frac{1}{A(D)} \int_D \log|\det J|\mathrm{d}u\mathrm{d}v.$$

Given that $A(D)$ is a constant, we obtain the conclusion that in each monotonic region, the joint entropy of I_1, I_2 is proportional to the area of the mapped surface in the sample space (differing by constants and a logarithm operator). As claimed in [72], considering the whole mapping of S' as a collection of small pieces of monotonic regions, i.e. the crossing parallelograms of level sets of I_1 and I_2, the joint entropy can be seen as the summation of the areas of each of these parallelograms. And maximizing mutual information is related to minimizing this area summation.

Being a "compression," the image of S' may be intersected with itself, or even collapse to one-dimension (or a point when both I_1 and I_2 are constants). But the "true" area of this compressed surface can still be computed by segmenting it into small monotonic regions before compression. For example, consider a pixel value vector (ι_1, ι_2) appearing N times in the mapping S', i.e. there exist N points $(u_1, v_1), \ldots, (u_N, v_N)$ such that their images are all (ι_1, ι_2). During the compression (removal of the axes u and v), all these N points collapse into a single point: for each $i = 1, \ldots, N$, $(u_i, v_i, \iota_1, \iota_2)$ becomes (ι_1, ι_2). In this setting, consider a small neighborhood of this point: to compute the full area of this surface, we need to sum the area of the small parallelogram containing (ι_1, ι_2) N times. Intuitively, this is why the joint entropy can be computed through counting the numbers of points falling into each bin.

3) From MI to ISSRA: As we know, minimizing the joint entropy alone is not an ideal approach for image registration. Correspondingly, ISSA alone might not be a good enough objective function as well. MI works better than the joint entropy because the introduction of the two marginal entropies does not permit the final result to just focus on flat background regions. Driven by this intuition, we seek to minimize ISSA while maximizing the area of each image. Consider the full expression of MI:

$$\mathrm{MI} \quad = \quad \int_{\mathcal{I}} p_{I_1,I_2}(i_1, i_2) \log \frac{p_{I_1,I_2}(i_1, i_2)}{p_{I_1}(i_1)p_{I_2}(i_2)}\mathrm{d}i_1\mathrm{d}i_2. \tag{7}$$

In the monotonic region D, similar to the above derivation, we have

$$p_{I_1}(i_1) = \left[A(D)\sqrt{(\frac{\partial I_1}{\partial u})^2 + (\frac{\partial I_1}{\partial v})^2}\right]^{-1}$$

and

$$p_{I_2}(i_2) = \left[A(D)\sqrt{(\frac{\partial I_2}{\partial u})^2 + (\frac{\partial I_2}{\partial v})^2}\right]^{-1}.$$

(For more detailed derivations, please refer to [72]). Substituting these expressions into Eq. (7), we have

$$\mathrm{MI} \quad = \quad \frac{1}{A(D)} \int_\Omega \log \frac{L_1 L_2}{\det J}\mathrm{d}u\mathrm{d}v + \log A(D)$$

where

$$L_1 = \sqrt{(\frac{\partial I_1}{\partial u})^2 + (\frac{\partial I_1}{\partial v})^2}$$

and

$$L_2 = \sqrt{(\frac{\partial I_2}{\partial u})^2 + (\frac{\partial I_2}{\partial v})^2}.$$

From this expression for MI, we see that by introducing the entropies of each single image, we get the $L_1 L_2$ factor as the numerator of the integrand, where L_i is exactly the surface area of image I_i under the MI setting (without the first two dimensions for the mapping). By getting rid of the constants and the logarithm operator, taking the reciprocal of the integrand, and slightly changing the integrand to act as a regularization, we obtain ISSRA, the objective function used in this paper with the individual image areas introduced in a similar manner to MI.

In conclusion, by showing the above connection of ISSRA to MI, we reveal the essential relationship between them. This shows that it makes sense to treat ISSRA as a variant of MI (by changing the problematic mapping in MI to a homeomorphism—ISPS), which is well-defined in both the pairwise and groupwise cases and based on a better mapping model.

IV. Experiments

We show experimental results across different categories of registration problems in this section. In pairwise registration, comparisons are conducted against MI and normalized MI (NMI). In the groupwise case, comparisons are conducted against BGW-CG by Balci *et al.* ([5]), voxel-wise sum of squared differences (VSSD) [107], and conditional template entropy (CTE) [69]. All approaches were implemented in MATLAB® and optimizations were performed using KNITRO ([11]), in order to ensure fair comparison across all experiments.

As discussed in Sections 1 and 2, MI and its enhanced version, NMI are the most widely used approaches, which work especially well for multimodal pairwise registration. The effectiveness of ISSRA can be seen in the MI comparisons. Most of the groupwise approaches (including VSSD and CTE) convert a groupwise problem into summations of pairwise problems, while CG is a well-known approach that solves the problem in a "groupwise" manner. This explains our choice of CG in the groupwise case. However, since the original CG method ([39]) uses SIFT features instead of image pixels, we opted instead for a purely intensity-based version of CG to ensure a fair comparison. BGW-CG is a Parzen window version of CG which strictly speaking makes it similar to other groupwise approaches that use sums of pairwise objective functions. We decided to stick with it due to its CG provenance.

In the case of VSSD, it is claimed [107] that it is derived from Markov-CG where images are assumed to be independent. The work in [107] proposed accumulated pairwise estimates (APE) metrics and showed the relationship between VSSD and APE. Since BGW-CG is already an APE metric, we decided to have comparisons with VSSD, since it is related to BGW-CG. CTE is a new metric for multimodal, groupwise, non-rigid registration, but it is still in the form of summations of pairwise metrics. It introduces a template image computed through principal component analysis (PCA).

In affine registration, the reported errors are the L_2 norms of the differences between the ground-truth and obtained affine transformation parameters. In non-rigid registration, we provide the full, length and angle errors. The errors are computed from the differences between displacement vectors of the ground-truth and obtained results. The full errors correspond to L_2 norms of the error displacement vectors whereas the length and angle errors correspond to the length and angle differences respectively. While the full error reflects the quality of the registration, the length and angle errors are usually also reported in the non-rigid registration literature.

A. Pairwise Registration with Affine Transformations

We performed both monomodal and multimodal experiments in the pairwise affine case. Each experiment contains 10 groups of images, across which the affine parameters and the noise parameters (Gaussian noise standard deviation) vary. Each group contains 10 moving images, with pairwise registration executed on each with respect to a fixed image. The fixed image in the monomodal experiments is a randomly chosen human face image with the moving images generated from it using random affines and noise. As the group number increases, the parameters are sampled from Gaussian distributions with increasing standard deviation, so that the transformation and noise increases. The original image in the multimodal experiments is a randomly chosen MRI image slice from BrainWeb[1] [16]. We have its PD, T1 and T2 images available. In the affine multimodal experiments, the moving images are generated from T1 with the fixed image for registration being PD. The error bar plots of the results are shown in Fig. 7 and Fig. 8. Fig. 11a and Fig. 11b shows some registration examples and comparison with MI.

From the results shown in the figures, we see that in each category, ISSRA performed better than MI on average with ISSRA having more outliers when the difficulty of registration is very high. Empirically, when the error is over 0.1, the obtained image usually has large differences from the fixed image, which indicates a failed registration. From the error values of both MI and ISSRA, we observe that for all groups, both metrics worked well on average. This supports our contention that ISSRA is a competitive approach in pairwise affine

[1](http://brainweb.bic.mni.mcgill.ca/)

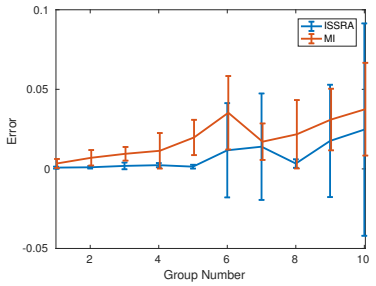


Figure 7: The error bar plot of monomodal pairwise registration with affine transformations. The errors are shown for ten groups of images. As the group number increases, both the affine transformations and the noise become larger.

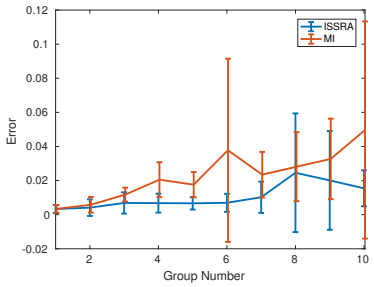


Figure 8: The error bar plot of multimodal pairwise registration with affine transformations. The errors are shown for ten groups of images. As the group number increases, both the affine transformations and the noise become larger.

registration. In most cases, we do not have to perform the global grid search for the initial values of the affine parameters.

B. Pairwise Registration with Non-rigid Transformations

We performed both monomodal and multimodal experiments in the pairwise non-rigid registration case. Each experiment contains 5 groups, each of which contains 5 moving images. All moving images were generated from an original MRI slice randomly chosen from BrainWeb with random thin-plate spline-based non-rigid transformations applied. As the group number increases, the transformation becomes larger. The error bar plots are shown in Fig. 9 and Fig. 10.

Empirically, a full displacement field error less than 3.0 indicates a successful registration. From the results, we observe that in the monomodal case, both NMI and ISSRA worked well for all groups. In the multimodal case, both worked well for the first four groups with Group 5 containing very difficult inputs. In either case, ISSRA performed better for almost all groups on average and with a smaller error standard deviations. The length error bar plots and angle error bar plots show that ISSRA tends to have lower length errors but higher angle errors than NMI. Though angle errors only indicate how image pixels are moved during transformation, and do not affect the registration quality much because of interpolation (if the length errors are less), it might be further improved by carefully choosing better regularization, which will be considered as a potential future work.

In general, the experimental results support our contention that ISSRA is competitive with NMI even in pairwise multimodal registration with non-rigid transformations for which NMI was specifically designed. In Fig. 11c and Fig. 11d, we show anecdotal comparisons between ISSRA and NMI.

C. Groupwise Registration with Affine Transformations

In the case of groupwise affine registration, we have two sets of experiments: monomodal with 10 images in each group with noise; multimodal with 3 images in each group with noise. Each experiment contains 10 groups. For each group, all moving images are registered together with the fixed images of each group. The original image is a randomly chosen MRI slice. Moving images were generated from the original images with randomly sampled affine parameters and noise. As the group number increases, the parameters have larger standard deviation, which implies larger transformations and noise. For the multimodal experiments, we used T1 as the fixed image, while PD and T2 images are used to generate moving images: we registered the PD and T2 images to the T1 image, regarding them as a group. The errors were computed in the same manner as in the pairwise affine case: we compute the errors of the parameters of each resulting image with the fixed image. Errors are shown in Tables I. and II.

From the results of the monomodal experiments, we see that both ISSRA and CG work well for most images as most of their errors are below 0.1. ISSRA is almost always better than CG, except for one obviously failed case in Group 7. In contrast, for VSSD and CTE, it is clear that when the transformation becomes larger, both methods failed starting from Group 7 and Group 5 respectively. This is due to the fact that both of these metrics are computed with respect to template images. For affine registration with very large transformations, it is easy to fall into local minima with the template images computed as a blank image. ISSRA and BGW-CG (which is an APE metric according to [107]) do not suffer from this deficiency. Fig. 11e shows a comparison example.

In the multimodal case, as expected, CG and VSSD do not always work, due to the nature of their objective functions. They sometimes work due to happenstance for

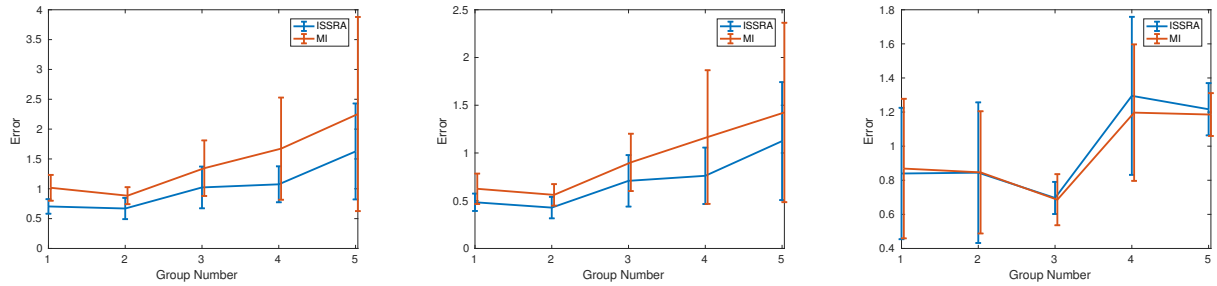


Figure 9: The error bar plot of monomodal pairwise registration with non-rigid transformations. **Left**: full error bar plot; **Mid**: length error bar plot; **Right**: angle error bar plot. The errors are shown for five groups of experiments. As the group number increases, the non-rigid transformations become larger.

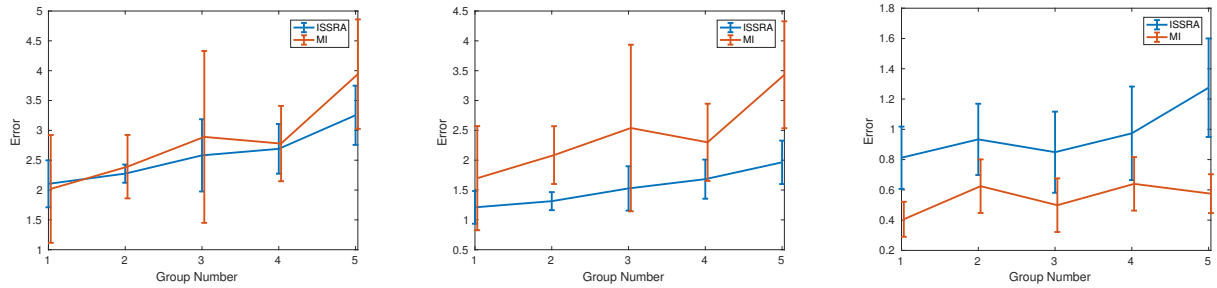


Figure 10: The error bar plot of multimodal pairwise registration with non-rigid transformations. **Left**: full error bar plot; **Mid**: length error bar plot; **Right**: angle error bar plot. The errors are shown for five groups of experiments. As the group number increases, the non-rigid transformations become larger.

Table I: Error table for groupwise monomodal registration with affine transformations and noise

Mean / Std ($\times 10^{-3}$)	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
ISSRA	**3.2/0.74**	**1.1/0.54**	**3.8/0.57**	**1.7/1.4**	**6.8/1.3**	**3.4/1.7**	14.2/1.5	**31.7/90.9**	**10.1/3.4**	**6.2/3.2**
CG	4.3/2.5	5.1/3.8	4.8/3.8	11.3/5.4	16.4/6.9	11.2/5.9	**12.2/3.9**	43.9/25.3	14.5/8.3	16.1/9.0
VSSD	3.6/2.8	5.3/5.2	8.3/3.5	11.3/6.4	21.6/4.8	6.8/2.8	396.5/169.3	101.0/63.4	212.6/95.1	366.4/109.7
CTE	7.3/4.1	21.9/36.2	16.9/5.2	48.1/57.2	101.8/63.3	72.8/74.8	52.5/20.6	115.9/66.7	219.7/132.5	233.1/154.6

Table II: Error table for groupwise multimodal registration with affine transformations and noise

Mean / Std ($\times 10^{-3}$)	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
ISSRA	**8.8/0.60**	**12.8/5.5**	**14.2/0.49**	**10.8/3.5**	**9.0/7.7**	**22.0/0.33**	**6.7/2.5**	**12.0/3.3**	43.9/14.2	**36.7/2.5**
CG	17.5/0.97	718.9/14.5	24.1/4.2	29.0/1.5	607.3/493.5	706.4/193.6	19.4/2.4	41.4/5.6	30.0/3.3	772.2/57.1
VSSD	33.9/6.2	14.4/2.5	16.5/5.1	386.2/202.1	20.0/1.1	626.8/444.0	610.9/20.5	191.2/34.4	311.7/6.3	353.4/203.6
CTE	20.2/0.58	18.7/0.17	19.4/4.4	20.8/4.6	26.5/0.86	53.4/42.9	60.2/20.1	27.4/5.9	**23.5/6.8**	60.1/17.8

several groups when the optimization approach found good local minima from randomly chosen initial conditions. ISSRA and CTE work well, and for most groups ISSRA is better. Fig. 11f shows a comparison example.

D. Groupwise Registration with Non-rigid Transformations

In the groupwise non-rigid case, we also provide both monomodal and multimodal experimental results. Each experiment contains 5 groups with each group of moving images registered simultaneously. The original image is an MRI slice, and thin-plate spline transformations were applied to generate all the moving images. As in the previous cases, the transformations become larger as the group number increases. In the monomodal case, each group has 6 T1 images while in the multimodal case, each group contains 3 T1 images and 2 PD images. The full errors are shown in Tables III and IV. (The length errors and angle errors are shown in the Supplemental Material.)

From the results, we observe that in the monomodal case, all these metrics worked well. ISSRA is signifi-

Table III: Error table for groupwise monomodal registration with non-rigid transformations

Mean / Std	Group 1	Group 2	Group 3	Group 4	Group 5
ISSRA - full errors	**0.95/0.05**	**1.16/0.25**	**1.23/0.32**	**1.43/0.57**	**2.21/1.18**
CG - full errors	1.02/0.17	1.69/0.56	2.11/0.63	1.88/0.52	2.55/1.18
VSSD - full errors	1.28/0.10	1.73/0.43	2.01/0.71	1.95/0.56	2.45/1.31
CTE - full errors	0.97/0.11	1.37/0.19	1.81/0.64	1.80/0.56	2.51/1.39

Table IV: Error table for groupwise multimodal registration with non-rigid transformations

Mean / Std	Group 1	Group 2	Group 3	Group 4	Group 5
ISSRA - full errors	**1.59/0.26**	**1.86/0.38**	**2.45/0.58**	**2.15/0.24**	**2.93/0.47**
CG - full errors	9.12/1.18	11.39/1.42	10.75/1.54	7.02/1.44	6.83/0.46
VSSD - full errors	12.41/1.71	10.64/1.24	10.81/1.03	9.94/1.63	6.52/0.48
CTE - full errors	1.86/0.24	2.49/0.41	3.43/0.53	2.95/0.62	4.46/1.22

cantly better for all groups. In the multimodal case, and similar to the affine experiments above, CG and VSSD do not work. ISSRA and CTE worked well in most groups (CTE did not work well for Group 5), and ISSRA results are significantly better. Fig. 11g and Fig. 11h show comparison examples.

E. Computational Efficiency of ISSRA

In each category of registration, we provide the average iteration time and average registration time for a pair / group of registrations for each metric. All comparisons were done with MATLAB® on a machine with an AMD FX®-8350 eight-core processor, 32 GB memory, and a 64-bit Ubuntu 14.04 operating system (please see Table V). In each entry, the numbers indicate the iteration / registration time. The time of each iteration reflects the practical computational time of each metric. The time shown for the registration reflects the time needed for convergence. For different experimental settings, the convergence time varied. Note that in the groupwise case, since monomodal and multimodal experiments have different group size (multimodal experiments have lesser number of images in each group), the computational time is shown for each experiment.

We observe that in the pairwise cases, ISSRA has better iteration time and registration time. This highlights the fact that estimating densities is slower than computing ISSRA. And ISSRA converges faster than MI. In the groupwise case, ISSRA is the slowest for both iteration time and registration time. This is because its computation is the most complex among the compared metrics, although the theoretical time complexity is the same. (Please note that other metrics have more failed cases where they converged earlier to local minima, especially CG and VSSD for multimodal cases.) These results indicate that more work needs to be done to improve the overall optimization algorithms in terms of speed while maintaining accuracy.

V. Conclusions

We highlight the three core contributions of this paper:

1. Higher dimensional mutual information based on the joint density function is not available, since the sample space of intensity vectors of multiple images (greater than 2 in 2D) has zero Lebesgue measure;

2. ISPS, a simple and powerful image model for a stack of images, is a fundamental approach, overcoming this core limitation;

3. ISSRA can be seen as the "mutual information" under the ISPS model, and it works for both pairwise and groupwise, monomodal and multimodal image registration.

These conclusions not only assert that ISSRA is a good registration approach, but also imply the power and versatility of ISPS. Although this paper mainly focused on image registration, the connection between MI and ISSRA implies the potential of ISPS in other application realms where MI plays an important role, like independent component analysis. Additionally, ISPS is so simple that we can apply it to cases where we have vectors defined on a 2D domain, like image features, color images, or even in image segmentation problems. Although the theory behind ISPS and ISSRA requires us to work with surfaces embedded in higher dimensional spaces, the resulting area computations are very simple and efficient. While this paper mainly focuses on the novel application of ISSRA to registration, future work will focus on improving ISSRA optimization, including decreasing the angle errors by better regularization. We also plan to integrate ISSRA with registration software suites such as SimpleElastix [46] to further improve the accuracy and speed of ISSRA. ISSRA can also be readily combined with deep learning and other regression schemes to register mapped images (derived from the original intensity images). Finally, we believe there are a variety of other applications of ISPS, suitable for future exploration.

Acknowledgments

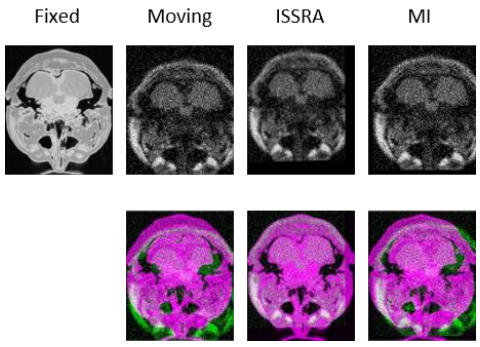
This work is partially supported by NSF IIS 1743050 to A.R. We acknowledge helpful conversations with John Corring, Manu Sethi and Yuan Zhou.

VI. Supplemental Material: A Longer List of Previous Work

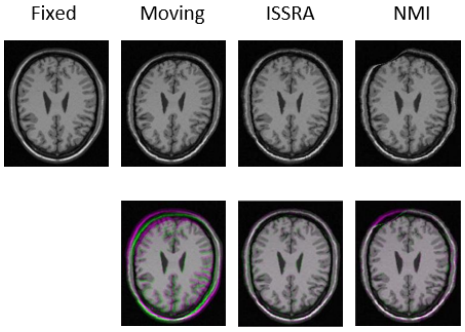
A. A brief history of image registration

As a classical image analysis problem, image registration has a long history. Some old and classic algorithms

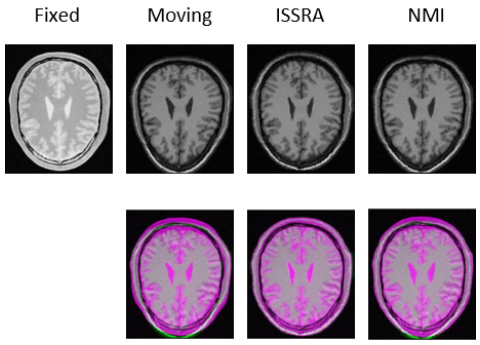
(a) Anecdotal example of pairwise monomodal registration with affine transformation: In this case, the difference images show that ISSRA outperforms MI.



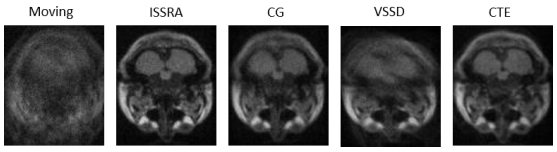
(b) Anecdotal pairwise multimodal registration with affine transformations: In this heavy noise case, ISSRA outperformed MI.



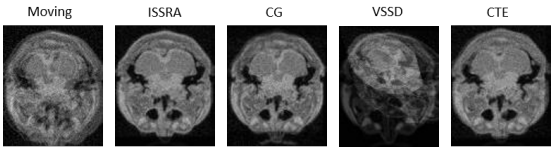
(c) Anecdotal example of pairwise monomodal registration with non-rigid transformations: Chosen to showcase a result where ISSRA has nearly no error, while MI failed in a small region with small displacement errors present.



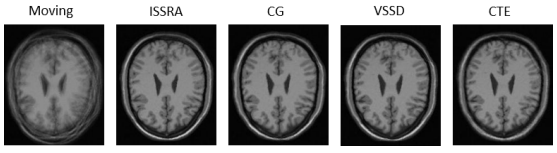
(d) Anecdotal example of pairwise multimodal registration with non-rigid transformations: ISSRA has better overall alignment while displacement errors are present in small regions. MI has better local alignment with worse overall error.



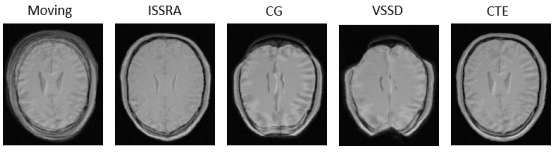
(e) Anecdotal example of groupwise monomodal registration with affine transformations: The original images were not aligned, so the average image is very blurred. ISSRA performs best, CTE is close to ISSRA, but we can observe blurred, local regions. CG is worse than both and VSSD does not work well, especially for the upper half.



(f) Anecdotal example of groupwise multimodal registration with affine transformation: The original images are a T1, T2 and PD. ISSRA shows a good alignment. The CG and CTE results are worse (this is an example where CG happened to fall into the correct local minimum). VSSD does not work.



(g) Anecdotal example of groupwise monomodal registration with non-rigid transformation: The images before registration lead to a blurred average image. In this example, for each approach, one image was not aligned very well, though it is harder to see the differences from the average image. In the Supplemental Material, we provide the piecewise registration results, where obvious differences can be observed.



(h) Anecdotal example of groupwise multimodal registration with non-rigid transformation: The original images before registration contain two T1 and two PD images, where the average image is blurred. The ISSRA result shows best overall alignment, but its local regions are not perfect. CTE has better local alignment, but the locations are heavily distorted compared to ISSRA. CG and VSSD do not work in this case.

Figure 11: Registration Examples. For pairwise examples, the first row shows the fixed images, moving images, ISSRA results and MI results; the second row shows the corresponding difference images with the fixed images. For groupwise examples, we provide the average images of the group before registration and the registration results of each approach (ISSRA, CG, VSSD, and CTE).

Table V: Computational Time

Iter (sec) / Registration (h)	Pairwise Affine	Pairwise Non-rigid	Groupwise Affine		Groupwise Non-rigid	
			Monomodal	Multimodal	Monomodal	Multimodal
ISSRA	**0.27/0.0055**	**10.5/0.43**	12.1/0.31	0.61/0.047	160.2/8.6	128.6/3.8
MI / NMI	1.05/0.029	34.9/1.0	-	-	-	-
CG	-	-	7.6/0.16	**0.31/0.016**	26.5/1.4	19.7/0.77
VSSD	-	-	**4.6/0.12**	0.31/0.019	**23.2/1.3**	**17.6/0.73**
CTE	-	-	6.4/0.09	0.49/0.025	63.0/0.9	52.9/0.56

and approaches can be traced back to as early as the 1970's [6], [59]. The work in [10] is a good survey on image registration for the pre-Internet (before 1995) period. Two main approaches were popular during that period: intensity correlation and landmark matching. Correlation methods use correlations of intensities between two images as the metric for optimization [6], [70], [43], while landmark approaches select specific points as landmarks and compute metrics (like distances of intensities or features) only based on the landmarks [30], [73], [100], [4], [64], [48]. As most landmark approaches use features at the landmark points, this category of image registration approaches is classified in the feature-based approach class.

Starting from 1995, Mutual Information (MI) has become one of the most important approaches in image registration [105], [109], [56]. Before MI, most intensity-based approaches took the sum of squared differences (SSD) at corresponding points in two images as the objective function. The main drawback here is the assumption that the intensities of registered images are the same, hence this is deficient in the multimodal setting. MI computes the joint entropy of two images and circumvents computing the difference of intensities—a novel and very powerful development in both monomodal and multimodal problems, and even for noisy images. Although the work in [56] discussed MI for intensity based, feature based, landmark based, and surface-based approaches, subsequent work mostly extended intensity-based approaches, and thus MI is normally considered as an intensity-based image registration approach.

After the introduction of MI as a registration measure, many articles followed. These extended MI in various ways, applied MI to different related problems, or discussed the nature of MI [99]. The works in [57], [68] are two good surveys for the period when MI was most popular. Among them, the work in [94] improved MI to a normalized version (NMI) while the work in [12] extended it to a modified entropy by introducing densities related to overlap regions to improve overlap invariance. The works in [120], [71], [72] tried to estimate the joint density of images using different approaches. Finally, the works in [121] developed new metrics related to MI while the work in [61] validated the capability and efficiency of MI on medical images.

During this period, there also exist some papers criticizing MI, like the work in [76] which pointed out that MI does not make use of neighborhood information, and the work in [81] which criticized MI for its location independence property. And there are also papers trying to mend these drawbacks of MI, like the work in [42] which put forward a new concept—local MI that is computed at each pixel via a neighborhood of that pixel while [67] combined MI with gradient information and uses the gradients of pixels as weights for MI.

There are also approaches using entropies other than the traditional Shannon entropy for computing MI, like the work in [52] on the Jensen-Shannon divergence, the work in [34] using the Renyi entropy, and the work in [75] using the cumulative residual entropy which replaced the joint density of images by cumulative probability distribution functions. The latter is useful to help move beyond the differential entropy concept but does not directly address the zero measure problem of an image stack.

In 2006, Congealing, another milestone in the history of image registration was put forth [49], [63]—mainly designed for groupwise image registration. Congealing computes the pixel stack entropy of an image stack at each corresponding image location from the group of images to be registered and uses this as the metric for registration. In the original paper, the authors only used Congealing on binary images (MNIST) and claimed that Congealing on gray-scale images did not work very well. In the following year, the work in [39] extended Congealing to natural images by applying it to feature-based registration: they divided SIFT features of each image at every pixel location into different clusters, and use the cluster number to compute the stack entropy. A following paper, [38], extends their idea by using a deep neural network to learn the features instead of simply adopting the SIFT features. By these means, Congealing moved in the direction of feature-based image registration. However, there are also papers claiming that Congealing can be directly used on gray scale intensities, like the works in [5], [123], [93].

Along with MI and Congealing, there are certainly other new and different approaches, like the approach in [117] using region-based methods to do registration and segmentation. The methods in [15], [32] use the Kullback-Leibler distance, the work in [33] makes use of intensity gradients for multi-modal registration, and finally [55], [84] introduce neural networks into the realm of image registration. Besides these various approaches, there are also other papers based on novel image models, like the work in [26] which considers images as sparse approximations of geometric functions. The transformation manifold of images is considered in [106]. The best survey papers on image registration (before the deep learning era) can be found in [91], [65], [114].

B. Categories of registration: Perspectives and taxonomies

Image registration problems can be categorized using different perspectives: (i) intensity-based vs. feature-based when information is extracted from images, (ii) pairwise vs. groupwise when the number of images is increased, (iii) affine vs. non-rigid when the type of transformation is considered, (iv) monomodal vs. multimodal when the type of images are taken into account etc. While intensity-based, pairwise, affine, and monomodal registrations are the default categories, to which most articles in the previous subsection belong, feature-based, non-rigid, groupwise, and multimodal approaches will be particularly discussed in this subsection. And of course, the type of optimization is also an important categorization criterion. However, since in our work, the method of optimization is not a focus, we will not discuss previous work using this categorization. The work in [47] is a good comparison paper on different optimization methods.

1) Feature-based registration: As stated above, feature-based approaches are not the focus in this work. Although feature-based approaches made an impact in the history of image registration [86], [113], [83], and all the newest DNN-based approaches are feature-based [37], they will not be discussed and compared in this paper.

2) Non-rigid registration: In the early years, most papers only discussed affine registration, or even simpler, rigid registration or just rotation. The works in [20], [27], [96], [19], [98] are some of the early papers on non-rigid registration problems. Most of these papers are also on 3D registration, since non-rigid image registration was mostly applied on 3D brain images in the early era. These methods adopted various models for non-rigid transformation, like active contours, local affine deformation, shape, and viscous fluid models [14], etc. The work in [96] is one of the earliest to adopt B-splines as the deformation model with some of the followers [80], [78] making improvements in terms of accuracy or speed. A very useful non-rigid registration software suite is available in [46]. Thin-plate splines, such as radial basis function models are also popular approaches: [79], [74], [28]. Additionally, there exist other models, like spatial frequency basis functions [3], Fourier series [13], shape models [66], and velocity fields [21], etc.

Most older work, like [3], [79], [40], use difference of intensities or their functions (like squares, or weighted sums) as the metrics. In the post-MI era, papers like [1], [47] started to apply MI to non-rigid registration. And the work in [36] discussed the well-posedness of MI as a similarity measure for non-rigid problems. There are also papers applying Congealing to groupwise non-rigid registration [5]. Finally, there are also some papers applying non-rigid registration techniques to other related fields, like optical flow motion estimation [104], or deformation construction [82].

3) Groupwise registration: Groupwise registration is relatively new compared with other registration problems. The work in [9] tried to apply NMI to groupwise registration. However, because of the complexity of computing high-dimensional histograms for joint density estimation of a group of images, they used the summation of pairwise joint densities between each image and the reference image as the objective function. In fact, as we stated in this paper, the difficulty in the high-dimensional histogram is not its computational complexity, but the infeasibility of estimating a high-dimensional density from a 2-dimensional sample space. The work in [122] is another paper that tried to compute high-dimensional histograms for groupwise registration. In fact, other than Congealing, most papers, like [9], [41], [8] used some techniques to convert the groupwise problem to an aggregation of pairwise problems, or even directly used the summation of differences between each pair of images [102], [101].

As we stated above, Congealing [49], [39], [63] is a milestone in groupwise registration. Stack entropy is an idea that is essentially different from summation of differences between pairs of images, or summation of other pairwise metrics. After it was proposed, papers like [18], [58], [5], were proposed trying to improve Congealing and apply it to more complicated registration cases. The work in [17] presented an objective function similar to congealing, but adopted the Renyi entropy to solve groupwise nonrigid registration for a sequence of images whose intensity vectors can be considered sparse. And finally the work in [38] led Congealing to the feature-based registration world.

Recently, even more papers on groupwise registration have appeared. However, most of them still use summation of squares of differences or distances between

images as the metric [60], [95], [2], or use some other, more appropriate feature-based metric [113], [37], [111], [112]. The work in [119] introduced a new idea based on geodesic distance on image manifolds, but its nature was still summation of distances. The work in [108] applied clustering on images to break a groupwise registration problem into a set of pairwise problems and then applied pairwise strategies like MI. The work in [92] still tried to estimate the high-dimensional joint entropy. And the work in [115] used the language of L_2 functions, but the nature of their objective function was still differences between images and the average image as the reference.

4) Modality: As we discussed above, the issue of modality came from medical image registration, especially brain images. The work in [110] is an early and classic paper which proposed a metric that is the sum of standard deviations of pixel values of one modality corresponding to one specific pixel value (or region) of the other modality. It computes the weighted sum of all these standard deviations across all regions of an MRI image, and registered PET images to the MRI images by minimizing this metric. Shortly thereafter, as is widely known, MI was proposed as the most classical approach for multi-modal registration. Obviously, most of the approaches based on differences of intensities cannot be applied to multi-modal problems, as well as any metrics that measures the distances between intensities, like correlation, or most of the groupwise registration approaches discussed in the above subsection. Other than MI, there do exist some following methods, like [33], [35], that tried to solve multi-modal registration problems using techniques like direct gradient matching or some functions based on changes in neighboring pixels (patch distance). However, these methods based on difference of gradients are usually noise-sensitive, and must do pre-processing like heavy Gaussian blurring.

C. Deep learning-based image registration

At the present time, with the increasing popularity of deep learning, many papers apply deep neural networks (DNN) to image registration. These approaches mainly fall into three categories: feature learning, deformation learning, and metric learning.

Feature learning approaches adopt deep neural networks (DNN) to learn features, and use traditional registration approaches (typically feature-based) to do the registration work. The work in [38] adopted restricted Boltzmann machines (CRBM) and convolutional neural networks (CNN) to train features for a group of images, and then applied CG as the metric for registration. The work in [111] and [112] use CNN to learn grid-wise features and applies traditional metrics for registration.

Deformation learning approaches directly learn deformation parameters or non-parametric models from training images to do the registration, with traditional registration metrics not necessarily required. They may use metrics, or directly use distances between features or intermediate layer outputs (labels etc.) as the optimization goal. The approach in [51] learns voxel-to-voxel spatial transformations using fully convolutional networks (FCN) from pairs of images as inputs, and optimize traditional registration metrics. The work in [62] computes features from X-ray images and trains hierarchical regressors using CNN regression models to learn the relation between the features and the transformation. In [88], the transformation parameters are learned using a CNN by directly using intensity images instead of features and non-linear regression to obtain rigid parameters. The surface matching model is adopted in [77] to register segmented regions of interests for cardiac MRI images, by using a CNN to learn the mapping functions from template grid to the surfaces. And the work in [116] uses well-aligned examples to train encoder-decoder structured DNNs in order to learn initial velocity momenta for a pair of input images, and then use the deformations as input to an LDDMMbased registration framework ([7]).

Metric learning approaches train entire DNNs as metrics to replace traditional metrics. The work in [87] uses well-aligned images as training data to a CNN in order to directly train the network to obtain a similarity metric. Here, patches are classified into categories of being similar or different, and later used for registration.

There are other approaches that do not fall into these categories, like the work in [22] which proposed an unsupervised CNN that extracts features from input images and uses regression to obtain deformation parameters. Later, it computes similarity metrics and provides it as feedback for the network, to gradually improve the registration. Finally, the agent-based modeling approach in [53] is unusual in adopting a sequence-of-actions point of view.

To conclude, most deep learning based approaches either improve feature-based approaches by learning features from DNNs, or directly learn deformations from image inputs. These ideas are different from traditional intensity-based registration approaches that focus on metrics, therefore not discussed and compared in this paper. However, most of the DNN approaches actually require similarity metrics, either for features (like feature Euclidean distance), or for intensities (like sum of squared differences, or MI). As a novel metric for both groupwise and pairwise registration, ISSRA—proposed in this paper—has the potential to serve as the similarity metrics for deep learning-based registration. Therefore, combining ISSRA with deep learning based approaches can be seen as a potential future direction.

D. Other related previous work

We are not the first to consider a stack of images as a parametric surface. The works in [89], [45], [44] also proposed models based on this idea. They used this idea to represent color images (each channel of which is a parametric function), or even movies and 3D medical images in which case each movie or 3D image is a parametric function with three variables. However, their models are mainly applied to image denoising and enhancement and not to image registration, and there are no previous works discussing the relationship between the parametric surface model and MI.

We are not the first to use image function graphs as surfaces and consider their areas, either. The work in [85] also used the surface idea and areas as objective functions to do image matching and even classification. However, the key difference between the two approaches is that our images are under the ISPS model, while this work considered two images as two surfaces, and used the product of distances between the two surfaces and their areas as the objective. The work in [54] is another image matching paper that includes the area concept, but similar to [85], they did not consider a stack of images as one single surface and minimize their "joint area." The work in [50] also used metric tensors for image matching, but their metric tensors were computed based on the transformation model. And the work in [97] also uses the image function graph idea, starting from the volume form of the surface of the image graph giving rise to a symmetric objective function but did not consider the ISPS approach.

VII. Supplemental Material: Proofs of Propositions

Proposition 6. *Supposing that I_i $(i = 1, 2, \ldots, N)$ is smooth, S is a diffeomorphism.*

Proof: Without loss of generality, we extend S from Ω to $\mathbb{R}^2$ by extending each I_i smoothly (by interpolation) to $\mathbb{R}^2$ so that $I_i(u, v) = 0$ for $(u, v) \in \mathbb{R}^2 \backslash (\Omega \cup \Omega_\epsilon)$, for each $i \in \{1, \ldots, N\}$, where Ω_ϵ is a set outside Ω where I_i's values are interpolated so that I_i is smooth on $\mathbb{R}^2$.

(a) Since for $i = 1, 2, \ldots, N + 2$, $x_i(u, v)$ is smooth, S is smooth.

(b) S is a bijection, because we can find $S^{-1}(y_1, y_2, \ldots, y_{N+2}) = (y_1, y_2)$ for each point $(y_1, y_2, \ldots, y_{N+2}) \in S(\mathbb{R}^2)$, and of course S is onto $S(\mathbb{R}^2)$.

(c) S^{-1} is also smooth.

Hence, S is a diffeomorphism from $\mathbb{R}^2$ to $S(\mathbb{R}^2)$. ■

Corollary 7. *$S(\mathbb{R}^2)$ is a 2-dimensional submanifold of $\mathbb{R}^{N+2}$.*

Proof: Omitted. ■

Proposition 8. *$\mathrm{CT} \geq 0$, and the "=" holds iff $\frac{\partial I_i}{\partial u}\frac{\partial I_j}{\partial v} = \frac{\partial I_j}{\partial u}\frac{\partial I_i}{\partial v}$ for each $i, j = 1, \ldots, N$.*

Proof:

$$\begin{aligned}\mathrm{CT} &= \sum_{i=1}^{N}\sum_{j=1}^{N}(\frac{\partial I_i}{\partial u})^2(\frac{\partial I_j}{\partial v})^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial I_i}{\partial u}\frac{\partial I_i}{\partial v}\frac{\partial I_j}{\partial u}\frac{\partial I_j}{\partial v} \\ &= \sum_{i<j}(\frac{\partial I_i}{\partial u}\frac{\partial I_j}{\partial v} - \frac{\partial I_j}{\partial u}\frac{\partial I_i}{\partial v})^2 \geq 0\end{aligned}$$

■

Proposition 9. *RAE reaches its global minimum at a point (u, v) when $I_1(u, v) = I_2(u, v) = \cdots = I_N(u, v)$. And*

$$\min_{I_1,\ldots,I_N} \mathrm{RAE}(u, v) = 1$$

Proof: First, we need to prove that the square of the denominator of RAE is no larger than the square of the numerator of RAE, i.e.

$$\begin{aligned}&\textstyle\prod_{i=1}^{N} \sqrt[N]{1 + N(\frac{\partial I_i}{\partial u})^2 + N(\frac{\partial I_i}{\partial v})^2} \\ \leq\ & 1 + \textstyle\sum_{i=1}^{N}(\frac{\partial I_i}{\partial u})^2 + \sum_{i=1}^{N}(\frac{\partial I_i}{\partial v})^2 + \mathrm{CT}\end{aligned}$$

By the AM-GM inequality (inequality of arithmetic and geometric means) we have,

$$\begin{aligned}&\textstyle\prod_{i=1}^{N} \sqrt[N]{1 + N(\frac{\partial I_i}{\partial u})^2 + N(\frac{\partial I_i}{\partial v})^2} \\ \leq\ & \textstyle\sum_{i=1}^{N} \frac{1+N(\frac{\partial I_i}{\partial u})^2+N(\frac{\partial I_i}{\partial v})^2}{N} \\ =\ & 1 + \textstyle\sum_{i=1}^{N}(\frac{\partial I_i}{\partial u})^2 + \sum_{i=1}^{N}(\frac{\partial I_i}{\partial v})^2\end{aligned}$$

Hence, it suffices to prove that $\mathrm{CT} \geq 0$, which holds by Proposition 3. Therefore, we proved that $\mathrm{RAE} \geq 1$. Additionally, by Proposition 3, we know that, when $I_1 = \cdots = I_N$, $\mathrm{CT} = 0$, and RAE reaches its global minimum. ■

Proposition 10. *The Lebesgue measure of $S'(\Omega)$ is zero, i.e. $m(S'(\Omega)) = 0$.*

Proof: As we know, the domain $\Omega \subset \mathbb{R}^2$ is bounded and closed. Since I_1, I_2, I_3 are differentiable, the mapping S' is differentiable, and thus it is uniformly continuous. Therefore, $\forall \epsilon > 0$, $\exists \delta > 0$ such that, if $\forall (u_1, v_1), (u_2, v_2) \in \Omega$, $d((u_1, v_1), (u_2, v_2)) < \delta$, we have $d(S'((u_1, v_1)), S'((u_2, v_2))) < \epsilon$. (*)

Since Ω is compact, for any cover of Ω: $\Omega \subseteq \bigcup_{(u,v)\in\Omega} B((u, v), \delta)$ where $B((u, v), \delta)$ is the ball centered at (u, v) with radius δ, there exists a finite cover such that $\Omega \subseteq \bigcup_{i=1}^{n} B((u_i, v_i), \delta)$. Then we have $S'(\Omega) \subseteq \bigcup_{i=1}^{n} S'(B((u_i, v_i), \delta))$.

By argument (*), for each ball's image $S'(B((u_i, v_i), \delta))$, we have $S'(B((u_i, v_i), \delta)) \subseteq B(S'((u_i, v_i)), \epsilon)$. Then we have that

$S'(\Omega) \subseteq \bigcup_{i=1}^{n} B(S'((u_i, v_i)), \epsilon)$. Therefore, $m(S'(\Omega)) \leq m(\bigcup_{i=1}^{n} B(S'((u_i, v_i)), \epsilon))) = \frac{4}{3}\pi n\epsilon^3$, i.e. $m(S'(\Omega)) = 0$. ■

VIII. Supplemental Material: Experimental Details

All experiments are done on a machine with an AMD FX®-8350 eight-core processor, 32 GB memory, and 64-bit Ubuntu 14.04 operating system. The algorithms are implemented and executed on MATLAB® 2016b. The optimization toolbox is Knitro ([11]), and the optimization algorithm is gradient descent, where the gradients are user-provided and numerically computed.

For the initial values of parameters, in affine experiments (the six affine transformation parameters), for each image, one iteration of global search on all of its six affine parameters were done before the gradient descent procedure, and these initial values were kept for all approaches; in non-rigid experiments (the B-spline coefficients of each control point), all initial values were set to a small perturbation of 0, to avoid 0 being a local minimum.

For the optimization to be better handled, we restricted the parameter boundaries as follows: For affine experiments, among the six parameters, a and d are bounded from 0.5 to 1.5, b and c are bounded from -1.57 to 1.57, and e and f are bounded from -20 to 20, where these parameters a, b, c, d, e, f are defined from the affine transformation matrix

$$T = \begin{pmatrix} a & c & e \\ b & d & f \\ 0 & 0 & 1 \end{pmatrix}.$$

For non-rigid experiments, each parameter is bounded so that the displacement at each point does not exceed 1/5 of the size of the image. For non-rigid experiments, only displacement vectors at the foreground locations were taken into consideration.

IX. Supplemental Material: More experimental results

We also have affine pairwise monomodal and multimodal registration results without noise (please see Figs. 12, 13).

We have two other sets of affine groupwise registration results. One is for the monomodal case, where each group contains three images. The original image is a random face image, and the moving images were generated from it with random affine parameters and noise (please see Table VI). The other is the multimodal case where no noise was applied (please see Table VII).

For the groupwise non-rigid experiments, we provide the length and angle error tables for reference. These errors were computed for the experiments shown in Section 4.4 in the paper (please see Tables III-VI). We observe that though ISSRA usually has better full errors, it tends to have worse angle errors, similar to the pairwise case when comparing with NMI.

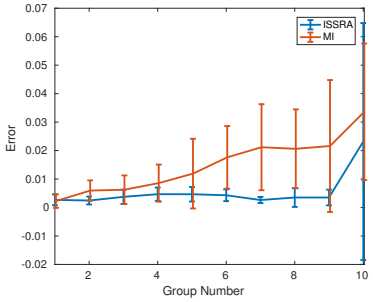


Figure 12: Error bar plot of monomodal pairwise registration with affine transformations and no noise. The errors are shown for ten groups of images. As the group number increases, the affine transformations become larger.

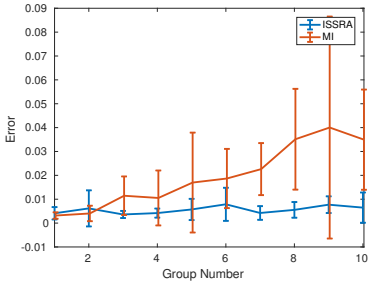


Figure 13: Error bar plot of multimodal pairwise registration with affine transformations and no noise. The errors are shown for ten groups of images. As the group number increases, both the affine transformations and noise become larger.

Fig. 14 shows registration examples for the groupwise monomodal non-rigid cases (please see Fig. 11(g) in the original paper). Since the average image did not show large differences among the approaches, in the following figure we show each image after registration.

References

[1] A. Andronache, M. von Siebenthal, G. Székely, and P. Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical Image Analysis*, 12(1):3–15, 2008.

[2] I. Arganda-Carreras, C. Sorzano, P. Thévenaz, A. Muñoz-Barrutia, J. Kybic, R. Marabini, J. Carazo, and C. Ortiz-de Solorzano. Non-rigid consistent registration of 2D image sequences. *Physics in Medicine and Biology*, 55(20):6215, 2010.

[3] J. Ashburner, K. J. Friston, et al. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266, 1999.

[4] H. S. Baird. *Model-based image matching using location*. ACM Distiniguished Dissertation. The MIT Press, first edition, 1985.

Table VI: Error table for groupwise monomodal registrations with affine transformations and noise

Mean / Std ($\times 10^{-3}$)	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
ISSRA	**0.50/0.19**	**1.3/0.60**	**2.6/2.0**	**4.1/1.5**	**5.3/0.49**	**7.2/0.99**	**9.1/2.6**	13.1/3.9	6.0/0.83	**14.1/2.4**
CG	4.1/1.5	2.9/0.76	6.6/1.1	8.5/1.6	11.8/0.90	28.9/26.6	17.4/0.98	**7.2/6.0**	**5.6/3.2**	22.5/1.8
VSSD	4.1/1.5	3.1/0.45	7.2/1.2	18.5/2.0	13.6/0.69	26.3/26.8	20.3/1.9	7.5/5.5	6.2/1.9	24.0/3.0
CTE	5.5/4.2	8.4/8.3	8.2/1.9	52.8/52.7	39.4/55.3	26.2/11.8	74.7/102.5	70.2/41.8	38.2/23.9	64.2/9.5

Table VII: Error table for groupwise multimodal registrations with affine transformations and no noise

Mean / Std ($\times 10^{-3}$)	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
ISSRA	**9.0/2.9**	**5.9/1.8**	**10.1/3.5**	15.0/0.96	15.7/0.32	**2.8/3.2**	**4.0/4.4**	**6.3/4.2**	**7.3/0.31**	**9.8/6.5**
CG	10.2/5.7	591.2/115.0	500.9/125.3	20.4/5.7	18.6/2.4	35.3/0.28	46.8/43.6	40.6/15.3	782.9/192.0	673.8/47.6
VSSD	16.5/1.5	10.3/3.9	19.9/6.4	**10.8/3.1**	**15.6/5.0**	43.7/10.1	636.8/561.2	251.8/74.3	299.8/156.2	276.7/104.9
CTE	18.3/1.9	8.1/6.8	18.0/17.3	44.0/5.5	19.0/11.0	32.0/1.2	21.8/0.98	46.9/38.4	52.0/5.5	38.2/32.1

Table VIII: Length error table for groupwise monomodal registrations with non-rigid transformations

Mean / Std	Group 1	Group 2	Group 3	Group 4	Group 5
ISSRA - length errors	0.60/0.06	**0.79/0.25**	**0.84/0.32**	**0.94/0.57**	**1.46/1.18**
CG - length errors	**0.59/0.11**	0.99/0.31	1.36/0.52	1.18/0.37	1.56/0.93
VSSD - length errors	0.76/0.09	1.01/0.26	1.33/0.59	1.27/0.45	1.51/1.00
CTE - length errors	0.64/0.09	0.95/0.20	1.30/0.66	1.22/0.37	1.78/1.17

Table IX: Angle error table for groupwise monomodal registrations with non-rigid transformations

Mean / Std	Group 1	Group 2	Group 3	Group 4	Group 5
ISSRA - angle errors	0.72/0.18	0.73/0.24	1.23/0.39	1.08/0.42	1.13/0.09
CG - angle errors	0.74/0.19	0.73/0.21	1.17/0.35	1.11/0.42	1.13/0.10
VSSD - angle errors	0.73/0.17	0.81/0.18	1.18/0.28	**1.06/0.40**	1.15/0.09
CTE - angle errors	**0.65/0.22**	**0.72/0.22**	**1.09/0.30**	1.08/0.54	**1.09/0.06**

Table X: Length error table for groupwise multimodal registrations with non-rigid transformations

Mean / Std	Group 1	Group 2	Group 3	Group 4	Group 5
ISSRA - length errors	**0.95/0.14**	**1.17/0.28**	**1.63/0.66**	**1.34/0.23**	**2.06/0.60**
CG - length errors	6.99/0.62	8.35/0.65	7.01//1.90	3.63/0.73	3.43/0.69
VSSD - length errors	9.87/1.49	7.87/0.47	6.93/1.50	6.06/0.86	3.22/0.69
CTE - length errors	1.34/0.21	1.80/0.47	2.50/0.82	2.18/0.49	2.71/0.73

Table XI: Angle error table for groupwise multimodal registrations with non-rigid transformations

Mean / Std	Group 1	Group 2	Group 3	Group 4	Group 5
ISSRA - angle errors	0.76/0.23	0.82/0.12	0.88/0.17	0.88/0.41	0.92/0.15
CG - angle errors	1.39/0.27	1.47/0.20	1.32/0.21	1.48/0.15	1.09/0.18
VSSD - angle errors	1.56/0.17	1.53/0.24	1.28/0.18	1.75/0.20	1.03/0.16
CTE - angle errors	**0.41/0.12**	**0.45/0.14**	**0.47/0.16**	**0.59/0.26**	**0.80/0.16**

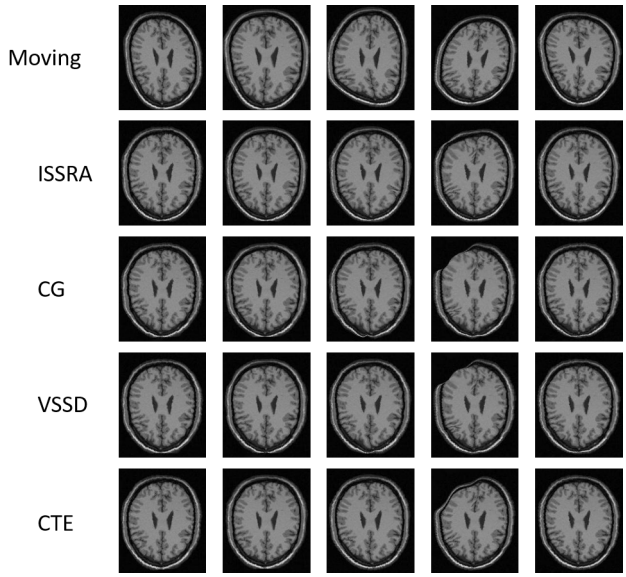


Figure 14: Registration examples for groupwise monomodal non-rigid experiments

[5] S. K. Balci, P. Golland, and W. Wells. Non-rigid groupwise registration using B-spline deformation model. *Open Source and Open Data for MICCAI*, pages 105–121, 2007.

[6] D. I. Barnea and H. F. Silverman. A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, 100(2):179–186, 1972.

[7] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.

[8] K. K. Bhatia, J. Hajnal, A. Hammers, and D. Rueckert. Similarity metrics for groupwise non-rigid registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 544–552. Springer, 2007.

[9] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *IEEE International Symposium on Biomedical Imaging*, pages 908–911. IEEE, 2004.

[10] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 24(4):325–376, 1992.

[11] R. H. Byrd, J. Nocedal, and R. A. Waltz. KNITRO: An integrated package for nonlinear optimization. In *Large-scale Nonlinear Optimization*, pages 35–59. Springer, 2006.

[12] N. D. Cahill, J. A. Schnabel, J. A. Noble, and D. J. Hawkes. Revisiting overlap invariance in medical image alignment. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8. IEEE, 2008.

[13] G. Christensen. Consistent linear-elastic transformations for image matching. In *Information Processing in Medical Imaging*, pages 224–237. Springer, 1999.

[14] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, 1996.

[15] A. C. Chung, W. M. Wells, A. Norbash, and W. E. L. Grimson. Multi-modal image registration by minimising Kullback-Leibler distance. In *Medical Image Computing and Computer-Assisted Intervention*, pages 525–532. Springer, 2002.

[16] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):425, 1997.

[17] L. Cordero-Grande, S. Merino-Caviedes, S. Aja-Fernández, and C. Alberola-López. Groupwise elastic registration by a new sparsity-promoting metric: Application to the alignment of cardiac magnetic resonance perfusion images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2638–2650, 2013.

[18] M. Cox, S. Sridharan, S. Lucey, and J. Cohn. Least squares congealing for unsupervised alignment of images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[19] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding*, 66(2):207–222, 1997.

[20] C. Davatzikos and J. L. Prince. Brain image registration based on curve mapping. In *IEEE Workshop on Biomedical Image Analysis*, pages 245–254. IEEE, 1994.

[21] M. De Craene, O. Camara, B. H. Bijnens, and A. F. Frangi. Large diffeomorphic FFD registration for motion and strain quantification from 3D-US sequences. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 437–446. Springer, 2009.

[22] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212. Springer, 2017.

[23] M. Ding and G. Fan. Generalized sum of Gaussians for real-time human pose tracking from a single depth sensor. In *IEEE Winter Conference on Applications of Computer Vision*, pages 47–54. IEEE, 2015.

[24] M. Ding and G. Fan. Articulated and generalized Gaussian kernel correlation for human pose estimation. *IEEE Transactions on Image Processing*, 25(2):776–789, 2016.

[25] K. A. Eppenhof, M. W. Lafarge, P. Moeskops, M. Veta, and J. P. Pluim. Deformable image registration using convolutional neural networks. In *SPIE Medical Imaging: Image Processing*. International Society for Optics and Photonics (SPIE 10574), 2018.

[26] A. Fawzi and P. Frossard. Image registration with sparse approximations in parametric dictionaries. *SIAM Journal on Imaging Sciences*, 6(4):2370–2403, 2013.

[27] J. Feldmar and N. Ayache. Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, 18(2):99–119, 1996.

[28] M. Fornefett, K. Rohr, and H. S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing*, 19(1):87–96, 2001.

[29] S. Ge, G. Fan, and M. Ding. Non-rigid point set registration with global-local topology preservation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 245–251. IEEE, 2014.

[30] A. Goshtasby. Registration of images with geometric distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):60–64, 1988.

[31] B. H. Guan, J. Corring, M. Sethi, S. Ranka, and A. Rangarajan. Image stack surface area minimization for groupwise and multimodal affine registration. In *International Conference on Pattern Recognition (ICPR)*, pages 4196–4201. IEEE, 2016.

[32] C. Guetter, C. Xu, F. Sauer, and J. Hornegger. Learning based non-rigid multi-modal image registration using Kullback-Leibler divergence. *Medical Image Computing and Computer-Assisted Intervention*, pages 255–262, 2005.

[33] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Medical Image Computing and Computer-Assisted Intervention*, pages 726–733, 2006.

[34] Y. He, A. B. Hamza, and H. Krim. A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing*, 51(5):1211–1220, 2003.

[35] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, 2012.

[36] G. Hermosillo and O. Faugeras. Well-posedness of two nonrigid multimodal image registration methods. *SIAM Journal on Applied Mathematics*, 64(5):1550–1587, 2004.

[37] S. Hu, L. Wei, Y. Gao, Y. Guo, G. Wu, and D. Shen. Learning-based deformable image registration for infant MR images in the first year of life. *Medical Physics*, 44(1):158–170, 2017.

[38] G. Huang, M. Mattar, H. Lee, and E. G. Learned-Miller. Learning to align from scratch. In *Advances in Neural Information Processing Systems*, pages 764–772, 2012.

[39] G. B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *IEEE International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[40] H. J. Johnson and G. E. Christensen. Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461, 2002.

[41] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

[42] B. Karaçali. Information theoretic deformable registration using local image information. *International Journal of Computer Vision*, 72(3):219–237, 2007.

[43] B. Kim, K. A. Frey, S. Mukhopadhyay, B. D. Ross, and C. R. Meyer. Co-registration of MRI and autoradiography of rat brain in three-dimensions following automatic reconstruction of 2D data set. In *Computer Vision, Virtual Reality and Robotics in Medicine*, pages 262–266. Springer, 1995.

[44] R. Kimmel. Demosaicing: image reconstruction from color CCD samples. *IEEE Transactions on Image Processing*, 8(9):1221–1228, 1999.

[45] R. Kimmel, R. Malladi, and N. Sochen. Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, 2000.

[46] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[47] S. Klein, M. Staring, and J. P. Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Transactions on Image Processing*, 16(12):2879–2890, 2007.

[48] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson. Object recognition by affine invariant matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 335–344. IEEE, 1988.

[49] E. G. Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):236–250, 2006.

[50] N. Lepore, C. Brun, Y.-Y. Chou, M.-C. Chiang, R. A. Dutton, K. M. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, et al. Generalized tensor-based morphometry of hiv/aids using multivariate statistics on deformation tensors. *IEEE Transactions on Medical Imaging*, 27(1):129–141, 2008.

[51] H. Li and Y. Fan. Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv preprint arXiv:1709.00799*, 2017.

[52] R. Liao, C. Guetter, C. Xu, Y. Sun, A. Khamene, and F. Sauer. Learning-based 2D/3D rigid registration using Jensen-Shannon divergence for image-guided surgery. In *International Workshop on Medical Imaging and Virtual Reality*, pages 228–235. Springer, 2006.

[53] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu. An artificial agent for robust image registration. In *AAAI*, pages 4168–4175, 2017.

[54] N. Litke, M. Droske, M. Rumpf, and P. Schröder. An image processing approach to surface matching. In *Symposium on Geometry Processing*, volume 255, pages 207–216. Citeseer, 2005.

[55] H. Liu, J. Yan, and D. Zhang. Three-dimensional surface registration: A neural network strategy. *Neurocomputing*, 70(1):597–602, 2006.

[56] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

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

[58] M. A. Mattar, A. R. Hanson, and E. G. Learned-Miller. Unsupervised joint alignment and clustering using Bayesian nonparametrics. *arXiv preprint arXiv:1210.4892*, 2012.

[59] C. D. McGillem and M. Svedlow. Image registration error variance as a measure of overlay quality. *IEEE Transactions on Geoscience Electronics*, 14(1):44–49, 1976.

[60] C. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen. Nonrigid registration of dynamic medical imaging data using nD+ t B-splines and a groupwise optimization approach. *Medical Image Analysis*, 15(2):238–249, 2011.

[61] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195–206, 1997.

[62] S. Miao, Z. J. Wang, Y. Zheng, and R. Liao. Real-time 2D/3D registration via CNN regression. In *IEEE International Symposium on Biomedical Imaging*, pages 1430–1434. IEEE, 2016.

[63] E. G. Miller, N. E. Matsakis, and P. A. Viola. Learning from one example through shared densities on transforms. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 464–471. IEEE, 2000.

[64] H. Ogawa. Labeled point pattern matching by Delaunay triangulation and maximal cliques. *Pattern Recognition*, 19(1):35–40, 1986.

[65] F. P. Oliveira and J. M. R. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, 2014.

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

[67] J. P. Pluim, J. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. In *Medical Image Computing and Computer-Assisted Intervention*, pages 452–461. Springer, 2000.

[68] J. P. Pluim, J. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[69] M. Polfliet, S. Klein, W. Huizinga, M. M. Paulides, W. J. Niessen, and J. Vandemeulebroucke. Intrasubject multimodal groupwise registration with the conditional template entropy. *Medical Image Analysis*, 46:15–25, 2018.

[70] T. Radcliffe, R. Rajapakshe, and S. Shalev. Pseudocorrelation: A fast, robust, absolute, grey-level image alignment algorithm. *Medical Physics*, 21(6):761–769, 1994.

[71] A. Rajwade, A. Banerjee, and A. Rangarajan. New method of probability density estimation with application to mutual information based image registration. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1769–1776. IEEE, 2006.

[72] A. Rajwade, A. Banerjee, and A. Rangarajan. Probability density estimation using isocontours and isosurfaces: Application to information-theoretic image registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):475–491, 2009.

[73] S. Ranade and A. Rosenfeld. Point pattern matching by relaxation. *Pattern Recognition*, 12(4):269–275, 1980.

[74] A. Rangarajan, H. Chui, and E. Mjolsness. A new distance measure for non-rigid image matching. In *EMMCVPR*, pages 237–252. Springer, 1999.

[75] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang. Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory*, 50(6):1220–1228, 2004.

[76] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. *Medical Image Computing and Computer-Assisted Interventation*, pages 1115–1124, 1998.

[77] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec. SVF-Net: Learning deformable image registration using shape matching. In *Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer, 2017.

[78] T. Rohlfing and C. R. Maurer. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*, 7(1):16–25, 2003.

[79] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging*, 20(6):526–534, 2001.

[80] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers. Diffeomorphic registration using B-splines. In *Medical Image Computing and Computer-Assisted Intervention*, pages 702–709. Springer, 2006.

[81] D. Rueckert, M. Clarkson, D. Hill, and D. J. Hawkes. Non-rigid registration using higher-order mutual information. In *Proc. SPIE*, volume 3979, pages 439–447, 2000.

[82] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 77–84. Springer, 2001.

[83] G. L. Scott and H. C. Longuet-Higgins. An algorithm for associating the features of two images. *Proceedings of the Royal Society of London B: Biological Sciences*, 244(1309):21–26, 1991.

[84] M. Sdika. A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. *IEEE Transactions on Medical Imaging*, 27(2):271–281, 2008.

[85] D. Seo, J. Ho, and B. Vemuri. Matching and classification of images using the space of image graphs. In *Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pages 99–110, 2011.

[86] D. Shen and C. Davatzikos. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

[87] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis. A deep metric for multimodal registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 10–18. Springer, 2016.

[88] J. Sloan, K. Goatman, and J. Siebert. Learning rigid image

registration-utilizing convolutional neural networks for medical image registration. 2018.

[89] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318, 1998.

[90] S. Sommer, M. Nielsen, S. Darkner, and X. Pennec. Higher-order momentum distributions and locally affine LDDMM registration. *SIAM Journal on Imaging Sciences*, 6(1):341–367, 2013.

[91] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.

[92] Ž. Spiclin, B. Likar, and F. Pernus. Groupwise registration of multimodal images by an efficient joint entropy minimization scheme. *IEEE Transactions on Image Processing*, 21(5):2546–2558, 2012.

[93] M. Storer, M. Urschler, and H. Bischof. Intensity-based congealing for unsupervised joint image alignment. In *International Conference on Pattern Recognition*, pages 1473–1476. IEEE, 2010.

[94] C. Studholme, D. L. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[95] H. Sundar, H. Litt, and D. Shen. Estimating myocardial motion by 4D image warping. *Pattern Recognition*, 42(11):2514–2526, 2009.

[96] R. Szeliski and S. Lavallée. Matching 3-D anatomical surfaces with non-rigid deformations using octree-splines. *International Journal of Computer Vision*, 18(2):171–186, 1996.

[97] H. D. Tagare, D. Groisser, and O. Skrinjar. Symmetric non-rigid registration: A geometric theory and some numerical techniques. *Journal of Mathematical Imaging and Vision*, 34(1):61–88, 2009.

[98] H. D. Tagare, D. O'Shea, and A. Rangarajan. A geometric criterion for shape-based non-rigid correspondence. In *IEEE International Conference on Computer Vision*, pages 434–439. IEEE, 1995.

[99] H. D. Tagare and M. Rao. Why does mutual-information work for image registration? A deterministic explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1286–1296, 2015.

[100] J. Ton and A. K. Jain. Registering Landsat images by point matching. *IEEE Transactions on Geoscience and Remote Sensing*, 27(5):642–651, 1989.

[101] A. Vedaldi, G. Guidi, and S. Soatto. Joint data alignment up to (lossy) transformations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[102] A. Vedaldi and S. Soatto. A complexity-distortion approach to joint pattern alignment. In *Advances in Neural Information Processing Systems*, pages 1425–1432, 2007.

[103] B. Vemuri, J. Ye, Y. Chen, and C. Leonard. A level-set based approach to image registration. In *Mathematical Methods in Biomedical Image Analysis*, pages 86–93. IEEE, 2000.

[104] B. C. Vemuri, S. Huang, S. Sahni, C. M. Leonard, C. Mohr, R. Gilmore, and J. Fitzsimmons. An efficient motion estimator with application to medical image registration. *Medical Image Analysis*, 2(1):79–98, 1998.

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

[106] E. Vural and P. Frossard. Analysis of image registration with tangent distance. *SIAM Journal on Imaging Sciences*, 7(4):2860–2915, 2014.

[107] C. Wachinger and N. Navab. Simultaneous registration of multiple images: Similarity metrics and efficient optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1221–1233, 2013.

[108] Q. Wang, L. Chen, P.-T. Yap, G. Wu, and D. Shen. Groupwise registration based on hierarchical image clustering and atlas synthesis. *Human Brain Mapping*, 31(8):1128–1140, 2010.

[109] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[110] R. P. Woods, J. C. Mazziotta, S. R. Cherry, et al. MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17:536–536, 1993.

[111] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen. Unsupervised deep feature learning for deformable registration of MR brain images. In *Medical Image Computing and Computer-Assisted Intervention*, pages 649–656. Springer, 2013.

[112] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Transactions on Biomedical Engineering*, 63(7):1505–1516, 2016.

[113] G. Wu, P.-T. Yap, Q. Wang, and D. Shen. Groupwise registration from exemplar to group mean: extending HAMMER to groupwise registration. In *IEEE International Symposium on Biomedical Imaging*, pages 396–399. IEEE, 2010.

[114] M. V. Wyawahare, P. M. Patil, H. K. Abhyankar, et al. Image registration techniques: an overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2(3):11–28, 2009.

[115] Q. Xie, S. Kurtek, E. Klassen, G. E. Christensen, and A. Srivastava. Metric-based pairwise and multiple image registration. In *European Conference on Computer Vision*, pages 236–250. Springer, 2014.

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

[117] A. J. Yezzi and S. Soatto. Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2):153–167, 2003.

[118] Z. Yi, C. Zhiguo, and X. Yang. Multi-spectral remote image registration based on SIFT. *Electronics Letters*, 44(2):107–108, 2008.

[119] S. Ying, G. Wu, Q. Wang, and D. Shen. Hierarchical unbiased graph shrinkage (hugs): a novel groupwise registration for large data set. *NeuroImage*, 84:626–638, 2014.

[120] J. Zhang and A. Rangarajan. Bayesian multimodality non-rigid image registration via conditional density estimation. In *Information Processing in Medical Imaging*, pages 499–511. Springer, 2003.

[121] J. Zhang and A. Rangarajan. Affine image registration using a new information metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 848–855. IEEE, 2004.

[122] J. Zhang and A. Rangarajan. Multimodality image registration using an extensible information metric and high dimensional histogramming. In *Information Processing in Medical Imaging*, pages 725–737. Springer, 2005.

[123] L. Zollei. *A unified information theoretic framework for pair- and group-wise registration of medical images*. PhD thesis, MIT, 2006.