A new efficient EM-ICP algorithm for non-linear registration of 3D point sets

Benoît Combès, Sylvain Prima

A new efficient EM-ICP algorithm for non-linear registration of 3D point sets

Benoit Combès, Sylvain Prima

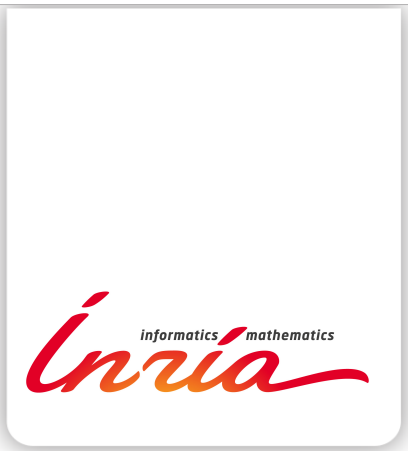
RESEARCH REPORT

N° 7853

Decembre 2011

Project-Team Visages

ISSN 0249-6399 ISRN INRIA/RR--7853--FR+ENG



A new efficient EM-ICP algorithm for non-linear registration of 3D point sets

Benoit Combès, Sylvain Prima

Project-Team Visages

Research Report n° 7853 — Decembre 2011 — 33 pages

Abstract: In this paper, we present a new method for non-linear pairwise registration of 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 points of the second set displaced by a deformation. Next we perform *maximum a posteriori* estimation of the parameters (which include the unknown transformation) of this model using the expectation-maximisation algorithm. Compared to other methods using the same "EM-ICP" paradigm/framework, we propose three key modifications leading to an efficient algorithm allowing for fast registration of large point sets: 1) symmetrisation of the point-to-point correspondences; 2) specification of priors on these correspondences using differential geometry; 3) efficient encoding of deformations using the RKHS theory and the Fourier analysis. The resulting algorithm is efficient and is able to register large data sets. We evaluate the added value of the modifications and compare our method to the state-of-the-art CPD algorithm on synthetic data.

Key-words: point sets, surface, non-linear registration, alignment, ICP, EM-ICP

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Un nouvel algorithme EM-ICP efficace pour le recalage non-linéaire de nuages de points tridimensionnels

Résumé : Dans cet article, nous présentons une nouvelle méthode pour le recalage non-linéaire de deux nuages de points. Dans cette méthode, nous considérons les points du premier nuage comme la réalisation d'un mélange de gaussiennes dont les centres sont les points du second ensemble déplacés par une déformation. Ensuite, nous estimons cette déformation, sur laquelle nous fixons un a priori, selon le principe du maximum a posteriori en utilisant l'algorithme "expectation-maximisation". Par rapport aux autres méthodes qui utilisent un paradigme similaire, nous proposons de: 1) symétriser le processus de correspondance entre les points des deux nuages, 2) spécifier des a priori sur les correspondances en utilisant des outils de la géométrie différentielle et 3) caractériser la déformation à estimer en utilisant la théorie des espaces de Hilbert à noyaux reproduisants et l'analyse de Fourier. L'algorithme résultant est relativement efficace et permet de recaler des nuages de points de grandes tailles. Enfin, nous évaluons l'impact de ces modifications puis nous comparons notre méthode à une méthode de l'état de l'art.

Mots-clés : nuages de points, surface, recalage non-linéaire, alignement, ICP, EM-ICP

Contents

1 Introduction **4**

2 Previous works **5**
2.1 The four classes 6
2.1.1 Point sets as modal matrices 6
2.1.2 Point sets as level set functions 6
2.1.3 Point sets as Schwartz distributions 7
2.1.4 Point sets as mixture models 7
2.2 Deformation models 8

3 Non-linear registration as a statistical inference problem **8**
3.1 General formulation 9
3.2 Robustifying the estimator and reducing the computational burden 11
3.3 Energetic interpretation 12

4 Symmetrising the matching process **12**
4.1 Minimisation with respect to T 13

5 Adding priors **15**
5.1 Designing π 15
5.1.1 Designing π using labels 15
5.1.2 Designing π using descriptors 16
5.2 Implementation 16

6 Solving the M-step **17**
6.1 Approximation problems in RKHS 18
6.2 Choosing a kernel 19
6.3 Why using frequencies? 19
6.4 Efficient choices 20
6.5 M-step in a nutshell 21
6.6 Note on vectorial kernels 21

7 Related algorithms **22**
7.1 The TPS-RPM algorithm 22
7.2 The CPD algorithm 22
7.3 Our method (Reg4) 23

8 Validation & Results **23**
8.1 Validation 23
8.1.1 Ground truth data 23
8.1.2 Results 24

9 Conclusions and perspectives **27**
9.1 Conclusions 27
9.2 Perspectives 28

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.* Some 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. Structured (meshes) or unstructured point sets are the most generic way for such a purpose, 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.

Our **first contribution** is to explicit these intermediate representations in Section 2 and to provide a comprehensive taxonomy of existing methods, in which we identify four broad classes depending on whether the point sets are considered as modal matrices (Section 2.1.1), level set functions (Section 2.1.2), Schwartz distributions (Section 2.1.3) or mixtures of probability density functions (pdfs) (Section 2.1.4). Within each of these classes, we also identify the models for non-linear deformations (ii) (Section 2.2).

Among these four classes and the methods therein, an especially attractive approach is to consider the two point sets as, respectively, a Gaussian and a Dirac mixture model, and to use the Kullback-Leibler divergence as a similarity metric between the two mixtures. In this case, it can be shown that registering the two datasets consists of a maximum a posteriori (MAP) problem (or equivalently, to a penalised likelihood problem), where the points of the first set can be seen as the draws of a Gaussian mixture model (GMM) whose centres are the points of the second dataset. Considering the variance of the Gaussian mixture model as fixed and known, the only unknown parameter to be estimated is the non-linear transformation, and this 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 correspondences between the point sets (as, in practice, they do no exist) by indirectly introducing probabilities of correspondence. It also allows one to deal with outlying data in an efficient way.

Our **second contribution** (Section 3), in addition to showing that using

two specific mixture models and one specific similarity metric allows one to see the registration problem as one of classical statistical inference, is (i) to show that when using the classification likelihood rather than the likelihood, the expectation-classification-maximisation (ECM) rather than the EM algorithm, and rigid-body transformations, the resulting registration algorithm is simply the ICP [5], while (ii) when using the likelihood, the EM algorithm and rigid-body (resp. TPS) transformations, the resulting registration algorithm is simply the EM-ICP [19] (resp. TPS-RPM [9]). Some other deformation models have been recently tested by other authors in the same framework (termed *EM-ICP* later on for the sake of simplicity), most notably the coherent point drift (CPD) [33] and articulated deformations [23].

As previously mentioned, this EM-ICP framework is very attractive and effective. However, to the best of our knowledge, only few efforts have been done to study its intrinsic limitations. First, one observes that the derivation of the MAP principle leads to an asymmetric formulation (note: which is also seen from the Kullback-Leibler divergence). 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. Second, 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, both the E-step and the M-step are very time and memory consuming when dealing with large datasets.

Our **third contribution** is to propose efficient solutions for each of these three problems. In Section 4, we propose to tackle problems due to the above-mentioned asymmetry of the MAP formulation. For this purpose, we notice 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 and we propose to modify it to make it symmetric. For this purpose, we introduce a second, column stochastic, match matrix within this criterion. 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 one to significantly improve 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 (RKHS) theory and on the Fourier analysis. It consists in building efficient regularisers leading to a closed-form solution based on sparse linear algebra.

In Section 7, we discuss the merits of our algorithm with respect to close works: the TPS-RPM [9] and the CPD [34] 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 Previous works

Throughout Section 2, we focus on registration methods having no strong priors on the structures to register (*e.g.* topological [55, 43, 57, 36], tessellation constraints [36]) or on the expected deformations (*e.g.* articulated motion [31],

isometric deformations [24]).

In Section 2.1, we propose a comprehensive four-class taxonomy of point set representations underlying most non-linear registration methods of the literature. 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 Section 2.2.

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 (in a sense to be later defined).

2.1 The four classes

2.1.1 Point sets as modal matrices

The first authors to propose the use of the spectral theory to align two point sets were Shapiro and Brady [42] and Scott and Longuet-Higgins [41]. The two methods are slightly different. For a sake of simplicity, we focus on the method proposed by Shapiro and Brady (and on 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 of X and Y. The modal matrix of a point set X is computed by (i) building a $\text{card}(X) \times \text{card}(X)$ symmetric proximity matrix $G_{jk} = \exp(-||x_j - x_k||/(2\sigma^2))$ (σ measures the variance associated to the error in inter-point distances), (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 almost invariant under the expected transformations shape descriptor. As a result, the estimation of the point-to-point correspondences is performed by comparing rows of both modal matrices. This can be done either by applying a "best one rule" [42] or by building correspondence probabilities [8, 46] (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 a deformation given known correspondences and the estimation of correspondences between points (given updated positions of the point set X) [8, 46].

2.1.2 Point sets as level set functions

Lüthi and colleagues [28, 12] proposed to represent each point set to register as the zeroth level of a signed distance 3D function I, *i.e.* for X:

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

Note that to consider such a representation, one must assume 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 [29]. Albrecht and colleagues proposed to use the demons algorithm [47] and modify it in order to incorporate the mean curvature images of I_X and I_Y (in addition to their intensity values) to guide the registration.

2.1.3 Point sets as Schwartz distributions

The framework of diffeomorphic matching of distributions was developed by Glaunes and colleagues [18]. 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 $\mathbb{R}^3$. 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 [2], this distance $||. - .||_{\mathcal{I}^*}$ can be easily evaluated and differentiated with respect to T. The transformation T can then be estimated using a gradient-descent minimisation. This work was later extended to other types of distributions [49].

2.1.4 Point sets as mixture models

To our knowledge, Wells [53] and Moss and Hancock [32] were the firsts to propose to use a probabilistic formulation of the point set registration problem. Indeed, considering 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:

$$f_T(z) = 1/N \sum_k p_f(T(x_k) - z)$$

and

$$g(z) = 1/M \sum_j p_g(y_j - z)$$

p_f and p_g being two pdfs. Then the registration problem is considered as a problem of minimisation of 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 [51, 52]. The three types of functions yield to a similar interpretation of the divergence measures, and below we focus on the methods using pdfs for the sake of simplicity. In this context, pdfs

are usually considered as isotropic Gaussians or Diracs, and several divergence measures have been investigated. In particular:

• Wang and colleagues [15] proposed to model f_T and g as GMMs and to use the Jensen-Shannon (JS) 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 colleagues [25] 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 [3]. Several works have followed this track [37, 50]. Note that when modelling T as a isometric transformation, minimising the L_2 distance between the two GMMs is equivalent to maximise their correlation [38, 48].

• 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 $T(X)$. As will be seen in Section 3, 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 parameters specific to the optimisation strategy and to achieve a monotonic convergence to a (at least) local maximum of the posterior distribution. Several algorithms (*e.g.* [9, 33]) emerged from this paradigm and are commonly considered as state-of-the-art methods. This is the track we follow in this paper, and which is extensively outlined in Section 3.

2.2 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 a displacement field with R as the scalar thin plate spline (TPS) [6], the coherent point drift (CPD) [56] regulariser or as a smooth velocity field with R as a scalar TPS [7], as a scalar CPD [18] or as a scalar Laplacian [26] regulariser. Each of these deformation models exhibit different regularisation properties and ease of implementation.

3 Non-linear registration as a statistical inference problem

As noted earlier, (i) modelling the two points sets as a Dirac and a Gaussian mixture model and (ii) using the Kullback-Leibler divergence is equivalent to (i) seeing the points of the first set as the draws of a Gaussian mixture model whose centres are the points of the second set and (ii) using the MAP/penalised likelihood principle. In Section 3.1, we make one step back and we start for even simpler hypotheses on the sets to register to show how incremental modifications to these hypotheses and adequate optimisation tools of classical statistical inference lead to respectively, the ICP, the EM-ICP and the TPS-RPM/CPD algorithms. In Section 3.2, we build on this statistical framework to outline a

generic non-linear registration algorithm which exhibits increased speed and robustness towards outliers, while keeping good convergence properties. Finally, in Section 3.3, we notice that this algorithm actually consists of the alternated minimisation of an energetic criterion, in which each term can be simply interpreted. Some critical limits of this algorithm will be exposed and tackled in Sections 4, 5 and 6.

3.1 General formulation

If one considers the surface 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 sample y_j has been drawn independently from any one of $N = \text{card}(X)$ possible 3-variate normal distributions with centres (means) $T(x_k)$ and covariance matrices $\sigma^2 I$ (with $\sigma > 0$ unknown, but fixed).

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 $\text{card}(X)$ possible distributions from which y_j has been drawn, and (ii) to estimate the parameters of these $\text{card}(X)$ 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 and especially the maximum likelihood principle and the EM algorithm to solve the registration problem. Three different paradigms have been especially followed in this context [30]. Let us introduce some notations first:

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

$\forall j \in 1...M$, $\forall k \in 1... \text{card}(X)$, $z_{jk} = 1$ iff y_j comes 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 criterion CL [40]:

$$CL = \prod_{y_j \in Y} \prod_{x_k \in X} [\psi_k(y_j;T)]^{z_{jk}} \tag{2}$$

The maximisation is typically performed by the Classification EM (CEM) algorithm [16], which can be shown to find an at least local maximum of the criterion 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 modeled as a rigid-body transformation, this algorithm is nothing else than the popular Iterative Closest Point (ICP) algorithm [5]. Note that this algorithm does not depend on σ.

In the Maximisation 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 [11]:

$$L = \prod_{y_j \in Y} \sum_{x_k \in X} \pi_{jk} \psi_k(y_j; T) \tag{3}$$

In essence, the prior π_{jk} conveys the probability that the point y_j comes from the distribution ψ_k without knowing anything else. The likelihood L can be maximised by using the popular EM algorithm, which converges to an at least local maximum of L [13]. 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 (which is not to be estimated in this framework) acts as a scale parameter. The problem of estimating σ (or even a different σ_k for each of the pdfs $\psi_k(.,T)$) is not investigated in this work. It can be given an initial value and decreased throughout the iterations for improved performances. When T is modeled as a rigid-body transformation and priors π_{jk} are chosen to be uniform, this algorithm is nothing else than the EM-ICP algorithm [19].

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 [20]. 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 using the MAP principle (also termed penalised ML):

$$pL = \prod_{y_j \in Y} \sum_{x_k \in X} \pi_{jk} \psi_k(y_j; T) p(T) \tag{4}$$

The EM algorithm then writes:

Algo Reg1: 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: $\tilde{T} = \arg\min_T \sum_{j,k} \tilde{A}_{jk} ||y_j - T(x_k)||^2 + \alpha R(T)$

We call Reg1 the resulting algorithm. In the following, this algorithm will be improved gradually and will yield to 3 other algorithms that we will call

Reg2, Reg3 and Reg4. The performances of these algorithms will be compared in Section 8.

On can see that, in essence, the parameter α weighs the relative influence of the regularity term $R(T)$ and the "distance" term $\sum_{j,k} A_{jk}||y_j - T(x_k)||^2$. Note that one can consider a similar adaptation for the CML approach.

Interpretation:

Intuitively, the ML approach is a *fuzzy* version of the CML. It appears clearly from the iterative formulas of both algorithms 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. Note that the ML formulation followed by the EM algorithm leads to the same iterative formulas that would have resulted from the addition of a barrier function on the indicator variables in the ICP criterion [9].

3.2 Robustifying the estimator and reducing the computational burden

In practice, Reg1 suffers from outliers *i.e.* points of X having no satisfactory correspondence in Y. To alleviate this problem, one can consider $\psi_k(.;T)$ as a truncated Gaussian pdf with cut-off distance $\delta > 0$. This modification has beneficial effects on the properties of the algorithm. Indeed, it allows one : (i) to reduce drastically the computational burden of the E-step (by the use of a kd-tree [4]), (ii) to reduce the impact of points of Y having no correct correspondence in X in the estimation of the optimal T (*i.e.* to improve the robustness of the criterion) and iii) to increase the convergence speed of the overall algorithm. The following algorithm can be shown to converge to an at least local maximum of the new (truncated) criterion:

Algo Reg2: Robust NL-EM-ICP

E-step: initialise $A = (A_{jk})$ as the null matrix
for each $x_k \in X$;
  $\mathcal{S} = \{y_j \in Y \text{ such that } ||y_j - \tilde{T}(x_k)||^2 < \delta\}$
(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 $y_j \notin S$; A_{jk} is left equal to 0.

for each $y_j \in Y$;
  if $\sum_i A_{ji} \neq 0$
    $\tilde{p}_j = 1$
    for each $x_k \in X$, $\tilde{A}_{jk} = A_{jk} / \sum_i A_{ji}$
  else
    $\tilde{p}_j = 0$

M-step: $\tilde{T} = \arg\min_T \sum_{j,k} \tilde{p}_j \tilde{A}_{jk} ||y_j - T(x_k)||^2 + \alpha R(T)$

This introduces a set (p_j) of binary variables such that $\forall j$, p_j is null if point y_j is considered as an outlier (thus has no correspondence) and is equal to one else.

3.3 Energetic interpretation

It can be shown that Reg2 is an iterative alternated minimisation (over A and T) of the following criterion:

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

with $\forall j$, $\sum_k A_{jk} = 1$ and where $\rho_\delta : r \longmapsto r$ if $r < \delta$ and δ else (δ corresponds to the cut-off distance of the truncated pdf of Section 3.2). To simplify the notation, we introduce $\alpha^{'} = 2\sigma^2\alpha$. Note that the ρ_δ function could be advantageously replaced by a smoother differentiable function (*e.g.* a Leclerc function [17]). However, we prefer to stay as close as possible to our probabilistic formulation in order to keep the probabilistic interpretation of the different parameters (particularly of the match matrix A).

It can be given an energetic interpretation of $\mathcal{E}1$ whose 3 terms represent respectively:

- a data-attachment term (where the quadratic cost function is replaced by a robust cost function),
- a barrier function allowing 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.

4 Symmetrising 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 is given by x_c where $c = \arg\max_k A_{jk}$. This leads to many-to-one correspondences between points of X and Y. In particular, there is no direct constraint to enforce a point of X to have a correspondence in Y. This makes the algorithm unable to achieve a proper matching in some specific configurations. It is particularly enlightening to consider the case when surfaces are far from each other (example on Fig. 1).

To alleviate this problem, Rangarajan and colleagues [35] 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 [45] 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 Gaussian pdf (as in Section 3.2). In practice, this limits its application to small point sets.

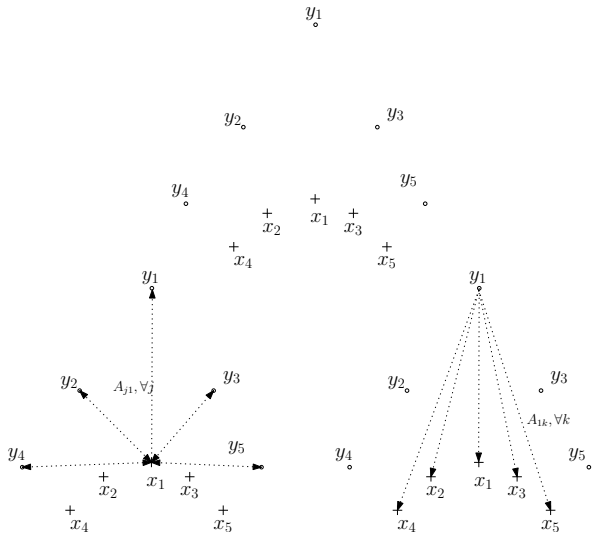


Figure 1: **Effect of the asymmetric normalisation constraint on A.** The "correct" matches are $(x_l, y_l)_{l=1...5}$. One considers T as the identity. From left to right and top to bottom: (i) two point sets X and Y, (ii) distances involved in the computation of A_{11} when registering Y on X. (iii) distances involved in the computation of A_{11} when registering X on Y. As an example, if one considers $\sigma = 1$, then in case (ii) $A_{11} = 0.01$ and points x_1 and y_1 have only little chance to be 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. This leads to the following criterion:

$$\begin{aligned} & \mathcal{E}2(Y, T(X), A, B) = \\ & \sum_{j,k} A_{jk} \rho_\delta(||y_j - T(x_k)||^2) + 2\sigma^2 \sum_{j,k} A_{jk} \log(A_{jk}) \\ + \quad & \sum_{j,k} B_{jk} \rho_\delta(||y_j - T(x_k)||^2) + 2\sigma^2 \sum_{j,k} B_{jk} \log(B_{jk}) \\ + \quad & \alpha^{'} R(T), \end{aligned} \tag{6}$$

with $\forall j$, $\sum_k A_{jk} = 1$ and $\forall k$, $\sum_j B_{jk} = 1$. Notice that if one considers the matrix $C = (\frac{(A_{jk}+B_{jk})}{\text{card}(X)+\text{card}(Y)})$, one can show that $\forall j, k$ $C_{jk} \geq 0$ and $\sum_{j,k} C_{jk} = 1$.

4.1 Minimisation with respect to T

While the minimisation of $\mathcal{E}2(Y, T(X), A, B)$ with respect to A and B when T is fixed (E-step) is straightforward and of low complexity (using the strategy proposed in Section 3.2), its minimisation with respect to T (M-step) generally consists in solving a numerical system of size proportional to $\text{card}(Y) \times \text{card}(X)$. When dealing with large point sets X and Y, this can have severe effects on

the computational and memory usage. In this section, we propose to reduce the complexity of this minimisation from $\text{card}(Y) \times \text{card}(X)$ to $\text{card}(X)$.

For this purpose, we now consider that T is represented as the initial position plus a displacement field: $T(x_k) = x_k + t(x_k)$ and R is a regulariser on t. Then, the M-step consists of:

M-step:
$\tilde{t} = \arg\min_t \sum_{j,k} (\tilde{p}_j \tilde{A}_{jk} + \tilde{q}_k \tilde{B}_{jk}) ||y_j - x_k - t(x_k)||^2 + \alpha^{'} R(t)$

where, analogously to (p_j), (q_k) are the binary variables associated to the matrix B.

The derivative of $\mathcal{E}_M = \sum_{j,k} (\tilde{p}_j \tilde{A}_{jk} + \tilde{q}_k \tilde{B}_{jk}) ||y_j - x_k - t(x_k)||^2 + \alpha^{'} R(t)$ with respect to $t(x_k)$ is:

$$\frac{\partial \mathcal{E}_M}{\partial t(x_k)} = -2 \left(\sum_j \tilde{p}_j \tilde{A}_{jk} y_j + \sum_j \tilde{q}_k \tilde{B}_{jk} y_j \right)$$

$$+2 \left((\sum_j \tilde{p}_j \tilde{A}_{jk} + \sum_j \tilde{q}_k \tilde{B}_{jk})(x_k + t(x_k)) \right) + \alpha^{'} \frac{\partial L(t(x_k))}{\partial t(x_k)}$$

Calling $\tilde{B}_{.k} = \sum_j \tilde{q}_k \tilde{B}_{jk}$ and $\tilde{A}_{.k} = \sum_j \tilde{p}_j \tilde{A}_{jk}$, this gives:

$$\frac{\partial \mathcal{E}_M}{\partial t(x_k)} = -2 \sum_j (\tilde{p}_j \tilde{A}_{jk} + \tilde{q}_k \tilde{B}_{jk}) y_j + 2(\tilde{A}_{.k} + \tilde{B}_{.k})(x_k + t(x_k))$$

$$+\alpha^{'} \frac{\partial L(t(x_k))}{\partial t(x_k)}$$

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

$$\sum_k (\tilde{A}_{.k} + \tilde{B}_{.k}) \left|\left| \frac{\sum_j (\tilde{p}_j \tilde{A}_{jk} + \tilde{q}_k \tilde{B}_{jk}) y_j}{\tilde{A}_{.k} + \tilde{B}_{.k}} - x_k - t(x_k) \right|\right|^2$$
$$+\alpha^{'} R(t) \tag{7}$$

Thus, it is equivalent to minimise the expressions 6 and 7 with respect to t. However, the first minimisation consists in solving a system of size $\mathcal{O}(\text{card}(X) \times \text{card}(Y))$ whereas the second consists in solving a system of size $\mathcal{O}(\text{card}(X))$. Then the overall algorithm to minimise $\mathcal{E}2$ can be expressed as:

Algo Reg3: Symmetric robust EM-ICP

E-step:
compute $\tilde{A}$; $\tilde{B}$; $(\tilde{p}_j)$; $(\tilde{q}_k)$ (using kd-trees)
$\forall x_k \in X$,
  compute $\tilde{A}_{.k} + \tilde{B}_{.k}$,
  compute $\tilde{y}_k = \sum_j (\tilde{p}_j \tilde{A}_{jk} + \tilde{q}_k \tilde{B}_{jk}) y_j / (\tilde{A}_{.k} + \tilde{B}_{.k})$
M-step: solve the approximation problem:
$\arg\min_T \sum_k (\tilde{A}_{.k} + \tilde{B}_{.k}) ||\tilde{y}_k - x_k - t(x_k)||^2 + \alpha^{'} L(t)$

5 Adding priors

The computation of the match matrices A and B is essentially based on the spatial distance between the points $x_k + t(x_k)$ 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 deformation 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 (*e.g.* [14]).

On the other hand, a more generic and simple 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 deformation t. By specifying relevant priors π_{jk}s, we introduce additional information on matches allowing to compute reliable posteriors even for a bad initial estimate of the deformation.

5.1 Designing π

In practice, we choose to design $\pi = (\pi_{jk})$ such that $\pi_{jk} \propto \exp(-\beta c(y_j, x_k))$ where $c : Y \times X \rightarrow \mathbb{R}^+$ 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 $||y_j - x_k - t(x_k)||$ during the E-step.

The equivalent criterion is:

$$\begin{aligned} & \mathcal{E}3(Y, T(X), A, B) = \\ & \sum_{j,k} (A_{jk} + B_{jk}) \rho_\delta(||y_j - x_k - t(x_k)||^2 + \beta c(y_j, x_k)) \\ + \quad & 2\sigma^2 \sum_{j,k} A_{jk} \log(A_{jk}) + 2\sigma^2 \sum_{j,k} B_{jk} \log(B_{jk}) + \alpha' R(t) \end{aligned} \tag{8}$$

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 two approaches to build c.

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 view allows to use pairs of landmarks in the registration process. One could also use crest lines extracted from both point sets [22] as they constitute salient features. Each point is given a label depending on whether it belongs to a crest

line or not. Then, we define $c_{crest}(y_j, x_k) = 0$ if y_j and x_k have the same label and $c_{crest}(y_j, x_k) = penalty$ else.

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 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:

- The shape index $d(x) = sh(x)$ [27] 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 fit 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 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 neighbors 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 $d(x) = cu(x)$ [27] that specifies the amount of curvature and that is invariant to rigid-body transformations. We compute it using the same techniques that we used for shape index:
$cu(x) = \sqrt{(a+b+c)^2 + (a-b)^2}$.

- The (normalised) total geodesic distance $d(x) = tgd(x)$ [1] 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 set and the Dijkstra's algorithm.

5.2 Implementation

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

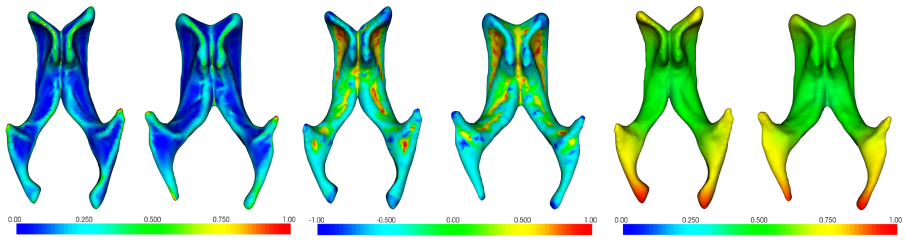


Figure 2: **Mapping of descriptor values on two different brain ventricles:** From left to right: curvedness, shape index and total geodesic distance on two lateral ventricles. Homologous anatomical structures yield qualitatively the same descriptor values.

Algo Reg4: Symmetric robust EM-ICP with priors

E-step:
initialise A and B to the null matrix
for each $x_k \in X$;
  $\mathcal{S} = \{y_j \in Y \text{ such that } ||y_j - x_k - \tilde{t}(x_k)||^2 < \delta\}$
(using a kd-tree)
  for each $y_j \in S$;
    if $||y_j - x_k - \tilde{t}(x_k)||^2/(2\sigma^2) + \beta c(y_j, x_k) \leq \delta$
    $A_{jk} = \exp(-(||y_j - \tilde{T}(x_k)||^2/(2\sigma^2) + \beta c(y_j, x_k)))$
  for each $y_j \notin S$; A_{jk} is left equal to 0.
$B = A$
normalise A in rows and B in columns
and compute $(\tilde{p}_j)$ and $(\tilde{q}_k)$ (see Section 3)
for each $x_k \in X$, compute $\tilde{A}_{.k} + \tilde{B}_{.k}$ and $\tilde{y}_k$ (see Section 4.1)
  compute $\tilde{y}_k = \sum_j (\tilde{p}_j \tilde{A}_{jk} + \tilde{q}_k \tilde{B}_{jk}) y_j / (\tilde{A}_{.k} + \tilde{B}_{.k})$
M-step: solve the approximation problem:
$\arg\min_t \sum_k (\tilde{A}_{.k} + \tilde{B}_{.k}) ||\tilde{y}_k - x_k - t(x_k)||^2 + \alpha' R(t)$

6 Solving the M-step

Solving the M-step is highly conditioned by the choice of the regulariser R. Several regularisers have been proposed in the literature and the most commonly used is probably the Thin Plate Spline (TPS) regulariser. It 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 also implies computational and memory usage that limit its application to point sets of small size. In this section, we focus on building a tractable (in terms of minimisation strategy) and powerful (in terms of reliability of the model) regulariser R.

To alleviate complicated notations, we restate the M-step (Reg2 and Reg3) in the following general form:

$$\tilde{t} = \arg\min_t \sum_k C_k ||\tilde{y}_k - x_k - t(x_k)||^2 + \alpha' R(t), \tag{9}$$

where $C_k = \tilde{A}_{.,k} + \tilde{B}_{.,k}$.

In this section, we devise a new solution for the M-step based on the Reproducing Kernel Hilbert Space (RKHS) theory and the Fourier analysis. More specifically, we design $R(t)$ as a function of the frequencies of t (allowing to tune it efficiently) and 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

We consider that t belongs to a space of admissible solutions $\mathcal{H}$ that we span using a positive definite kernel (pdk) k on $\Omega \subset I\!R^3$ ($i.e\, \forall (q_0, \ldots, q_N) \in \Omega^N, \forall (\alpha_0, \ldots, \alpha_N) \in I\!R^+, \sum_{i,j} \alpha_i \alpha_j k(q_i, q_j) \geq 0$):

$$\mathcal{H} = \{f | f(.) = \sum_{i=0}^{\infty} k(q_i, .) w_i, w_i \in I\!R^3, q_i \in \Omega; ||f||_{\mathcal{H}} < \infty\}^C$$

where $\mathcal{S}^C$ denotes the completion of the set $\mathcal{S}$ and where $\Omega \subset I\!R^3$. This space is endowed with the inner product: $< f, h >_{\mathcal{H}} = \sum_{i,j=0}^{\infty} w_i^T k(q_i, q_j) w_j$. Under these conditions, the space $\mathcal{H}$ is a Hilbert space with reproducing kernel k (or more compactly a RKHS) [2]. Then we assume that $t \in \mathcal{H}$ and define our regulariser $R(t)$ as $||t||_{\mathcal{H}}$:

$$\tilde{t} = \arg\min_{f \in \mathcal{H}} \sum_k C_k ||\tilde{y}_k - x_k - f(x_k)||^2 + \alpha^{'} ||f||_{\mathcal{H}}. \tag{10}$$

One of the key advantage of RKHS is that one can show [39] that the values taken by the solution $\tilde{t}$ can be expressed as $\tilde{t}(x) = \sum_{i=1}^{N} k(x_i, x) w_i$. As a consequence, one can rewrite problem 10 as:

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

$$+\alpha^{'} \sum_{i,j=1,..,N} w_i^T k(x_j, x_i) w_j.$$

Pragmatically, we replaced the estimation of a function $\tilde{t}$ belonging to a space of functions defined all over the space by the estimation of $3 \times \text{card}(X)$ scalar values.

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

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

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 $M \times M$ diagonal matrix formed by the C_k values. Now, the challenge is to choose a kernel corresponding to a relevant regulariser.

6.2 Choosing a kernel

In order to design a suitable kernel k, one can use an interesting relationship with Fourier-based regularisers. Let $\forall f$ square-integrable,

$$R(f = (f_1, f_2, f_3)^T) = R(f_1) + R(f_2) + R(f_3)$$

where for $i = 1, 2, 3$,

$$R(f_i) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} |f_i^*(\omega)|^2 \phi^*(||\omega||/b)^{-1} d\omega,$$

where * is the Fourier transform operator, $\phi : I\!R \rightarrow I\!R$ is an integrable function and b is a real positive rescaling factor. Intuitively, the function R integrates the spectral densities of f_i over the frequencies ω weighted by a factor $\phi^*(||\omega||/b)^{-1}$ (itself depending on ω).

Let $\mathcal{F} = \{f : I\!R^3 \rightarrow I\!R^3 | R(f) < \infty\}$. Interestingly, one can state that if the function $(q_i, q_j) \mapsto \phi(||q_i - q_j||)$ is a pdk then $\mathcal{F}$ is a RKHS whose reproducing kernel is given by $k(q_i, q_j) = b \times \phi(b \times ||q_i - q_j||)$ and such that $\forall f \in \mathcal{F}$, $||f||_{\mathcal{F}} = R(f)$ [54, 44]. This dual view is convenient as it allows one to design a wide variety of efficient regularisers directly into the Fourier domain.

In order to design an efficient regulariser, one must choose ϕ^{*-1} as a high-pass filter. This way, high frequencies of the deformation will be drastically penalised whereas low frequencies will only be penalised a little. Thus, a natural choice consists in choosing $\phi^*_{[0,\infty]}$ as a monotonically decreasing function. 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 $\phi^*(||\omega||/b)$ is null are forbidden. The two parameters α' and b allow to handle the regularisation properties: α' 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). More precisely, b can be seen as a parameter contracting (resp. dilating) the kernel function $\phi^*(.)$ and thus decreasing (resp. increasing) the range of admissible frequencies. Figure 3 shows the influence of b and α' when approximating a noisy 2D field when choosing ϕ as the Wu kernel [54].

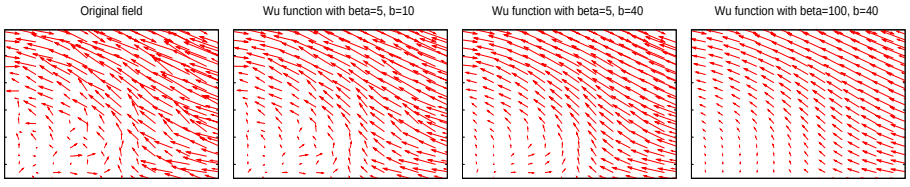


Figure 3: **Effect of parameters α' and b on the approximation of a noisy field.**

6.3 Why using frequencies?

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 high frequences penalisation. Thus, tuning the parameter b allows to introduce a multiscale approach by first trying to capture a global deformation and then, if needed, local deviations from this global deformation. This view allows one to adapt our algorithm to different applications needing either a fine registration (*e.g.* automatic labeling of substructures) or a more global registration (*e.g.* statistical shape analysis) of two structures. Figure 4 illustrates the influence of different kernels and different scale parameters b on the regularisers. Figure 5 illustrates that modifying $\alpha^{'}$ allows one to characterise different scales of deformations linking two structures.

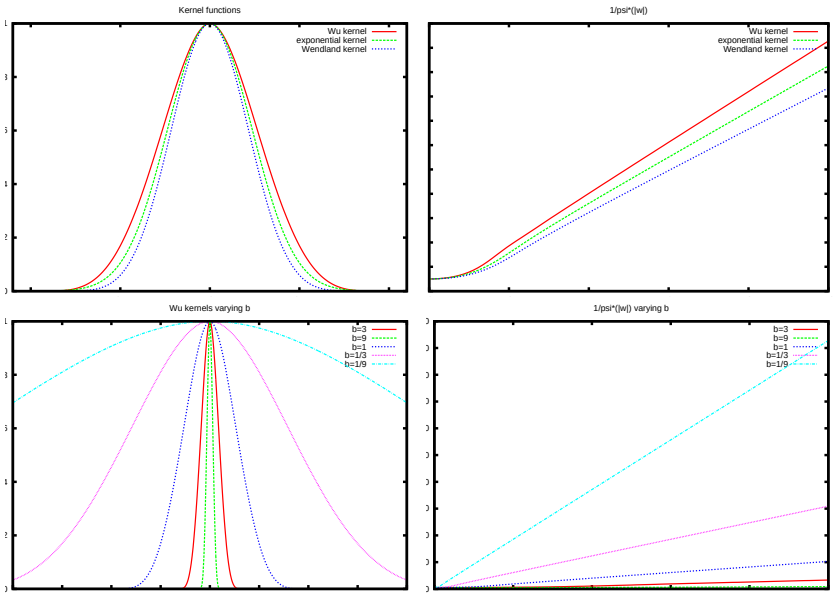


Figure 4: **Different positive definite kernels ϕ and their associated functions $\phi^*(||.||/b)^{-1}$.** From left to right and top to bottom: i) 2D plot of three kernels (b=1), ii) 2D plot of the values of $\phi^*(||.||)^{-1}$ for the three above-mentioned kernels, iii) 2D plot of $\phi(||.|| \times b)$ for the Wu kernel with different b values, iv) 2D plot of $\phi^*(||\omega||/b)^{-1}$ for the Wu kernel with different b values.

6.4 Efficient choices

Although we proposed a closed-form solution for the approximation problem, it consists in solving a $M \times M$ system of equations. This can be problematic in terms of memory usage and of computational time when M increases. Suppose that we choose a compactly supported pdk (*i.e.* there exists $r > 0$ such that, $\forall x, y$ such that $||y-x|| > r$; $k(x, y) = 0$), then i) $D(P)K+\alpha^{'}I$ is a sparse matrix that can be computed using a kd-tree and ii) computing W consists in solving a sparse system. Some compactly supported pdk corresponding to low-pass filters have been proposed in the literature (such as Wendland, Wu or Buhmann pdks). Moreover, techniques to generate a wide variety of them have been proposed [54]. Alternatively one can use a highly decreasing function and approximate it by zeroying all its values over a given threshold. We experimentally found the

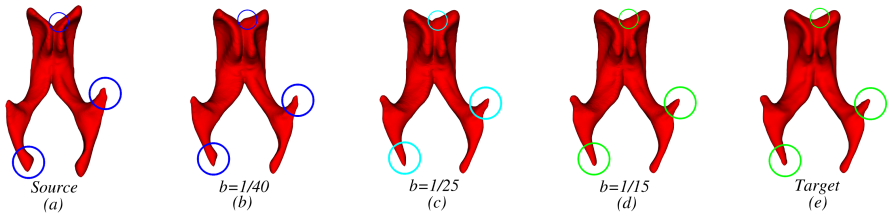


Figure 5: **Influence of b on the registration.** We register the source ventricle (a) on the target ventricle (e) by increasing b throughout the iterations of the overall EM-ICP algorithm (without modifying the other parameters). Intermediate registration results for a given b are represented between the source and the target (b), (c) and (d). For a small 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 increases, the small discriminative patterns tend to fit the target (green circles). Light blue circles indicate intermediate configurations of the patterns.

compact support pdk of Wu ($\phi_{2,3}$) as the one providing the best results and we will use it in the following. Note that the top-right plot in Figure 5 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.

6.5 M-step in a nutshell

M-step:
for each $x_k \in X$
  $\mathcal{S}_k = \{x_i \in X \text{ such that } ||x_k - x_i||^2 < b\}$
(using a kd-tree)
  for each x_i in $\mathcal{S}_k$
    $K(k,i) = b \times \phi(b \times ||x_k - x_i||) \times C_k$
  $K(k,k) = K(k,k) + \alpha'$
preconditioning of K (using sparse algebra)
solve $KW^1 = D(P)[X^1 - Y^1]$ (using sparse algebra)
solve $KW^2 = D(P)[X^2 - Y^2]$ (using sparse algebra)
solve $KW^3 = D(P)[X^3 - Y^3]$ (using sparse algebra)

where X^1, X^2 and X^3 are respectively the vectors extracted from the first, second and third columns of matrix X (the same for Y and W).

6.6 Note on vectorial kernels

By building a scalar kernel k, we consider each component of the deformation field independently ($R(t) = R(t_1) + R(t_2) + R(t_3)$). In fact, the previous results stay true when designing vectorial positive definite kernels. In our case, the evaluation $k(.,.)$ would take its values in the set of the 3 by 3 matrices, the generated space would be such that $\mathcal{H} = \{f | f(.) = \sum_{i=0}^{\infty} k(q_i,.)w_i, w_i \in I\!R^3, q_i \in \Omega; ||f||_{\mathcal{H}} < \infty\}^C$ and the dot product would write $< f, h >_{\mathcal{H}} =$

$\sum_{i,j=0}^{\infty} w_i^T k(q_i, q_j) w_j$. This way, the corresponding Fourier-based stabiliser would write:

$$R(f) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \overline{f^*(\omega)} \phi^*(\omega/b)^{-1} f^*(\omega) d\omega$$

With this formulation the cross frequencies (and indirectly cross derivatives) can be penalised. We restrict our study to scalar kernels.

7 Related algorithms

Several recent algorithms are close to four algorithms (Reg 1,2,3,4) we proposed. Especially, TPS-RPM [10] and CPD algorithms [33] are theoretically very close to ours. 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 function and (iv) the resulting implementation choices.

In this section, we briefly present these two algorithms, their advantages, limitations and compare them to our method. More quantitative results will be given in Section 8.

7.1 The TPS-RPM algorithm

In the TPS-RPM algorithm [10], 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 a "virtual" centroid located at the barycentre of X that is associated to a large variance. The points of X having no satisfying correspondence in Y will have a high probability of correspondence with this virtual centroid. As mentioned by the authors themselves, this solution is very debatable. Indeed, it both supposes that matching a point with the centroid of the point set does not affect the estimation of the transformation and that outliers are more probably located close to the barycentre of the point set. The regulariser R amounts to solve the classical TPS regulariser for which the M-step consists of solving 3 dense systems of equations of size $(\text{card}(X) \times \text{card}(Y))^2$. This problem is heuristically reduced to 3 dense systems of equations of size $\text{card}(X)^2$ (which is still computationally demanding). The overall resulting algorithm is limited to small point sets (typically a few hundreds).

7.2 The CPD algorithm

In the CPD algorithm [33], the match matrix is dealt asymmetrical (*i.e.* simply row stochastic). The robustness is achieved by modeling 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 fixing the number of expected outliers (that is generally unknown). Note that the parameter σ is estimated at the beginning of each iteration. R is the coherent point drift (CPD) 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 different matrices involved

in the computations are evaluated efficiently using the Fast Gauss transform [21] (that does not allow one 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 structure of X (that is fixed throughout the iterations of the algorithm).

In essence, one notices that for both TPS-RPM and CPD algorithms, the implementation choices are performed using heuristics/approximations without taking care of their real effects on the minimisation scheme. In particular one can wonder how the approximations performed in the M-step change the nature of the regularisation.

7.3 Our method (Reg4)

As opposed to TPS-RPM and CPD algorithms, we do not use any heuristic neither in the E-step nor in the M-step and the criterion $\mathcal{E}3$ is properly minimised (no approximation is made). The estimation of the correspondences is dealt symmetrically and the rejection of outliers is performed using a robust cost function parametrised by a single cut-off distance parameter. The E-step is implemented efficiently using kd-trees. The M-step is reformulated as an approximation problem of size N that is solved efficiently using sparse linear algebra and a kd-tree.

Table 1 sums-up the characteristics of the three above-mentioned algorithms.

	efficiency	exact/approx	robustness	symmetric	minimisation	non-linear metrics
TPS-RPM	**intractable**	**approx**	**distance to centroid**	symmetric	EM-like	easy to incorporate
CPD	very efficient	**approx**	% of outliers	**asymmetric**	EM-like	**no**
Reg4	efficient	exact	cut-off	symmetric	EM-like	yes

Table 1: Summarising the characteristics of TPS-RPM, CPD and Reg4.

8 Validation & Results

In this section, we assess the different improvements we proposed in this paper and compare their performances with that of the recent CPD algorithm. Note that, due too high memory usage and time complexity, we did not manage to register any of our data with Reg1 and with the TPS-RPM algorithm. For Reg4, the cost function c is built as: $c(y_j, x_k) = |sh(x_k) - sh(y_j)| + |cu(x_k) - cu(y_j)| + |tgd(x_k) - tgd(y_j)|$ (we do not use labels).

8.1 Validation

8.1.1 Ground truth data

We propose to register caudate nuclei (1000 points, Figure 6) and brain ventricles (7000 points, Figure 7) and bony labyrinths (8000 points, Figure 7). Notice that each of these structures has not the same geometrical complexity. We generate pairs of ground truth data by deforming a given point set using:

- randomly generated one local smooth deformations of the type: $t(x) = x + K \times G_v(x - x_c)n_x$, 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.

- a randomly generated non-linear transformations using the widely used TPS and CPD deformations,

- random removal of given quantities of adjacent points (that we fix to $0.1 \times N$).

Figures 6 and 7 illustrate the generation of pairs of ground truth data. Then, we register the original and deformed point sets and compute the overall residual distance between the known correspondences between the two point sets:

$$endPt(Y, T(X)) = \frac{1}{\text{card}(X)} \sum_{k \in [1,\ldots,\text{card}(X)]} ||y_k - x_k - t(x_k)||^2$$

and the overall Barron error between the transformation fields:

$$barron(t_1, t_2) = \frac{1}{\text{card}(X)} \sum_{k \in [1,\ldots,\text{card}(X)]} \arccos\left(\frac{(t_1(x_k))^T}{||t_1(x_k)||} \frac{t_2(x_k)}{||t_2(x_k)||}\right)$$

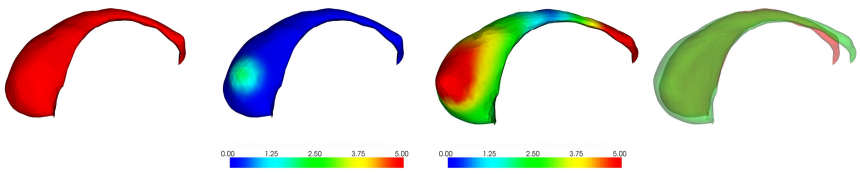


Figure 6: **Illustration of the generation of the ground truth data.** From left to right: i) original data, ii) we generate a random local deformation (the resulting distances between the corresponding points is mapped), iii) we generate a random global deformation (TPS) (the resulting distances between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

8.1.2 Results

To investigate the added value of different improvements we proposed in this article, we consider the average scores obtained by the algorithms Reg2 (Section 3.2), Reg3 (Section 4) and Reg4 (Section 5) over the registration of 100 pairs of ground truth data for each of the three data sets. Moreover, we perform the same experiments using the CPD algorithm. Table 2 indicates the mean and maximal end-point and Barron errors obtained for each of these 4 algorithms and for the different datasets. Table 3 indicates the mean run-time for each

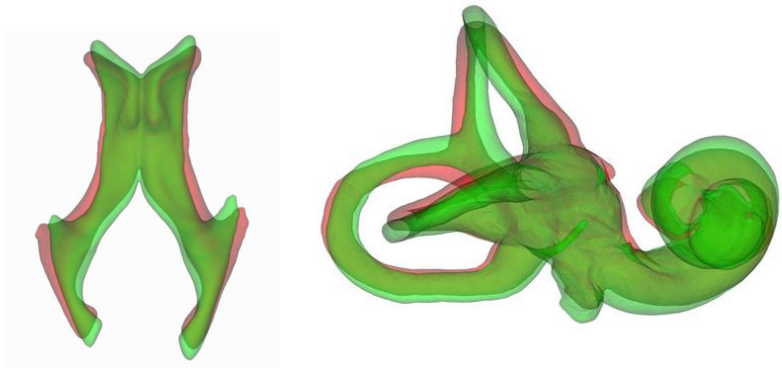


Figure 7: **Examples of pairs of ground truth data:** From left to right ventricles and osseous labyrinths

algorithm and for the different datasets. In the following, we summarise these results and give illustrating examples.

• Influence of the symmetrisation of the correspondences (Reg2 vs Reg3):

We point out two main conclusions when comparing results from Reg2 and Reg3. **First**, symmetrising the correspondences decreases the registration error significantly (in average from a factor of about 2.8 for the end-point error). In practice, Reg2 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 Reg3. This is illustrated in Figures 8 and 9. **Second**, one observes that for Reg2 the location of the registration errors depends on what point set is used as the template. For Reg3, this effect is considerably reduced. This is illustrated in Figure 10. Notice that there is a substantial difference of run-time between Reg2 and Reg3 which is mainly due to $N \times M$ nearest points searchings in each E-step for Reg3 instead of N for Reg2.

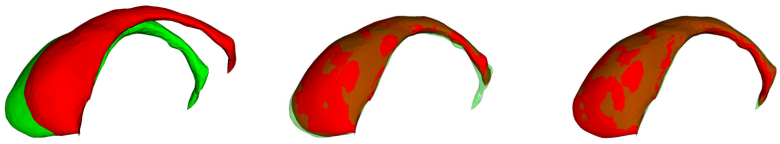


Figure 8: **Effect of the symmetrisation :** From left to right : i) initial alignment of two caudate nuclei, ii) alignment obtained using Reg2 and iii) alignment obtained using Reg3. The asymmetric formulation leads to registration errors close to the extrema of the head and of the tail.

• Influence of the priors (Reg3 vs Reg4):

In average, Reg3 gives results that are quite close than Reg2. However, one can notices that it decreases the value of the worse result importantly: by

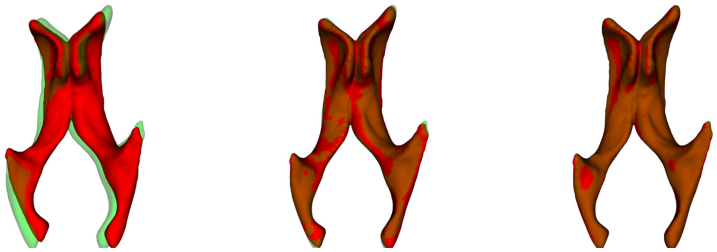


Figure 9: **Effect of the symmetrisation:** From left to right : i) initial alignment of two brain ventricles, ii) alignment obtained using Reg2 and iii) alignment obtained using Reg3. The asymmetric formulation leads to registration errors close to the horns.

forcing correspondences (independently of the initial estimate of T) it leads the registration even for a bad initial alignment. In particular, one observes that the adding of priors does not improve the results when the initial alignment is of good qualities.

• Comparison with a state-of-the-art method (Reg2, Reg3 and Reg4 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 Reg2. This observation is very interesting as it indicates that CPD, that does not deal with the correspondence symmetrically, manages to tackle this major flaw 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 Reg4 and Reg3. We notice that this effect is much larger for the Barron error, which can be due to the approximation performed during the M-step of the CPD algorithm. Figures 11 and 12 give illustrative examples of the registration error obtained with Reg4 and CPD.

	caudate		ventricle		ear	
	mean/max end-pt	mean/max Barron	mean/max end-pt	mean/max Barron	mean/max end-pt	mean/max Barron
Reg2	0.49/2.43	13.65/32.90	0.78/3.29	28.21/37.76	0.23/0.48	5.89/8.11
Reg3	0.22/1.14	3.60/18.69	0.20/0.67	1.78/9.65	0.09/0.17	3.22/6.14
Reg4	0.21/0.89	3.65/12.75	0.20/0.51	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 for different methods and different datasets

	caudate	ventricle	ear
Reg2	26s	11min	9min
Reg3	40s	16min	11min
Reg4	47s	18min	14min
CPD	35s	13min	6min

Table 3: Mean execution time for different methods and different datasets on a standard personal computer.

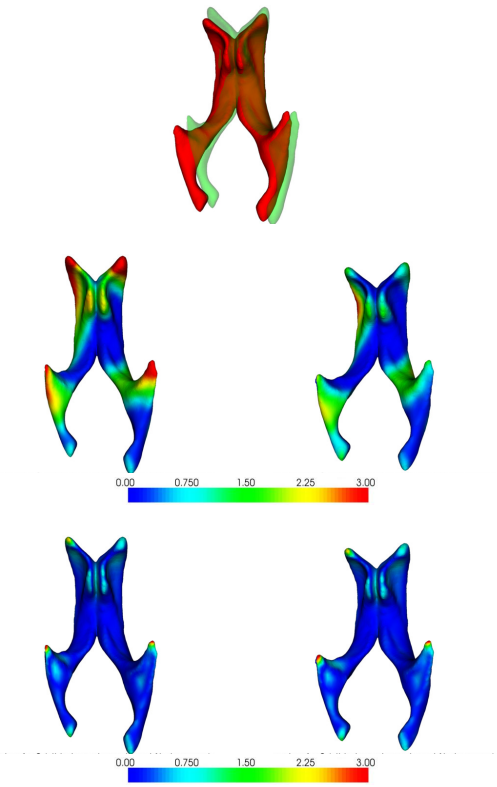


Figure 10: **Effect of the symmetrisation :** From left to right and top to bottom: i) two misaligned pairs of lateral ventricles A and B; ii) residual point to point error when registering A to B and iii) B to A with Reg2; iv) residual end-point errors when registering A to B and v) B to A with Reg3. The asymmetric formulation leads to registration errors close to the horns. The location of these errors depends on what surface is used as the template (second row). When symmetrising the correspondence, the order-dependent registration error is reduced and the overall registration quality is visually and quantitatively improved (third row).

9 Conclusions and perspectives

9.1 Conclusions

After briefly presenting a comprehensive classification of point sets non-linear registration methods, we proposed to study the limits of the EM-ICP algorithm. 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 of the EM-ICP algorithm, we proposed a new algorithm for the non-linear registration of large point sets. More specifically, we considered three drawbacks of the original algorithm and we proposed efficient and original solutions to handle them. In particular, we reduced the asymmetry of the matching process, we proposed to add alignment-invariant

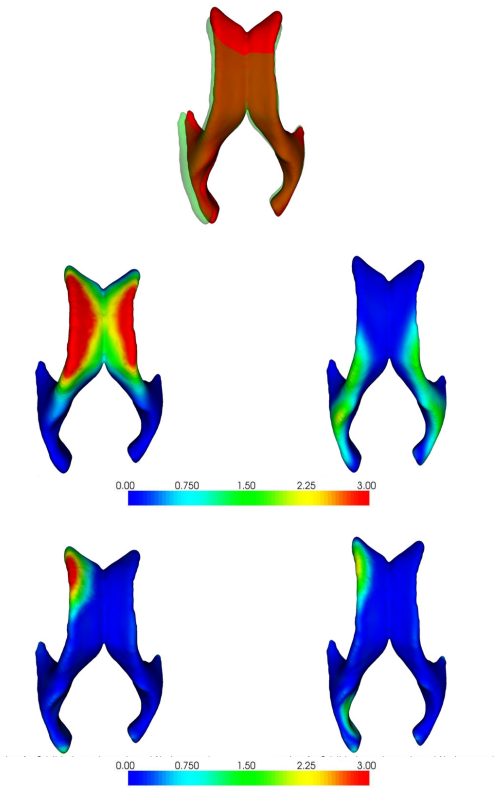


Figure 11: **Reg4 vs CPD :** From left to right and top to bottom: i) two misaligned pairs of lateral ventricles A and B (back view); ii) residual point to point error when registering A to B and iii) B to A with the CPD; iv) residual point to point error when registering A to B and v) B to A with Reg4. For CPD, the location of the errors depends on what surface is used as the template. Moreover the overall error is higher for CPD than for Reg4.

information into the criterion to improve the capture range and proposed a regulariser leading to efficient solutions for implementing the M-step. We investigated the added value of these modifications on synthetical data. Particularly, we observed that i) symmetrising the matching process greatly decreases the registration error and makes the result of the registration less dependent from the choice of the source/target point sets and ii) adding priors reduces the registration errors when the initial alignment is bad. Then, we showed that the complete algorithm outperforms the state-of-the-art CPD algorithm.

9.2 Perspectives

Our methodology relies on modelling point sets as mixtures of pdf. An interesting work would consist in trying to fit mixture models directly on the point sets (or on the surfaces) by estimating both means and (anisotropic) variances of a sum of Gaussian laws. Then the registration could be done on the fitted models. We think that such an approach will provide less sampling-dependent

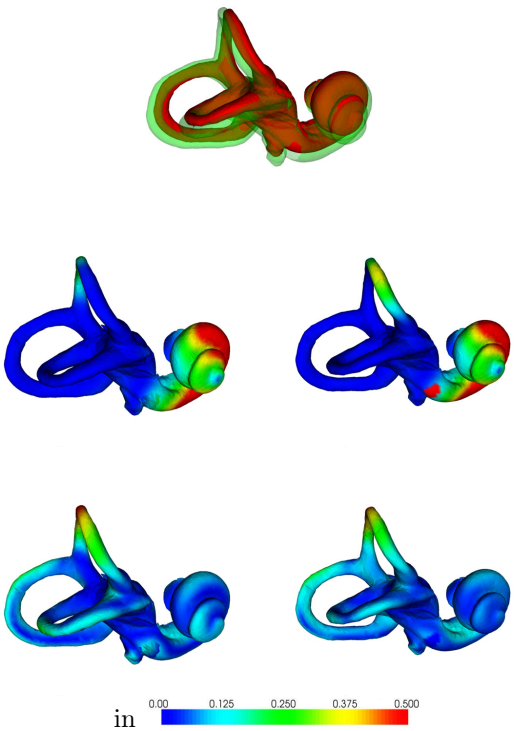


Figure 12: **Reg4 vs CPD :** Same experiments as Figure 11, with bony labyrinths. The conclusions are the same as for Figure 11.

results.

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 will be of great interest.

References

[1] D. Aouada, S. Feng, and H. Krim. Statistical Analysis of the Global Geodesic Function for 3D Object Classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing.*, volume 1, pages 645–648, 2007.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, May 1950.

[3] A. Basu, I.R. Harri, N. L. Hjort, and M.C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.

[4] J.L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[5] P.J. Besl and N.D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, December 1992.

[6] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 11(6):567–585, june 1989.

[7] V. Camion and L. Younes. Geodesic interpolating splines. In *International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 513–527, 2001.

[8] M. Carcassoni and E. R. Hancock. Spectral correspondence for deformed point-set matching. In *International Workshop on Articulated Motion and Deformable Objects*, pages 120–132, 2000.

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

[10] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 44–51, 2000.

[11] N.E. Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56(3):463–474, 1969.

[12] A. Dedner, M. Lüthi, T. Albrecht, and T. Vetter. Curvature guided level set registration using adaptive finite elements. In *IEEE Conference on Pattern recognition*, pages 527–536, 2007.

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

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

[15] F.Wang, B.C. Vemuri, A. Rangarajan, and S. J. Eisenschenk. Simultaneous nonrigid registration of multiple point sets and atlas construction. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 30(11):2011–2022, 2008.

[16] G. Celeux and G. Govaert. A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics and Data Analysis*, 14(3):315–332, 1992.

[17] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(3):367–383, 1992.

[18] J. Glaunes, A. Trouvé, and L. Younes. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In *IEEE Conference on computer vision and pattern recognition*, pages 712–718, 2004.

[19] S. Granger and X. Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *European Conference on Computer Vision*, volume 2353, pages 418–432, 2002.

[20] P.J. Green. On Use of the EM for Penalized Likelihood Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):443–452, 1990.

[21] L. Greengard and J. Strain. The fast gauss transform. *SIAM Journal of Sci. Stat. Comput.*, 12(1):79–94, 1991.

[22] S. Gumhold, X. Wang, and R. Macleod. Feature Extraction from Point Clouds. In *In Proceedings of the 10th International Meshing Roundtable*, pages 293–305, 2001.

[23] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang. Rigid and articulated point registration with expectation conditional maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):587–602, arch 2010.

[24] Q. Huang, B. Adams, M. Wicke, and L.J. Guibas. Non-rigid registration under isometric deformations. *Computer Graphics Forum*, 27(5):1449–1457, 2008.

[25] B. Jian and B.C. Vemuri. A robust algorithm for point set registration using mixture of gaussians. In *IEEE International Conference on Computer Vision*, volume 2, pages 1246–1251, October 2005.

[26] S.C. Joshi and M.I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, august 2000.

[27] J.J Koenderink and A.J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–564, October 1992.

[28] M. Lüthi, T. Albrecht, and T. Vetter. Curvature guided surface registration using level sets. In *Congress on Computer Assisted Radiology and Surgery*, 2007.

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

[30] F. H. C. Marriott. Separating Mixtures of Normal Distributions. *Biometrics*, 31(3):767–769, 1975.

[31] D. Mateus, R.P. Horaud, D. Knossow, F. Cuzzolin, and E. Boyer. Articulated shape matching using laplacian eigenfunctions and unsupervised point registration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[32] S. Moss and E. Hancock. Image registration with shape mixtures. In *Image Analysis and Processing*, volume 1311, pages 172–179, 1997.

[33] A. Myronenko and X. Song. Point-Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, May 2009.

[34] A. Myronenko, X. Song, and M. Carreira-Perpinan. Non-rigid point set registration: Coherent point drift. In *Advances in Neural Information Processing Systems*, pages 1009–1016, 2007.

[35] A. Rangarajan, H. Chui, E. Mjolsness, S. Pappu, L. Davachi, P. Goldman-Rakic, and J. Duncan. A robust point-matching algorithm for autoradiograph alignment. *Medical Image Analysis*, 1(4):379–398, September 1997.

[36] Steven M. Robbins. *Anatomical standardization of the human brain in euclidean 3-space and on the cortical 2-manifold*. PhD thesis, McGill University, 2004.

[37] A.S. Roy, A. Gopinath, and A. Rangarajan. Deformable density matching for 3D non-rigid registration of shapes. *Medical Image Computind anf Computer Assisted Intervention*, 10:942–9, 2007.

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

[39] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. In *Annual Conference on Computational Learning Theory*, pages 416–426, 2001.

[40] A.J. Scott and M.J. Symons. Clustering Methods Based on Likelihood Ratio Criteria. *Biometrics*, 27(2):387–397, 1971.

[41] G.L. Scott and H.C. Longuet-Higgins. An algorithm for associating the features of two images. *Proceedings. of Biological sciences*, 244:21–26, April 1991.

[42] L.S. Shapiro and J.M. Brady. Feature-based correspondence: an eigenvector approach. *Image and Vision Computing*, 10(5):283 – 288, 1992.

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

[44] T. Sidlofov. Existence and uniqueness of minimization problems with Fourier based stabilizers. In *Compstat*, Prague, 2004.

[45] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics*, 35(2):876–879, 1964.

[46] T. Tang and A. Chung. Non-rigid image registration using graph-cuts. In *Medical Image Computing and Computer-Assisted Intervention*, volume 4791, 2007.

[47] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243 – 260, 1998.

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

[49] M. Vaillant and J. Glaunès. Surface matching via currents. In *Proceedings of Information Processing in Medical Imaging*, pages 381–392, 2005.

[50] F. Wang, B. Vemuri, and T. Syeda-Mahmood. Generalized L2-Divergence and its application to shape alignment. In *IEEE Information Processing in Medical Imaging*, pages 227–238, 2009.

[51] F. Wang, B.C. Vemuri, and A. Rangarajan. Groupwise point pattern registration using a novel CDF-based Jensen-Shannon divergence. In *IEEE conference on Computer Vision and Pattern Recognition*, pages 1283–1288, 2006.

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

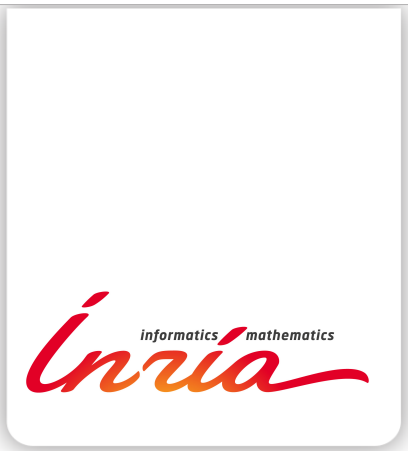
[53] M. William Wells, III. Statistical approaches to feature-based object recognition. *International Journal of Computer Vision*, 21(1-2):63–98, 1997.

[54] H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics, 2005.

[55] B.T.T. Yeo, M.R. Sabuncu, T. Vercauteren, N. Ayache, B. Fischl, and P. Golland. Spherical demons: Fast diffeomorphic landmark-free surface registration. *IEEE Transactions on Medical Imaging*, 29(3):650 –668, march 2010.

[56] A.L. Yuille and N.M. Grzywacz. A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision*, 3(2):155–175, june 1989.

[57] G. Zou, J. Hua, and O. Muzik. Non-rigid surface registration using spherical Thin-Plate splines. In *Medical Image Computing and Computer-Assisted Intervention*, pages 367–374, 2007.



RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399