This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

Fast and Robust Symmetric Image Registration Based on Distances Combining Intensity and Spatial Information

Johan Öfverstedt, Joakim Lindblad, *Member, IEEE,* and Nataša Sladoje, *Member, IEEE*

***Abstract*—Intensity-based image registration approaches rely on similarity measures to guide the search for geometric correspondences with high affinity between images. The properties of the used measure are vital for the robustness and accuracy of the registration. In this study a symmetric, intensity interpolation-free, affine registration framework based on a combination of intensity and spatial information is proposed. The excellent performance of the framework is demonstrated on a combination of synthetic tests, recovering known transformations in the presence of noise, and real applications in biomedical and medical image registration, for both 2D and 3D images. The method exhibits greater robustness and higher accuracy than similarity measures in common use, when inserted into a standard gradient-based registration framework available as part of the open source Insight Segmentation and Registration Toolkit (ITK). The method is also empirically shown to have a low computational cost, making it practical for real applications. Source code is available.**

***Index Terms*—Image registration, set distance, gradient methods, optimization, cost function, iterative algorithms, fuzzy sets, magnetic resonance imaging, transmission electron microscopy.**

I. Introduction

IMAGE registration [1]–[4] is the process of establishing correspondences between images and a reference space, such that the contents of the images have a high degree of affinity in the reference space. An example of such correspondence is a mapping of an image (often referred to as *floating* image) of a brain to a reference space of another image (often referred to as *reference* image) of a brain where their important structures are well co-localized. There are two main categories of approaches for image registration: *feature-based* methods extract a set of feature points between which a correspondence is found, whereas *intensity-based* methods use the voxel-values directly, and evaluate candidate mappings based on a similarity measure (affinity). There are also two main categories of transformation models: linear (which include, as special cases, rigid, similarity, and affine transformations), and non-linear (deformable). The combination of differentiable transformation models and differentiable similarity measures enables the use of gradient-based local optimization methods.

Medical and biomedical image registration, [4]–[6], is an important branch of general image registration and a lot of effort has been invested over the last decades to refine the tools and techniques, [2]. Although a majority of the recent research has been devoted to non-linear registration techniques, the most prevalent registration method used in the clinic is still linear registration. In a number of situations, the deformations allowed by non-linear registration can be difficult to evaluate and may affect reliability of diagnosis, [2]; hence, physicians may prefer a more constrained rigid or affine alignment. Considering their wide usage as fundamental tools, improvement of rigid and affine registration in terms of performance and reliability is highly relevant in practice.

Feature-based image registration is dependent on the ability of the feature extraction method to locate distinct points of interest appearing in both (all) images. Feature-extractors (e.g. SIFT [7]) typically presuppose the existence and relevance of specific local characteristics such as edges, corners and other salient features; if no, or too few, such distinct points are found, the registration will fail. This is often the case in medical and biomedical applications, [8], [9], where intensity-based registration, therefore, tends to be the method of choice. Figure 1 shows an illustrative example of a biomedical application where feature-based method fails, whereas intensity-based method can be successful.

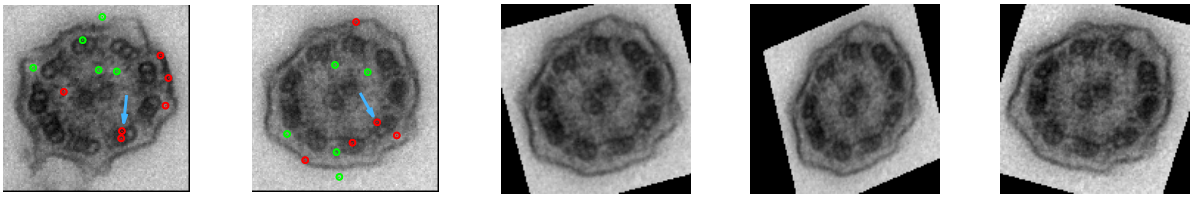
Intensity-based registration is, in general, formulated as a non-convex optimization problem. The similarity measures commonly used as optimization criteria typically exhibit a high number of local optima [10], [11]; a count which tends to rapidly increase under noisy conditions. A small region of attraction of a global optimum imposes that the starting position has to be set very close to the optimal solution for it to be found by an optimizer. This leads to reliability challenges for automated solutions.

In this study we develop a registration framework based on a family of symmetric distance measures, proposed in [11], which combine intensity and spatial information in a single measure. These measures have been shown to be characterized by smooth distance surfaces with significantly fewer local minima than the commonly used intensity-based measures, when studied in the context of template matching and object recognition. In this work we demonstrate that slightly modified versions of these distance measures can be successfully used for fast and robust affine image registration. By differentiating the distance measure we are able to use efficient gradient-based optimization. The proposed method outperforms the commonly used similarity measures in both synthetic and real scenarios of medical and biomedical registration tasks, which

The authors are with the Centre for Image Analysis, Department of Information Technology, Uppsala University, Uppsala, Sweden. Lindblad and Sladoje are also with Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia. E-mail: {johan.ofverstedt,joakim.lindblad,natasa.sladoje}@it.uu.se

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing



(a) Reference Image (b) Floating Image (c) FB Rigid (Fail) (d) FB Affine (Fail) (e) IB, Rigid+Affine

Fig. 1. Illustrative example of a biomedical registration task where a widely used feature-based (FB) method (SIFT, as implemented in FIJI-plugin Linear Stack Alignment[2]) fails, while the proposed intensity-based (IB) method (Sec. V-F1) performs well. Green points in (a) and (b) are incorrectly detected as having a match, and red points do not have a match. The feature-extractor fails to detect points corresponding to the relevant structures (one approximately correct match, indicated with arrows, can be found manually), and both the central rings and the outer rings are misaligned.

we confirm by (i) landmark-based evaluation on transmission electron microscopy (TEM) images of cilia [12], with the aim of improving multi-image super-resolution reconstruction, as well as (ii) evaluation on the task of atlas-based segmentation of magnetic resonance (MR) images of brain, on the LPBA40-dataset [13].

Intensity interpolation is typically a required tool in the context of intensity-based registration performed with commonly used similarity measures since the sought transformation (and intermediate candidates) is likely to map points to regions outside of the regular grid. Treating the reference and floating images differently in terms of the interpolation introduces a significant source of asymmetry [14] and may lead to success or failure of a registration task depending on which image is selected as *reference* and which is *floating*. Our proposed approach requires no off-grid intensity values, and is interpolation-free in terms of intensities; empirical tests confirm that it is highly symmetric in practice.

Noting that intensity-based image registration can be computationally demanding, we also include a study of execution time of (i) isolated distance and gradient computations through micro-benchmarks, and (ii) entire image registration tasks. We observe that the proposed measure is fast to compute in comparison with the implementations of the measures existing in the ITK-framework [14]. The registration framework, with the proposed measure, is implemented in C++/ITK, and its source code is available[3].

II. Preliminaries and Previous Work

A. Images as Fuzzy Sets

First we recall a few basic concepts related to fuzzy sets [15], a theoretical framework where gray-scale images are conveniently represented.

A *fuzzy set* $\mathcal{S}$ on a reference set $X_\mathcal{S}$ is a set of ordered pairs, $\mathcal{S} = \{(x, \mu_\mathcal{S}(x)) : x \in X_\mathcal{S}\}$, where $\mu_\mathcal{S} \colon X_\mathcal{S} \to [0, 1]$ is the *membership function* of $\mathcal{S}$. Where there is no risk for confusion, we equate the set and its membership function and let $\mathcal{S}(x)$ be equivalent to $\mu_\mathcal{S}(x)$

A gray-scale image can directly be interpreted as a spatial fuzzy set by rescaling the valid intensity range to $[0, 1]$. We assume, w.l.o.g., that the images to be registered have an intensity range $[0, 1]$ and we directly interpret them as fuzzy sets defined on a reference set which is the image domain, and is in most cases a subset of $\mathbb{Z}^n$. We use the terms *image* and *fuzzy set* interchangeably in this text.

A crisp set $C \subseteq X_C$ (a binary image) is a special case of a fuzzy set, with its characteristic function as membership function

$$\mu_C(x) = \begin{cases} 1, & \text{for } x \in C \\ 0, & \text{for } x \notin C\,. \end{cases} \tag{1}$$

The *height* of a fuzzy set $S \subseteq X_\mathcal{S}$ is $h(\mathcal{S}) = \max_{x \in X_\mathcal{S}} \mu_\mathcal{S}(x)$. The *complement* $\overline{\mathcal{S}}$ of a fuzzy set $\mathcal{S}$ is $\overline{\mathcal{S}} = \{(x, 1 - \mu_\mathcal{S}(x)) : x \in X_\mathcal{S}\}$. An α-*cut* of a fuzzy set $\mathcal{S}$ is a crisp set defined as ${}^\alpha\mathcal{S} = \{x \in X_\mathcal{S} : \mu_\mathcal{S}(x) \geq \alpha\}$, i.e., a thresholded image.

Let p be an element of the reference set $X_\mathcal{S}$. A *fuzzy point* p (also called a fuzzy *singleton*) defined at $p \in X_\mathcal{S}$ with height $h(p)$, is defined by a membership function

$$\mu_p(x) = \begin{cases} h(p), & \text{for } x = p \\ 0, & \text{for } x \neq p\,. \end{cases} \tag{2}$$

B. Intensity-Based Registration and Point-Wise Distances

Intensity-based registration is a general approach to image registration defined as a minimization process, where a distance measure between the intensities of overlapping points (or regions) is used as optimization criterion. Given a distance measure d and a set of valid transformations Ω, intensity-based registration of two images $\mathcal{A}$ (floating) and $\mathcal{B}$ (reference) can be formulated as the optimization problem,

$$\hat{T} = \arg\min_{T \in \Omega} d(T(\mathcal{A}), \mathcal{B}), \tag{3}$$

where $T(\mathcal{A})$ denotes a valid transform of image $\mathcal{A}$ into the reference space of image $\mathcal{B}$.

Intensity-based similarity/distance measures which are most commonly used for image registration are Sum of Squared Differences (SSD) [16], Pearson Correlation Coefficient (PCC) and Mutual Information (MI) [17]. These measures are point-based, i.e. they are functions of the intensities of points belonging to the overlapping regions of the two compared sets. Their evaluation, therefore, typically requires interpolation of image intensities.

For two images P and Q defined on a common reference set $X_{P,Q}$ of overlapping points, these measures are defined as

$$\mathrm{SSD}(P, Q) = \sum_{v \in X_{P,Q}} (P(v) - Q(v))^2, \tag{4}$$

[2]imagej.net/Linear_Stack_Alignment_with_SIFT

[3]Source code available from www.github.com/MIDA-group

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

$$\mathrm{PCC}(P,Q) = \frac{\sum\limits_{v\in X_{P,Q}} (P(v)-\mathrm{avg}(P))(Q(v)-\mathrm{avg}(Q))}{\sqrt{\sum\limits_{v\in X_{P,Q}} (P(v)-\mathrm{avg}(P))^2}\sqrt{\sum\limits_{v\in X_{P,Q}} (Q(v)-\mathrm{avg}(Q))^2}} \tag{5}$$

and

$$\mathrm{MI}(P,Q) = H_P + H_Q - H_{P,Q}. \tag{6}$$

In (5) $\mathrm{avg}(P)$, and $\mathrm{avg}(Q)$ denote means of the resp. intensity distributions over the evaluated region. In (6) the (joint and marginal) entropies H_P, H_Q and $H_{P,Q}$ of the image intensity distributions P and Q are defined in terms of the estimated probability p of a randomly selected point v having intensities $P(v)$, $Q(v)$, as

$$H_P = -\sum_{v\in X_{P,Q}} p(P(v)) \log(p(P(v)))\,, \tag{7}$$

and

$$H_{P,Q} = -\sum_{v\in X_{P,Q}} p(P(v),Q(v)) \log(p(P(v),Q(v)))\,. \tag{8}$$

Intensity-based registration, as formulated in (3), is, in general, a non-convex optimization problem with a large number of local optima, especially for the commonly used point-based measures (SSD, PCC, and MI). To try to overcome this optimization challenge, a resolution-pyramid-scheme is normally used [18], [19], where smoothed low resolution images are first registered, followed by registration of images with increasing resolution and decreasing degree of smoothing, using the transform obtained from the previous stage as starting position (so-called coarse-to-fine approach).

C. Distances Combining Intensity and Spatial Information

While the distances of Sec. II-B only rely on intensities of overlapping points, the distances presented in this section incorporate also spatial information of non-overlapping points. For such spatial relations, we consider distances between two points, between a point and a set, and between two sets. The most commonly used point-to-point distance is the Euclidean distance, denoted d_E.

Given a point-to-point distance $d(a,b)$, the common crisp point-to-set distance between a point a and a set B is

$$d(a,B) = \inf_{b\in B} d(a,b)\,. \tag{9}$$

Closely related to the crisp point-to-set distance is the (external) distance transform of a crisp set $B \subseteq X_B$ (with point-to-point distance d) which is defined as

$$\mathrm{DT}[B](x) = \min_{y\in B}\{d(x,y)\}\,. \tag{10}$$

Taking into the consideration the intensity, or equivalently, the *height* of a fuzzy point, the *fuzzy point-to-set inwards distance* d^α, based on integration over α-cuts [11], between a fuzzy point p and a fuzzy set $\mathcal{S}$, is defined as

$$d^\alpha(p,\mathcal{S}) = \int_0^{h(p)} d(p,{}^\alpha\mathcal{S})\,\mathrm{d}\alpha\,, \tag{11}$$

where d is a point-to-set distance defined on crisp sets. The *complement distance* [20] of a fuzzy point-to-set distance d is

$$\overline{d}(p,\mathcal{S}) = d(\overline{p},\overline{\mathcal{S}})\,. \tag{12}$$

The *fuzzy point-to-set bidirectional distance* $\vec{d}^\alpha$ is

$$\vec{d}^\alpha(p,\mathcal{S}) = d^\alpha(p,\mathcal{S}) + \overline{d}^\alpha(p,\mathcal{S})\,. \tag{13}$$

For an arbitrary point-to-set distance d, Sum of Minimal Distances (SMD) [21] defines a set-to-set distance as

$$d_{\mathrm{SMD}}(A,B) = \frac{1}{2}\Big(\sum_{a\in A} d(a,B) + \sum_{b\in B} d(b,A)\Big)\,. \tag{14}$$

A weighted version can be defined [11], which may be useful if a priori information about relative importance of contributions of different points to the overall distance is available:

$$d_{\mathrm{wSMD}}(A,B;w_A,w_B) = \frac{1}{2}\Big(\sum_{a\in A} w_A(a)d(a,B) + \sum_{b\in B} w_B(b)d(b,A)\Big)\,. \tag{15}$$

Inserting distances (11) or (13) in (14) or (15) provides extensions of the SMD family of distances to fuzzy sets [11]. We refer to them as d^α_{SMD}, $\vec{d}^\alpha_{\mathrm{SMD}}$, d^α_{wSMD} and $\vec{d}^\alpha_{\mathrm{wSMD}}$.

It has been observed for fuzzy set distances [22] in general, and for distances based on (11) and (13) in particular, that distances between sets with empty α-cuts may give infinite or ill-defined distances. We follow a previous study and introduce a parameter $d_{\mathrm{MAX}} \in \mathbb{R}_{\geq 0}$, [23], to limit the underlying crisp point-to-set distance. This has a double benefit of (i) reducing the effect of outliers and (ii) making the distances well defined also for images with empty α-cuts.

Distances based on Optimal Mass Transport (OMT), such as the Wasserstein distance, also combine intensity and spatial information, and are widely studied and used in image processing [24]. The OMT can be framed as a linear programming optimization problem, which is solvable in $\mathcal{O}(N^3)$ [25]. This is intractable for most realistic image processing scenarios, and approximations are typically considered [25], [26]. It is possible to incorporate these distances in image registration frameworks, but to the best of our knowledge, this has only been done for non-linear (deformable) registration, and has been shown to be very computationally demanding [27], [28]. We performed a preliminary study of OMT-based methods using the formulation in [26], and observed both very high computational demands and noisy distance landscapes. In absence of a complete registration framework for linear registration based on OMT, this family of measures is excluded from the empirical part of this study.

D. Transformations, Interpolation, and Symmetry

Linear transformations relate points in one space to another via application of a linear function. A transformation is rigid if only rotations and translations are permitted, and affine if

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

shearing and reflections are also permitted. Affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ can be expressed as matrix multiplication,

$$Tx = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & t_1 \\ a_{21} & a_{22} & \dots & a_{2n} & t_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & t_n \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}. \tag{16}$$

Linear transformations can, in general, transform points sampled on an image grid to positions outside of the grid, hence an interpolator is required for obtaining the image intensity at the transformed point's location. Interpolation is a large source of error, bias, and a significant contributing factor of asymmetry in intensity-based registration, [14]. Commonly, interpolation is only required for one of the two images, where sampling (for optimization) is done from the grid of the other image space; hence, the two images are treated asymmetrically, yielding distinct similarity surfaces (over the transformation parameters) depending on which image is taken as reference. This can cause a registration task to succeed or fail, depending on the registration direction.

E. Optimization

Registration with a differentiable distance measure as objective function enables the use of gradient-based optimization algorithms, which typically are significantly more efficient than derivative-free algorithms for local iterative optimization. An effective and commonly used subset of gradient-based algorithms are the stochastic gradient descent methods [29], which consider a random subset of the points in each optimization iteration, incurring a two-fold benefit: utilizing randomness to escape shallow local optima in the implicit distance surface, while also decreasing the computational work required per iteration. The size of the random subset is usually given as a fraction of the total number of points, and denoted as the sampling fraction. Approximation of the cost function by random subset sampling (where a new random subset of points is chosen in every iteration) has been, in previous studies, [17], [30], shown to perform well for intensity-based registration.

III. Proposed Image Registration Framework

A. Distances

To extend the family of distance measures given by (15), to be suitable for registration, optionally with random subsampling optimization methods [30], we here define a new related family of distance measures.

Definition 1 (Asymmetric average minimal distance)**.** *Given fuzzy set $\mathcal{A}$ on a reference set $X_\mathcal{A} \subset \mathbb{R}^n$, fuzzy set $\mathcal{B}$ on reference set $X_\mathcal{B} \subset \mathbb{R}^n$, and a weight function $w_A \colon X_\mathcal{A} \to \mathbb{R}_{\geq 0}$, the* Asymmetric average minimal distance *from $\mathcal{A}$ to $\mathcal{B}$, is*

$$\underrightarrow{d}_{AMD}(\mathcal{A}, \mathcal{B}; w_A) = \frac{1}{\sum\limits_{x \in X_\mathcal{A}} w_A(x)} \sum_{x \in X_\mathcal{A}} w_A(x) d(\mathcal{A}(x), \mathcal{B}) \,. \tag{17}$$

We consider point-to-set distances defined by (11) or (13).

Building on the asymmetric distance, we formulate a symmetric distance as follows:

Definition 2 (Average minimal distance)**.** *Given fuzzy set $\mathcal{A}$ on reference set $X_\mathcal{A} \subset \mathbb{R}^n$, fuzzy set $\mathcal{B}$ on reference set $X_\mathcal{B} \subset \mathbb{R}^n$, weight functions $w_A \colon X_\mathcal{A} \to \mathbb{R}_{\geq 0}$ and $w_B \colon X_\mathcal{B} \to \mathbb{R}_{\geq 0}$, the* Average minimal distance *between $\mathcal{A}$ and $\mathcal{B}$, is*

$$d_{AMD}(\mathcal{A}, \mathcal{B}; w_A, w_B) = \frac{1}{2}\left(\underrightarrow{d}_{AMD}(\mathcal{A}, \mathcal{B}; w_A) + \underrightarrow{d}_{AMD}(\mathcal{B}, \mathcal{A}; w_B) \right). \tag{18}$$

In the context of image registration, we utilize $\underrightarrow{d}_{\text{AMD}}$ to express a (weighted) distance between transformed fuzzy points $T(\mathcal{A}(x))$, and the image $\mathcal{B}$, where the transformation of a fuzzy point $\mathcal{A}(x) = \{(x, \mu_\mathcal{A}(x))\}$ is given by the transformation of the reference point x:

$$T(\mathcal{A}(x)) = \{(T(x), \mu_\mathcal{A}(x))\} \,. \tag{19}$$

To reflect the bounded image domain, only the transformed points falling on a predefined region $M_B \subset \mathbb{R}^n$ are considered. Note that, when $\mathcal{A}$ and $\mathcal{B}$ are digital images, $X_\mathcal{A}$ and $X_\mathcal{B}$ are typically subsets of $\mathbb{Z}^n$ and the transformed points $T(x)|_{x \in X_\mathcal{A}}$ do not necessarily coincide with the points of the reference set $X_\mathcal{B}$; an illustrative example is given in Fig. 2.

We, therefore, provide the following definitions suited for the task of image registration:

Definition 3 (Asymmetric average minimal distance for image registration)**.** *Given fuzzy set $\mathcal{A}$ on reference set $X_\mathcal{A} \subset \mathbb{R}^n$, fuzzy set $\mathcal{B}$ on $X_\mathcal{B} \subset \mathbb{R}^n$, a weight function $w_A \colon X_\mathcal{A} \to \mathbb{R}_{\geq 0}$, and a crisp subset (mask) $M_B \subset \mathbb{R}^n$, the* Asymmetric average minimal distance for image registration *from $\mathcal{A}$ to $\mathcal{B}$, parameterized by a transformation $T \colon X_\mathcal{A} \to \mathbb{R}^n$, is*

$$\underrightarrow{d}^R_{AMD}(\mathcal{A}, \mathcal{B}; T, w_A, M_B) = \frac{1}{\sum\limits_{x \in \hat{X}} w_A(x)} \sum_{x \in \hat{X}} w_A(x) d(T(\mathcal{A}(x)), \mathcal{B}) \quad \text{where } \hat{X} = \{x : x \in X_\mathcal{A} \wedge T(x) \in M_B\} \,. \tag{20}$$

Definition 4 (Average minimal distance for image registration)**.** *Given fuzzy set $\mathcal{A}$ on reference set $X_\mathcal{A}$, fuzzy set $\mathcal{B}$ on $X_\mathcal{B}$, weight functions $w_A \colon X_\mathcal{A} \to \mathbb{R}_{\geq 0}$ and $w_B \colon X_\mathcal{B} \to \mathbb{R}_{\geq 0}$, and crisp subsets (masks) $M_A, M_B \subset \mathbb{R}^n$, the* Average minimal distance for image registration *between $\mathcal{A}$ and $\mathcal{B}$, parameterized by an invertible transformation $T \colon \mathbb{R}^n \to \mathbb{R}^n$, with inverse T^{-1}, is defined as*

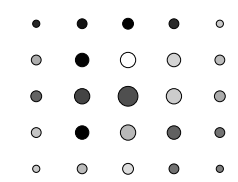
$$d^R_{AMD}(\mathcal{A}, \mathcal{B}; T, w_A, w_B, M_A, M_B) = \frac{1}{2}\left(\underrightarrow{d}^R_{AMD}(\mathcal{A}, \mathcal{B}; T, w_A, M_B) + \underrightarrow{d}^R_{AMD}(\mathcal{B}, \mathcal{A}; T^{-1}, w_B, M_A) \right). \tag{21}$$

The distance d^R_{AMD} is based on full sampling, taking into account all points in the two sets which have non-zero weights, as long as they are transformed to points inside the mask associated with the other set. To reduce the computational cost of the distance and, in addition, to enable random iterative sampling, we propose an approximate version of d^R_{AMD}:

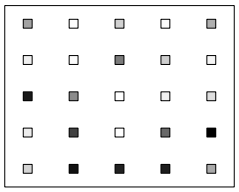
Definition 5 (Subsampled average minimal distance for image registration)**.** *Given fuzzy set $\mathcal{A}$ on reference set $X_\mathcal{A}$, fuzzy set $\mathcal{B}$ on $X_\mathcal{B}$, weight functions $w_A \colon X_\mathcal{A} \to \mathbb{R}_{\geq 0}$ and*

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

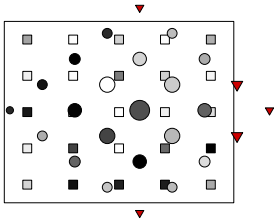
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing



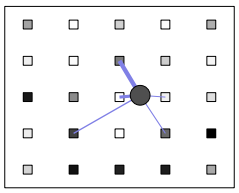
(a) $\mathcal{A}$ with weights w_A (radius).



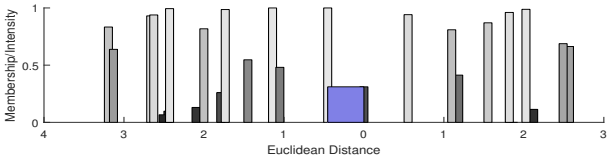
(b) $\mathcal{B}$ with mask M_B (surrounding rectangle).



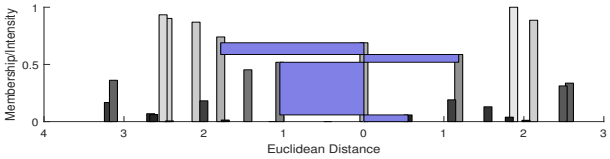
(c) $\mathcal{A}$ mapped into the space of $\mathcal{B}$. Triangles mark points mapped outside M_B, and thus discarded.



(d) Distance contribution graph for $\vec{d}^{\alpha}(T(\mathcal{A}(x)), \mathcal{B})$; where x is central point of $\mathcal{A}$.



(e) Inwards part of $\vec{d}^{\alpha}$: $d^{\alpha}(\mathcal{A}(x), \mathcal{B})$ point-to-set distance.



(f) Complement part of $\vec{d}^{\alpha}$: $d^{\alpha}(\overline{\mathcal{A}(x)}, \overline{\mathcal{B}})$ point-to-set distance.

Fig. 2. Illustration of the Asymmetric average minimal distance for image registration. (a) Source set $\mathcal{A}$ (radius represents associated weight; gray-level represents membership). (b) Target set $\mathcal{B}$ with associated mask M_B. (c) The transformed (by rotation and translation) $\mathcal{A}$ on top of $\mathcal{B}$. (d) Illustration of the contributions to the point-to-set distance $\vec{d}^{\alpha}$ by the central point of $\mathcal{A}$. Thickness of lines show the α-integrated height contributed by each point in $\mathcal{B}$. (e-f) The inwards and complement parts of $\vec{d}^{\alpha}$ visualized as 1D graphs, where the x-axis is the Euclidean distance (in 2D) from the mid-point and the points at the left and right side (of the origin) respectively are the points on the left and right side of the mid-point $T(\mathcal{A}(x))$ in (d).

$w_B \colon \mathcal{B} \to \mathbb{R}_{\geq 0}$, *and crisp subsets (masks)* $M_A, M_B \subset \mathbb{R}^n$, *the* Subsampled average minimal distance for image registration *between* $\mathcal{A}$ *and* $\mathcal{B}$, *parameterized by an invertible transformation* $T \colon \mathbb{R}^n \to \mathbb{R}^n$, *with inverse* T^{-1}, *and crisp sets* $P_A \subseteq X_{\mathcal{A}}$ *and* $P_B \subseteq X_{\mathcal{B}}$, *is defined as*

$$\tilde{d}^R_{AMD}(\mathcal{A}, \mathcal{B}; \mathcal{P}_A, \mathcal{P}_B, T, w_A, w_B, M_A, M_B) = \frac{1}{2}\big(\underrightarrow{d}^R_{AMD}(\mathcal{A} \cap P_A, \mathcal{B}; T, w_A, M_B) + \underrightarrow{d}^R_{AMD}(\mathcal{B} \cap P_B, \mathcal{A}; T^{-1}, w_B, M_A)\big)\,. \tag{22}$$

Inserting (11) or (13) as point-to-set distance in (20), and hence indirectly in (21) and (22), provides extensions of this family of distances to the α-cut-based distances, which we denote $\underrightarrow{d}^R_{\alpha\text{AMD}}$, $\underrightarrow{\tilde{d}}^R_{\alpha\text{AMD}}$, $d^R_{\alpha\text{AMD}}$, $\vec{d}^R_{\alpha\text{AMD}}$, $\tilde{d}^R_{\alpha\text{AMD}}$ and $\tilde{\vec{d}}^R_{\alpha\text{AMD}}$.

Normalization of the weights of the sampled points, introduced through Def. 1, renders the magnitude of the distance (and subsequently its derivatives) invariant to the size and aggregated weight of the sets or of the sampled subsets. Since the normalization is done separately from $\mathcal{A}$ to $\mathcal{B}$ and from $\mathcal{B}$ to $\mathcal{A}$, both directions are weighted equally even if the total weights of the point subsets from the two sets are different. This normalization can simplify the process of choosing e.g. step-length of various optimization methods, and makes it more likely that default hyper-parameter values can be found and reused across different applications.

B. Registration

We propose to utilize symmetric distances $\vec{d}_{\alpha\text{AMD}}$ and $\tilde{\vec{d}}_{\alpha\text{AMD}}$ as cost functions in (3) to define concrete image registration methods. Inserting $\vec{d}^R_{\alpha\text{AMD}}$ into (3) we obtain

$$\hat{T} = \arg\min_{T \in \Omega} \vec{d}^R_{\alpha\text{AMD}}(\mathcal{A}, \mathcal{B}; T, w_A, w_B, M_A, M_B)\,. \tag{23}$$

For the case of subset sampling with sets P_A and P_B, registration is defined as

$$\hat{T} = \arg\min_{T \in \Omega} \tilde{\vec{d}}^R_{\alpha\text{AMD}}(\mathcal{A}, \mathcal{B}; P_A, P_B, T, w_A, w_B, M_A, M_B). \tag{24}$$

By selection of new random subsets P_A and P_B in each iteration, various stochastic gradient descent optimization methods can be realized.

To solve the optimization problems stated in (23) and (24) with efficient gradient-based optimization methods, the partial derivatives of the distance measures with respect to the transformation parameters of T are required.

C. Gradients

The derivative of (9), the crisp point-to-set distance measure $d(T(x), S)$ (in n-dimensional space), with respect to parameters T_i of the transformation T applied to a point $x \in X$, yielding $y = T(x)$, can be written as

$$\frac{\partial d}{\partial T_i} = \sum_{j=1}^{n} \frac{\partial d}{\partial y_j} \frac{\partial y_j}{\partial T_i}. \tag{25}$$

The values $\frac{\partial d}{\partial y_j}$ are the components (partial derivatives) of the gradient $\nabla d(y, S)$ of the point-to-set distance in point $y \in Y \subset \mathbb{R}^n$, and are not dependent on the transformation model.

The gradient of the fuzzy point-to-set distance measure (11) is given by the integral over α-cuts, of gradients of the (crisp) point-to-set distances:

$$\nabla d(\mathcal{X}, \mathcal{S}) = \int_0^{h(\mathcal{X})} \nabla d(x, {}^{\alpha}\mathcal{S})\,\mathrm{d}\alpha\,. \tag{26}$$

D. Algorithms for Digital Images on Rectangular Grids

The distances and gradients can be computed efficiently for the special case of digital images on rectangular grids. For images quantized to $\ell \in \mathcal{N}_{>0}$ non-zero discrete α-levels the integrals in (11) and (26) are suitably replaced by sums. The number of quantization levels is typically taken to be a small constant; a choice of $\ell = 7$ non-zero equally spaced α-levels

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

has previously shown to provide a good trade-off between performance, speed and noise-sensitivity [11], and we keep it for all experiments.

We need a discrete operator to approximate the gradient of $d(x, S)$ for a set S defined on a rectangular grid with spacing $s \in \mathbb{R}^n_{>0}$. We propose to use the following difference operator providing a discrete approximation of $\nabla d(x, S)$:

$$\Delta d(x) = \gamma_x(\delta_1[d](x), \ldots, \delta_n[d](x)), \tag{27}$$

where

$$\delta_i[d](x) = \tfrac{1}{2s_i}(d(x + s_i u_i, S) - d(x - s_i u_i, S)), \tag{28}$$

γ_x is an indicator function,

$$\gamma_x = \begin{cases} 1, & \text{for } d(x,S) \neq 0 \\ 0, & \text{for } d(x,S) = 0, \end{cases} \tag{29}$$

and u_i is the unit-vector along the i-th dimension.

The indicator function γ_x causes the operator $\Delta[S](x)$ to be zero-valued for points included in S (i.e., where the distance transform is zero-valued). This prevents discretization issues along set boundaries, where the standard central difference operator yields non-zero gradients, which can cause the measure to overshoot a potential voxel-perfect overlap.

Algorithm 1 Distance and Gradient Maps

Input: Image $\mathcal{A}$. Mask M_A. Set of α-levels $\{\alpha_1, \ldots, \alpha_\ell\}$.
Output: Stack of pre-computed distance sums $D[0 \ldots \ell]$, and corresponding discrete gradient approximations $G[0 \ldots \ell]$.

procedure $\Delta\alpha$DT($\mathcal{A}, M_A, (\alpha_1, \ldots, \alpha_\ell), d_{\text{MAX}}$)
  $\alpha_0 \leftarrow 0$, $D[0] \leftarrow \mathbf{0}$, $G[0] \leftarrow (\mathbf{0}, \ldots, \mathbf{0})$
  for $i = 1$ to ℓ **do**
    $\text{DT}_i \leftarrow \text{DT}[^{\alpha_i}\mathcal{A} \cap M_A]$ ▷ *Compute DT of α-cut.*
    $DT_i \leftarrow \min(DT_i, d_{MAX})$
    for all $v \in X_{\mathcal{A}}$ ***do***
      $D[i](v) \leftarrow (\alpha_i - \alpha_{i-1})DT_i(v) + D_{i-1}(v)$
      $G[i](v) \leftarrow (\alpha_i - \alpha_{i-1})\Delta DT_i(v) + G_{i-1}(v)$
    end for
  end for
  return D, G
end procedure

procedure $\Delta\alpha$DT_BD($\mathcal{A}, M_A, (\alpha_1, \ldots, \alpha_\ell), d_{MAX}$)
  $D, G \leftarrow \Delta\alpha DT(\mathcal{A}, M_A, (\alpha_1, \ldots, \alpha_\ell), d_{MAX})$
  $\overline{D}, \overline{G} \leftarrow \Delta\alpha DT(\overline{\mathcal{A}}, M_A, (1 - \alpha_\ell, \ldots, 1 - \alpha_1), d_{MAX})$
  for $i = 0$ *to* ℓ ***do***
    for all $v \in X_{\mathcal{A}}$ ***do***
      $D[i](v) \leftarrow D[i](v) + \overline{D}[\ell - i](v)$
      $G[i](v) \leftarrow G[i](v) + \overline{G}[\ell - i](v)$
    end for
  end for
  return D, G
end procedure

By creating tables for the distance and gradient sums for each image as a pre-processing step, using either of the procedures in Alg. 1 ($\Delta\alpha$DT for inwards distances and $\Delta\alpha$DT_BD for bidirectional distances), the distance and gradient may then be readily computed with Alg. 2. $|T|$ denotes the number of parameters of the transformation, which is 6 for two-dimensional (2D) affine transformations, and 12 for three-dimensional (3D) affine transformations.

Algorithm 2 Point-to-Set Distance and its Gradient w.r.t. T

Input: Fuzzy point $\mathcal{A}(v)$ and associated weight $w_A(v)$, fuzzy set B and associated (binary) mask M_B, distances $D[0 \ldots \ell]$ and gradients $G[0 \ldots \ell]$; transformation T.
Output: Point-to-set distance d, derivatives $\frac{\Delta d}{\Delta T}$, weight w.

procedure D_PTS($\mathcal{A}(v), \mathcal{B}; T, w_A(v), M_B, D_B, G_B$)
  $\hat{v} \leftarrow T(v)$
  if $\hat{v} \in M_B$ **then**
    $h \leftarrow \text{QUANTIZE}(\mu_{\mathcal{A}}(v))$
    $d, \mathbf{g} \leftarrow \text{INTERPOLATE}(D_B[h], G_B[h], \hat{v})$
    for $i = 1$ to $|T|$ **do**
      $\frac{\Delta d}{\Delta T_i} \leftarrow \sum_{j=1}^{n} \frac{\partial \hat{v}_j}{\partial T_i}\mathbf{g}_j$
    end for
    return $w_A(v)d, w_A(v)\frac{\Delta d}{\Delta T}, w_A(v)$
  else
    return $0, \mathbf{0}, 0$ ▷ *Zero dist., grad., and weight.*
  end if
end procedure

The procedures in Alg. 1 have linear run-time complexity $\mathcal{O}((\ell+1)\,|X_{\mathcal{A}}|)$, achieved by using a linear-time algorithm for computing the distance transform (e.g. [31]) in line 4 of Alg. 1. The space complexity of the algorithm is $\mathcal{O}((\ell+1)\,|X_{\mathcal{A}}|)$ and the D, G structures must remain in memory to enable fast lookup in Alg. 2. Figure 3 shows an example of the distance and gradient of a sample α-level. Alg. 2 computes the point-to-set distance and gradient w.r.t. the transformation using the pre-computed tables and has run-time complexity $\mathcal{O}(|T|\,n)$ thus being independent of ℓ and the size of $\mathcal{A}$ and $\mathcal{B}$.

Algorithm 3 Symmetric Registration

Input: Fuzzy sets $\mathcal{A}, \mathcal{B}$, with binary masks M_A, M_B, and weight functions w_A, w_B, initial transformation guess T. No. of α-levels ℓ, distance saturation d_{MAX}, step-lengths $\lambda[1 \ldots N]$, and iteration count N are hyper-parameters.
Output: Symmetric set distance d ($\hat{d}^R_{\alpha\text{AMD}}$), estimated transformation $\hat{T}$.

$D_A, G_A \leftarrow \Delta\alpha\text{DT_BD}(\mathcal{A}, M_A, (\alpha_1, \ldots, \alpha_\ell), d_{\text{MAX}})$
$D_B, G_B \leftarrow \Delta\alpha\text{DT_BD}(\mathcal{B}, M_B, (\alpha_1, \ldots, \alpha_\ell), d_{\text{MAX}})$
for $i = 1$ to N **do**
  $d_1, \frac{\Delta d_1}{\Delta T}, w_1 \leftarrow \ldots$
    $\sum_{v \in X_{\mathcal{A}}} \text{D_PTS}(\mathcal{A}(v), \mathcal{B}; T, w_A(v), M_B, D_B, G_B)$
  $d_2, \frac{\Delta d_2}{\Delta T^{-1}}, w_2 \leftarrow \ldots$
    $\sum_{v \in X_{\mathcal{B}}} \text{D_PTS}(\mathcal{B}(v), \mathcal{A}; T^{-1}, w_B(v), M_A, D_A, G_A)$
  $T \leftarrow T - \frac{\lambda[i]}{2}\left(\frac{1}{w_1}\frac{\Delta d_1}{\Delta T} + \frac{1}{w_2}\frac{\Delta T^{-1}}{\Delta T}\frac{\Delta d_2}{\Delta T^{-1}}\right)$
end for
$d \leftarrow \frac{1}{2}(\frac{1}{w_1}d_1 + \frac{1}{w_2}d_2)$ ▷ *Output final distance value.*
$\hat{T} \leftarrow T$ ▷ *Output final transformation estimate.*

Algorithm 3 performs a complete registration given two images, their binary masks, weight functions, and an initial

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

transformation. Algorithm 3 completes N full iterations, however other termination criteria may be beneficial (see Sec. IV). The registration consists of pre-processing, followed by a loop where the symmetric set distance and derivatives are computed and T is updated. $\frac{\Delta T^{-1}}{\Delta T}$, in line 6 of Alg. 3 denotes a matrix $\left[\frac{\partial T_j^{-1}}{\partial T_i}\right]$ of partial derivatives of the parameters of the inverse transform T^{-1} w.r.t. the parameters of the forward transform T. This matrix relates the computed partial derivatives $\frac{\Delta d_2}{\Delta T^{-1}}$ with the parametrization of the forward transform.

The overall run-time complexity is $\mathcal{O}(N\,|T|\,n(|X_{\mathcal{A}}| + |X_{\mathcal{B}}|) + (\ell+1)(|X_{\mathcal{A}}| + |X_{\mathcal{B}}|))$. Practical choices for N tend to be in the range $[1000, 3000]$, depending on hyper-parameters (e.g. λ), and distance in parameter-space between starting position and the global optimum. The evaluation in Sec. V confirms empirically that convergence, according to (31) or (32), tends to be reached after 1000 to 3000 iterations, using an optimizer with a decaying λ.

The QUANTIZE procedure in Alg. 2 takes the membership of point v, $\mu_{\mathcal{A}}(v)$, and gives the index i of the minimal α-level $(\alpha_1, \ldots, \alpha_\ell)$ for which $\mu_{\mathcal{A}}(v) \geq \alpha_i$. If the membership is below all α-levels, the index is 0. For ℓ equally spaced α-levels, the quantization can be expressed as

$$\text{QUANTIZE}(\mu_{\mathcal{A}}(v)) = \lfloor \ell \mu_{\mathcal{A}}(v) + 0.5 \rfloor. \tag{30}$$

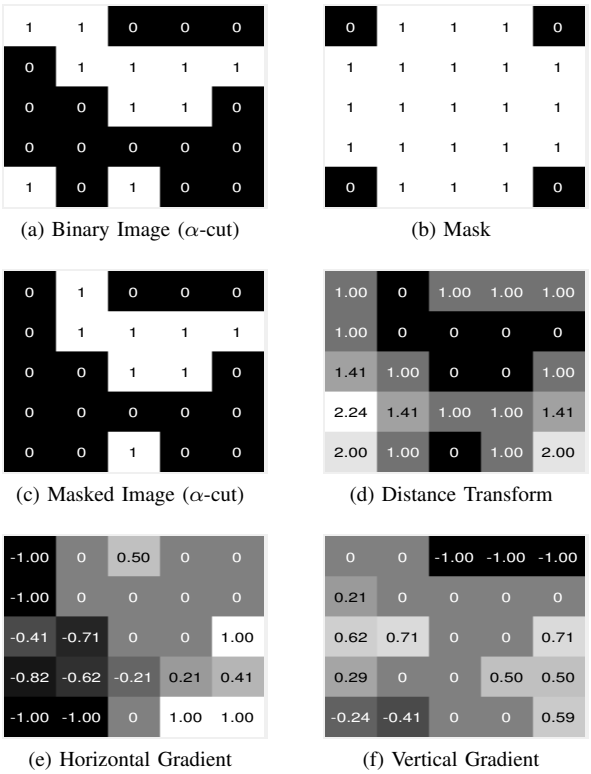
The INTERPOLATE procedure in Alg. 2 computes the value of the discrete functions D and G in point $\hat{v}$ which may not be on the grid due to application of T. There are many interpolation schemes proposed in the literature, but we suggest that either nearest neighbor (for maximal speed) or linear interpolation (for higher accuracy) are used here, since the distance and gradient fields are smooth. By linearity of integration and summation, nearest neighbor and linear interpolation may be performed on the pre-processed D and G and yield the same result as if each level was interpolated before integration, allowing efficient computation. The (discretized) measure does not require intensity interpolation; the interpolation operates on distances and gradients only.

IV. Implementation

We implemented the proposed distance measure and registration method in the open-source Insight Segmentation and Registration Toolkit (ITK) [14]. We chose this particular software framework because it

- enables the use of an existing optimization framework,
- allows for a fair comparison against well written, tested, and widely used implementations of reference similarity measures, with support for resolution-pyramids,
- supports anisotropic/scaled voxels in relevant algorithms,
- facilitates reproducible evaluation,
- makes the proposed measure easily accessible for others.

The built-in ITK optimizer we have used for the registration tools and all the evaluation is *RegularStepGradientDescentOptimizerv4*. This is an optimizer based on gradient descent, with an initial step-length λ, and a relaxation factor which reduces the used step-length gradually as the direction of the gradient changes, in order to enable convergence with high accuracy. In addition to a maximum number of iterations N,



(a) Binary Image (α-cut) (b) Mask (c) Masked Image (α-cut) (d) Distance Transform (e) Horizontal Gradient (f) Vertical Gradient

Fig. 3. (a) Example α-cut in a small 2D image. (b) Binary mask. (c) α-cut after masking. (d) (Euclidean) Distance transform (for (c)). (e-f) Gradient approximation of the distance transform ΔDT.

two termination criteria are used: (i) a gradient magnitude threshold (GMT),

$$\sqrt{\frac{\partial d}{\partial T_1}^2 + \ldots + \frac{\partial d}{\partial T_{|T|}}^2} < \text{GMT}, \tag{31}$$

and (ii) a minimum step-length (MSL),

$$\lambda r < \text{MSL}, \tag{32}$$

where r is the current relaxation coefficient. We use default values of 0.0001 for both of these criteria. A relaxation factor of 0.99 is used for all experiments, since it performed well in preliminary tests; in this study we are willing to trade some (potential) additional iterations for better robustness. To maximize utilization of the limited number of α-levels, images are normalized before registration to make sure that they are within the valid $[0, 1]$ interval. We use the following robust normalization technique: Let $P_\rho(X)$ denote the ρ-th percentile of image X with respect to image intensities,

$$\text{NORM}_\rho(X) = \max\left[0, \min\left[1, \frac{(X - P_\rho(X))}{P_{1-\rho}(X) - P_\rho(X)}\right]\right]. \tag{33}$$

V. Performance Analysis

We evaluate performance of the proposed method, both for 2D and 3D images, in two different scenarios; (i) we perform a statistical study on synthetically generated images, where we seek to recover known transformations and measure the registration error by comparing the ground truth locations of known landmarks with the corresponding registered ones; (ii) we apply the proposed framework to real image analysis tasks:

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

landmark-based evaluation of registration of TEM images in 2D, and atlas-based segmentation evaluation of 3D MR images of brain.

To compare the proposed measure and registration method against the most commonly used alternative methods and similarity measures, we select the widely used ITK implementations of optimization framework and similarity measures (SSD, PCC and MI) as the baseline of intensity-based registration accuracy. Note that the PCC measure is denoted Normalized Cross Correlation (NCC) in the ITK framework.

All experiments are performed on a workstation with a 6-core Intel i7-6800K, 3.4GHz processor with 48GB of RAM and 15MB cache. The operating system used is Ubuntu 16.04 LTS. The compiler used to build the framework is g++ version 5.4.0 (20160609). Version 4.9 of the ITK-framework is used for testing and evaluation.

A. Datasets

One biomedical 2D dataset and one medical 3D dataset are used for the evaluation.

1) TEM Images of Cilia (2D): The dataset of 20 images of cilia [12] is acquired with the MiniTEM[4] imaging system. Each image is isotropic of size 129×129 pixels, with a pixel-size of a few *nm*. An example is shown in Fig. 1. A particular challenge is the near-rotational symmetry of the object: 9 pairs of rings are located around a central pair of rings, which gives 9 plausible solutions for a registration problem. The alignment of the central pair can be taken into special consideration to reduce the number of solutions to two. The dataset comes with a set of 20 landmarks per image, indicating the position of each of the relevant structures to be detected and analysed – 20 rings (2 in the center and 18 in a circle around the center). The landmarks are produced by a domain expert and are only used for evaluation of the registrations.

2) LPBA40 (3D): LPBA40 [13] is a publicly available dataset of 40 3D images of brains of a diverse set of healthy individuals, acquired with MRI. The images are anisotropic, of size $256 \times 124 \times 256$ voxels with voxel-size $0.86 \times 1.5 \times 0.86\,mm^3$. The dataset comes with segmentations of the brains into 56 distinct regions marked by a medical expert, which are used in this study as ground-truth for evaluation. LPBA40 includes two atlases: first 20 out of 40 MR images of brain in the dataset are used to generate one brain atlas by *Symmetric Groupwise Normalization* (SyGN) [32]; another atlas is created analogously, from the last 20 brains in the dataset. The atlases contain both a synthesized MR image and the fused label category in all the voxels, as well as a whole brain mask which may be used for brain extraction.

B. Evaluation Criteria

We evaluate accuracy and robustness of the registration methods in presence of noise, their robustness w.r.t. change of roles of reference and floating image (symmetry), and their speed. We quantify the performance of the observed frameworks in terms of the following quality measures:

[4]MiniTEM imaging system is developed by Vironova AB.

1) Average Error measure (AE): The registration result is quantified as the mean Euclidean distance between the sets of corresponding image corner landmarks L_{R} and $T(L_{\mathrm{F}})$ in the reference image space, after transformation of the floating image corner landmarks L_{F}, where $|L_{\mathrm{R}}|$ is the number of landmarks (4 in 2D; 8 in 3D). The quality measure is defined as

$$\mathrm{AE}(T; L_{\mathrm{R}}, L_{\mathrm{F}}) = \frac{1}{|L_{\mathrm{R}}|} \sum_{i=1}^{|L_{\mathrm{R}}|} d_E(L_{\mathrm{R}}(i), T(L_{\mathrm{F}}(i))). \quad (34)$$

A slight variation of this measure, the Average Minimal Error (AME), is used in the real task of cilia registration:

$$\mathrm{AME}(T; L_{\mathrm{R}}, L_{\mathrm{F}}) = \frac{1}{|L_{\mathrm{R}}|} \sum_{i=1}^{|L_{\mathrm{R}}|} \min_{x \in L_{\mathrm{F}}} d_E(L_R(i), T(x)) . \quad (35)$$

For the central pair, the error is simply $\mathrm{AME}_{\mathrm{CP}} = \mathrm{AME}$, whereas for the outer rings we utilize the knowledge that an odd (even) landmark should be matched with an odd (even) landmark of the other image. The error function for the outer rings, [12], is therefore defined as:

$$\begin{aligned} &\mathrm{AME}_{\mathrm{Outer}}(T; L_{\mathrm{R}}^{\mathrm{Odd}}, L_{\mathrm{F}}^{\mathrm{Odd}}, L_{\mathrm{R}}^{\mathrm{Even}}, L_{\mathrm{F}}^{\mathrm{Even}}) = \\ &\tfrac{1}{2}(\mathrm{AME}(T; L_{\mathrm{R}}^{\mathrm{Odd}}, L_{\mathrm{F}}^{\mathrm{Odd}}) + \mathrm{AME}(T; L_{\mathrm{R}}^{\mathrm{Even}}, L_{\mathrm{F}}^{\mathrm{Even}})) . \end{aligned} \quad (36)$$

2) Success Rate (SR): A registration is considered successful if its AE is below one voxel(pixel). Success rate (SR) at a given AE value corresponds to the ratio of successful registrations (w.r.t. the set of performed ones).

3) Symmetric Success Rate (SymSR): is defined as the ratio of performed registrations which are successful (i.e., $AE \leq 1$) in both directions, i.e., when the roles of reference and floating image are exchanged.

4) Inverse Consistency Error (ICE), [33]: Given a set of interest $X_A \subseteq A$, the transformations $T_{\mathrm{AB}} \colon A \to B$, and $T_{\mathrm{BA}} \colon B \to A$, the ICE of this pair of transformations is

$$\mathrm{ICE}(T_{\mathrm{AB}}, T_{\mathrm{BA}}; X_A) = \frac{1}{|X_A|} \sum_{x \in X_A} d_E(T_{\mathrm{BA}}(T_{\mathrm{AB}}(x)), x). \quad (37)$$

We compute ICE considering all the points of the reference image for each of the cases where Symmetric Success is observed ($AE \leq 1$ in both directions).

5) Jaccard Index for segmentation evaluation: For two binary sets, R_1 and R_2, the Jaccard Index is defined as

$$J(R_1, R_2) = \frac{|(R_1 \cap R_2)|}{|(R_1 \cup R_2)|}. \quad (38)$$

6) Execution Time: We evaluate (i) the execution times required for one iteration in the registration procedure, i.e., times needed to compute the distance (similarity) measure and its derivatives, with full sampling, and in full image resolution, between two distinct images from the same set, as well as (ii) the execution time for complete registrations.

C. Parameter Tuning

The distance measure and optimization method have a number of parameters which must be properly chosen. Synthetic tests indicated that the following values lead to good optimization performance: three pyramid levels with downsampling

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

factors $(4, 2, 1)$ and Gaussian smoothing $\sigma = (5.0, 3.0, 0.0)$, max 3000 iterations per level and an initial step-length $\lambda = 0.5$. The number of α-levels used is $\ell = 7$, which has shown to provide a reasonable trade-off between computational costs, sensitivity to significant variations in intensity and robustness to noise [11]. The optimal value for ℓ is application-dependent; in essentially all observed cases, $\ell > 1$ (non-crisp) outperforms a crisp (binarized) representation. Normalization percentile is normally 5%. This same parameter setting, if not stated differently, is used in all the tests, on both synthetic and real data.

D. Synthetic Tests

A synthetic evaluation framework is used to evaluate the performance of the proposed method, and to compare it with standard tools based on SSD, PCC, and MI, in a controlled environment. For this evaluation, we construct sets of transformed versions of a reference image and add (a new instance of) Gaussian noise to each generated image. The transformations are selected at random from a multivariate uniform distribution of rotations measured in degrees (1 angle for 2D images and 3 Euler angles for 3D images) and translations measured in fractions of the original image size.

1) 2D TEM Images of Cilia: Three sets of transformed images are built based on image Nr. 1 in the observed dataset, by applying on it the following three groups of transformations: *Small*, containing compositions of translations of up to 10% of image size (in any direction) and rotations by up to $10°$; *Medium*, containing compositions of translations and rotations such that at least one of the parameters exceeds the range of Small, and falls within $10 - 20\%$ of image size of translation (in at least one direction), or $10 - 20°$ of rotation; and *Large*, containing compositions of translations and rotations such that at least one of the parameters exceeds the range of Medium, and falls within $20 - 30\%$ of image size of translation (in at least one direction), or $20 - 30°$ of rotation. The transformed images are also corrupted by additive Gaussian noise, from $\mathcal{N}(0, 0.1^2)$ ($\sigma = 0.1$, corresponding to a PSNR ≈ 20 dB). Each group of transformations is applied 1000 times, and the resulting images are registered to image Nr. 1, each time corrupted by a new instance of Gaussian noise.

To evaluate symmetry, we performed 1000 registrations of images transformed by randomly selected translations of up to 30% of image size, and rotations by up to $30°$, and corrupted by additive noise from $\mathcal{N}(0, 0.1^2)$. Each of the registrations were performed twice, with exchanged roles of reference image and floating image.

Intensity-based registration with gradient-descent optimization can be computationally demanding, requiring the distance function and its derivative for each iteration of the optimization procedure. The time to compute the distance and derivatives is directly proportional to the number of sampled points. We, therefore, evaluate influence of the sampling fraction on registration success, observing registrations after *Small* transformations and added noise (with $\sigma = 0.1$), over a range of sampling fractions. For each evaluated sampling fraction, 1000 registrations are performed and SR and AE are computed for successful registrations ($AE \leq 1$). No resolution pyramids are used for these tests.

TABLE I
REGISTRATION OF SYNTHETIC 2D IMAGES OF CILIA. THE TABLES SHOW SUCCESS RATE (SR), AVERAGE ERROR (AE) OF SUCCESSFUL REGISTRATIONS, SYMMETRIC SUCCESS RATE (SYMSR), AVERAGE INVERSE CONSISTENCY ERROR (ICE) AND AVERAGE RUNTIME FOR THE REGISTRATION WITH COMPLETE SAMPLING (A) AND WITH RANDOM SUBSAMPLING (B). SUCCESSFUL REGISTRATIONS ($AE \leq 1$) OF TRANSFORMATIONS UP TO (AND INCLUDING) LARGE, ARE CONSIDERED.

Measure	SR	AE	SymSR	ICE	Time (s)
SSD	0.536	0.3086	0.313	0.2424	17.3
PCC	0.363	0.3413	0.249	0.4227	20.8
MI	0.440	0.3495	0.251	0.4518	18.3
$\bar{d}^R_{\alpha\text{AMD}}$	**1.000**	**0.1295**	**1.000**	**0.0023**	**3.49**

(a) Full sampling

Measure	SR	AE	SymSR	ICE	Time (s)
SSD	0.367	0.6270	0.186	0.5260	1.761
PCC	0.299	0.6364	0.152	0.5676	2.171
MI	0.283	0.6219	0.068	0.5974	2.083
$\tilde{d}^R_{\alpha\text{AMD}}$	**1.000**	**0.1311**	**1.000**	**0.0193**	**0.834**

(b) 0.1 sampling fraction

2) 3D MR Images of Brain: Three sets of transformed images are built based on image Nr. 1 in the observed dataset, by applying to it the following three groups of transformations: *Small*, containing compositions of translations of up to 10% of image size (in any direction) and rotations by up to $10°$ (around each of the rotation axes); *Medium*, containing compositions of translations and rotations such that at least one of the parameters exceeds the range of Small, and falls within $10 - 15\%$ of image size of translation (in at least one direction), or $10 - 15°$ of rotation (around at least one rotation axes); and *Large*, containing compositions of translations and rotations such that at least one of the parameters exceeds the range of Medium, and falls within $15 - 20\%$ of image size of translation (in at least one direction), or $15 - 20°$ of rotation (around at least one rotation axes). The transformed images are also corrupted by additive Gaussian noise, from $\mathcal{N}(0, 0.1^2)$. Each group of transformations is applied 200 times, and the resulting images are registered to image Nr. 1, each time corrupted by a new instance of Gaussian noise.

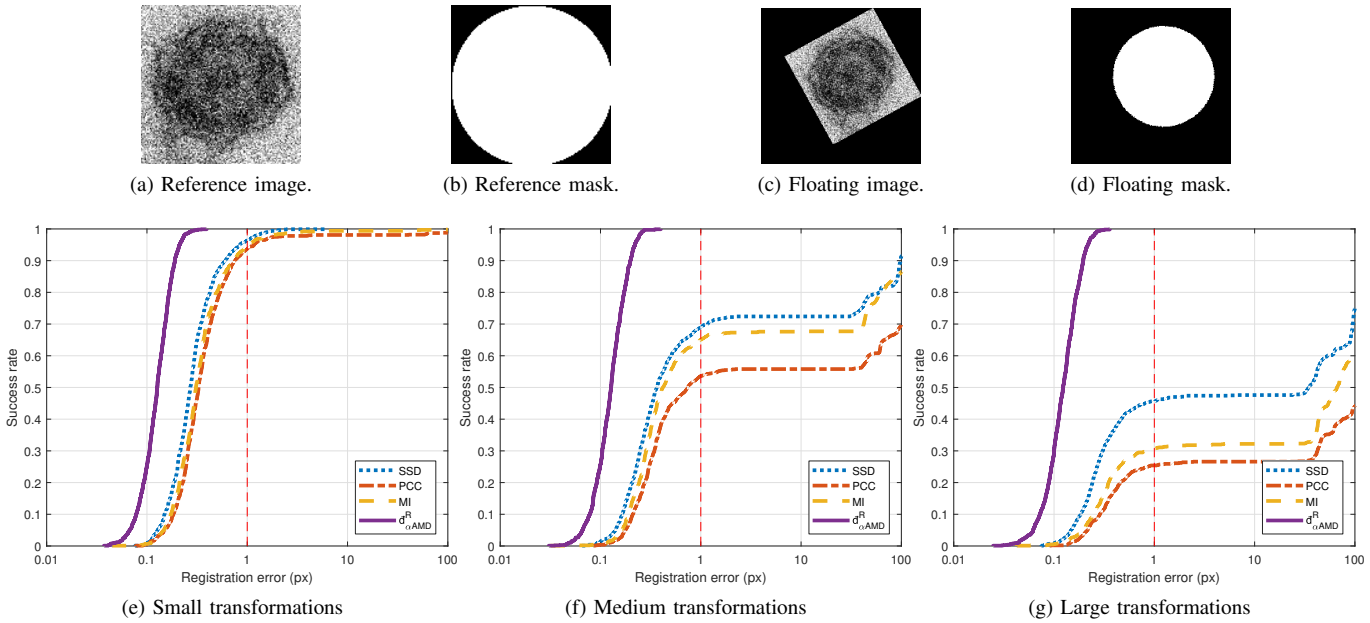
E. Results of Synthetic Tests

1) 2D TEM Images of Cilia: Figure 4 shows the distributions of registration errors (AE), for the three transformation classes. Superiority of the proposed measure, and the corresponding registration framework, is particularly clear for Medium and Large transformations; it reaches a 100% success rate, with subpixel accuracy, whereas the competitors not only exhibit considerably lower accuracy, but also much lower success rate, i.e., they completely fail in a large number of cases.

Overall registration performance is summarized in Table I, for complete sampling (a), and for random sampling of 10% of the points (b). The proposed method has 100% success rate and also 100% symmetric success rate. The other observed measures exhibit much lower success rate and poor symmetry

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing



(a) Reference image. (b) Reference mask. (c) Floating image. (d) Floating mask.

(e) Small transformations (f) Medium transformations (g) Large transformations

Fig. 4. Registration error for 2D TEM images of cilia with Gaussian noise of $\sigma = 0.1$ added, for three observed transformation classes. (a-d) Examples of reference-floating image pair with corresponding masks. (e-g) Cumulative histograms of the fraction of registrations with registration error AE below a given value (left and up is better). The red vertical line shows the chosen threshold for success, $AE \leq 1$.

scores; the second best, SSD, succeeds in 54% of the cases, and succeeds symmetrically in only 31% of the cases. The registration error for successful registrations is considerably smaller for the proposed method, while the execution time is considerably lower. The reduced sampling fraction in (b) has a small impact on the proposed method while substantially degrading the performance of the other measures.

Figure 5 shows registration performance for varying sampling fractions; Small transformations, in presence of noise ($\sigma = 0.1$) are considered. We observe that the registration performance flattens and stabilizes at approximately 0.01 sampling fraction (1% of the points). We conclude that previous findings of [17], [30], suggesting that random subsampling provides good performance even with very small sampling fractions, apply well for the proposed measure.

2) 3D MR Images of Brain: Figure 6 shows the observed distributions of registration errors (AE) for the three transformation classes, and clearly confirms that the proposed method is robust and with high performance, even for larger transformations, while the magnitude of the transformation has a substantial negative effect on the performance of the other observed measures.

Figure 7 presents bar plots corresponding to the performed synthetic tests on the LPBA40-dataset, consisting of 200 registrations of images after up to (and including) Large transformations (with additive Gaussian noise, $\mathcal{N}(0, 0.1^2)$). Successful registrations ($AE \leq 1$) are observed. Here as well, the proposed method delivers 100% success rate, whereas the second best, SSD, succeeds in only 33% of the cases. The registration error for successful registrations is the smallest for the proposed method. We observe a relative increase in execution time of the proposed registration framework in 3D case, where it is slightly slower than the other measures.

TABLE II
TIME ANALYSIS OF DISTANCE (SIMILARITY) VALUE AND DERIVATIVE COMPUTATIONS FOR A FULL RESOLUTION IMAGE, REPEATED TO GENERATE STATISTICS. BOLD MARKS THE FASTEST MEASURE IN EACH CATEGORY (2D AND 3D). THE 2D IMAGES ARE OF SIZE 1600×1278, AND THE 3D IMAGES ARE OF SIZE $256 \times 124 \times 256$.

Measure	Cilia (2D) [s]		Brain (3D) [s]	
	Mean	Std.dev.	Mean	Std.dev.
$\vec{d}^R_{\alpha\text{AMD}}$	**0.270**	0.010	**3.116**	0.098
SSD	0.718	0.036	4.782	0.066
PCC	1.191	0.026	8.562	0.002
MI	0.890	0.025	5.699	0.002

3) Execution Time Analysis: The number of iterations required for convergence of the optimization (registration) typically range from 1000 to 3000. Measures SSD, PCC and MI use cubic spline interpolation. Lookups from the distance maps for $\vec{d}^R_{\alpha\text{AMD}}$ are done using linear interpolation. Table II shows the mean (and standard deviation) execution time of one iteration, which includes computation of the measures and their derivatives, repeated 1000 times for 2D, and 50 times for 3D affine image registrations. We observe that the proposed measure is the fastest per iteration both in 2D and 3D. Note that these execution time measurements exclude preprocessing.

F. Evaluation on Real Applications

1) Registration of Cilia: Registration of multiple cilia instances detected in a single TEM sample, for enhancement of diagnostically relevant sub-structures, requires a pixel-accurate and robust method which is able to overcome the challenges posed by the near-rotational symmetry of a cilium. At most two of the possible solutions properly align the central pair, which is vital for a successful reconstruction.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

We compare the performance of the proposed method with reported results of a previous study [12] which uses intensity-based registration with PCC as similarity measure. We follow the general protocol described in [12] and perform, as a first step, a multi-start rigid registration (parameterized by angle θ in radians, and translation $\mathbf{t} = (t_x, t_y)$), followed, in a second step, by affine registration initiated by the best (lowest final distance) registration of the 9 rigid ones.

No resolution pyramids are used since they were observed to interfere with the multi-start approach (by facilitating large movements). The registrations are performed in full resolution, without stochastic subsampling. For the rigid registration we use a small circular binary mask with radius of 24 pixels, positioned in the center, combined with a squared circular Hann window function. The affine registration is performed using a circular binary mask with radius of 52 pixels; the mask removes the outside background and the outer plasma which is not helpful in guiding the registration. No additional weight-mask is used for the affine registration. Step length 0.1 was used for the rigid and 0.5 for the affine registration. We use $\ell = 7$. Normalization percentile is set to 0% for the rigid stage and 1% for the affine stage.

A feature-based approach is also included in this performance evaluation. The SIFT feature-detector [7], with RANSAC [34] as model fitting and correspondence point filtering method, as implemented in FIJI, is evaluated with both rigid and affine transformation models. The tests are performed with, and without, circular masks (as described above), and with systematically varied parameter settings (using grid search): initial Gaussian blur tested with values in the range $[0.4, 2.4]$, with steps of 0.4; feature descriptor size tested with $\{1, 2, 4, 6, 8\}$; steps per scale octave tested with $\{1, 2, 3, 4, 5\}$. The other available parameters are set to their default values, since we observed insensitivity to those parameters in our preliminary tests.

2) Atlas-based Segmentation (LPBA40): In [35], a protocol for evaluation of distance/similarity measures in the context of image registration was proposed. The protocol starts with affine registration, for which results are reported, and then proceeds to deformable registration. Since this study focuses on the development of an affine (linear) registration framework based on the proposed distance measure, we compare with the reported affine-only performance; an improved affine registration is of great significance since a very high correlation between the performance of the affine registration and that of the subsequent deformable registration has been established.

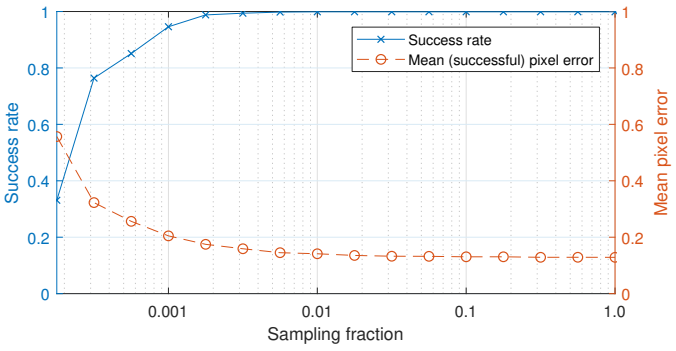


Fig. 5. (Left/Blue) SR for registrations of cilia images, and (Right/Red) AE of the successful registrations, as functions of sampling fraction for the proposed method. Both measures improve (almost) monotonically with sampling fraction and flatten out after approximately 0.01.

TABLE III
REGISTRATION OF CILIA: PERFORMANCE OF THE PROPOSED METHOD COMPARED TO REFERENCE RESULTS, SHOWN AS THE 'MEAN (STD-DEV)' OF THE REGISTRATION ERROR (IN PIXELS) W.R.T. THE CONSIDERED SETS OF LANDMARKS FOR THE 19 REGISTRATIONS. 'R' DENOTES RIGID, 'A' DENOTES AFFINE AND 'D' DIENOTES DEFORMABLE REGISTRATION. BOLD MARKS THE SMALLEST ERROR FOR EACH SET OF LANDMARKS.

Method		Registration Error		
Measure	Transform	Central Pair	Outer	All
-	Identity	4.32 (0.78)	5.75 (3.49)	5.61 (3.09)
PCC [12]	R	2.6 (1.5)	-	-
	R+A	-	3.27 (1.75)	-
	R+A+D	2.79 (1.84)	2.30 (1.80)	2.35 (1.82)
$\tilde{d}^R_{\alpha\text{AMD}}$	R	2.65 (0.83)	6.49 (2.64)	6.10 (2.41)
	R+A	**2.03 (1.04)**	**1.64 (0.36)**	**1.68 (0.29)**

We start from the two atlases created utilizing the Advanced Neuroimaging Tools (ANTs) registration software suite and the open-source evaluation script provided in the reference study [35]. We utilize the atlas created using Mutual Information since that is the one found in [35] to be best performing and is used as the basis for the whole deformable registration study. Two-fold cross validation is utilized; the first atlas is registered to the last 20 brain images and the second atlas is registered to the first 20 brains, hence all registrations are done with brains that did not contribute to the creation of the atlas.

The multi-label segmentations defined by the atlas are transformed using the transformation parameters found during the registration and compared to the ground-truth segmentations for each brain. The Jaccard Index [36] is calculated per region, as well as for the entire brain mask.

For the proposed method based on $\tilde{d}^R_{\alpha\text{AMD}}$ we use $\ell = 7$, normalization percentiles 5%, $N = 3000$, 0.05 sampling fraction, and circular Hann windows as weight-masks.

G. Results of Real Applications

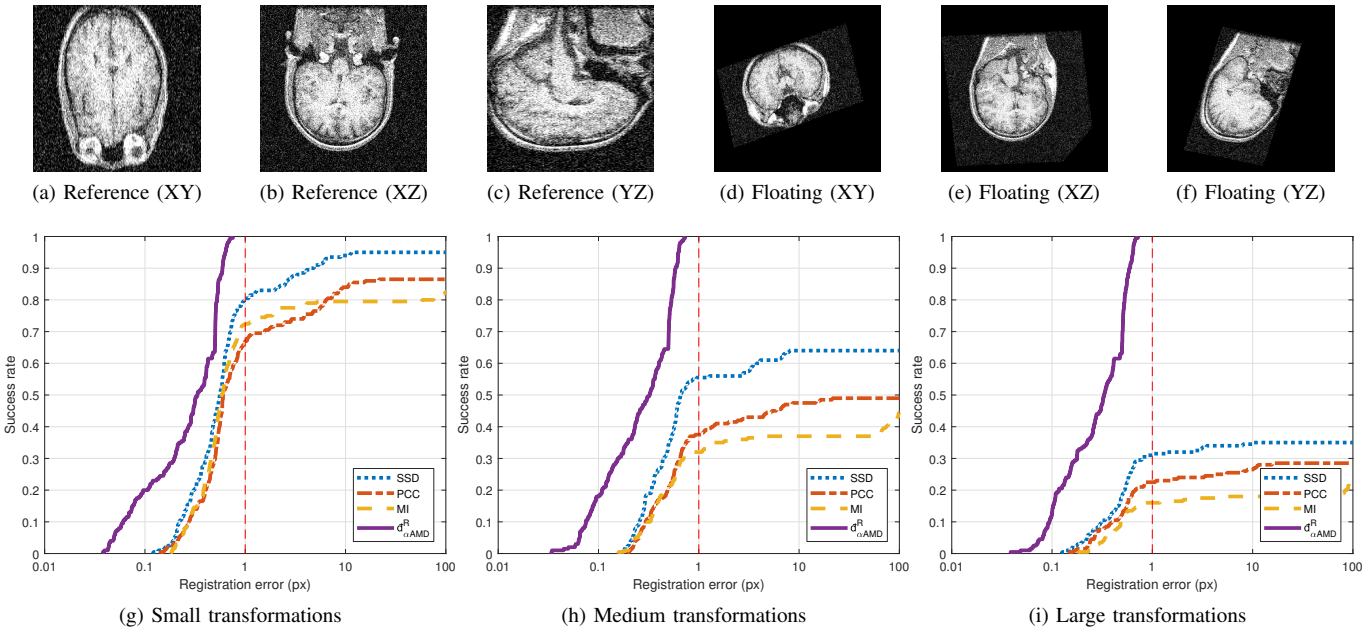
1) Results of Registration of Cilia: Performance of the proposed method, together with the best previously published results, are shown in Tab. III. The table shows the mean and standard deviation of registration error (AME, in pixels) of the 19 registrations, for the three considered sets of landmarks: the Central pair, the Outer rings, and All (1+9) ring pairs. 'R' denotes rigid; 'A' denotes affine; and 'D' denotes deformable registration.

The original study includes deformable registration as a final stage, after the rigid and affine steps. Here presented framework based on $\tilde{d}^R_{\alpha\text{AMD}}$ includes linear (rigid and affine), but not deformable registration. However, as results included in Tab. III confirm, the proposed method outperforms the previous state-of-the-art, even if using only rigid and affine registrations.

We note that with only rigid registration we improve the alignment of the central pair while degrading the alignment of the outer rings. After the affine registration, the alignment of

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

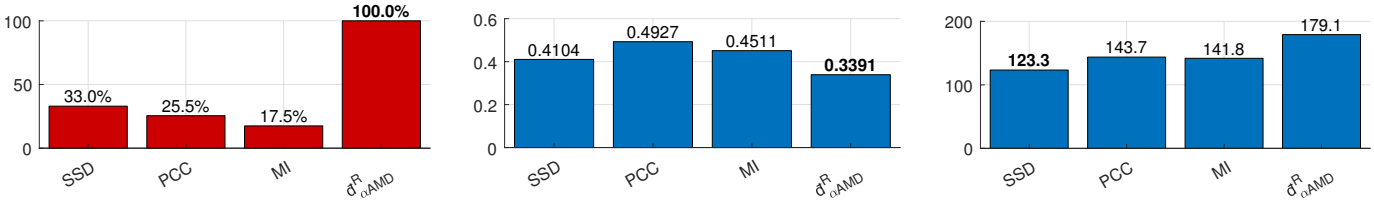
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing



(a) Reference (XY) (b) Reference (XZ) (c) Reference (YZ) (d) Floating (XY) (e) Floating (XZ) (f) Floating (YZ)

(g) Small transformations (h) Medium transformations (i) Large transformations

Fig. 6. Registration error for 3D MR images of brain with Gaussian noise, $\sigma = 0.1$, added, for three observed transformation magnitudes classes. (a-f) Example of reference-floating image pair in slices along each major axis. (g-i) Cumulative histograms of the fraction of registrations with registration error AE below a given value (left and up is better). The red vertical line shows the chosen threshold of success, $AE \leq 1$.



(a) Success-rate (%) (b) Mean error (px) (c) Execution time (s)

Fig. 7. Results of synthetic registration of 3D brain images from the LPBA40 dataset. The plots show the (a) success-rate (SR), (b) mean error (ME) for successful registrations and (c) the average runtime in seconds for the registration with random subsampling with 0.01 sampling fraction. (a) Higher is better. (b-c) Lower is better. Bold marks the best result w.r.t. each statistic.

the central pair is improved further, plausibly due to the less constrained transformation model of affine compared to rigid, and we observe that the alignment of the outer rings and the total alignment are improved substantially.

The feature-based method is omitted from Tab. III due to complete failure on all 19 image registration tasks, both with rigid and affine transformations; either too few matching points were detected, or the ones found resulted in large erroneous transformations. One such failed registration example is illustrated in Fig. 1.

2) Results of Atlas-Based Segmentation of Brains: Table IV shows results of atlas-based brain segmentation. The Mean Jaccard Index is computed for each of the brain regions, for $\tilde{d}^R_{\alpha\text{AMD}}$ and MI, with affine registration as reported in [35]. For $\tilde{d}^R_{\alpha\text{AMD}}$, mean and std. dev. are displayed; for the comparative results (MIAff, [35]), only mean was reported.

We observe that for the whole brain mask, for the aggregated overlap, and for 43 out of the 56 distinct regions, the proposed measure outperforms the reported performance obtained with the MI metric; MI was the best performing measure out of the three evaluated in [35].

VI. Discussion

Compared to the traditional similarity measures (SSD, PCC, MI), the proposed measure and associated registration method require substantial amounts of memory to store the auxiliary data-structures. A single 3D registration of two MR images of brains may require approximately 4GB of working memory with a reasonable set of parameters; contemporary machines for high-end data processing typically have a lot more memory than 4GB, but this requirement can affect how many registrations can be performed in parallel on a single machine.

VII. Conclusion

In this study we have adapted a family of distance measures [11] to gradient descent based image registration, for 2D and 3D images. We have shown that such an extension is feasible and that the very good performance of the measures observed previously for object recognition and template matching, and their property of a large catchment basin for local optimization, also hold in the context of registration. This has been shown by evaluating the method in four main ways: (i) on synthetic tests,

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

TABLE IV
RESULTS OF ATLAS-BASED BRAIN SEGMENTATION. THE TABLE SHOWS THE MEAN JACCARD INDEX FOR EACH OF THE BRAIN REGIONS FOR $\tilde{d}^R_{\alpha\text{AMD}}$ AND MUTUAL INFORMATION WITH AFFINE REGISTRATION AS REPORTED IN [35]. FOR $\tilde{d}^R_{\alpha\text{AMD}}$, MEAN AND STD. DEV. ARE DISPLAYED; FOR THE COMPARATIVE RESULTS (MIAFF), ONLY MEAN WAS REPORTED.

LPBA40 Label	$\tilde{d}^R_{\alpha\text{AMD}}$	MIAff [35]
All_LPBA_Data	**0.595** ± 0.0187	0.554
Brain	**0.922** ± 0.0082	0.905
21_L_superior_frontal_gyrus	0.690 ± 0.0351	**0.708**
22_R_superior_frontal_gyrus	0.683 ± 0.0322	**0.748**
23_L_middle_frontal_gyrus	**0.672** ± 0.0345	0.536
24_R_middle_frontal_gyrus	**0.663** ± 0.0451	0.513
25_L_inferior_frontal_gyrus	**0.590** ± 0.0467	0.569
26_R_inferior_frontal_gyrus	**0.591** ± 0.0562	0.550
27_L_precentral_gyrus	**0.560** ± 0.0579	0.503
28_R_precentral_gyrus	**0.550** ± 0.0584	0.508
29_L_middle_orbitofrontal_gyrus	**0.529** ± 0.0723	0.505
30_R_middle_orbitofrontal_gyrus	**0.522** ± 0.0569	0.484
31_L_lateral_orbitofrontal_gyrus	0.434 ± 0.0828	**0.551**
32_R_lateral_orbitofrontal_gyrus	0.421 ± 0.0846	**0.564**
33_L_gyrus_rectus	**0.507** ± 0.0571	0.503
34_R_gyrus_rectus	**0.536** ± 0.0712	0.485
41_L_postcentral_gyrus	0.479 ± 0.0739	**0.490**
42_R_postcentral_gyrus	**0.474** ± 0.0653	0.463
43_L_superior_parietal_gyrus	**0.563** ± 0.0509	0.470
44_R_superior_parietal_gyrus	**0.569** ± 0.0532	0.470
45_L_supramarginal_gyrus	0.502 ± 0.0719	**0.504**
46_R_supramarginal_gyrus	**0.510** ± 0.0702	0.463
47_L_angular_gyrus	**0.509** ± 0.0782	0.506
48_R_angular_gyrus	**0.520** ± 0.0527	0.472
49_L_precuneus	0.525 ± 0.0613	**0.546**
50_R_precuneus	**0.541** ± 0.0610	0.540
61_L_superior_occipital_gyrus	**0.424** ± 0.0825	0.413
62_R_superior_occipital_gyrus	**0.409** ± 0.0665	0.399
63_L_middle_occipital_gyrus	**0.516** ± 0.0686	0.421
64_R_middle_occipital_gyrus	**0.512** ± 0.0573	0.397
65_L_inferior_occipital_gyrus	0.448 ± 0.0967	**0.484**
66_R_inferior_occipital_gyrus	0.451 ± 0.0914	**0.492**
67_L_cuneus	**0.442** ± 0.1087	0.372
68_R_cuneus	**0.445** ± 0.0889	0.388
81_L_superior_temporal_gyrus	**0.574** ± 0.0478	0.514
82_R_superior_temporal_gyrus	**0.586** ± 0.0446	0.498
83_L_middle_temporal_gyrus	**0.513** ± 0.0580	0.481
84_R_middle_temporal_gyrus	**0.540** ± 0.0495	0.473
85_L_inferior_temporal_gyrus	**0.509** ± 0.0601	0.462
86_R_inferior_temporal_gyrus	**0.534** ± 0.0572	0.460
87_L_parahippocampal_gyrus	0.546 ± 0.0743	**0.556**
88_R_parahippocampal_gyrus	0.535 ± 0.0630	**0.544**
89_L_lingual_gyrus	**0.519** ± 0.0943	0.421
90_R_lingual_gyrus	**0.541** ± 0.0678	0.420
91_L_fusiform_gyrus	**0.542** ± 0.0801	0.488
92_R_fusiform_gyrus	**0.548** ± 0.0646	0.453
101_L_insular_gyrus	**0.625** ± 0.0504	0.378
102_R_insular_gyrus	**0.611** ± 0.0639	0.420
121_L_cingulate_gyrus	**0.553** ± 0.0504	0.491
122_R_cingulate_gyrus	**0.545** ± 0.0588	0.508
161_L_caudate	**0.583** ± 0.1002	0.494
162_R_caudate	**0.574** ± 0.0952	0.502
163_L_putamen	**0.626** ± 0.0581	0.559
164_R_putamen	**0.636** ± 0.0668	0.561
165_L_hippocampus	0.603 ± 0.0836	**0.633**
166_R_hippocampus	0.604 ± 0.0810	**0.643**
181_cerebullum	**0.813** ± 0.0416	0.659
182_brainstem	**0.778** ± 0.0447	0.660

(ii) execution time measurement, (iii) registration of TEM-images of cilia for multi-image super-resolution reconstruction, and (iv) atlas-based segmentation with annotated MR brain images. We observe that the proposed method provides outstanding performance for intensity-based affine registration in terms of robustness, accuracy and symmetry. It is also faster or similar in speed to the commonly used measures, which allows its practical applications. The framework developed in this study operates on single-layer (e.g. gray-scale) images, but can be extended to multi-layer images such as color images, either by considering a linear sum of distances, or more sophisticated methods based on simultaneous presence or absence of membership in the multiple layers [23], [37]. Future work includes extending the measures to non-linear (deformable), as well as multi-modal registration.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ida-Maria Sintorn for providing the cilia dataset and the landmarks used for the evaluation in Sec. V-F1. This work is supported by VINNOVA, MedTech4Health grants 2016-02329 and 2017-02447, Swedish Research Council grants 2015-05878 and 2017-04385, and the Ministry of Education, Science, and Techn. Development of the Republic of Serbia (proj. ON174008 and III44006).

REFERENCES

[1] J. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1 – 36, 1998.
[2] M. A. Viergever, J. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. Pluim, "A survey of medical image registration – under review," *Medical Image Analysis*, vol. 33, pp. 140 – 144, 2016.
[3] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.
[4] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review," *Computer methods in biomechanics and biomedical engineering*, vol. 17, no. 2, pp. 73–93, 2014.
[5] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.
[6] S. Matl, R. Brosig, M. Baust, N. Navab, and S. Demirci, "Vascular image registration techniques: a living review," *Medical image analysis*, vol. 35, pp. 1–17, 2017.
[7] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2. IEEE, 1999, pp. 1150–1157.
[8] B. Berkels, P. Binev, D. A. Blom, W. Dahmen, R. C. Sharpley, and T. Vogt, "Optimized imaging using non-rigid registration," *Ultramicroscopy*, vol. 138, pp. 46 – 56, 2014.
[9] B. Fischer and J. Modersitzki, "Ill-posed medicine - an introduction to image registration," *Inverse Problems*, vol. 24, no. 3, p. 034008, 2008.
[10] D. Skerl, B. Likar, and F. Pernus, "A protocol for evaluation of similarity measures for rigid registration," *IEEE Transactions on Medical Imaging*, vol. 25, no. 6, pp. 779–791, June 2006.
[11] J. Lindblad and N. Sladoje, "Linear time distances between fuzzy sets with applications to pattern matching and classification," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 126–136, 2014.
[12] A. Suveer, N. Sladoje, J. Lindblad, A. Dragomir, and I.-M. Sintorn, "Enhancement of cilia sub-structures by multiple instance registration and super-resolution reconstruction," in *Scandinavian Conference on Image Analysis*. Springer, 2017, pp. 362–374.
[13] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga, "Construction of a 3D probabilistic atlas of human cortical structures," *Neuroimage*, vol. 39, no. 3, pp. 1064–1080, 2008.
[14] B. B. Avants, N. J. Tustison, M. Stauffer, G. Song, B. Wu, and J. C. Gee, "The Insight ToolKit image registration framework," *Frontiers in neuroinformatics*, vol. 8, p. 44, 2014.
[15] L. A. Zadeh, "Information and control," *Fuzzy sets*, vol. 8, no. 3, pp. 338–353, 1965.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2019.2899947, IEEE Transactions on Image Processing

[16] J. V. Hajnal, N. Saeed, E. J. Soar, A. Oatridge, I. R. Young, and G. M. Bydder, "A registration and interpolation procedure for subvoxel matching of serially acquired MR images," *Journal of computer assisted tomography*, vol. 19, no. 2, pp. 289–296, 1995.

[17] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.

[18] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[19] A. Rosenfeld, *Multiresolution image processing and analysis*. Springer Science & Business Media, 2013, vol. 12.

[20] J. Lindblad, V. Curic, and N. Sladoje, "On set distances and their application to image registration," in *Proceedings International Symposium on Image and Signal Processing and Analysis*, Sept 2009, pp. 449–454.

[21] T. Eiter and H. Mannila, "Distance measures for point sets and their computation," *Acta Informatica*, vol. 34, no. 2, pp. 109–133, 1997.

[22] P. Brass, "On the nonexistence of hausdorff-like metrics for fuzzy sets," *Pattern Recognition Letters*, vol. 23, no. 1-3, pp. 39–43, 2002.

[23] J. Öfverstedt, N. Sladoje, and J. Lindblad, "Distance between vector-valued fuzzy sets based on intersection decomposition with applications in object detection," in *Int. Symp. on Mathematical Morphology and Its Applications to Signal and Image Process.* Springer, 2017, pp. 395–407.

[24] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.

[25] W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, and G. K. Rohde, "A linear optimal transportation framework for quantifying and visualizing variations in sets of images," *International journal of computer vision*, vol. 101, no. 2, pp. 254–269, 2013.

[26] O. Pele and M. Werman, "Fast and robust earth mover's distances." in *ICCV*, vol. 9, 2009, pp. 460–467.

[27] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *International Journal of computer vision*, vol. 60, no. 3, pp. 225–240, 2004.

[28] T. ur Rehman, E. Haber, G. Pryor, J. Melonakos, and A. Tannenbaum, "3d nonrigid registration via optimal mass transport on the gpu," *Medical image analysis*, vol. 13, no. 6, pp. 931–940, 2009.

[29] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[30] 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*, vol. 16, no. 12, pp. 2879–2890, 2007.

[31] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265–270, 2003.

[32] B. B. Avants, P. Yushkevich, J. Pluta, D. Minkoff, M. Korczykowski, J. Detre, and J. C. Gee, "The optimal template effect in hippocampus studies of diseased populations," *Neuroimage*, vol. 49, no. 3, pp. 2457–2466, 2010.

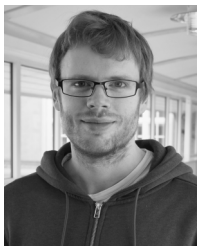
[33] G. E. Christensen and H. J. Johnson, "Consistent image registration," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 568–582, 2001.

[34] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[35] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.

[36] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, p. 29, 2015.

[37] N. Sladoje and J. Lindblad, "Distance between vector-valued representations of objects in images with application in object detection and classification," in *International Workshop on Combinatorial Image Analysis*. Springer, 2017, pp. 243–255.



Johan Öfverstedt received the B.Sc. and M.Sc. degrees in computer science in 2012 and 2017 respectively, and is currently pursuing a Ph.D. degree in computerized image processing at Centre for Image Analysis, Uppsala University, Sweden, since 2017. His research interests include distance measures, image registration, object recognition, and optimization.

Joakim Lindblad received the M.Sc. in engineering physics and Ph.D. in computerized image analysis from Uppsala University, Sweden, in 1997 and 2003, respectively. He is currently Researcher at the Centre for Image Analysis, Uppsala University, Sweden; Senior Research Associate at Mathematical Institute of the Serbian Academy of Sciences and Arts, Serbia; and Head of Research at Topgolf Sweden AB, Stockholm, Sweden. His research interests include development of general and robust methods for image processing and analysis.

Nataša Sladoje received the B.Sc. and M.Sc. degrees in mathematics from the Faculty of Science, University of Novi Sad, Serbia, in 1992 and 1998, respectively, and the Ph.D. degree in computerized image analysis from the Centre for Image Analysis, Swedish University of Agricultural Sciences, Uppsala, Sweden, in 2005. She is a Senior Lecturer at the Centre for Image Analysis, Uppsala University, Sweden. Her research interests include theoretical development of image analysis methods with applications in biomedicine and medicine.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.