An efficient EM-ICP algorithm for non-linear registration of large 3D point sets

Benoit Combès, Sylvain Prima

Computer Vision and Image Understanding
journal homepage: www.elsevier.com

An efficient EM-ICP algorithm for non-linear registration of large 3D point sets

Benoit Combès[a], Sylvain Prima[b,**]

[a] *Visages Team at Inria Rennes - Bretagne Atlantique, Campus de Beaulieu, F-35042 Rennes, France*
[b] *Serpico Team at Inria Rennes - Bretagne Atlantique, Campus de Beaulieu, F-35042 Rennes, France*

ABSTRACT

In this paper, we present a new method for non-linear pairwise registration of 3D point sets. In this method, we consider the points of the first set as the draws of a Gaussian mixture model whose centres are the displaced points of the second set. Next we perform a maximum a posteriori estimation of the parameters (which include the unknown transformation) of this model using the expectation-maximisation (EM) algorithm. Compared to other methods using the same "EM-ICP" framework, we propose four key modifications leading to an efficient algorithm allowing for fast registration of large 3D point sets: 1) truncation of the cost function; 2) symmetrisation of the point-to-point correspondences; 3) specification of priors on these correspondences using differential geometry; 4) efficient encoding of deformations using the RKHS theory and the Fourier analysis. We evaluate the added value of these modifications and compare our method to the state-of-the-art CPD algorithm on real and simulated data.

1. Introduction

Non-linear registration (or alignment) is the process of estimating and applying a geometrical transformation to a first dataset to superpose it on a second dataset, so as to make the homologous objects/structures (or parts/subsets thereof) in both sets coincide. The need for automated registration methods is common to many fields such as computer vision, medical image analysis, biometrics, *etc*. Applications include the analysis of movements in videos, the assessment of tumour growth in longitudinal brain MRI datasets, the recognition and indexing of shapes, *etc*.

In practice, a registration method implicitly assumes the choice of (i) a way to represent the structures to register (*e.g.* grey level images, surfaces, point sets, *etc*.), (ii) a model to explicit the nature of the expected deformations (or movements) and (iii) a metric to specify what registration/alignment means.

As for the first point (i), a particularly convenient way to focus on specific objects/structures is to first isolate them from the rest of the image/video by segmentation and then use their outline surface to represent them; this outline surface can also be directly estimated through the use of direct 3D acquisition devices (*e.g.* laser scanners). Structured (meshes) or unstructured point sets are the most generic way for such a representation, and this is what we are interested in here. In this context, numerous methods have been proposed in the literature for non-linear registration of point sets. It turns out that before specifying choices for points (ii) and (iii) above, most of these methods (often implicitly) resort to intermediate mathematical representations of the two point sets to register.

An especially attractive approach is to consider the points of the first set as the draws of a Gaussian mixture model (GMM) whose centres are the points of the second dataset. In this case, registering the two datasets can be cast into a maximum a posteriori (MAP) problem (or equivalently, into a penalised likelihood problem), where the only unknown parameter to be estimated (if one considers the variance of the GMM as fixed and known) is the non-linear transformation, which can be done efficiently using the expectation-maximisation (EM) algorithm. The estimation of the transformation then boils down to a simple iterative estimation of fuzzy point-to-point correspondences (called *matching* later on) between the two sets (E-step) (encoded in what is often termed the *match matrix*) and of the non-linear transformation (M-step) in turn. Apart from its simplicity, such a scheme also allows one to partly alleviate the complicated problem of defining binary point-to-point corre-

** Corresponding author: Tel.: +33-2-99-84-73-59;
e-mail: sylvain.prima@inria.com (Sylvain Prima)

spondences between the point sets (as, in practice, they do not necessarily exist, first of all because the two datasets are usually of different sizes) by indirectly introducing probabilities of correspondence; this can be critical, for instance, to build relevant shape models (Hufnagel et al., 2008).

This framework has been extensively used in the literature, as shown in Section 2, often without being identified as such. Most notably, if one uses the classification likelihood rather than the penalised likelihood, the expectation-classification-maximisation (ECM) rather than the EM algorithm, and rigid-body transformations, the resulting registration algorithm is simply the ICP (Besl and McKay, 1992). If one uses the likelihood, the EM algorithm and rigid-body transformations, the resulting registration algorithm is the EM-ICP (Granger and Pennec, 2002). If one uses the penalised likelihood, the EM algorithm and TPS (thin plate splines) transformations (resp. the motion coherence theory), the resulting registration algorithm is the TPS-RPM (resp. the CPD) or, to be more precise, a simplified version thereof (Chui et al., 2003; Myronenko and Song, 2010).

Despite its simplicity and intuitiveness, this framework (which we later call *EM-ICP* for the sake of clarity) is undermined by several intrinsic limitations that have not been extensively studied and dealt with, to the best of our knowledge. First, given two point sets X and Y, building the match matrix in the E-step involves the evaluation of $\text{card}(X) \times \text{card}(Y)$ values, which is prohibitive when dealing with high-resolution data, typically involved in medical image analysis or computer graphics; in addition, potential outliers can make some components of the matrix meaningless, and severely affect the subsequent estimation of the deformation. Second, one can see that the derivation of the MAP principle leads to an asymmetric algorithmic formulation. In particular, in this framework, the match matrix arises as a row stochastic matrix (leading to many-to-one correspondences). This asymmetric formulation makes the algorithm unable to achieve a good point-to-point matching in specific cases and makes the choice of source and target sets critical. Third, the overall iterative scheme exhibits a monotonic convergence that leads to a local maximum of the MAP criterion and thus can provide a bad estimate of the deformation when a bad initialisation is provided. Finally, the M-step is very time and memory consuming when dealing with large datasets.

Our **fourfold contribution** is to propose efficient solutions for each of these four problems. We start by noting that the derived EM algorithm can be seen as an iterative alternated minimisation (over the match matrix and the unknown transformation) of an (energetic) criterion composed of a data-attachment term and a regularisation term. In Section 3, we propose to truncate the quadratic function in the data- attachment term using a cut-off distance, which drastically reduces the computational cost of the E-step while making it robust to outliers. In Section 4, we propose to tackle problems due to the asymmetry of the MAP formulation by introducing a second, column stochastic, match matrix within the energetic criterion to make it more symmetric, albeit without a straightforward statistical interpretation. This modification, only changes the E-step and improves the estimation of correspondences. In Section 5, we show how to specify priors on the two match matrices with only minor changes to the optimisation algorithm. These priors based on local and global shape descriptors allow for significant improvement of the capture range of the algorithm. In Section 6, we devise an efficient solution for the M-step that stands on the Reproducing Kernel Hilbert Space theory and on the Fourier analysis. It consists in building efficient regularisers leading to a closed-form solution that can be implemented using sparse linear algebra. In Section 7, we discuss the merits of our algorithm with respect to the two most popular algorithms using the same framework: the TPS-RPM (Chui et al., 2003) and the CPD (Myronenko and Song, 2010) algorithms. Finally, in Section 8, we evaluate the added value of our improvements and compare our method with the state-of-the-art CDP algorithm.

2. Non-linear registration as a statistical inference problem

In Appendix A we provide a comprehensive (although not exhaustive) taxonomy of existing methods for non-linear registration of point sets. This taxonomy is based on the intermediate representations used to model the points sets, and we identify four broad classes depending on whether the sets are considered as modal matrices, level set functions, Schwartz distributions or mixtures of probability density functions (pdfs); within each of these classes, we identify some popular similarity metrics between the intermediate representations. We then give a short list of some popular choices of models for non-linear deformations.

Among these methods, to date the most popular approach to find T best superposing points sets X and Y has been to view the point set Y as a noised version of $T(X)$, with a simple model of isotropic Gaussian noise on data $T(X)$. A simple way to formulate this viewpoint is to assume that each of the M samples y_j ($M = \text{card}(Y)$) has been drawn independently from any one of the $N = \text{card}(X)$ possible 3-variate normal distributions with centres (means) $T(x_k)$ and covariance matrices $\sigma^2 I$ (with $\sigma > 0$).

This way, the registration problem becomes one of statistical inference, whose challenge is (i) to find the *label* of each point y_j, *i.e.* the one out of N possible distributions from which y_j has been drawn, and (ii) to estimate the parameters of these N distributions. The connection between registration and statistical inference becomes clear when one realises that (i) actually amounts to match each point y_j in Y with a point x_k in X, while (ii) simply consists in computing T given these matches.

This viewpoint is extremely fruitful, as it allows one to refer to classical optimisation techniques to solve the registration problem. Three different paradigms have been especially followed in this context (Marriott, 1975; Combès and Prima, 2009). Let us introduce some notations first:

$\forall k \in 1, \ldots, N, \psi_k(.;T) = \mathcal{N}(T(x_k), \sigma^2 I)$

$\forall j \in 1, \ldots, M, \forall k \in 1, \ldots, N, z_{jk} = 1$ iff y_j is drawn from $\psi_k(.;T)$

In the Classification Maximum Likelihood (CML) approach, T is considered as a fixed unknown parameter and one

tries to find the indicator variables z_{jk} and the transformation T so as to maximise the classification likelihood CL (Scott and Symons, 1971):

$$CL = \prod_{j=1}^{M} \prod_{k=1}^{N} \left[\psi_k(y_j; T)\right]^{z_{jk}} \tag{1}$$

The maximisation is typically performed by the Classification EM (CEM) algorithm (G. Celeux and G. Govaert, 1992), which can be shown to find an at least local maximum of CL and proceeds as follows, in an iterative way, starting from an initial value $\tilde{T}$:

EC-step: for each j, $\tilde{z}_{jk} = 1$ iff k maximises $\psi_k(y_j; \tilde{T})$
M-step:
$\tilde{T} = \arg\min_T \sum_{j,k} \tilde{z}_{jk} \|y_j - T(x_k)\|^2$

In other words, the Expectation-Classification (EC) step consists in matching each point y_j of Y with the closest point in $\tilde{T}(X)$, while the Maximisation (M) step consists in computing the transformation best superposing these pairs of matched points. When T is modelled as a rigid-body transformation, this algorithm is nothing else than the popular Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992). Note that this algorithm does not depend on σ.

In the Maximum Likelihood (ML) approach, the indicator values z_{jk} are no longer considered as unknown quantities to estimate, but rather as hidden/unobservable variables. This is actually a drastic and fundamental change of viewpoint, as the focus is no longer on assigning each y_j to one of the distributions ψ_k but rather on estimating the parameters of the Gaussian mixture made of these distributions. If we involve priors π_{jk} on the indicator variables ($\forall j, k,\ 0 < \pi_{jk} < 1$, and $\forall j,\ \sum_k \pi_{jk} = 1$), the likelihood then simply writes (Day, 1969):

$$L = \prod_{j=1}^{M} \sum_{k=1}^{N} \pi_{jk} \psi_k(y_j; T) \tag{2}$$

In essence, the prior π_{jk} conveys the probability that the point y_j is drawn from the distribution ψ_k regardless of T. The likelihood L can be maximised by using the popular EM algorithm, which converges to an at least local maximum of L (Dempster et al., 1977). If we consider the priors π_{jk} as known beforehand and if we introduce the notation A_{jk} as the posterior probability of the hidden indicator variable z_{jk} to be equal to 1, the EM algorithm writes:

E-step: $\tilde{A}_{jk} = \frac{\pi_{jk} \exp\left[-\|y_j - \tilde{T}(x_k)\|^2/(2\sigma^2)\right]}{\sum_i \pi_{ji} \exp\left[-\|y_j - \tilde{T}(x_i)\|^2/(2\sigma^2)\right]}$
M-step: $\tilde{T} = \arg\min_T \sum_{j,k} \tilde{A}_{jk} \|y_j - T(x_k)\|^2$

$A = (A_{jk})$ is termed the match matrix and is a row stochastic matrix. The parameter σ acts as a scale parameter. When T is modelled as a rigid-body transformation and priors π_{jk} are chosen to be uniform, this algorithm is nothing else than the EM-ICP algorithm (σ being given an initial value and decreased throughout the iterations for an improved capture range) (Granger and Pennec, 2002). Intuitively, the ML approach is a *fuzzy* version of the CML. It appears clearly from the iterative formulas of the CEM and the EM algorithm that the classification likelihood is an "all-or-nothing" version of the likelihood, leading to a "bumpier" and harder-to-maximise criterion, something that is well known by those who are familiar with the ICP algorithm.

In the Maximum A Posteriori (MAP) approach, instead of simply considering T as a fixed unknown parameter of the pdfs $\psi_k(., T)$, one can consider it as a random variable on which priors (acting as regularisers on T) can be easily specified. Then, the ML estimation can be easily turned into a MAP problem with only slight modifications to the optimisation scheme, as shown by Green (Green, 1990). If $p(T)$ is a prior of the form $p(T) \propto \exp(-\alpha R(T))$ (*i.e.* a Gibbs prior) then the optimal deformation can be found by maximising the posterior (also termed penalised likelihood):

$$PL = \left[\prod_{j=1}^{M} \sum_{k=1}^{N} \pi_{jk} \psi_k(y_j; T)\right] p(T) \tag{3}$$

The EM algorithm then writes:

Algo Reg0: NL-EM-ICP

E-step: $\tilde{A}_{jk} = \frac{\pi_{jk} \exp\left[-\|y_j - \tilde{T}(x_k)\|^2/(2\sigma^2)\right]}{\sum_i \pi_{ji} \exp\left[-\|y_j - \tilde{T}(x_i)\|^2/(2\sigma^2)\right]}$
M-step:
$\tilde{T} = \arg\min_T \sum_{j,k} \tilde{A}_{jk} \|y_j - T(x_k)\|^2 + 2\sigma^2 \alpha R(T)$

Actually, it can be shown that this algorithm is simply an iterative alternated minimisation (over A and T) of the following criterion:

$$\begin{aligned} \mathcal{E}0(X, Y, T, A) &= \sum_{j,k} A_{jk} (\|y_j - T(x_k)\|^2 - 2\sigma^2 \log \pi_{jk}) \\ &+ 2\sigma^2 \sum_{j,k} A_{jk} \log(A_{jk}) + 2\sigma^2 \alpha R(T), \end{aligned} \tag{4}$$

with $\forall j, k,\ 0 < A_{jk} < 1$ and $\forall j,\ \sum_k A_{jk} = 1$. It can be given an energetic interpretation of $\mathcal{E}0$, whose three terms represent respectively:

- a data-attachment term (modulated by prior information about potential matches),
- a barrier function allowing one to control the fuzziness of A (the higher σ^2, the greater the fuzziness); in practice, this term convexifies the criterion (barrier functions are widely used in the context of combinatorial optimisation),
- a regularisation term.

One can see that, in essence, the parameter α weighs the relative influence of the regularisation term $R(T)$ and of the data-attachment term $\sum_{j,k} A_{jk} \|y_j - T(x_k)\|^2$. Starting from either the statistical or the (equivalent) energetic view, a number of state-of-the-art algorithms (Chui et al., 2003; Guo and Rangarajan, 2009; Myronenko and Song, 2010; Horaud et al., 2011; Hermans et al., 2011; Tao and Sun, 2014; Nguyen and

Wu, 2016) have been proposed to tackle some of the limitations of this framework, which mostly include: i) the deleterious influence of outliers; ii) the asymmetrical matching process; iii) the local convergence of the algorithm. In Sections 3, 4 and 5, we detail these limitations, we outline the propositions that have been made in the literature to overcome these, and we propose and justify original solutions of our own for the same purpose. These successive improvements lead us to gradually modify the original algorithm (termed Reg0) to propose the algorithms Reg1, Reg2 and Reg3, the latter including all the proposed modifications.

3. Reducing the computational burden and robustifying the estimator

Reg0 suffers from outliers *i.e.* points of Y having no satisfactory correspondence in X. To alleviate this problem, several solutions have been proposed, such as adding an extra Gaussian with large variance (Chui et al., 2003), or a uniform component (Myronenko and Song, 2010; Tao and Sun, 2014) to the Gaussian mixture, or replacing the Gaussians in the mixture by heavy-tailed distributions, for instance Student ones (Nguyen and Wu, 2016), or introducing weights in the likelihood function (Hermans et al., 2011; Tao and Sun, 2014). Here we propose to truncate the quadratic function in the data-attachment term using a cut-off distance $\delta > 0$. If one temporarily drops the priors π_{jk} for the sake of clarity of exposition (which amounts to say that for all j, k, $\pi_{jk} = 1/N$; we reintroduce them in Section 5 where we show how to design them in a meaningful way), the corresponding term becomes a constant which can be removed from the criterion to minimise, which we redefine as:

$$\begin{aligned} \mathcal{E}1(X, Y, T, A) &= \sum_{j,k} A_{jk} \rho_\delta(\|y_j - T(x_k)\|^2) \\ &+ 2\sigma^2 \sum_{j,k} A_{jk} \log(A_{jk}) + 2\sigma^2 \alpha R(T), \end{aligned} \tag{5}$$

where $\rho_\delta : r \longmapsto r$ if $r < \delta$ and $r \longmapsto \delta$ else. In practice, it allows one to eliminate the points of X having no correct correspondence in Y from the estimation of the optimal T (*i.e.* to improve the robustness of the criterion). Note that the ρ_δ function could be advantageously replaced by a smoother differentiable function (*e.g.* a Leclerc function (Geman and Reynolds, 1992)). However, using this truncated function allows one to keep a useful probabilistic interpretation for the different parameters (and particularly for the match matrix A). More importantly, compared to the other abovementioned strategies, this amounts to ignore matches between highly distant point which, thanks to a kd-tree subdivision of the space (Bentley, 1975), drastically reduces the computational burden of the E-step. We modify the E-step and the M-step so that the algorithm successively decreases the new (truncated) criterion with respect to A and T. In practice, given that the data-attachment term does not depend on points x_k and y_j for which $\|y_j - \tilde{T}(x_k)\|^2 > \delta$, this can be simply achieved by setting A_{jk} to 0 for all pairs of such points. This gives the following algorithm:

Algo Reg1: Robust NL-EM-ICP

E-step:
Initialise A as the null matrix
for each $x_k \in X$;
  $\mathcal{S} = \{y_j \in Y \,;\, \|y_j - \tilde{T}(x_k)\|^2 < \delta\}$ (using a kd-tree)
  for each $y_j \in S$
    $A_{jk} = \exp(-\|y_j - \tilde{T}(x_k)\|^2/(2\sigma^2))$
for each $y_j \in Y$;
  if $\sum_i A_{ji} \neq 0$
    for each $x_k \in X$, $\tilde{A}_{jk} = A_{jk}/\sum_i A_{ji}$

M-step: solve the approximation problem (see Sec. 6):
$\tilde{T} = \arg\min_T \sum_{j,k} \tilde{A}_{jk} \|y_j - T(x_k)\|^2 + 2\sigma^2 \alpha R(T)$

Together with the implementation choices we later make to solve the M-step (see Section 6), this algorithm allows us to deal with large point sets.

Note that here and in the remainder of the paper we continue to use the terms "E-step" and "M-step" for the sake of simplicity, while we are fully aware that, from this moment on, our minimisation problem can no longer be seen as a MAP problem, and that our two-step algorithm is no longer an EM algorithm.

4. Symmetrising the matching process

4.1. A symmetric formulation for the matching process

A particularly undesirable property of $\mathcal{E}1$ is the asymmetric constraint $\forall j, \sum_k A_{jk} = 1$ (*i.e.* A is row stochastic). In practice, on the basis of the MAP principle, for a given match matrix A, the correspondence in X of a point $y_j \in Y$ is given by x_c where $c = \arg\max_k A_{jk}$. This leads to many-to-one correspondences between X and Y: nothing prevents a same point in X to be matched with several different points in Y, and nothing enforces each point of X to have a correspondence in Y. This makes the algorithm unable to achieve a proper matching in some specific configurations and makes the choice of source and target sets critical. It is particularly enlightening to consider the case when surfaces are far from each other (example on Fig. 1, where we show that the matrix A highly depends on which is the source and which is the target point set).

To alleviate this problem, Rangarajan *et al.* (Rangarajan et al., 1997; Chui et al., 2003) proposed to impose the matrix A to be doubly stochastic (*i.e.* $\forall k, \sum_j A_{jk} = 1$ and $\forall j, \sum_k A_{jk} = 1$) instead of simply row stochastic. With this new constraint on A, the E-step has no longer a simple solution. As a consequence, they approximate the optimal solution for A by performing a Sinkhorn normalisation (Sinkhorn, 1964) on the original (*i.e.* row-normalised) match matrix A at the end of the E-step. However, this empirical method is not applicable to matrices having null entries and thus cannot be applied when using a truncated quadratic function (as in Section 3). In practice, this limits its application to small point sets.

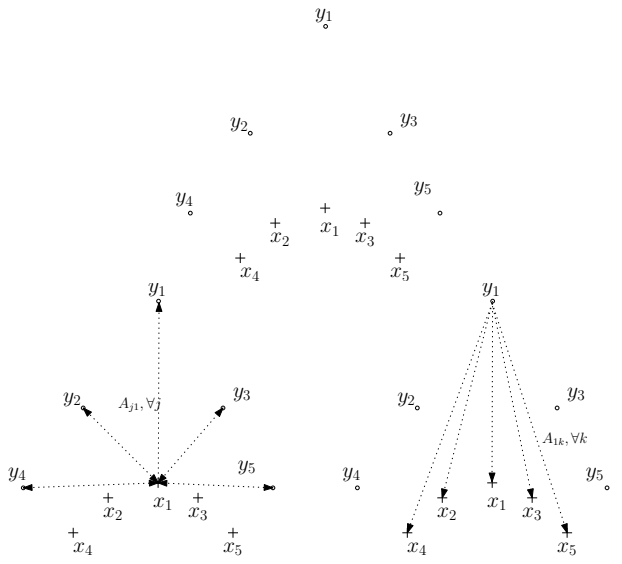


Fig. 1: **Effect of the asymmetric normalisation constraint on A.** The "correct" matches are $(x_l, y_l)_{l=1...5}$. The example here is a typical initial E-step, in which T is given as the identity (and $\sigma = 1$). From left to right and top to bottom: (i) two point sets X and Y, (ii) distances involved in the computation of A_{11} when registering Y on X (X is the target point set). (iii) distances involved in the computation of A_{11} when registering X on Y (Y is the target point set, as is the case in the present paper). In case (ii) $A_{11} = 0.01$ and points x_1 and y_1 have only little chance to be ever matched at the end of the overall process whereas in case (iii) $A_{11} = 0.45$.

As an alternative, we propose to modify the criterion $\mathcal{E}1$ (Eq. 5) by introducing a new match matrix B in this criterion, B being column stochastic, and a real number $0 \leq \gamma \leq 1$. We propose to define $\mathcal{E}2$ as:

$$\begin{aligned} &\mathcal{E}2(X, Y, T, A, B, \gamma) = \\ &\sum_{j,k} (\gamma A_{jk} + (1-\gamma) B_{jk}) \rho_\delta(\|y_j - T(x_k)\|^2) \\ + \quad &2\sigma^2 \sum_{j,k} (\gamma A_{jk} \log(A_{jk}) + (1-\gamma) B_{jk} \log(B_{jk})) \\ + \quad &2\sigma^2 \alpha R(T), \end{aligned} \tag{6}$$

with $\forall j$, $\sum_k A_{jk} = 1$ and $\forall k$, $\sum_j B_{jk} = 1$. When $\gamma = 1$, only the match matrix A is used, and one simply gets $\mathcal{E}2(X, Y, T, A, B, 1) = \mathcal{E}1(X, Y, T, A)$. The algorithm is thus simply the robust MAP algorithm just described. When $\gamma = 0$, only the match matrix B is used, and the opposite of the situation just described happens: nothing prevents a same point in Y to be matched with several different points in X, and nothing enforces each point in Y to have a correspondent in X. When $\gamma = 1/2$, the two match matrices are dealt with in the exact same way, and thus the choice of the target and source point sets becomes essentially unimportant (although T is still defined from X to Y), and the correspondences are obtained in a much "fairer" way. The criterion $\mathcal{E}2$ can then be simply rewritten as:

$$\begin{aligned} &\mathcal{E}2(X, Y, T, A, B, 1/2) = \\ &\sum_{j,k} (A_{jk} + B_{jk}) \rho_\delta(\|y_j - T(x_k)\|^2)/2 \\ + \quad &\sigma^2 \sum_{j,k} (A_{jk} \log(A_{jk}) + B_{jk} \log(B_{jk})) \\ + \quad &2\sigma^2 \alpha R(T) \end{aligned} \tag{7}$$

This alternative to the Sinkhorn normalisation makes it possible to use a truncated quadratic function in the criterion for increased robustness. As is the case with the Sinkhorn normalisation, though, as soon as the match matrix is modified, the criterion can no longer be given a probabilistic interpretation in a straightforward way.

4.2. Minimisation and implementation

The minimisation of $\mathcal{E}2(X, Y, T, A, B, 1/2)$ with respect to A and B when T is fixed (E-step) is straightforward and of low complexity (using the strategy proposed in Section 3 to build estimates of the match matrices $\tilde{A}$ and $\tilde{B}$). The overall algorithm to minimise $\mathcal{E}2$ can be expressed as:

Algo Reg2: Symmetric robust NL-EM-ICP

E-step:
compute $\tilde{A}$; $\tilde{B}$ (using kd-trees as in **Algo Reg1)**
M-step: solve the approximation problem (see Sec. 6):
$\arg\min_T \sum_{j,k} (\tilde{A}_{jk} + \tilde{B}_{jk}) \|y_j - T(x_k)\|^2/2 + 2\sigma^2 \alpha R(T)$

5. Adding priors

The computation of the match matrices A and B is essentially based on the spatial distance between the points $T(x_k)$ and y_j. This is unsatisfactory for two reasons. First, this distance is highly conditioned by the previous estimation of T, which in turn depends on the previous estimation of A_{jk} and B_{jk} and so on. This chicken-and-egg problem limits the capture range of the algorithm, which is likely to converge to a bad solution if no good initial transformation T is given. Second, in many applications it is difficult to design a physical model R capturing the expected deformation between two structures. Thus, the global maximiser of $\mathcal{E}2$ is likely not to be realistic.

Some efforts have been made to include richer information in the matching process in addition to the classical spatial distance between points, *e.g.* the similarity of the normals at points x_k and y_j. Such approaches assume that one can compute how the normals evolve when the surface is deformed. This generally results in adding non-linear terms to the M-step and leads to an intractable minimisation strategy (see *e.g.* the work by Feldmar and Ayache (Feldmar and Ayache, 1996)).

On the other hand, a more generic and simpler method consists in specifying an *a priori* probability π_{jk} between the points x_k and y_j to be matched, that we suppose to be independent of 1) the spatial proximity between the points of the two point sets and 2) the unknown transformation T. By specifying relevant priors π_{jk}s, we introduce additional information on matches allowing one to compute reliable posteriors even for a bad initial estimate of the deformation.

5.1. Designing π

An intuitive idea is to design $\pi = (\pi_{jk})$ such that $\pi_{jk} \propto \exp(-\beta c(y_j, x_k))$ where $c : Y \times X \to \boldsymbol{R}^+$ conveys the cost of matching points y_j and x_k, independently of T. The parameter $\beta > 0$ weighs the influence of π_{jk} over $\|y_j - T(x_k)\|$ during the E-step. In practice, to keep the algorithm robust to outliers and efficient, that leads us to propose the following modified criterion:

$$\begin{aligned} & \mathcal{E}3(X, Y, T, A, B, 1/2) = \\ & \sum_{j,k} (A_{jk} + B_{jk}) \rho_\delta(\|y_j - T(x_k)\|^2 + 2\sigma^2 \beta c(y_j, x_k))/2 \\ + \quad & \sigma^2 \sum_{j,k} (A_{jk} \log(A_{jk}) + B_{jk} \log(B_{jk})) \\ + \quad & 2\sigma^2 \alpha R(T) \end{aligned} \tag{8}$$

with $\forall j, \sum_k A_{jk} = 1$ and $\forall k, \sum_j B_{jk} = 1$.

Depending on the information to encode (discrete labels or continuous descriptors), we propose to build c as follows.

5.1.1. Designing π using labels

The cost function c can be computed via the comparison between labels of points (*e.g.* cortical sulci/gyri for brain registration). We define: $c(y_j, x_k) = 0$ if points y_j and x_k have compatible labels and $c(y_j, x_k) = penalty > 0$ else. In particular, this allows one to use pairs of landmarks in the registration process. One could also use crest lines extracted from both point sets (Gumhold et al., 2001) as they constitute salient features. Each point is given a label depending on whether it belongs to a crest line or not.

5.1.2. Designing π using descriptors

The cost function c can be computed via the comparison between continuous values (or vectors) $d(.)$ describing the point set around the considered points. To account for potential inaccuracies on $d(.)$, we define the measure as: $c_d(y_j, x_k) = 0$ if $\|d(y_j) - d(x_k)\| < \tau$ and $c_d(y_j, x_k) = penalty > 0$ else.

Then we choose $d(.)$ among the numerous local/global shape descriptions designed in the literature. In our context, one expects the descriptors to be:

- invariant to a certain class of transformations;
- robust to noise;
- robust to small distortions.

Among them, we choose to use (example on Fig. 2):

- The shape index $d(x) = sh(x)$ (Koenderink and van Doorn, 1992) that describes the local shape irrespective of the scale and that is invariant to similarities. To achieve robustness to noise and small distortions, we compute it by fitting a quadratic surface in the neighbourhood of the considered point. The fitting is performed by (i) computing a unit normal at point x, (ii) defining a local coordinate system (where the z-axis lies along the unit normal) and (iii) fitting a quadratic surface of the type $au^2 + buv + cv^2$ in the least-squares sense using the neighbours of x. The shape index can then be expressed as a function of a, b and c:

$$sh(x) = -(2/\pi) \arctan\left(\frac{2(a+c) + 2b^2}{2|a-c|}\right)$$

- The curvedness $d(x) = cu(x)$ (Koenderink and van Doorn, 1992) that specifies the amount of curvature and that is invariant to rigid-body transformations. We compute it using the same techniques as we used for the shape index:

$$cu(x) = \sqrt{(a+b+c)^2 + (a-b)^2}$$

- The (normalised) total geodesic distance $d(x) = tgd(x)$ (Aouada et al., 2007) that is invariant to non-elastic deformations. This distance is defined as:

$$tgd(x) = \sum_j \frac{d_g(x, x_j)}{\max_j \sum_k d_g(x_j, x_k)},$$

where $d_g(x_j, x_k)$ is the geodesic distance between x_j and x_k. It is computed efficiently using a graph representation of the (tessellated) point sets and the Dijkstra's algorithm.

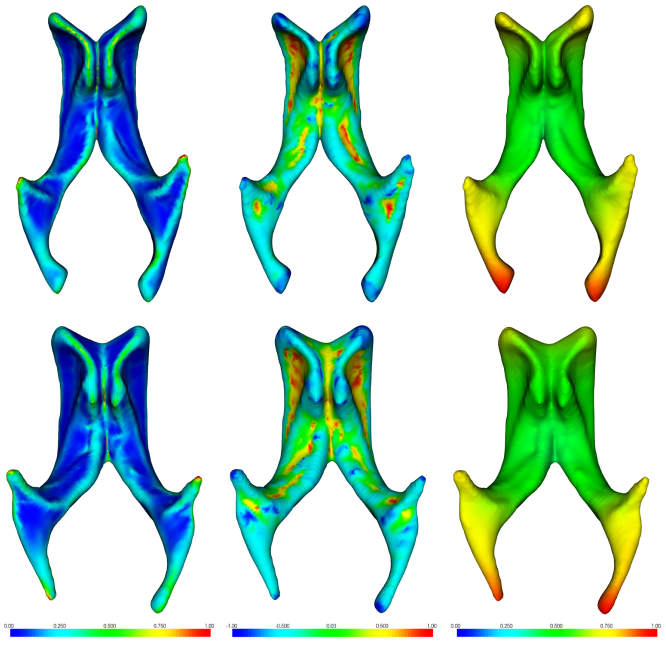


Fig. 2: **Mapping of descriptor values on two different pairs of brain lateral ventricles.** For each pair (the first pair on the first line, the second pair below), we display, from left to right: the curvedness, the shape index and the total geodesic distance. Homologous anatomical structures yield qualitatively the same descriptor values in each pair.

5.2. Minimisation and implementation

In practice, adding the cost function c only changes the way the matrices A and B are computed (E-step). We propose the following efficient algorithm for this purpose:

Algo Reg3: Symmetric robust NL-EM-ICP with priors

E-step:
initialise A and B to the null matrix
for each $x_k \in X$;
  $\mathcal{S} = \{y_j \in Y \,;\, \|y_j - \tilde{T}(x_k)\|^2 < \delta\}$ (using a kd-tree)
  for each $y_j \in \mathcal{S}$;
    if $\|y_j - \tilde{T}(x_k)\|^2 + 2\sigma^2\beta c(y_j, x_k) < \delta$
      $A_{jk} = \exp(-(\|y_j - \tilde{T}(x_k)\|^2/2\sigma^2 + \beta c(y_j, x_k)))$
$B = A$
normalise A in rows and B in columns
M-step: solve the approximation problem (see Sec. 6):
$\arg\min_T \sum_{j,k}(\tilde{A}_{jk} + \tilde{B}_{jk})\|y_j - T(x_k)\|^2/2 + 2\sigma^2\alpha R(T)$

6. Solving the M-step

Solving the M-step is highly conditioned by the choice of the transformation T and the regulariser R. Several proposals have been made in the literature. A first solution consists in modelling the transformation as the integration of a time-dependent velocity field regularised through R, with R being defined as a scalar TPS (Camion and Younes, 2001), as a scalar MCT (Glaunès et al., 2004) or as a scalar Laplacian (Joshi and Miller, 2000) regulariser. This results in a diffeomorphic transformation, adequate to deal with large deformations, but computationally demanding and thus restricted to small point sets. An alternative consists in modelling the transformation as a simple regularised dispacement field. Although such an approach does not ensure invertibility and preservation of topology, it generally leads to more tractable methods. A popular model consists in defining R as the scalar thin plate spline (TPS) (Bookstein, 1989), which has the main advantages to exhibit a closed-form solution for the approximation problem, and to be justified by a physical interpretation. However, this choice implies high computational and memory load that also limit its application to point sets of small size. Alternatively, in this section, while following this small displacement model, we focus on building a tractable (in terms of minimisation strategy) and powerful (in terms of reliability of the model) regulariser R allowing one to deal with large point sets (typically more than 1000 points).

For this purpose, we consider in the rest of this paper that T is represented as the initial position plus a displacement field: $T(x_k) = x_k + t(x_k)$ and that R is a regulariser on t. If one further notes that the M-step corresponding to $\mathcal{E}1$, $\mathcal{E}2$ and $\mathcal{E}3$ (and even $\mathcal{E}0$), which entails a double summation over j and k can actually be simply rewritten using a single summation over k, allowing for a more efficient implementation (we show that in Appendix C for $\mathcal{E}1$), one can restate it in the following general form:

$$\tilde{t} = \arg\min_t \sum_k \tilde{C}_k\|\tilde{y}_k - x_k - t(x_k)\|^2 + \kappa R(t), \qquad (9)$$

where (calling $\tilde{A}_{.k} = \sum_j \tilde{A}_{jk}$ and $\tilde{B}_{.k} = \sum_j \tilde{B}_{jk}$):

- $\tilde{C}_k = \tilde{A}_{.k}$, $\tilde{y}_k = \sum_j \tilde{A}_{jk} y_j / \tilde{A}_{.k}$ and $\kappa = 2\sigma^2\alpha$ for Reg1,
- $\tilde{C}_k = (\tilde{A}_{.k} + \tilde{B}_{.k})/2$, $\tilde{y}_k = \sum_j (\tilde{A}_{jk} + \tilde{B}_{jk}) y_j / (\tilde{A}_{.k} + \tilde{B}_{.k})$ and $\kappa = 2\sigma^2\alpha$ for Reg2 and Reg3.

Along the lines followed by Myronenko and Song (Myronenko and Song, 2010), we devise a new solution for the M-step based on the Reproducing Kernel Hilbert Space (RKHS) theory and the Fourier analysis (Schölkopf et al., 2001; Šidlofová, 2004; Wendland, 2004). More specifically, we design $R(t)$ as a function of the frequencies of $x \mapsto t(x)$ (which allows for an efficient tuning) and we propose a closed-form solution for the M-step that can be implemented very efficiently using sparse linear algebra. Moreover, the properties of the regulariser R depend on the choice of a kernel that can be easily modified to fit the applications.

6.1. Approximation problems in RKHS

For a reason that will soon become clear, we choose to look for the optimal displacement field $x \mapsto \tilde{t}(x)$ inside the reproducing kernel Hilbert space (RKHS) uniquely defined by a matrix-valued (3×3) kernel k, which we will later define in Section 6.2. This kernel must be defined such that:

- $\forall z_i, z_j \in \boldsymbol{R}^3$, $k(z_j, z_i) = k(z_i, z_j)^T$, and that
- $\forall M \in \mathbb{N}, \forall z_1, \ldots, z_M \in \boldsymbol{R}^3, \forall \alpha_1, \ldots, \alpha_M \in \boldsymbol{R}^3$, $\sum_{i,j=1}^{M} \alpha_i^T k(z_i, z_j)\alpha_j \geq 0$.

The resulting RKHS, that we call $\mathcal{H}$, can be shown (Aronszajn, 1950; Younes, 2001) to be the set of functions $t_i : x \mapsto \sum_{i=0}^{\infty} k(z_i, x)w_i$ with $w_i \in \boldsymbol{R}^3$, $z_i \in \boldsymbol{R}^3$ and $\|t_i\|_{\mathcal{H}} < \infty$, the norm $\|.\|_{\mathcal{H}}$ being defined using the following inner product $< ., . >_{\mathcal{H}}$:

$$< t_i, t_j >_{\mathcal{H}} = \sum_{i,j=0}^{\infty} w_i^T k(z_i, z_j) w_j$$

We then define our regulariser $R(t)$ as $\|t\|_{\mathcal{H}}$; the minimisation problem (9) can thus be rewritten as:

$$\tilde{t} = \arg\min_{t \in \mathcal{H}} \sum_{k=1,..,N} \tilde{C}_k\|\tilde{y}_k - x_k - t(x_k)\|^2 + \kappa\|t\|_{\mathcal{H}} \qquad (10)$$

The key advantage of enforcing $\tilde{t}$ to be inside a RKHS is that, given the specific form of the criterion to minimise, one can show that the optimal solution $\tilde{t}$ can be expressed as $\tilde{t} : x \mapsto \sum_{i=1}^{N} k(x_i, x)w_i$, *i.e.* that $\tilde{t}$ only and entirely depends on the kernel k and the finite point set X (Schölkopf et al., 2001). As a consequence, one can rewrite the problem (10) as:

$$(\tilde{w}) = \arg\min_{(w)} \sum_{k=1,..,N} \tilde{C}_k\|\tilde{y}_k - x_k - \sum_{i=1,..,N} k(x_i, x_k)w_i\|^2$$

$$+\kappa \sum_{i,j=1,..,N} w_i^T k(x_i, x_j) w_j,$$

where $N = \mathrm{card}(X)$. Pragmatically, this way we replace the estimation of a function $\tilde{t}$ belonging to a space of functions defined all over $\boldsymbol{R}^3$ by the estimation of the $3 \times N$ scalar values.

Vanishing the derivatives gives a linear system whose solution can be expressed in a closed-form as:

$$W = (D(C)K + \kappa I)^{-1} D(C)[Y - X],$$

where $X = [x_1, ..., x_N]^T$, $Y = [\tilde{y}_1, ..., \tilde{y}_N]^T$, $W = [w_1, ..., w_N]^T$, $K = (k(x_i, x_j))_{i,j}$ is the $N \times N$ matrix associated to kernel k and expressing the geometrical configuration of the point set X and $D(C)$ is the $N \times N$ diagonal matrix formed by the $\tilde{C}_k$ values. Now, the challenge is to choose a kernel corresponding to a relevant regulariser.

6.2. Choosing a kernel

The displacement field t is a function $\mathbb{R}^3 \to \mathbb{R}^3$. For the sake of simplicity, let us consider for the moment the set of functions $f : \mathbb{R} \to \mathbb{R}$; an interesting regulariser for such functions can be proposed, provided they are regular enough, based on the Fourier transform:

$$R(f) = \int_{-\infty}^{\infty} |\hat{f}(x)|^2 / \hat{\phi}(bx) dx, \tag{11}$$

where $\hat{}$ is the Fourier transform, $\hat{\phi}$ is a positive function, regular enough, and tending to zero as $|x| \to \infty$, and b is a real positive rescaling factor (Girosi et al., 1995). Thus, in this functional, $1/\hat{\phi}$ is akin to a high-pass filter: high frequencies of the deformation will be drastically penalised whereas low frequencies will only be penalised a little. The most important element that characterises its influence on the regularisation is the way it decreases, which indicates the amount of penalisation with respect to frequencies. Particularly, the frequencies for which $\hat{\phi}(bx)$ is null are forbidden. Interestingly, the functions f for which $R(f) < \infty$ is a RKHS whose kernel is $(x_i, x_j) \mapsto 1/b \cdot \phi(|x_i - x_j|/b)$. Coming back to our displacement field t, that leads us to define the matrix-valued kernel k as:

$$\forall z_i, z_j \in \boldsymbol{R}^3, k(z_i, z_j) = 1/b \cdot \phi(\|z_i - z_j\|/b) \cdot I_3,$$

where I_3 is the identity matrix in $\mathbb{R}^3$. The two parameters κ and b allow one to handle the regularisation properties: κ is a quantitative parameter (it indicates the amount of smoothness) whereas b is more qualitative (in a way, it defines what the term "smoothness" means).

Designing R as a function penalising t in terms of its spatial frequencies can be of great interest. Roughly speaking, high frequencies of t concern details and local changes of the deformation field whereas its low frequencies concern the global aspects of the deformation. For a given kernel k, the larger the b value, the more drastic the penalisation of the high frequencies. Thus, tuning the parameter b allows one to devise a multiscale approach by first trying to capture a global deformation and then, if needed, local deviations from this global deformation. This allows our algorithm to be tuned for different applications needing either a fine registration (*e.g.* automatic labelling of brain substructures) or a more global registration (*e.g.* statistical shape analysis) of two point sets. Figure 3 illustrates the influence of different kernels and different scale parameters b on the regularisers. Figure 4 shows the influence of b and κ when approximating a noisy 2D field when choosing ϕ as the Wu kernel (Buhmann, 1998). Figure 5 illustrates that modifying b allows one to characterise different scales of deformations linking two point sets.

6.3. Efficient choices

Although we proposed a closed-form solution for the approximation problem, it consists in solving $N \times N$ linear systems of equations. This can be problematic in terms of memory usage and of computational time when N increases. This is why we propose to choose a compactly supported kernel (*i.e.* without loss of generality, $\forall x, y$ such that $\|y - x\| > 1$, $k(x, y) = 0$), then i) $R = D(C)K + \kappa I$ is a sparse matrix that can be computed efficiently using a kd-tree (Bentley, 1975) and ii) computing W consists in solving a sparse linear system, for which efficient solutions exist. Some compactly supported kernels corresponding to low-pass filters have been proposed in the literature (such as Wendland, Wu or Buhmann kernels (Buhmann, 1998)). Moreover, techniques to generate a wide variety of them have been proposed. We experimentally found the compact support kernel of Wu ($\phi_{2,3}$) as the one providing the best results and we use it in the following. Note that the top-right plot in Figure 3 shows that the Wu kernel penalises high frequencies "faster" than the exponential or Wendland kernels (for a given b). More experiments will be needed to evaluate the respective characteristics of the different possible kernels.

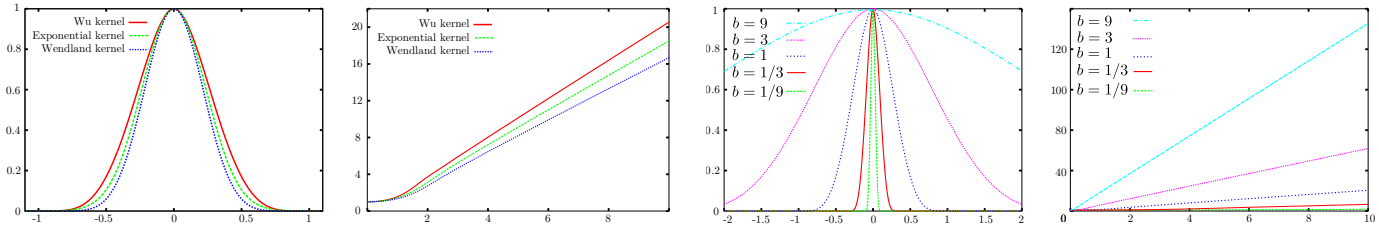
6.4. M-step in a nutshell

M-step:
initialise R as the null matrix
for each $x_k \in X$;
  $\mathcal{S}_k = \{x_i \in X$ such that $\|x_k - x_i\| < b\}$
(using a kd-tree)
  for each x_i in $\mathcal{S}_k$;
    $R(k, i) = 1/b \cdot \phi(\|x_k - x_i\|/b) \cdot \tilde{C}_k$
  $R(k, k) = R(k, k) + \kappa$
preconditioning of K (using sparse algebra)
solve $RW^1 = D(C)[Y^1 - X^1]$ (using sparse algebra)
solve $RW^2 = D(C)[Y^2 - X^2]$ (using sparse algebra)
solve $RW^3 = D(C)[Y^3 - X^3]$ (using sparse algebra)

where X^1, X^2 and X^3 are the vectors extracted from the first, second and third columns of matrix X respectively (the same for Y and W). Note that most of the quantities involved in the computation of R can be computed beforehand.

7. Related algorithms

Several recent algorithms are theoretically related to Reg3, most notably the TPS-RPM (Chui et al., 2003) and CPD algorithms (Myronenko and Song, 2010) which are probably the most popular algorithms to register point sets. To better characterise our approach compared to these state-of-the-art methods, we sum-up the characteristics of Reg3 compared to CPD and TPS-RPM in Table 1 and describe their advantages and limitations in the core of the section. The main differences consist in (i) the way the matching step (E-step) is symmetrised, (ii) the choice of the regulariser R, (iii) the choice of the robustness



(i) Some kernel functions ϕ (ii) $\hat{\phi}(\|.\|/b)^{-1}$ for different kernel (iii) ϕ as the Wu kernel with different values for b (iv) $\hat{\phi}(\|.\|/b)^{-1}$ (ϕ being the Wu kernel) with different values for b

Fig. 3: **Different positive definite kernels ϕ and their associated functions $\hat{\phi}(\|.\|/b)^{-1}$.** From left to right and top to bottom: i) 2D plot of three kernels (b=1), ii) 2D plot of the values of $\hat{\phi}(\|.\|)^{-1}$ for the Wu, the exponential and the Wendland kernels, iii) 2D plot of $\phi(\|.\| \cdot b)$ for the Wu kernel with different b values, iv) 2D plot of $\hat{\phi}(\|\omega\|/b)^{-1}$ for the Wu kernel with different b values.

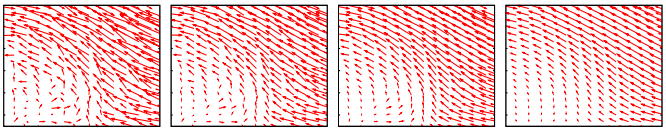


Fig. 4: **Effect of parameters κ and b on the approximation of a 2D noisy field.** From left to right: i) original field, ii) regularised field with ϕ as a Wu kernel, $\kappa = 5$ and $b = 10$, iii) regularised field with ϕ as a Wu kernel, $\kappa = 5$ and $b = 40$ and iv) regularised field with ϕ as a Wu kernel, $\kappa = 100$ and $b = 40$.

function and (iv) the resulting implementation choices. Quantitative results are given in Section 8.

7.1. The TPS-RPM algorithm

In the TPS-RPM algorithm (Chui et al., 2003), the match matrix is imposed to be doubly stochastic. The resulting E-step is solved heuristically by first computing a row stochastic matrix and then applying a Sinkhorn normalisation on this matrix. Robustness is achieved by introducing two "virtual" centroids located at the barycentre of X and of Y. The variances associated to the location of these centroids are fixed to high values. This way, the points of X (resp. Y) having no satisfying correspondence in Y (resp. X) will have a high probability of correspondence with the virtual centroid of Y (resp. X) that is obviously chosen to be absent in the data-attachment term and thus does not influence the estimation of the transformation T (M-step). Moreover a penalisation term is added to the overall criterion to avoid matching too many points with the virtual centroid. As mentioned by the authors themselves, this solution is debatable as it supposes that outliers are most probably located close to the barycentre of the point set. The variance parameter σ^2 (called temperature in their formalism) is initialised to a fixed large value and decreased by dividing it by a constant factor greater than 1 after each iteration.

The regulariser R is chosen as the TPS regulariser, which leads to an approximation problem (M-step) involving three dense $(N \times M)^2$ matrices. Due to its intractability, this problem is heuristically converted into three sub-problems, involving a dense square N^2 matrix, whose large size limits the resulting algorithm to small point sets (typically a few hundreds).

7.2. The CPD algorithm

In the CPD algorithm (Myronenko and Song, 2010), the match matrix is dealt with asymmetrically (*i.e.* is simply row stochastic). Robustness is achieved by modelling the point set $T(X)$ as the weighted sum of a Gaussian mixture and of a uniform distribution (instead of a simple Gaussian mixture). The relative weights of these two elements allow for tuning the number of expected outliers (that is generally unknown). The variance parameter σ is estimated at the beginning of each iteration of the algorithm. R is the MCT regulariser that is a particular case of the Fourier-based regulariser introduced in Section 6.1 where the kernel k is chosen as a Gaussian function: $k(x, y) = \exp(-\|x - y\|^2/b)$. The matrices involved in the computations are evaluated efficiently using the fast Gauss transform (Greengard and Strain, 1991) (which makes it impossible to add a cost function c in the data-attachment term as we did in Section 5 to add some prior information) and the solving of the M-step is accelerated by the precomputation of a low-rank approximation of the large matrix (called K in Section 6.1) representing the spatial structure of X (that is fixed throughout the iterations of the algorithm).

7.3. Our method (Reg3)

For both TPS-RPM and CPD algorithms, the implementation choices are performed using heuristics/approximations. In particular one can wonder how the approximations performed in the M-step change the nature of the regularisation.

As opposed to the latter, we do not use any heuristic either in the E-step or in the M-step and the criterion $\mathcal{E}3$ is properly minimised (no approximation is made). The correspondences are dealt with symmetrically and the rejection of outliers is performed using a robust cost function parameterised by a single cut-off distance parameter.

The E-step is implemented efficiently using a kd-tree (Bentley, 1975). The M-step is reformulated as three approximation problems involving sparse N^2 matrices; these problems are solved efficiently using sparse linear algebra and a kd-tree.

8. Validation & results

In this section, we investigate the added value of the different improvements we proposed in this article by considering average scores obtained by the algorithms Reg1 (Section 3), Reg2

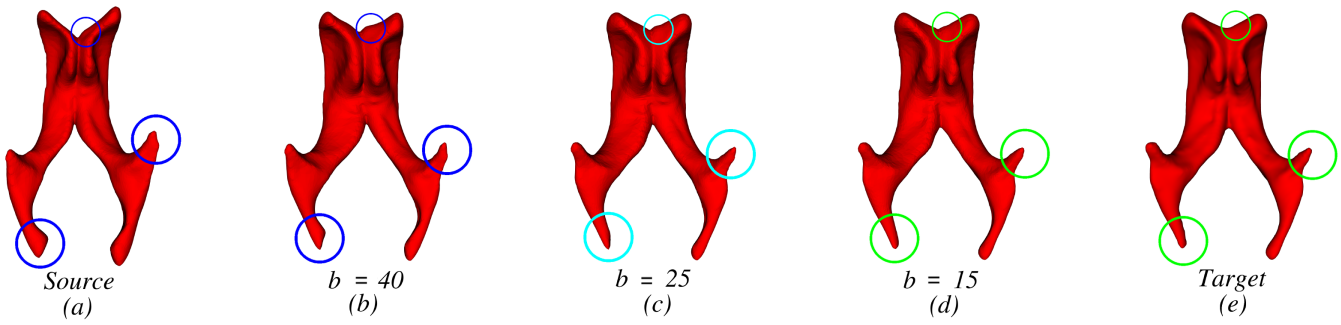


Fig. 5: **Influence of b on the registration.** We register the source ventricle (a) on the target ventricle (e) by decreasing b throughout the iterations of the overall EM-ICP algorithm (without modifying the other parameters). Intermediate registration results for a given b are represented between the source and the target (b), (c) and (d). For a large b, the source is only globally transformed towards the target and small patterns (dark blue circles) are left unchanged (as the penalisation does not allow such a deformation). When b decreases, the small discriminative patterns tend to fit the target (green circles). Light blue circles indicate intermediate configurations of the patterns.

	efficiency	exact/approx	robustness	matching process	minimisation	priors
TPS-RPM	**intractable**	**approx**	**distance to centroid**	symmetric	EM-like	easy to incorporate
CPD	very efficient	**approx**	% of outliers	**asymmetric**	EM-like	**no**
Reg3	efficient	exact	cut-off distance	symmetric	EM-like	yes

Table 1: Summarising the characteristics of TPS-RPM, CPD and Reg3. The characteristics written in boldface stress the potential weaknesses of the algorithms.

(Section 4) and Reg3 (Section 5) over the registration of many pairs of point sets for which we know the real deformation and the real point-to-point correspondences. The way we build this "ground truth" dataset and the way we measure the registration error is outlined in Section 8.1, while we summarise our results in Section 8.2.

Moreover, we perform the same experiments using the CPD algorithm (`https://sites.google.com/site/myronenko/research/cpd`). Note that, due too high memory and CPU usage, we did not manage to register any of our data with Reg0 and with the TPS-RPM algorithm (`https://www.cise.ufl.edu/~anand/students/chui/tps-rpm.html`).

For Reg3, the cost function c is built as: $c(y_j, x_k) = |sh(x_k) - sh(y_j)| + |cu(x_k) - cu(y_j)| + |tgd(x_k) - tgd(y_j)|$. Moreover, for all the following experiments we set:

- $\sigma^2 = \sigma^2_{init}/2$ every 10 iterations from $\sigma^2_{init} = 0.03 \cdot d$ (where d is the highest of the two maximal interpoint distances in X and Y, that is, the highest of the two diameters) with a lower limit equal to $\sigma^2_{init}/8$;
- $\delta = \delta_{init}/2$ every 10 iterations from $\delta_{init} = 0.2 \cdot d$ with a lower limit equal to $\delta_{init}/8$;
- $\kappa = 50$;
- $b = 0.2 \cdot d$;
- $\tau = 20\%$ and $penalty = 0.05 \cdot d$;
- the maximal number of iterations to 40.

For CPD, we use identical parameters for β and b and set the ratio of outliers to its ground truth values.

8.1. Ground truth data

We build the ground truth dataset using three anatomical structures which have been previously segmented from MR or CT images. These structures are: a brain caudate nucleus (mesh of 1,000 points, Figure 6), a pair of brain lateral ventricles (mesh of 7,000 points, Figure 7, left) and a bony labyrinth (mesh of 8,000 points, Figure 7, right). We thank Prof. José Braga (who is with AMIS, UMR 5288 CNRS-Université de Toulouse and with the Evolutionary Studies Institute, University of the Witwatersrand) for segmenting the bony labyrinth. We chose these structures, which are between about 2 and 7 centimetres in length, because they do not have the same geometrical complexity: caudate nuclei and brain ventricles are usually relatively smooth, while bony labyrinths are very convoluted due to the intricate shape of the cochlea and of the semi-circular canals. We modify and deform each of these three meshes 100 times, as follows:

- we apply a randomly generated local smooth deformation of the type: $x + KG_v(x - x_c)n_x$, where x_c is the centre of the deformations (randomly chosen on the point sets), n_x is the normal vector at point x, G_v is a 3D non-normalised Gaussian function of variance v^2 and K is the deformation strength;
- we apply a randomly generated non-linear transformation using either the TPS or the MCT regulariser;
- we perform a random removal of a given quantity of adjacent points (up to $0.1N$).

The generated displacement field is called $\hat{t}$. Figure 6 illustrates how one modified/deformed caudate nucleus is generated from the original mesh, and Figure 7 shows the superposition

of the original and one of the modified/deformed meshes for the ventricles and the labyrinth. Overall, we thus generate $3 \times 100 = 300$ pairs of surfaces for which we know both the real deformation and the real point-to-point correspondences. Then, we register the original and modified/deformed point sets; the recovered displacement field is called $\tilde{t}$. Then we compute:

- the overall residual distance between the known correspondences between the two point sets (end-point error):

$$end\text{-}pt(\hat{t}, \tilde{t}) = \frac{1}{N} \sum_{k=1,\dots,N} \|\hat{t}(x_k) - \tilde{t}(x_k)\|^2$$

- the average angular (later called "Barron") error between the displacement fields (Barron et al., 1994):

$$barron(\hat{t}, \tilde{t}) = \frac{1}{N} \sum_{k=1,\dots,N} \arccos\left(\frac{\hat{t}(x_k)^T}{\|\hat{t}(x_k)\|} \frac{\tilde{t}(x_k)}{\|\tilde{t}(x_k)\|} \right)$$

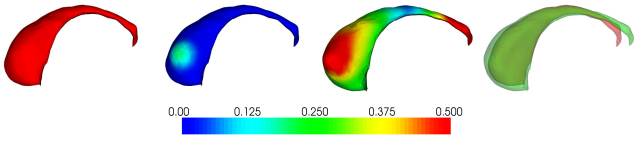


Fig. 6: **Illustration of the generation of the ground truth data.** From left to right: i) original data, ii) we generate a random local deformation (the resulting distances between the corresponding points is mapped), iii) we generate a random global deformation (TPS) (the resulting distances between the corresponding points is mapped) and iv) superimposition of the original (red) and deformed (green) data.

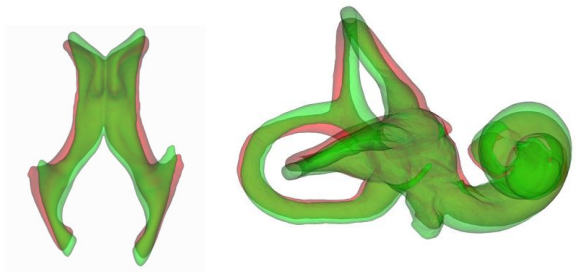


Fig. 7: **Examples of pairs of ground truth data:** Left: lateral ventricles. Right: osseous labyrinths.

8.2. Results

Table 2 indicates the mean and maximal end-point and Barron errors obtained for each of the four studied algorithms (Reg1, Reg2, Reg3 and CPD) and for the different datasets (that is, 100 pairs of meshes for each of the three anatomical structures). Table 3 indicates the mean run-time for each algorithm and for the different datasets. In the following, we summarise these results and give illustrating examples.

• Influence of the symmetrisation of the correspondences (Reg1 vs Reg2):

We point out two main conclusions when comparing results from Reg1 and Reg2. **First**, symmetrising the correspondences greatly decreases the registration error (almost threefold reduction on average for the end-point error). In practice, Reg1 is particularly unsuited when the correspondences between the source and the target are ambiguous (as shown previously in Figure 1) whereas it is not the case for Reg2. This is illustrated in Figures 8 and 9. **Second**, one observes that for Reg1 the location of the registration errors depends on which point set is used as the template. For Reg2, this effect is considerably reduced. This is illustrated in Figure 10.

Fig. 8: **Effect of the symmetrisation:** From left to right: i) initial alignment of two brain caudate nuclei, ii) alignment obtained using Reg1 and iii) alignment obtained using Reg2. The asymmetric formulation of Reg1 leads to registration errors close to the extrema of the head and of the tail.

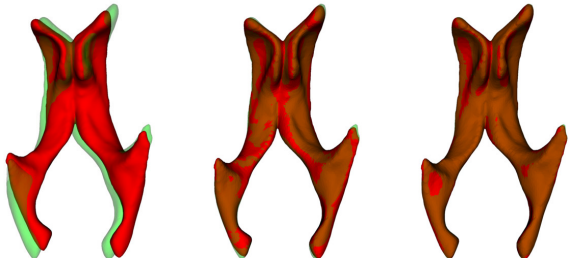


Fig. 9: **Effect of the symmetrisation:** From left to right : i) initial alignment of two brain ventricles, ii) alignment obtained using Reg1 and iii) alignment obtained using Reg2. The asymmetric formulation of Reg1 leads to registration errors close to the horns.

• Influence of the priors (Reg2 vs Reg3):

On average, Reg3 provides results that are close to those of Reg2. However, it decreases the value of the worse result importantly (max error in Table 2): by enforcing correspondences using priors (independently of the initial estimate of T) it allows a correct registration even for a bad initial alignment.

• Comparison with a state-of-the-art method (Reg1, Reg2 and Reg3 vs CPD):

We compare the results obtained with our methods with the one obtained with the CPD algorithm. We point out two main conclusions. **First**, CPD exhibits better results than Reg1. It indicates that CPD, that does not deal with the correspondences symmetrically, manages to tackle limitations potential linked to this drawback in another way. This could be due to the fact that CPD estimates a variance parameter instead of fixing it empirically. **Second**, we observe that the end-point error and the Barron error are higher for CPD than for Reg3 and Reg2. We notice that this effect is much larger for the Barron error, which can be due to the approximation performed during the M-step

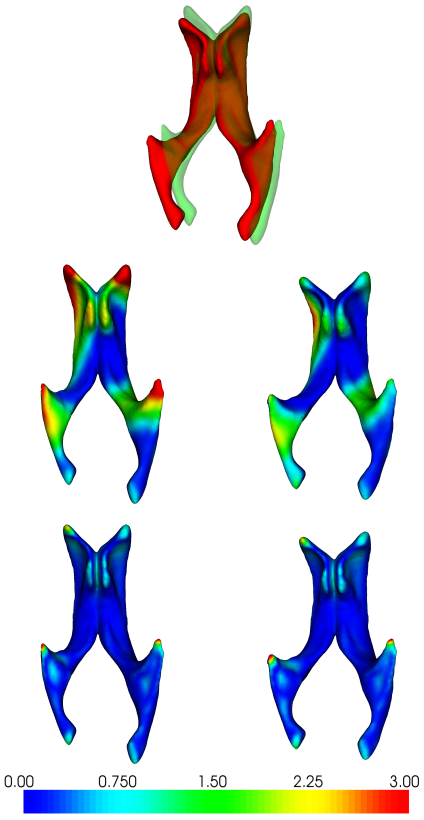


Fig. 10: **Effect of the symmetrisation:** From left to right and top to bottom: i) two misaligned pairs of lateral ventricles A and B viewed from below; ii) residual end-point error when registering A to B and iii) B to A with Reg1; iv) residual end-point errors when registering A to B and v) B to A with Reg2. The asymmetric formulation of Reg2 leads to registration errors close to the horns of the ventricles. The location of these errors depends on which surface is used as the template (second row). When symmetrising the matching, the order-dependent registration error is reduced and the overall registration quality is visually and quantitatively improved (third row).

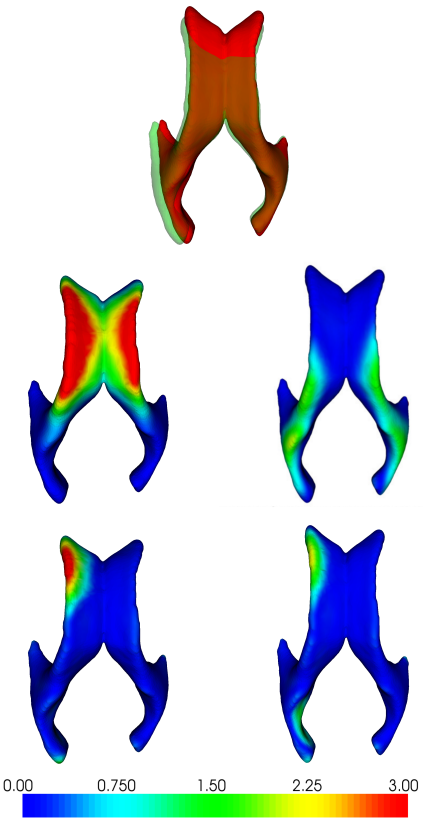


Fig. 11: **Reg3 vs CPD :** From left to right and top to bottom: i) two misaligned pairs of lateral ventricles A and B viewed from above; ii) residual end-point error when registering A to B and iii) B to A with CPD; iv) residual end-point error when registering A to B and v) B to A with Reg3. For CPD, the location of the errors depends more on which surface is used as the template. Moreover the overall error is higher for CPD than for Reg3.

of the CPD algorithm. Figures 11 and 12 give illustrative examples of the registration error obtained with Reg3 and CPD.

• Real data:

In Figures 13, 14 and 15, we display registration results obtained with Reg3 on pairs of anatomical structures belonging to two different patients.

9. Conclusions and perspectives

9.1. Conclusions

In this paper, we proposed to study the limits of the EM-ICP approach. For this purpose, we first formulated the algorithm in a self-contained manner in which we underlined its relationship with other classical algorithms. Then, starting from an energetic formulation, we proposed a new algorithm for the non-linear registration of large point sets. More specifically, we considered four drawbacks of the original algorithm and we proposed efficient and original solutions to handle them. In particular, we proposed to truncate the function in the data-attachment term, to reduce the asymmetry of the matching process, to add alignment-invariant information into the criterion and to devise a regulariser leading to efficient solutions to implement the M-step. We investigated the added value of these modifications on synthetic data. Particularly, we observed that i) symmetrising the matching process greatly decreases the registration error and makes the result of the registration less dependent on the choice of the source/target point sets and ii) adding priors reduces the registration errors when the point sets are initially far from each other. Then, we showed that the complete algorithm Reg3 outperforms the state-of-the-art CPD algorithm in terms of accuracy while being slower.

9.2. Perspectives

Our methodology relies on modelling point sets as mixtures of pdfs. An interesting work would consist in trying to fit mixture models directly on the point sets by estimating both means and (anisotropic) variances associated to the sum of Gaussians. Then the registration could be done on the fitted models (see *e.g.* (Roy et al., 2007)).

As previously shown, Fourier-based regularisers offer a simple and suitable way to compute the deformations linking the point sets at different scales. The design of an elaborated and well-grounded strategy to compute/analyse these different deformations could be of great interest.

	caudate nuclei		ventricles		labyrinths	
	mean/max end-pt	mean/max Barron	mean/max end-pt	mean/max Barron	mean/max end-pt	mean/max Barron
Reg1	0.49/2.43	13.65/32.90	0.78/3.29	28.21/37.76	0.23/0.48	5.89/8.11
Reg2	0.22/1.14	3.60/18.69	0.20/0.67	1.78/9.65	0.09/0.17	3.22/6.14
Reg3	**0.21/0.89**	**3.65/12.75**	**0.18/0.53**	**1.81/6.73**	**0.09/0.14**	**3.27/5.04**
CPD	0.35/1.39	6.88/28.22	0.40/0.91	5.00/12.92	0.12/0.20	4.05/7.31

Table 2: Statistics on registration errors (in mm) for different methods and different datasets

	caudate nuclei	ventricles	labyrinths
Reg1	26s	11min	9min
Reg2	40s	16min	11min
Reg3	47s	18min	14min
CPD	35s	13min	6min

Table 3: Mean run time for different methods and different datasets on a standard personal computer (Intel Core Duo T7700 at 2.4GHz with 2GB Ram).

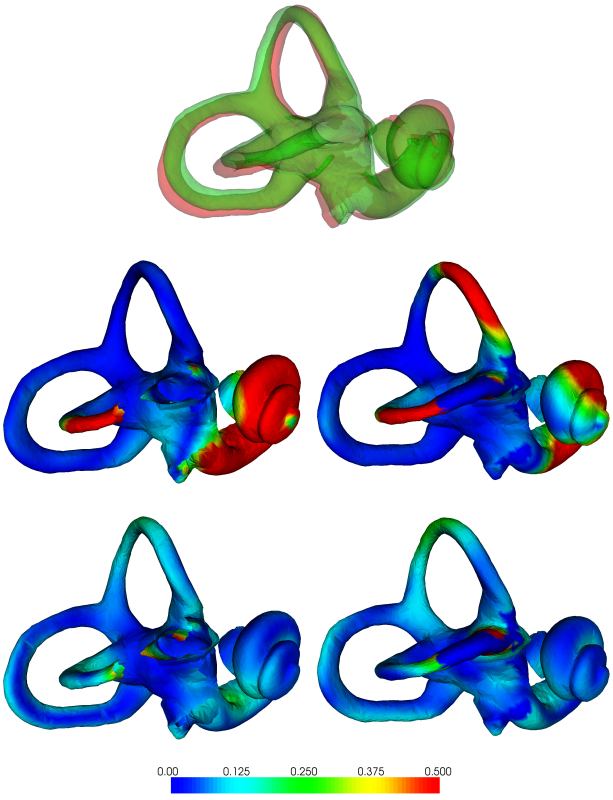


Fig. 12: **Reg3 vs CPD :** Same experiments as Figure 11 with bony labyrinths. Note the missing part of the lateral canal, which has been removed to simulate outliers. The conclusions are the same as for Figure 11.

Fig. 13: **Registration of two brain caudate nuclei segmented from MRI data using Reg3:** From left to right: i) initial alignment, ii) registration of the green caudate on the red one iii) registration of the red caudate on the green one.

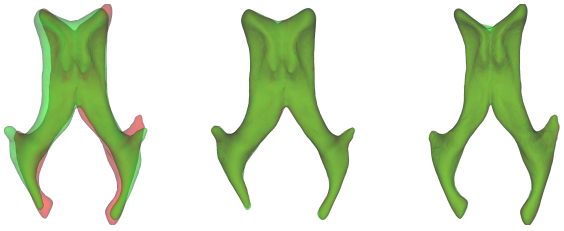


Fig. 14: **Registration of two brain ventricles segmented from MRI data using Reg3:** From left to right: i) initial alignment, ii) registration of the green ventricle on the red one iii) registration of the red ventricle on the green one.

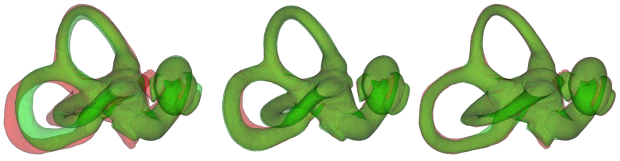


Fig. 15: **Registration of two bony labyrinths segmented from CT data using Reg3:** From left to right: i) initial alignment, ii) registration of the green bony labyrinth on the red one iii) registration of the red bony labyrinth on the green one.

Another track to follow would be to "symmetrise" the transformation instead of (or in addition to) the matching procedure as proposed here. In a previous work (Combès and Prima, 2010), we made preliminary steps towards this direction, by proposing to estimate simultaneously both the backward and forward transformations T^X and T^Y best superposing the two point sets X and Y, and enforcing these transformations to be consistent, in the sense that $T^X \circ T^Y$ was enforced to be as close to identity as possible. While appealing, we found this framework to suffer from several limitations, both theoretical (some terms had to be dropped in the E-step for the minimisation to stay tractable) and computational (two approximation problems of larger size had to be solved at each M-step, limiting the method to smaller point sets). In practice, we found this consistency constraint on the transformation to be largely redundant with that provided by the symmetrisation of the matching pro-

cess (*i.e.* through the use of two match matrices). Experiments showed us that the registration error was not reduced in practice when adding this consistency term. In some specific cases, it even seemed to limit the capture range of the algorithm, leading to poor registration results. Future work would be needed to understand why; it may be that constraints on the transformation make it "too rigid" and does not allow for the target and source images to get sufficiently close to each other for the point-to-point matching to provide correct correspondences at any time throughout the iterations.

Appendix A. A taxonomy of methods for non-linear registration of point sets

Throughout this section, which can be read independently of the rest of the article, we focus on registration methods having no strong priors on the structures to register (*e.g.* topological (Yeo et al., 2010; Shen et al., 2007; Zou et al., 2007; Robbins, 2004), tessellation constraints (Robbins, 2004)) or on the expected deformations (*e.g.* articulated motion (Mateus et al., 2008), isometric deformations (Huang et al., 2008)).

In Appendix A.1, we propose a comprehensive four-class taxonomy of point set representations underlying most non-linear registration methods of the literature and we describe the similarity measures ensuing from these representations. The choice of a transformation model is largely independent of the point set representation, and we briefly outline some of the most often used models in Appendix A.2. For further and more exhaustive references on point set registration, we direct the reader to two recent reviews (Tam et al., 2013; Maiseli et al., 2017).

Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_M\}$ be two point sets. Let T be a non-linear deformation. The goal of the following methods is to find T that best superposes $T(X)$ on Y.

Appendix A.1. The four classes

Appendix A.1.1. Point sets as modal matrices

The first authors to propose the use of the spectral theory to align two point sets were Scott and Longuet-Higgins (Scott and Longuet-Higgins, 1991) and Shapiro and Brady (Shapiro and Brady, 1992). The two methods are slightly different. For the sake of simplicity, we focus on the method proposed by the latter (and on some of its extensions).

The principle consists in building a *modal matrix* for each point set and then in using these matrices to establish correspondences between points. The modal matrix of a point set X is computed by (i) building a $N \times N$ symmetric proximity matrix $G_{jk} = \exp(-\|x_j - x_k\|/(2\sigma^2))$ (σ controls the degree of interaction between points), (ii) performing a SVD decomposition of $G = VDV^T$ where D contains the (positive) eigenvalues of G in a decreasing order. The matrix V is the modal matrix. Note that G (thus V) is invariant under rigid-body transformations. Each row of V is associated with one of the points of X whereas each column measures how the points of X are distributed among the different eigenmodes of G. Once one has computed the modal matrices V_X and V_Y of X and Y, the strategy consists in considering these measures as a shape descriptor almost invariant under the transformation T. As a result, the estimation of the point-to-point correspondences is performed by comparing rows of V_X and V_Y. This can be done either by applying a "best one rule" (Shapiro and Brady, 1992) or by building correspondence probabilities (Carcassoni and Hancock, 2000; Tang et al., 2007) (when both point sets do not have the same number of points, the larger modal matrix is truncated). This basic matching procedure can be embedded in an iterative scheme involving the estimation of the transformation T given known correspondences and the estimation of correspondences between $T(X)$ and Y given T (Carcassoni and Hancock, 2000; Tang et al., 2007).

Appendix A.1.2. Point sets as level set functions

Lüthi and colleagues (Lüthi et al., 2007; Dedner et al., 2007) proposed to represent each point set to register as the zeroth level of a signed distance 3D function I, *i.e.* for X:

$$I_X(x) = \begin{cases} \operatorname{dist}(x, X) & \text{if } x \in \mathit{outside}(X) \\ 0 & \text{if } x \in X \\ -\operatorname{dist}(x, X) & \text{if } x \in \mathit{inside}(X) \end{cases} \tag{A.1}$$

Note that to consider such a representation, one must assume the point sets X and Y to be structured as (closed) meshes.

The two 3D functions I_X and I_Y are then discretised on 3D grids, leading to 3D grey-level images, and T is computed as the deformation best superposing these grids. In other words, the original problem is tackled as a 3D grey-level image registration problem.

In essence, any iconic registration algorithm can then be used (Maintz and Viergever, 1998) to solve this problem. Albrecht and colleagues proposed to use the demons algorithm (Thirion, 1998) and modified it in order to incorporate the mean curvature images of I_X and I_Y (in addition to their intensity values) into the cost function to guide the registration.

Appendix A.1.3. Point sets as Schwartz distributions

The framework of diffeomorphic matching of distributions was developed by Glaunès and colleagues (Glaunès et al., 2004). The authors proposed to consider the point set X (and similarly for Y) as a weighted sum of Diracs (a Dirac being a discrete Schwartz distribution) localised at the points of X: $\nu_X = \sum_k \delta(x_k)$. They showed that the action of a diffeomorphic deformation T generated from the integration of a time-dependent smooth velocity field is simply $T\nu_X = \sum_k \delta(T(x_k)) = \nu_{T(X)}$. In order to compare two distributions (typically $\nu_{T(X)}$ and ν_Y), the authors noticed that all differences of distributions are contained in the dual $\mathcal{I}^*$ of an Hilbert space $\mathcal{I}$ containing continuous bounded functions on $\boldsymbol{R}^3$. As a result, the quantity $\|\nu_{T(X)} - \nu_Y\|_{\mathcal{I}^*}$ is used as a measure of distance between $T(X)$ and Y. By designing $\mathcal{I}$ as a reproducing kernel Hilbert space (Aronszajn, 1950), this distance $\|. - .\|_{\mathcal{I}^*}$ can be easily evaluated and differentiated with respect to T. The transformation T can then be estimated using *e.g.* a gradient-descent minimisation. This work was later extended to other types of distributions (Vaillant and Glaunès, 2005).

Appendix A.1.4. Point sets as mixture models

To our knowledge, Wells (Wells, 1997) and Moss and Hancock (Moss and Hancock, 1997) were the first to propose to use a probabilistic formulation of the point set registration problem. Indeed, if one considers the point set registration as the problem of estimating an optimal parameter T linking data corrupted by noise and outliers, the use of pdfs to model each point set appears natural. Numerous methods, directly or indirectly, rely on such a modelling.

One defines f_T and g as mixture models having $(T(x_k))_k$ and $(y_j)_j$ as centres:

$$f_T(z) = \frac{1}{N}\sum_k p_f(T(x_k) - z)$$

and

$$g(z) = \frac{1}{M}\sum_j p_g(y_j - z),$$

p_f and p_g being two pdfs. Then the registration problem consists in minimising a divergence between f_T and g with respect to T. Note that, in a similar way, Wang and collegues proposed to use the cumulative distribution functions or the cumulative residual entropy associated to the pdfs f_T and g rather than the pdfs themselves, arguing that the former are less sensitive to outlying points than the latter (Wang et al., 2006, 2010). The three types of functions lead to a similar interpretation and below we focus on the methods using pdfs for the sake of simplicity. In this context, p_f and p_g are usually considered as spherically-symmetrical (isotropic) Gaussians or Diracs, and several divergence measures have been investigated. In particular, Wang and colleagues (Wang et al., 2008) proposed to model f_T and g as GMMs and to use the Jensen-Shannon divergence (also termed total divergence to the average). The intrinsic inability of the minimum Jensen-Shannon divergence estimator to cope with outliers has stimulated the use of a divergence leading to more robust estimators for T. In particular, Jian and Vemuri (Jian and Vemuri, 2005) proposed to use the L_2 distance between two GMMs. The minimum L_2 distance estimator is known to be robust and can be shown to belong to the class of M-estimators (Basu et al., 1998). Several works have followed this track (Tsin and Kanade, 2004; Roy et al., 2007; Wang et al., 2009; Sandhu et al., 2009; Chen et al., 2009; Tustison et al., 2011; Ma et al., 2013).

Interestingly, minimising the Kullback-Leibler divergence between a mixture of Gaussians f_T and a mixture of Diracs g is equivalent to solving the MAP problem where the points of Y are considered as the draws of a mixture of Gaussians centred at the points of $T(X)$ (proof in Appendix B). This MAP problem can be solved using the EM algorithm. In addition to its simplicity, this algorithm has the advantage over gradient-based optimisation algorithms (such as the conjugate-gradient algorithm) not to need additional optimisation-related parameters to tune and to achieve a monotonic convergence to a (at least) local maximum of the posterior distribution under study. Several recently proposed algorithms (*e.g.* (Chui and Rangarajan, 2000; Guo and Rangarajan, 2009; Myronenko and Song, 2010)) are modifications of this MAP framework and are commonly considered as state-of-the-art methods. Recent modifications of the original approach include using a full (non isotropic) covariance matrix in the Gaussian pdf for improved modelling of the data (Horaud et al., 2011), introducing weights in the likelihood function (Hermans et al., 2011; Tao and Sun, 2014) or using a Student (Nguyen and Wu, 2016) pdf instead of a Gaussian one for improved robustness. This is within this context that we propose our own solutions in this paper.

Appendix A.2. Deformation models

For each class of methods (and each method), a given implementation is characterised by the choice of a deformation model T that mainly stands on a parametrisation for T and of a regulariser R on T. Common choices consist in considering T as either i) a displacement field with R as the scalar thin plate spline (TPS) (Bookstein, 1989), or the motion coherent theory (MCT) (Yuille and Grzywacz, 1989) regulariser or as ii) the integration of a time-dependent velocity field regularised through R, with R as a scalar TPS (Camion and Younes, 2001), as a scalar MCT (Glaunès et al., 2004) or as a scalar Laplacian (Joshi and Miller, 2000) regulariser.

Appendix B. Equivalence between Kullback-Leibler minimiser and maximum likelihood

$$KL(g\|f_T) = \int g(z)\log(g(z))dz - \int g(z)\log(f_T(z))dz,$$

the first term being a constant with respect to T, one can write (with f_T as a mixture of Gaussians and g as a mixture of Diracs):

$$\begin{aligned}
\arg\min_T KL(g\|f_T) &= -\arg\min_T \int g(z)\log(f_T(z))dz,\\
&= \arg\max_T \int \sum_{y_j\in Y}\delta(z-y_j)\log(f_T(z))dz,\\
&= \arg\max_T \sum_{y_j\in Y}\log(f_T(y_j)),\\
&= \arg\max_T \sum_{y_j\in Y}\log\left(\sum_{x_i\in X}\psi_i(y_j;T)\right),\\
&= \arg\max_T \prod_{y_j\in Y}\left(\sum_{x_i\in X}\psi_i(y_j;T)\right),
\end{aligned}$$

which corresponds to the maximum likelihood estimator of T.

Appendix C. Equivalence of the M-steps

The derivative of $\mathcal{E}_M = \sum_{j,k}\tilde{A}_{jk}\|y_j - x_k - t(x_k)\|^2 + 2\sigma^2\alpha R(t)$ with respect to $t(x_k)$ is:

$$\begin{aligned}
\frac{\partial\mathcal{E}_M}{\partial t(x_k)} &= -2\sum_j \tilde{A}_{jk}y_j + 2(\sum_j \tilde{A}_{jk})(x_k + t(x_k))\\
&\quad +2\sigma^2\alpha\frac{\partial R(t(x_k))}{\partial t(x_k)}
\end{aligned}$$

Calling $\tilde{A}_{.k} = \sum_j \tilde{A}_{jk}$, this gives:

$$\begin{aligned}\frac{\partial \mathcal{E}_M}{\partial t(x_k)} &= -2\sum_j \tilde{A}_{jk} y_j + 2\tilde{A}_{.k}(x_k + t(x_k)) \\ &\quad +2\sigma^2\alpha\frac{\partial R(t(x_k))}{\partial t(x_k)}\end{aligned}$$

which is the derivative with respect to $T(x_k)$ of:

$$\begin{aligned}\mathcal{E}'_M &= \sum_k \tilde{A}_{.k}\left\|\frac{\sum_j \tilde{A}_{jk} y_j}{\tilde{A}_{.k}} - x_k - t(x_k)\right\|^2 \\ &\quad +2\sigma^2\alpha R(T) \qquad \text{(C.1)}\end{aligned}$$

Thus minimising $\mathcal{E}_M$ is equivalent to minimise $\mathcal{E}'_M$.

References

Aouada, D., Feng, S., Krim, H., 2007. Statistical Analysis of the Global Geodesic Function for 3D Object Classification, in: IEEE International Conference on Acoustics, Speech and Signal Processing., pp. 645–648.

Aronszajn, N., 1950. Theory of reproducing kernels. Transactions of the American Mathematical Society 68, 337–404.

Barron, J., Fleet, D., Beauchemin, S., 1994. Performance of optical flow techniques. International Journal of Computer Vision 12, 43–77.

Basu, A., Harri, I., Hjort, N.L., Jones, M., 1998. Robust and efficient estimation by minimising a density power divergence. Biometrika 85, 549–559.

Bentley, J., 1975. Multidimensional binary search trees used for associative searching. Communications of the ACM 18, 509–517.

Besl, P., McKay, N., 1992. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14, 239–256.

Bookstein, F.L., 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transaction Pattern Analysis and Machine Intelligence 11, 567–585.

Buhmann, M.D., 1998. Radial functions on compact support. Proceedings of the Edinburgh Mathematical Society 41, 33–46.

Camion, V., Younes, L., 2001. Geodesic interpolating splines, in: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 513–527.

Carcassoni, M., Hancock, E.R., 2000. Spectral correspondence for deformed point-set matching, in: International Workshop on Articulated Motion and Deformable Objects, pp. 120–132.

Chen, T., Vemuri, B.C., Rangarajan, A., Eisenschenk, S.J., 2009. Group-wise point-set registration using a novel cdf-based havrda-charvát divergence. International Journal of Computer Vision 86, 111–124.

Chui, H., Rangarajan, A., 2000. A new algorithm for non-rigid point matching. IEEE Computer Vision and Pattern Recognition 2, 44–51.

Chui, H., Win, L., Schultz, R., Duncan, J.S., Rangarajan, A., 2003. A unified non-rigid feature registration method for brain mapping. Medical Image Analysis 7, 113–130.

Combès, B., Prima, S., 2009. Setting priors and enforcing constraints on matches for nonlinear registration of meshes, in: Medical Image Computing and Computer Assisted Interventions Conference, pp. 175–183.

Combès, B., Prima, S., 2010. An efficient em-icp algorithm for symmetric consistent non-linear registration of point sets, in: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010, Springer Berlin Heidelberg. pp. 594–601.

Day, N., 1969. Estimating the components of a mixture of normal distributions. Biometrika 56, 463–474.

Dedner, A., Lüthi, M., Albrecht, T., Vetter, T., 2007. Curvature guided level set registration using adaptive finite elements, in: IEEE Conference on Pattern recognition, pp. 527–536.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39, 1–38.

Feldmar, J., Ayache, N., 1996. Rigid, affine and locally affine registration of free-form surfaces. International Journal of Computer Vision 18, 99–119.

G. Celeux and G. Govaert, 1992. A classification EM algorithm for clustering and two stochastic versions. Computational Statistics and Data Analysis 14, 315–332.

Geman, D., Reynolds, G., 1992. Constrained restoration and the recovery of discontinuities. IEEE Transaction on Pattern Analysis and Machine Intelligence 14, 367–383.

Girosi, F., Jones, M., Poggio, T., 1995. Regularization theory and neural networks architectures. Neural Computation 7, 219–269. URL: `http://dx.doi.org/10.1162/neco.1995.7.2.219`.

Glaunès, J., Trouvé, A., Younes, L., 2004. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 712–718.

Granger, S., Pennec, X., 2002. Multi-scale EM-ICP: A fast and robust approach for surface registration, in: European Conference on Computer Vision, pp. 418–432.

Green, P., 1990. On Use of the EM for Penalized Likelihood Estimation. Journal of the Royal Statistical Society. Series B (Methodological) 52, 443–452.

Greengard, L., Strain, J., 1991. The fast Gauss transform. SIAM Journal of Sci. Stat. Comput. 12, 79–94.

Gumhold, S., Wang, X., Macleod, R., 2001. Feature Extraction from Point Clouds, in: 10th International Meshing Roundtable, pp. 293–305.

Guo, H., Rangarajan, A., 2009. Diffeomorphic point matching with applications in biomedical image registration. International Journal of Tomography and Statistics 15.

Hermans, J., Smeets, D., Vandermeulen, D., Suetens, P., 2011. Robust point set registration using em-icp with information-theoretically optimal outlier handling, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2465–2472.

Horaud, R., Forbes, F., Yguel, M., Dewaele, G., Zhang, J., 2011. Rigid and articulated point registration with expectation conditional maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 587–602.

Huang, Q., Adams, B., Wicke, M., Guibas, L., 2008. Non-rigid registration under isometric deformations. Computer Graphics Forum 27, 1449–1457.

Hufnagel, H., Pennec, X., Ehrhardt, J., Ayache, N., Handels, H., 2008. Generation of a statistical shape model with probabilistic point correspondences and the expectation maximization- iterative closest point algorithm. International Journal of Computer Assisted Radiology and Surgery 2, 265–273. URL: `https://doi.org/10.1007/s11548-007-0138-9`.

Jian, B., Vemuri, B., 2005. A robust algorithm for point set registration using mixture of Gaussians, in: IEEE International Conference on Computer Vision, pp. 1246–1251.

Joshi, S., Miller, M., 2000. Landmark matching via large deformation diffeomorphisms. IEEE Transactions on Image Processing 9, 1357–1370.

Koenderink, J., van Doorn, A., 1992. Surface shape and curvature scales. Image and Vision Computing 10, 557–564.

Lüthi, M., Albrecht, T., Vetter, T., 2007. Curvature guided surface registration using level sets, in: Congress on Computer Assisted Radiology and Surgery.

Ma, J., Zhao, J., Tian, J., Tu, Z., Yuille, A.L., 2013. Robust estimation of nonrigid transformation for point set registration, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2147–2154.

Maintz, J., Viergever, M., 1998. A survey of medical image registration. Medical Image Analysis 2, 1–36.

Maiseli, B., Gu, Y., Gao, H., 2017. Recent developments and trends in point set registration methods. Journal of Visual Communication and Image Representation 46, 95–06.

Marriott, F.H.C., 1975. Separating Mixtures of Normal Distributions. Biometrics 31, 767–769.

Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., Boyer, E., 2008. Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration, in: IEEE Conference on Computer Vision and Pattern Recognition.

Moss, S., Hancock, E., 1997. Image registration with shape mixtures, in: International Conference on Image Analysis and Processing, pp. 172–179.

Myronenko, A., Song, X., 2010. Point-Set Registration: Coherent Point Drift. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 2262–2275.

Nguyen, T.M., Wu, Q.M.J., 2016. Multiple kernel point set registration. IEEE Transactions on Medical Imaging 35, 1381–1394.

Rangarajan, A., Chui, H., Mjolsness, E., Pappu, S., Davachi, L., Goldman-

Rakic, P., Duncan, J., 1997. A robust point-matching algorithm for autoradiograph alignment. Medical Image Analysis 1, 379–398.

Robbins, S.M., 2004. Anatomical standardization of the human brain in euclidean 3-space and on the cortical 2-manifold. Ph.D. thesis. McGill University.

Roy, A., Gopinath, A., Rangarajan, A., 2007. Deformable density matching for 3D non-rigid registration of shapes, in: Medical Image Computing and Computer Assisted Intervention, pp. 942–949.

Sandhu, R., Dambreville, S., Tannenbaum, A., 2009. Point set registration via particle filtering and stochastic dynamics. IEEE Transactions on Pattern Analysis and Machine Intelligence 99.

Schölkopf, B., Herbrich, R., Smola, A.J., 2001. A generalized representer theorem, in: In Proceedings of the Annual Conference on Computational Learning Theory, pp. 416–426.

Scott, A., Symons, M., 1971. Clustering Methods Based on Likelihood Ratio Criteria. Biometrics 27, 387–397.

Scott, G., Longuet-Higgins, H., 1991. An algorithm for associating the features of two images. Proceedings: Biological Sciences 244, 21–26.

Shapiro, L., Brady, J., 1992. Feature-based correspondence: an eigenvector approach. Image and Vision Computing 10, 283–288.

Shen, L., Huang, H., Makedon, F., Saykin, A., 2007. Efficient Registration of 3D SPHARM Surfaces. Canadian Conference on Computer and Robot Vision , 81–88.

Sinkhorn, R., 1964. A relationship between arbitrary positive matrices and doubly stochastic matrices. The Annals of Mathematical Statistics 35, 876–879.

Tam, G.K.L., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X.F., Rosin, P.L., 2013. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. IEEE Transactions on Visualization and Computer Graphics 19, 1199–1217.

Tang, J., Liang, D., Wang, N., Jia, Z.H., 2007. Spectral correspondence using local similarity analysis, in: 2007 International Conference on Computational Intelligence and Security (CIS 2007), pp. 395–399.

Tao, W., Sun, K., 2014. Asymmetrical gauss mixture models for point sets matching, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1598–1605.

Thirion, J.P., 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2, 243 – 260.

Tsin, Y., Kanade, T., 2004. A correlation-based approach to robust point set registration. European conference on computer vision 3, 558–569.

Tustison, N.J., Awate, S.P., Song, G., Cook, T.S., Gee, J.C., 2011. Point Set Registration Using Havrda-Charvat-Tsallis Entropy Measures. IEEE Transactions on Medical Imaging 30, 451–460.

Vaillant, M., Glaunès, J., 2005. Surface matching via currents, in: Proceedings of Information Processing in Medical Imaging, pp. 381–392.

Šidlofová, T., 2004. Existence and uniqueness of minimization problems with fourier based stabilizers, in: Antoch, J. (Ed.), COMPSTAT 2004 — Proceedings in Computational Statistics, Physica-Verlag HD, Heidelberg. pp. 1853–1860.

Wang, F., Vemuri, B., Rangarajan, A., 2006. Groupwise point pattern registration using a novel CDF-based Jensen-Shannon divergence, in: IEEE conference on Computer Vision and Pattern Recognition, pp. 1283–1288.

Wang, F., Vemuri, B., Rangarajan, A., Eisenschenk, S.J., 2008. Simultaneous nonrigid registration of multiple point sets and atlas construction. IEEE Transactions on Pattern Analysis Machine Intelligence 30, 2011–2022.

Wang, F., Vemuri, B., Syeda-Mahmood, T., 2009. Generalized L2-Divergence and its application to shape alignment, in: IEEE Information Processing in Medical Imaging, pp. 227–238.

Wang, H., Mirota, D., Hager, G.D., 2010. A generalized kernel consensus-based robust estimator. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 178–184.

Wells, III, M.W., 1997. Statistical approaches to feature-based object recognition. International Journal of Computer Vision 21, 63–98.

Wendland, H., 2004. Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press. URL: http://dx.doi.org/10.1017/CBO9780511617539.

Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., Golland, P., 2010. Spherical demons: Fast diffeomorphic landmark-free surface registration. IEEE Transactions on Medical Imaging 29, 650 –668.

Younes, L., 2001. Deformation analysis for shape and image processing. Technical Report.

Yuille, A., Grzywacz, N., 1989. A mathematical analysis of the motion coherence theory. International Journal of Computer Vision 3, 155–175.

Zou, G., Hua, J., Muzik, O., 2007. Non-rigid surface registration using spherical Thin-Plate splines, in: Medical Image Computing and Computer-Assisted Intervention, pp. 367–374.