Defining the Pose of any 3D Rigid Object and an Associated Metric

Romain Brégier · **Frédéric Devernay** · **Laetitia Leyrit** · **James Crowley**

Abstract A pose of a rigid object is usually regarded as a rigid transformation, described by a translation and a rotation. In this article, we define a pose as a distinguishable static state of the considered object, and show that the usual identification of the pose space with the space of rigid transformations is abusive, as it is not adapted to objects with proper symmetries.

Based solely on geometric considerations, we propose a frame-invariant metric on the pose space, valid for any physical object, and requiring no arbitrary tuning. This distance can be evaluated efficiently thanks to a representation of poses within a low dimension Euclidean space, and enables to perform efficient neighborhood queries such as *radius searches* or *k-nearest neighbor searches* within a large set of poses using off-the-shelf methods. We lastly solve the problems of projection from the Euclidean space onto the pose space, and of pose averaging for this metric. The practical value of those theoretical developments is illustrated with an application of pose estimation of instances of a 3D rigid object given an input depth map, via a Mean Shift procedure.

Keywords pose · 3D rigid object · symmetry · distance · metric · average · rotation · $SE(3)$ · object recognition

1 Introduction

The rigid body model plays an important role in many technical and scientific fields, such as physics science, mechanical engineering, computer vision or 3D animation. Under the rigid body assumption, the static state of an object is referred to as a *pose*, and is often described in term of a *position* and an *orientation*.

Poses of a 3D rigid object are in general regarded as rigid transformations and the set of poses is identified with the set of rigid transformations $SE(3)$, the special Euclidean group. The Lie group structure of $SE(3)$ enables to exhibit the relative displacement of the object between two poses, and to define a distance between poses as the length of shortest motion performing this displacement. This identification is therefore particularly meaningful for applications where the motion of a rigid body is considered – such as motion planning (Sucan et al 2012) or object tracking (Tjaden et al 2016).

There are however applications in which motion considerations are irrelevant, and for which only a notion of *similarity* between poses is required. Pose estimation of instances of a rigid object within a scene is a good example, which can be performed e.g. via pose averaging, or modes seeking within a pose distribution in a Hough-like scheme, in order to deal with noisy data.

While motion-based applications typically consider only transitions between two poses and have been the subject of important research work over the years, those based on similarity have to deal with numerous poses at once, performing operations such as neighborhood queries – i.e. finding poses in a set of poses similar to a given one – or pose averaging, and have not gathered as much interest.

As a consequence, similarity measures suffering from major flaws are still used in practical applications for the sake of efficiency. Following the work of Fanelli et al (2011), Tejani et al (2014) base their state-of-the-art object detection and pose estimation method on an Euclidean distance between poses representations based on Euler angles. Such measure is fast to compute and enables the use of efficient tools developed for Euclidean spaces to perform the neighborhood queries required in their modes-seeking procedure,

R. Brégier, L. Leyrit
Siléane, Saint-Étienne, FRANCE
E-mail: r.bregier@sileane.com

R. Brégier, F. Devernay, J. Crowley
Inria Grenoble Rhône-Alpes, FRANCE

but it is not a distance. Euler angles representation notoriously suffers from border effects, singularities, and is dependent on the choice of frames orientation. Those issues may only have limited effects on the results announced by the authors, thanks to an appropriate choice of frames orientation and to the low variability of objects orientations within their datasets. Nonetheless, they cannot be avoided when dealing with the general case of poses having arbitrary orientation. Such an example expresses the lack of tools for dealing efficiently with large sets of poses.

There are moreover numerous cases where the pose of a rigid object cannot be identified to a single rigid transformation and therefore, for which existing results cannot be applied: such cases occur when dealing with objects showing some symmetry properties such as revolution objects or cuboids, and are actually pretty common among manufactured objects.

Given those observations, our goal in this paper is to provide a consistent and general framework for dealing with any kind of physically admissible rigid object in practical applications. To this end, we propose a pose definition valid for any bounded rigid object, through a characterization of its proper symmetries (section 2). We then propose a physically meaningful distance over the pose space (section 4), and show how poses can be represented in a relatively low dimension Euclidean space to enable fast distance computations and neighborhood queries (section 5). We expose how the pose averaging problem can be solved quite efficiently (section 8) for this metric thanks to a projection technique (section 7) and lastly we propose an example of application for the pose estimation of instances of a rigid object given a set of votes.

2 Pose definition

While the notion of pose of a rigid object is widely used, e.g. in robotics or computer vision, we have not found in the literature a general definition. We therefore propose the following one:

Definition 1 A *pose* of a 3D rigid object is a distinguishable static state of this object.

We will refer to the set of possible poses as *pose space* which we will denote $\mathscr{C}$ for consistency with the notion of *configuration space* in robotics literature.

2.1 Link between the pose space and $SE(3)$

The pose space is highly related to the group of rigid transformations $SE(3)$. Let us consider a rigid object, and $\mathscr{P}_0 \in \mathscr{C}$ an arbitrary reference pose of this one.

A rigid transformation applied on the object at its reference pose defines a static state of the object, i.e. a pose. In a similar way, a pose $\mathscr{P} \in \mathscr{C}$ of the object can be reached through a rigid displacement from the reference pose $\mathscr{P}_0$, and therefore $\mathscr{P}$ can be described completely by the rigid transformation corresponding to this displacement.

We will denote $\mathscr{P} \in \mathscr{C}$ and $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in SE(3)$ such a couple of pose and rigid transformation – with $\mathbf{R} \in SO(3)$ a rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ a translation vector. The transformation considered here is such that each point $\mathbf{x} \in \mathbb{R}^3$ linked to an object instance at reference pose $\mathscr{P}_0$ is transformed by $\mathbf{T}$ into the corresponding point $\mathbf{T}(\mathbf{x})$ of an instance at pose $\mathscr{P}$ as follows and such as depicted in figure 1:

$$\mathbf{T}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{1}$$

However, for a given pose, the corresponding rigid transformation is not necessarily unique and therefore the identification of $SE(3)$ with the pose space is in the general case incorrect. Objects — and especially manufactured ones — may indeed show some proper symmetry properties that make them invariant to some rigid motions.

2.2 Pose as equivalence class of $SE(3)$

Let $M \subset SE(3)$ be the set of rigid transformations representing the same pose as a rigid transformation $\mathbf{T}$. The set of relative rigid transformations $G = \{\mathbf{T}^{-1} \circ \mathbf{M}, \mathbf{M} \in M\}$ is by definition the set of rigid transformations that have no effect on the static state of the object, and this set therefore does not depend on the arbitrary transformation $\mathbf{T}$. It is moreover a subgroup of $SE(3)$. Indeed, combinations and inversions of such transformations can be applied to the object while still letting it unchanged, and the identity transformation has obviously no effects on the pose of the object. We will refer to the elements of this group as the *proper symmetries* of the object and to $G \subset SE(3)$ as the group of *proper symmetries* of the object.

Given a rigid transformation $\mathbf{T}$ defining a pose $\mathscr{P}$, we can therefore identify $\mathscr{P}$ to the following equivalence class $[\mathbf{T}] \subset SE(3)$, consisting in the combination of $\mathbf{T}$ with any rigid transformation that has no effect on the pose of the object:

$$\mathscr{P} = [\mathbf{T}] \triangleq \{\mathbf{T} \circ \mathbf{G}, \mathbf{G} \in G\}. \tag{2}$$

2.3 The group of proper symmetries

We propose here a classification of the potential groups of proper symmetries for a bounded object. This classification will be helpful to derive the practical results associated to our proposed metric. In this article, we consider only the

case of physically meaningful bounded objects. While models of infinite objects are commonly used e.g. for plane detection in 3D scene analysis, they do not correspond to actual physical objects and the definition of a suitable metric on the pose space of such objects would be very dependent on the application. As a consequence, all proper symmetries of a considered object necessarily have a common fixed point, and by choosing such a point as the origin of the object frame, we can consider the group of proper symmetries as a subgroup of the rotation group $SO(3)$.

Subgroups of $SO(3)$ are sometimes referred to as chiral point groups, and have been widely studied, notably in the context of crystallography. The interested reader is referred to Vainsthein (1994) for more insight on the theory of symmetry.

Ignoring the pathological case of infinite subgroups of $SO(3)$ that are not closed under the usual topology as they do not make sense physically, we will distinguish between five classes of groups of proper symmetries, synthesized in table 1. We consider three cases of infinite groups: spherical symmetry i.e. invariance by any rotation, revolution symmetry i.e. invariance by rotation along a given axis of any angle, which can be decomposed in two classes depending on whether or not the object is also invariant by reflection across a plane orthogonal to the revolution axis – to which we will refer to respectively as revolution symmetry with/without rotoreflection invariance. There are an infinite number of finite groups, therefore we will discuss those in a general manner. We will nevertheless distinguish the case of an object without proper symmetry – i.e., for which the group of proper symmetries G contains only the identity transformation – from the non trivial case as it is essential in our theoretical developments.

Note that potential indirect symmetries of the object – such as reflection symmetries – are not accounted for. This is due to the fact that we consider an oriented 3D space – e.g. through the right-hand rule – in which reflections are not physically feasible through rigid displacements. Revolution symmetry with rotoreflection invariance is nonetheless considered since it is a proper symmetry group: the reflection symmetry can indeed be generated by the introduction of a rotation invariance of 180° along an arbitrary axis orthogonal to the revolution axis.

2D objects can be considered in a similar manner and we distinguish between the cases listed in table 2.

3 Prior work on metrics over the pose space

We propose here a brief review of the recent work on metrics over the pose space of a rigid object. We consider only mathematical distances in our discussion – i.e. symmetric, positive-definite applications from $\mathscr{C} \times \mathscr{C}$ to $\mathbb{R}^+$ satisfying triangle inequality. The existing literature does not take into account potential proper symmetries of the object, and therefore, in this review, the pose space can be identified to the group of rigid transformations $SE(3)$.

3.1 Objectiveness

The identification of the pose space to $SE(3)$ is based on the choice of two arbitrary frames: a frame linked to the object – to which we will refer to as *object frame* – and a fixed *inertial frame* such as the object frame coincides with the inertial frame when the object is in the reference pose $\mathscr{P}_0$. For a distance to be well-defined, it should not depend on an arbitrary choice of those frames, a notion that Lin and Burdick (2000) formalize as *objectiveness* or *frame invariance*.

Among possible distances, geodesic distances have focused most interest and have been studied within the framework of Riemannian geometry on the Lie group $SE(3)$. Geodesic distances are well-suited for applications dealing with motions as they represent the minimum length of a motion to bring the object from one pose to an other. Park (1995) showed that there are no bi-invariant Riemannian metrics on $SE(3)$ – that is, invariant to any change of inertial frame (left invariance) and of object frame (right invariance). Chirikjian (2015) recently studied this question further and showed that while continuous bi-invariant metrics do not exist, there are continuous left-invariant distances that are invariant under right shifts by pure rotations.

3.2 Hyper-rotation approximation

Nonetheless, several authors have worked on an "approximate bi-invariant" metric (Purwar and Ge 2009) for $SE(3)$ through the mapping of rigid transformations to hyper-rotations of $SO(4)$, and the use of a bi-invariant metric on $SO(4)$. Techniques to perform such mapping have been proposed based on biquaternion representation (Etzel and McCarthy 1996) and polar decomposition (Larochelle et al 2007). Such transformation unfortunately requires a scaling for the translation part, which has to be set empirically depending on the application (Angeles 2006).

3.3 Decomposition into translation and rotation

Hopefully, while inertial frame invariance is necessary for the *objectiveness* of a metric, object frame invariance is not. Lin and Burdick (2000) indeed showed that a distance is objective if and only if it is independent of the choice of inertial frame, and transforms by a right shift in response to a change of object frame. Therefore, a method to define an objective

Table 1 Classification of the potential groups of proper symmetries for a 3D bounded physical object

Infinite groups			Finite groups	
Revolution symmetry				
				
Without rotoreflection invariance	With rotoreflection invariance	Spherical symmetry	No proper symmetry	Finite non trivial

Table 2 Classification of the potential groups of proper symmetries for a 2D bounded physical object.

		
No proper symmetry	Cyclic symmetry (finite non trivial)	Circular symmetry

metric consists in defining a left invariant distance considering a given object frame and always using this one, in order to avoid having to transform the distance expression.

When applying this technique, a frequent approach consists in splitting a pose into a position and an orientation part and to define a distance on $SE(3)$ based on frame invariant metrics on both $\mathbb{R}^3$ and the rotation group $SO(3)$. Those metrics can then be fused in the form of a weighted generalized mean, here written with two strictly positive scaling factors a and b and for an exponent $p \in [1,\infty]$:

$$\mathrm{d}(\mathbf{T}_1,\mathbf{T}_2) = \sqrt[p]{a\,\mathrm{d}_{\mathrm{rot}}(\mathbf{R}_1,\mathbf{R}_2)^p + b\,\mathrm{d}_{\mathrm{trans}}(\mathbf{t}_1,\mathbf{t}_2)^p}. \tag{3}$$

The Euclidean distance is the usual choice for measuring distances between different positions. Considering the usual Riemannian distance over $SO(3)$, a Riemannian distance over $SE(3)$ can be obtained by combining those together into (Park 1995):

$$d(\mathbf{T}_1,\mathbf{T}_2) = \sqrt{a\|\log(\mathbf{R}_1^{-1}\mathbf{R}_2)\|^2 + b\|\mathbf{t}_2-\mathbf{t}_1\|^2}. \tag{4}$$

This expression is particularly interesting in that the distance between orientations $\|\log(\mathbf{R}_1^{-1}\mathbf{R}_2)\|$ corresponds to the angle α of the relative rotation between the two, and because it can be evaluated quite easily e.g. from the following relations, using matrix or unit quaternion representations and respectively trace or inner product operators:

$$\mathrm{Tr}(\mathbf{R}_1^{-1}\mathbf{R}_2) = 2\cos(\alpha)+1 \tag{5}$$

$$1-\langle\mathbf{q_1}|\mathbf{q_2}\rangle^2 = \frac{1}{2}(1-\cos(\alpha)). \tag{6}$$

Without the Riemannian constraint, a large number of inertial frame invariant distances can be considered. Gupta (1997) notably proposed to consider a Froebenius distance for the rotation part, which also only depends on the angle of the relative rotation between the two poses:

$$\|\mathbf{R}_2-\mathbf{R}_1\|_F = 4|\sin(\alpha/2)|. \tag{7}$$

Similar properties are obtained considering the Euclidean distance between representations of antipodal pairs of unit quaternions:

$$\min\|\mathbf{q_2}\pm\mathbf{q_1}\| = 2|\sin(\alpha/4)|. \tag{8}$$

Merging position and orientation distances together requires setting the scaling factors a and b. The choice of those factors remains an heuristic issue and depends on the application. Setting the position weight b to 1 and the orientation weight a as the square of the maximum radius of the object (Di Gregorio 2008) in equation (4) can be a reasonable choice when considering an object frame at the center of the object in order to get an upper bound of the displacement of the object's points between two poses.

3.4 Geometric approaches

To avoid the need for arbitrary scaling factors, some distances are based on geometric properties of the object only. A possibility that we consider particularly interesting is to define a metric based on the distance between corresponding 3D points of instances of the object at these poses, such as depicted figure 1. Let μ be a density distribution relative to the object and $V = \int\mu(\mathbf{x})dv$ its integral over the whole object, we can formulate such a distance the following way considering an L^p norm:

$$\mathrm{d}(\mathbf{T}_1,\mathbf{T}_2) = \frac{1}{V}\left(\int\mu(\mathbf{x})\|\mathbf{T}_2\mathbf{x}-\mathbf{T}_1\mathbf{x}\|^p dv\right)^{\frac{1}{p}} \tag{9}$$

This expression has a strong physical meaning. It is by construction frame-invariant since its definition does not depend on a particular frame, and takes the shape of the object into account, without the need of arbitrary tuning. Martinez and Duffy (1995) suggest to consider the maximum

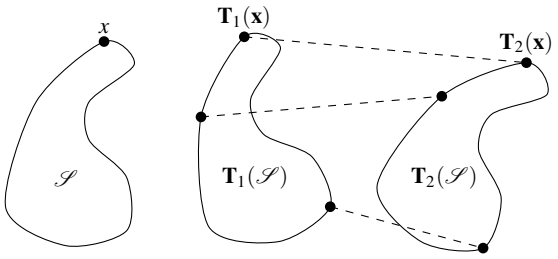


Fig. 1 Representation of corresponding points between instances at different poses of a rigid object without proper symmetries

displacement ($p = \infty$), and Hinterstoisser et al (2013) used the average displacement ($p = 1$) for a pose estimation evaluation. For the sake of tractability, those authors suggest to consider only some vertices of an object model since the integral has to be evaluated explicitly. Kazerounian and Rastegar (1992) on the other hand proposed to consider the integral of square displacements ($p = 2$) upon the whole object, and showed that it could actually be evaluated efficiently given the inertia matrix of the object. Chirikjian and Zhou (1998) later on improved this formulation and extended it for arbitrary affine transformations, showing that the distance could be evaluated as a weighted Froebenius norm.

Zefran and Kumar (1996) and Lin and Burdick (2000) independently proposed a Riemannian tensor being linked to the notion of kinetic energy and therefore taking into account the object properties without the need for some arbitrary tuning. Their tensor can actually be seen as a local equivalent of the distance of Kazerounian and Rastegar (1992). However, to the best of our knowledge there is no known closed-form expression for the resulting geodesic distance in the general case.

3.5 Local metric

Various others local parametrization methods exist which, by mapping locally the pose space to an Euclidean space enable to locally define distances. Those parametrizations are e.g. based on the representation of orientation with Euler angles, or the local stereographic projection of the pose space, identified to Study's quadric – an hypersurface embedded in $\mathbb{R}^7$ (Eberharter and Ravani 2004). In-depth discussion of this topic is out of the scope of this paper as we are interested in global distances.

4 Proposed metric

In this section, we propose a distance over the pose space of a 3D rigid bounded object, which can be considered as an extension of the work of Kazerounian and Rastegar (1992) and Chirikjian and Zhou (1998) to arbitrary bounded objects.

4.1 Formal definition

Let $\mathscr{S}$ be the set of points of the object at reference pose $\mathscr{P}_0 \in \mathscr{C}$, and μ a positive density distribution defined on $\mathscr{S}$. In order to be meaningful, the set of points of the object and its density distribution are assumed to be compatible with the proper symmetry properties of the object and exhibit those symmetry properties. Formally, we assume they verify $\mathbf{G}(\mathscr{S}) = \mathscr{S}$ and $\mu \circ \mathbf{G} = \mu$ for any proper symmetry $\mathbf{G} \in G$.

Definition 2 Let $\mathscr{P}_1, \mathscr{P}_2 \in \mathscr{C}$ be two poses and $\mathbf{T}_1, \mathbf{T}_2 \in SE(3)$ two rigid transformations whose equivalence classes are respectively identified to $\mathscr{P}_1$ and $\mathscr{P}_2$ given the reference pose, such as defined in (2). We define the distance between $\mathscr{P}_1$ and $\mathscr{P}_2$ as follows :

$$\mathrm{d}(\mathscr{P}_1, \mathscr{P}_2) \triangleq \min_{\mathbf{G}_1, \mathbf{G}_2 \in G} \mathrm{d}_{\mathrm{no_sym}}(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2),$$

with

$$\mathrm{d}_{\mathrm{no_sym}}(\mathbf{T}_1, \mathbf{T}_2) \triangleq \sqrt{\frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x}) \|\mathbf{T}_2(\mathbf{x}) - \mathbf{T}_1(\mathbf{x})\|^2 ds} \quad (10)$$

where $S \triangleq \int_{\mathscr{S}} \mu(\mathbf{x}) ds$.

This expression is well defined. The minimum in definition (10) is reached because of the compactness of the proper symmetry group G – as a closed subgroup of $SO(3)$ which itself is compact – and of the continuity of $\mathrm{d}_{\mathrm{no_sym}}$. Moreover, this definition is by construction independent of the choice of the rigid transformations $\mathbf{T}_1, \mathbf{T}_2$ identified to the poses considered. We verify easily that it satisfies the conditions of a distance definition: d is symmetric, positive-definite, and triangle inequality derives from the triangle inequality satisfied by $\mathrm{d}_{\mathrm{no_sym}}$, which is a direct consequence of Minkowski inequality.

4.2 Objectiveness

The proposed distance is by construction independent of the choice of some arbitrary frames as it admits a purely geometric interpretation, a point we discuss in subsection 4.4.

Definition 2 makes indeed no assumption on the choice of an object frame. The use of a reference pose (i.e. an inertial frame) in our formulation is only here for the sake of writability, and to exhibit the symmetry of the roles of the two poses considered. Indeed, the Euclidean distance between 3D points is invariant to isometries by definition of these, and in particular to any rigid transformation $\mathbf{T}_3^{-1} \in SE(3)$:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^3, \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{T}_3^{-1}(\mathbf{x}) - \mathbf{T}_3^{-1}(\mathbf{y})\| \quad (11)$$

Therefore, an arbitrary new reference pose $\mathscr{P}_3$ could be considered without any effect on the metric properties. Denot-

ing $\mathbf{T}_3$ a rigid transformation identified to $\mathscr{P}_3$ relatively to the old reference pose $\mathscr{P}_0$, we verify the independence of $\mathrm{d}_{\mathrm{no_sym}}$ from the choice of reference pose

$$\mathrm{d}_{\mathrm{no_sym}}(\mathbf{T}_1,\mathbf{T}_2) = \mathrm{d}_{\mathrm{no_sym}}(\mathbf{T}_3^{-1}\mathbf{T}_1,\mathbf{T}_3^{-1}\mathbf{T}_2), \tag{12}$$

and hence the independence of the general distance:

$$\mathrm{d}([\mathbf{T}_1],[\mathbf{T}_2]) = \mathrm{d}([\mathbf{T}_3^{-1}\mathbf{T}_1],[\mathbf{T}_3^{-1}\mathbf{T}_2]). \tag{13}$$

4.3 Object points

In typical applications, one is particularly interested in the positioning of the surface of the object. Therefore, in our illustrations, we consider as set of points $\mathscr{S}$ the surface of the object. The density function μ can be used to modulate the importance of the positioning of specific areas, but without additional information it is natural to consider an uniform weight $\mu = 1$.

4.4 Geometric interpretation

A picture being worth a thousand words, the reasoning we develop in this section is illustrated in figure 2 for the case of a 2D object with a rotation symmetry of $2\pi/3$: a flower with three petals.

As we discussed in section 2, a pose $\mathscr{P}_i \in \mathscr{C}$ can be identified to a set of rigid transformations $\{\mathbf{T}_i \circ \mathbf{G}, \mathbf{G} \in G\}$. Each of those transformations can themselves be identified with the pose of an object with identical characteristics to the considered one but with no proper symmetries – to which we will refer to as the equivalent object and which we depict on figure 2 with a grey petal (in order to break the symmetry of the initial object). A pose of the object can therefore be considered as a set of poses of the equivalent object – 3 in our example. Points of the equivalent object can be unambiguously put in correspondence between different poses – correspondences we represent by dashed segments on the figure. Therefore it is legitimate to define a distance between poses of the equivalent object based on the distance between such corresponding points. In this paper, we consider the RMS distance $\mathrm{d}_{\mathrm{no_sym}}$ as it enables efficient computations (see section 5 for more details):

$$\mathrm{d}^2_{\mathrm{no_sym}}(\mathbf{T}_1,\mathbf{T}_2) = \frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\|\mathbf{T}_2(\mathbf{x}) - \mathbf{T}_1(\mathbf{x})\|^2 ds. \tag{14}$$

The proposed distance between two poses of the object can then be defined as the minimum distance between each potential pair of poses for the equivalent object (3×3 combinations in our example):

$$\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{G}_1,\mathbf{G}_2 \in G} \mathrm{d}_{\mathrm{no_sym}}(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2). \tag{15}$$

An other and more intuitive interpretation is to consider our distance as a measure of the smallest displacement from one pose to the other. A displacement from a pose $\mathscr{P}_1$ to an other $\mathscr{P}_2$ is a relative transformation from a pose of the equivalent object corresponding to $\mathscr{P}_1$ to a pose of the equivalent object corresponding to $\mathscr{P}_2$, and the length of a displacement is measured via $\mathrm{d}_{\mathrm{no_sym}}$. Different pairs of poses of the equivalent object are actually linked by the same displacement – as can be observed on figure 2 where pairs of poses of the equivalent object being linked by the same transformation are highlighted by identical boxes. All displacements from one pose $\mathscr{P}_1$ to an other $\mathscr{P}_2$ are in fact considered when choosing an arbitrary pose of the equivalent object $\mathbf{T}_1$ for $\mathscr{P}_1$ and considering rigid transformations from $\mathbf{T}_1$ to the poses of the equivalent object corresponding to $\mathscr{P}_2$. Thanks to this, the distance between two poses can actually be computed considering the proper symmetries only for one pose:

Proposition 1 *For any poses $\mathscr{P}_1, \mathscr{P}_2 \in \mathscr{C}$ and $\mathbf{T}_1, \mathbf{T}_2 \in SE(3)$ two rigid transformations whose equivalency classes are respectively identified to $\mathscr{P}_1$ and $\mathscr{P}_2$ given the reference pose,*

$$\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{G}\in G} \mathrm{d}_{\mathrm{no_sym}}(\mathbf{T}_1, \mathbf{T}_2 \circ \mathbf{G}) \tag{16}$$

This formulation is simpler than the definition 2, however it breaks the symmetry of the roles of the two poses.

Proof Formally, expression (16) can be deduced from the distance definition (10) as follows. Given two proper symmetries $\mathbf{G}_1, \mathbf{G}_2 \in G$, one can perform the change of variables $x \leftarrow \mathbf{G}_1(x)$ and $\mathbf{G} \leftarrow \mathbf{G}_2 \circ \mathbf{G}_1^{-1}$ to write the following equality:

$$\begin{aligned}
&\mathrm{d}^2_{\mathrm{no_sym}}(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2) \\
&= \frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\|\mathbf{T}_2 \circ \mathbf{G}_2(\mathbf{x}) - \mathbf{T}_1 \circ \mathbf{G}_1(\mathbf{x})\|^2 ds \\
&= \frac{1}{S}\int_{\mathbf{G}_1(\mathscr{S})} \mu(\mathbf{G}_1^{-1}(\mathbf{x}))\|\mathbf{T}_2 \circ \mathbf{G}(\mathbf{x}) - \mathbf{T}_1(\mathbf{x})\|^2 ds
\end{aligned} \tag{17}$$

The symmetry of the object pointset and of its density ensures that $\mathbf{G}_1(\mathscr{S}) = \mathscr{S}$ and $\mu \circ \mathbf{G}_1^{-1} = \mu$, leading to the following result from which the conclusion is straightforward:

$$\begin{aligned}
&\mathrm{d}^2_{\mathrm{no_sym}}(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2) \\
&= \frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\|\mathbf{T}_2 \circ \mathbf{G}(\mathbf{x}) - \mathbf{T}_1(\mathbf{x})\|^2 ds \\
&= \mathrm{d}^2_{\mathrm{no_sym}}(\mathbf{T}_1, \mathbf{T}_2 \circ \mathbf{G}).
\end{aligned} \tag{18}$$

□

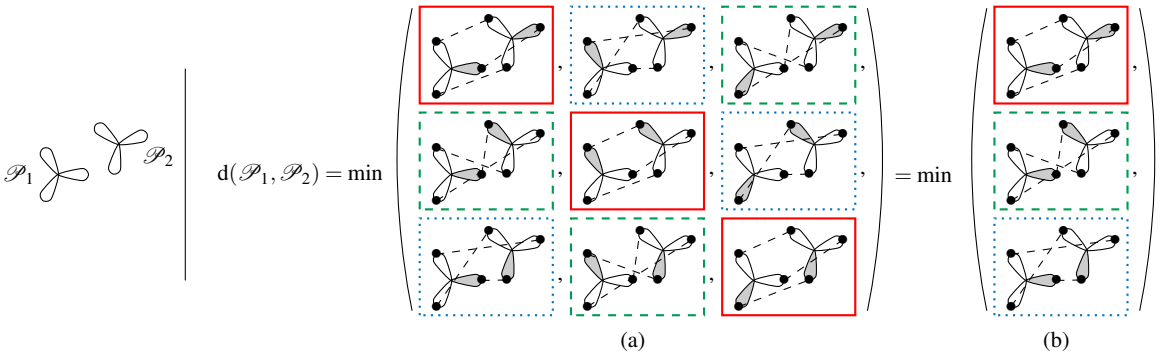


Fig. 2 Illustration of our proposed distance for a 2D object with a rotation symmetry of $2\pi/3$. (a) The distance between two poses is the minimum distance between two poses of an equivalent object without proper symmetry – here there are 3 possible poses of the equivalent object for each pose of the original object. The distance between poses of an object without proper symmetry corresponds to the RMS distance between corresponding object points (dashed segments). (b) Equivalently, the proposed distance can be considered as a measure of the smallest displacement from one pose to an other – here there are actually only 3 different displacements between those two poses (solid, dotted and dashed boxes).

5 Efficient distance computation

The distance definition 2 and the simpler expression of proposition 1 are of little practical use for actual applications as they contain a summation term over the set of points of the object and a minimization over its proper symmetry group, both sets being potentially infinite.

In this section, we show how our proposed distance can be evaluated efficiently. To this aim, we propose a representation of a pose $\mathscr{P}$ as a finite set of points $\mathscr{R}(\mathscr{P})$ of a relatively low dimension Euclidean space $\mathbb{R}^N$. We refer to an element of $\mathscr{R}(\mathscr{P})$ as a *representative* of $\mathscr{P}$, since a representative completely defines a pose (see section 7).

Within this representation framework, the distance between a pair of poses $\mathscr{P}_1, \mathscr{P}_2$ can be expressed as the minimum of Euclidean distance between their respective representatives,

$$\boxed{\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{p}_1\in\mathscr{R}(\mathscr{P}_1),\mathbf{p}_2\in\mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2-\mathbf{p}_1\|} \tag{19}$$

or equivalently, as the minimum Euclidean distance between a given representative for one pose and the representatives of the other, thanks to a reasoning similar to the one developed in the proof of proposition 1:

$$\boxed{\forall \mathbf{p}_1 \in \mathscr{R}(\mathscr{P}_1), \mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{p}_2\in\mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2-\mathbf{p}_1\|.} \tag{20}$$

The cardinal of $\mathscr{R}(\mathscr{P})$ is independent of the pose considered, depending solely from the class of proper symmetries of the considered object. We therefore denote it $|\mathscr{R}(\bullet)|$. For most classes – objects with no proper symmetry, revolution objects without rotoreflection invariance and spherical objects – a pose admits a single representative and in that case we will refer to it as $\mathscr{R}(\mathscr{P})$ by an abuse of notation. $\mathscr{R}$ can be considered in such case as an isometric embedding of $\mathscr{C}$ into the Euclidean space $\mathbb{R}^N$, and the distance between two poses simply corresponds to the Euclidean distance between their respective representatives.

A synthesis of our proposed expressions of pose representatives is provided in table 3 for 3D objects, and in table 4 for 2D ones.

5.1 Neighborhood query

The distance formulation (20) is of great practical value as it enables to perform efficient *radius search* and exact or approximate *k-nearest neighbors* queries within a large set of poses through the use of any off-the-shelf neighborhood query algorithms designed for Euclidean spaces. Neighborhood queries are useful for numerous problems, and we provide an example in section 10 where radius search is heavily used. Existing methods for neighborhood queries enable fast neighborhood retrieval within a set of points of a vector space compared to a brute-force approach consisting in computing the distance to every points of the set. They make use of a specific search structure – such as a grid or a kD-tree for example – adapted to the specific properties of the considered metric space. A review of those algorithms is out of the scope of this work, and we will only refer the interested reader to the well-known FLANN library (Muja and Lowe 2009) as a starting point.

Let S be a finite set of poses. We consider the pointset R consisting of the aggregation of all representatives of the poses of S:

$$R = \bigcup_{\mathscr{P}\in S} \mathscr{R}(\mathscr{P}). \tag{21}$$

From (20), given a query pose $\mathscr{Q}$ and one of its representatives $\mathbf{q} \in \mathscr{R}(\mathscr{Q})$, the poses of S that are closer to $\mathscr{Q}$ than a given distance δ are the poses that have a representative closer to $\mathbf{q}$ than δ using the Euclidean distance.

Table 3 Proposed representatives for a pose $\mathscr{P} = [(\mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^3)]$ of a 3D object depending on its proper symmetries.

Assumptions		
Center of mass of the object as origin of the object frame. For revolution objects, revolution axis as $\mathbf{e}_z$ axis of the object frame.		
Notations		
$\boldsymbol{\Lambda} \triangleq \left(\frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\mathbf{x}\mathbf{x}^\top ds\right)^{1/2}$		
and $\lambda \triangleq \sqrt{\lambda_r^2 + \lambda_z^2}$ for revolution objects where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_r, \lambda_r, \lambda_z)$		
Proper symmetry class	**Proper symmetry group** G	**Pose representatives** $\mathscr{R}(\mathscr{P})$
Spherical symmetry	$SO(3)$	$\mathbf{t} \in \mathbb{R}^3$
Revolution symmetry without rotoreflection invariance	$\{\mathbf{R}_z^\alpha \mid \alpha \in \mathbb{R}\}$	$(\lambda(\mathbf{R}\mathbf{e}_z)^\top, \mathbf{t}^\top)^\top \in \mathbb{R}^6$
Revolution symmetry with rotoreflection invariance	$\{\mathbf{R}_x^\delta \mathbf{R}_z^\alpha \mid \delta \in \{0,\pi\}, \alpha \in \mathbb{R}\}$	$\{(\pm\lambda(\mathbf{R}\mathbf{e}_z)^\top, \mathbf{t}^\top)^\top\} \subset \mathbb{R}^6$
No proper symmetry	$\{\mathbf{I}\}$	$(\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^\top, \mathbf{t}^\top)^\top \in \mathbb{R}^{12}$
Finite nontrivial	Finite	$\{(\mathrm{vec}(\mathbf{R}\mathbf{G}\boldsymbol{\Lambda})^\top, \mathbf{t}^\top)^\top \mid \mathbf{G} \in G\} \subset \mathbb{R}^{12}$

Table 4 Proposed representatives for a pose $\mathscr{P} = [(\theta \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^2)]$ of a 2D object depending on its proper symmetries.

Assumptions		
Center of mass of the object as origin of the object frame.		
Notations		
$\forall \alpha \in \mathbb{R}, e^{i\alpha} = (\cos(\alpha), \sin(\alpha))$, and $\lambda \triangleq \left(\frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\|\mathbf{x}\|^2 ds\right)^{1/2}$		
Proper symmetry class	**Proper symmetry group** G	**Pose representatives** $\mathscr{R}(\mathscr{P})$
Circular symmetry	$SO(2)$	$\mathbf{t} \in \mathbb{R}^2$
No proper symmetry	$\{\mathbf{I}\}$	$(\lambda e^{i\theta}, \mathbf{t}^\top)^\top \in \mathbb{R}^4$
Cyclic symmetry (order $n \in \mathbb{N}^*$)	$\{\mathbf{R}^{2k\pi/n} \mid k \in [\![0, n[\![\}$	$\left\{(\lambda e^{i(\theta+2k\pi/n)}, \mathbf{t}^\top)^\top \middle| k \in [\![0, n[\![\right\} \subset \mathbb{R}^4$

Such representatives can be retrieved efficiently through a standard *radius search* operation around $\mathbf{q}$ in R. The search for nearest neighbors can be performed in a similar fashion.

One should nevertheless be careful of potential duplicates with those operations, as a pose may have several representatives depending on the proper symmetries of the object. Nonetheless, the absence of duplicates is locally guaranteed around the query point $\mathbf{p}$ in an open ball of radius $T/2$, where T is defined further in definition 6.

5.2 Decomposition into translation and rotation parts

From this point on, we consider a direct orthonormal coordinates system $(\mathbf{O}, \mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$. As in section 2.3 on the proper symmetry group, we assume that the origin $\mathbf{O}$ of the object frame is an invariant point of the object with respect to its proper symmetries. As an example, it is assumed chosen at the center of the object for a spherical object, and on the revolution axis for a revolution object. Doing so, the proper symmetry group of the object can be considered as a group of rotations around the origin, and we therefore assimilate proper symmetries to rotation matrices. We exploit this property to develop the inner term of the expression (10) of the square distance into:

$$\begin{aligned}
&\|(\mathbf{T}_2 \circ \mathbf{G}_2)(\mathbf{x}) - (\mathbf{T}_1 \circ \mathbf{G}_1)(\mathbf{x})\|^2 \\
&= \|\mathbf{R}_2\mathbf{G}_2\mathbf{x} + \mathbf{t}_2 - (\mathbf{R}_1\mathbf{G}_1\mathbf{x} + \mathbf{t}_1)\|^2 \\
&= \|\mathbf{R}_2\mathbf{G}_2\mathbf{x} - \mathbf{R}_1\mathbf{G}_1\mathbf{x}\|^2 + \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\
&\quad + 2(\mathbf{t}_2 - \mathbf{t}_1)^\top(\mathbf{R}_2\mathbf{G}_2 - \mathbf{R}_1\mathbf{G}_1)\mathbf{x}.
\end{aligned} \tag{22}$$

We add the further constraint that the origin of the object frame is the center of mass of the object surface, i.e. $\int_{\mathscr{S}} \mu(\mathbf{x})\mathbf{x} ds = \mathbf{0}$. This constraint is compatible with the previous one because the center of mass is unique, and therefore has to be left unchanged by the proper symmetries of the object. Thanks to this choice, the last term of (22) disappears during the integration, and the square distance (10) can therefore be decomposed into a translation and a rotation part:

$$\begin{aligned}
\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\
&\quad + \underbrace{\min_{\mathbf{G}_1, \mathbf{G}_2 \in G} \frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\|\mathbf{R}_2\mathbf{G}_2\mathbf{x} - \mathbf{R}_1\mathbf{G}_1\mathbf{x}\|^2 ds}_{\mathrm{d}^2_{\mathrm{rot}}(\mathbf{R}_1, \mathbf{R}_2)}.
\end{aligned} \tag{23}$$

In the following sections, we show how this rotation part can be simplified, and how it leads to the notion of pose representatives.

5.3 Object with no proper symmetries

Let us consider the case of an object showing no proper symmetries. The proper symmetry group of this object is reduced to the identity rotation: $G = \{\mathbf{I}\}$, therefore the rotation part of the square distance (23) can be expressed as follows:

$$\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1, \mathscr{P}_2) = \int_{\mathscr{S}} \mu(\mathbf{x}) \|\mathbf{R}_2\mathbf{x} - \mathbf{R}_1\mathbf{x}\|^2 ds \tag{24}$$

Let us show how, for any given pose $\mathscr{P} \in \mathscr{C}$, one can define a representative $\mathscr{R}(\mathscr{P}) \in \mathbb{R}^{12}$ such that the distance between two poses in $\mathscr{C}$ is the same as the distance between their representatives in $\mathbb{R}^{12}$. Our approach is inspired by the work of Kazerounian and Rastegar (1992) and Chirikjian and Zhou (1998).

Let $\boldsymbol{\Lambda}$ be the symmetric positive semi-definite square root matrix of the covariance matrix of the object's weighted surface:

$$\boxed{\boldsymbol{\Lambda} \triangleq \left(\frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\mathbf{x}\mathbf{x}^\top ds\right)^{1/2}.} \tag{25}$$

$\boldsymbol{\Lambda}$ does not depend on the considered pose and can therefore be estimated once and for all for a given rigid object in a preprocessing step. We provide in appendix C formulas to compute $\boldsymbol{\Lambda}$ when $\mathscr{S}$ is the surface of a triangular mesh.

Rewriting the inner part of (24) with a trace operator,

$$\|\mathbf{R}_2\mathbf{x} - \mathbf{R}_1\mathbf{x}\|^2 = \mathrm{Tr}\left((\mathbf{R}_2 - \mathbf{R}_1)\mathbf{x}\mathbf{x}^\top(\mathbf{R}_2 - \mathbf{R}_1)^\top\right),$$

one can express the rotation part of the square distance in a closed form as a weighted Frobenius square distance between the rotation matrices:

$$\begin{aligned}\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1, \mathscr{P}_2) &= \mathrm{Tr}\left((\mathbf{R}_2 - \mathbf{R}_1)\boldsymbol{\Lambda}^2(\mathbf{R}_2 - \mathbf{R}_1)^\top\right) \\ &= \|\mathbf{R}_2\boldsymbol{\Lambda} - \mathbf{R}_1\boldsymbol{\Lambda}\|_F^2.\end{aligned} \tag{26}$$

Therefore, denoting vec the operator vectorizing columnwise a matrix into a column vector, we can define an isometry $\mathscr{R}$ from the pose space into the 12-dimensional Euclidian space:

$$\boxed{\begin{array}{l}\text{Object without proper symmetry:} \\ \begin{aligned}\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{R}_2\boldsymbol{\Lambda} - \mathbf{R}_1\boldsymbol{\Lambda}\|_F^2 + \|t_2 - t_1\|^2 \\ &= \|\mathscr{R}(\mathscr{P}_2) - \mathscr{R}(\mathscr{P}_1)\|^2\end{aligned} \\ \text{with } \mathscr{R}(\mathscr{P}) \triangleq \left(\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^\top, \mathbf{t}^\top\right)^\top \in \mathbb{R}^{12}.\end{array}} \tag{27}$$

The conversion from a pose represented in term of a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$ to its representative in $\mathbb{R}^{12}$ is direct, since it consists in a simple linear operation. If the object frame is moreover chosen aligned with the principal axes of the object, $\boldsymbol{\Lambda}$ is diagonal, making the computation of the pose representative even cheaper.

5.4 Revolution object without rotoreflection invariance

We now consider the case of a revolution object without rotoreflection invariance. As stated in section 5.2, we assume that the origin of the object frame corresponds to the center of mass of the object. Without loss of generality, we moreover assume that the axis $\mathbf{e}_z$ of the object frame is aligned with the revolution axis. A pose $\mathscr{P}$ is thus defined up to a rotation $\mathbf{R}_z^\phi$ along this $\mathbf{e}_z$ axis, and the proper symmetry group of the object is $G = \left\{\mathbf{R}_z^\phi \middle| \phi \in \mathbb{R}\right\}$ where ϕ is the angle of the considered rotation.

The simplification to get rid of the integral within the distance expression thanks to the introduction of the matrix $\boldsymbol{\Lambda}$ in section 5.3 is also valid here. Moreover, because $(\mathbf{O}, \mathbf{e_z})$ is the revolution axis of the object, $\boldsymbol{\Lambda}$ is necessarily diagonal and of the form

$$\boldsymbol{\Lambda} = \begin{pmatrix}\lambda_r & 0 & 0 \\ 0 & \lambda_r & 0 \\ 0 & 0 & \lambda_z\end{pmatrix} \tag{28}$$

with $\lambda_r, \lambda_z \in \mathbb{R}^+$.

Those facts enable us to express the rotation part of the distance as a simple scaled distance between the revolution axes seen as 3D vectors:

$$\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1, \mathscr{P}_2) = \lambda^2 \|\mathbf{R}_2\mathbf{e}_z - \mathbf{R}_1\mathbf{e}_z\|^2 \tag{29}$$

with $\lambda \triangleq \sqrt{\lambda_r^2 + \lambda_z^2}$. See appendix A for a proof of this result.

Therefore, similarly to what we proposed for an object without proper symmetry, we can consider a simple isometry $\mathscr{R}$ which associates to a pose of a revolution object without rotoreflection invariance, a 6D vector – consisting of the concatenation of the coordinates of its scaled revolution axis and of its position:

$$\boxed{\begin{array}{l}\text{Revolution object without rotoreflection invariance:} \\ \begin{aligned}\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{t}_2 - \mathbf{t}_1\|^2 + \lambda^2\|\mathbf{R}_2\mathbf{e}_z - \mathbf{R}_1\mathbf{e}_z\|^2 \\ &= \|\mathscr{R}(\mathscr{P}_2) - \mathscr{R}(\mathscr{P}_1)\|^2\end{aligned} \\ \text{with } \mathscr{R}(\mathscr{P}) \triangleq \left(\lambda(\mathbf{R}\mathbf{e}_z)^\top, \mathbf{t}^\top\right)^\top \in \mathbb{R}^6 \\ \text{where } \lambda = \sqrt{\lambda_r^2 + \lambda_z^2}.\end{array}} \tag{30}$$

5.5 Spherical object

We now consider the simpler case of an object with spherical symmetry. Choosing the center of the object as origin of the object frame, the proper symmetry group of the object is the whole rotation group $SO(3)$. The rotation part of the distance (23) can thus be rewritten as follows:

$$\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{R}_1, \mathbf{R}_2}\left(\frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\|\mathbf{R}_2\mathbf{x} - \mathbf{R}_1\mathbf{x}\|^2 ds\right). \tag{31}$$

This term is null, the minimum being reached for $\mathbf{R}_1 = \mathbf{R}_2$. Therefore, the pose space of a spherical object can be isometrically embedded into a $\mathbb{R}^3$ by representing a pose by the position of its center:

$$\boxed{\begin{aligned}&\text{Spherical object:}\\&\mathrm{d}^2(\mathscr{P}_1,\mathscr{P}_2) = \|\mathbf{t}_2-\mathbf{t}_1\|^2\\&\qquad\qquad = \|\mathscr{R}(\mathscr{P}_2)-\mathscr{R}(\mathscr{P}_1)\|^2\\&\text{With } \mathscr{R}(\mathscr{P}) = \mathbf{t}\in\mathbb{R}^3.\end{aligned}} \tag{32}$$

5.6 Revolution object with rotoreflection invariance

Let us consider the case of a revolution object with rotoreflection invariance, i.e. having a reflection symmetry with respect to a plane orthogonal to the revolution axis. With the same constraints on the choice of the object frame as for a revolution object without rotoreflection invariance, the proper symmetry group of such object can be written as follows:

$$G = \left\{\mathbf{R}_x^{\delta}\mathbf{R}_z^{\alpha} \mid \alpha\in\mathbb{R}, \delta\in\{0,\pi\}\right\}. \tag{33}$$

Therefore, the distance between two poses $\mathscr{P}_1, \mathscr{P}_2$ can be expressed as:

$$\min_{\delta_1,\delta_2,\phi_1,\phi_2} \mathrm{d}_{\text{no_sym}}\left((\mathbf{R}_1\mathbf{R}_x^{\delta_1}\mathbf{R}_z^{\phi_1},\mathbf{t}_1),(\mathbf{R}_2\mathbf{R}_x^{\delta_2}\mathbf{R}_z^{\phi_2},\mathbf{t}_2)\right). \tag{34}$$

We discussed in section 5.4 how to compute such an expression relatively to the symmetries along the revolution axis. Therefore, using result (30), our distance can be rewritten as the minimum Euclidean distance between 6D points, two being assigned to each pose:

$$\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\delta_1,\delta_2\in\{0,\pi\}} \|\mathbf{p}_2^{\delta_2}-\mathbf{p}_1^{\delta_1}\| \tag{35}$$

with $\mathbf{p}_i^{\delta} = \left(\lambda(\mathbf{R}_i\mathbf{R}_x^{\delta}\mathbf{e}_z)^\top,\mathbf{t}^\top\right)^\top \in \mathbb{R}^6$ the representatives of pose $\mathscr{P}_i$, for $i = 1,2$.

Simplifying the representative expression a little given that $\mathbf{R}_x^0\mathbf{e}_z = \mathbf{e}_z$ and $\mathbf{R}_x^{\pi}\mathbf{e}_z = -\mathbf{e}_z$, we see that a pose of a revolution object with rotoreflection invariance can be represented by two 6D vectors consisting of the concatenation of the coordinates of its scaled revolution axis and of its position, each potential orientation of the axis being taken into account by one representative:

$$\boxed{\begin{aligned}&\text{Revolution object with rotoreflection invariance:}\\&\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{p}_1\in\mathscr{R}(\mathscr{P}_1),\mathbf{p}_2\in\mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2-\mathbf{p}_1\|\\&\text{With } \mathscr{R}(\mathscr{P}) \triangleq \left\{\left(\pm\lambda(\mathbf{R}\mathbf{e}_z)^\top,\mathbf{t}^\top\right)^\top\right\} \subset \mathbb{R}^6.\end{aligned}} \tag{36}$$

5.7 Object with a nontrivial finite proper symmetry group

The last type of 3D object to deal with is the case of an object with a finite proper symmetry group G different from the identity. Examples of such object are proposed in sections 9 and 10. The proposed distance between two poses of such an object can be written as:

$$\min_{\mathbf{G}_1,\mathbf{G}_2\in G} \mathrm{d}_{\text{no_sym}}((\mathbf{R}_1\mathbf{G}_1,\mathbf{t}_1),(\mathbf{R}_2\mathbf{G}_2,\mathbf{t}_2)). \tag{37}$$

We showed in section 5.3 that the pose of an object without proper symmetry can be represented as a 12D point, such that the distance between two poses of such object corresponds to the Euclidean distance between their respective representatives. Therefore, it is straightforward to conclude that the pose of an object with a finite proper symmetry group can be represented by a finite set of 12D representative points, such that the distance between two poses corresponds to the minimum Euclidean distance between their respective representatives:

$$\boxed{\begin{aligned}&\text{Object with a nontrivial finite proper symmetry group:}\\&\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{p}_1\in\mathscr{R}(\mathscr{P}_1),\mathbf{p}_2\in\mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2-\mathbf{p}_1\|\\&\text{With } \mathscr{R}(\mathscr{P}) \triangleq \left\{\left(\mathrm{vec}(\mathbf{R}\mathbf{G}\boldsymbol{\Lambda})^\top,\mathbf{t}^\top\right)^\top \middle| \mathbf{G}\in G\right\} \subset \mathbb{R}^{12}.\end{aligned}} \tag{38}$$

5.8 2D object

The notion of pose representative can be extended to 2D objects as well. For the sake of conciseness, we will only discuss the case of a 2D object with no proper symmetry, as the reasoning is very similar to the one performed for 3D objects. The full list of proposed representatives is given in table 4.

The decomposition of the square distance between two poses in a translation and rotation terms (23) and the expression of the rotation part as a Frobenius norm (26) are still valid in the 2D case, and the distance expression can be even further simplified.

A 2D rotation matrix can be parametrized by an angle θ as follows:

$$\mathbf{R}^{\theta} = \begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}. \tag{39}$$

Introducing the elements of the covariance matrix

$$\boldsymbol{\Lambda}^2 = \begin{pmatrix}\lambda_{xx}^2 & \lambda_{xy}^2\\ \lambda_{xy}^2 & \lambda_{yy}^2\end{pmatrix}, \tag{40}$$

the rotation part can be simplified into

$$\begin{aligned} \mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1,\mathscr{P}_2) &= \mathrm{Tr}\left((\mathbf{R}^{\theta_2}-\mathbf{R}^{\theta_1})\boldsymbol{\Lambda}^2(\mathbf{R}^{\theta_2}-\mathbf{R}^{\theta_1})^\top\right) \\ &= (\lambda_{xx}^2+\lambda_{yy}^2)\|e^{i\theta_2}-e^{i\theta_1}\|^2 \end{aligned} \tag{41}$$

where $e^{i\theta} \triangleq (\cos(\theta),\sin(\theta))$. Therefore, we can include in our framework a 2D object without proper symmetry, and represent a pose of such object by a 4D vector, consisting of the concatenation of the coordinates of its scaled complex orientation and of its position:

2D object without proper symmetry:

$$\mathrm{d}(\mathscr{P}_1,\mathscr{P}_2) = \min_{\mathbf{p}_1\in\mathscr{R}(\mathscr{P}_1),\mathbf{p}_2\in\mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2-\mathbf{p}_1\|$$

$$\text{with } \mathscr{R}(\mathscr{P}) \triangleq \left(\lambda e^{i\theta},\mathbf{t}^\top\right)^\top \in \mathbb{R}^4 \tag{42}$$

$$\text{where } \lambda = \left(\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\|\mathbf{x}\|^2 ds\right)^{1/2}.$$

6 Symmetry within representatives

Objects with finite non trivial symmetry groups and revolution objects with rotoreflection invariance admit several representatives per pose. This multiplicity of representatives expresses the proper symmetries of the object that are not accounted for in the expression of a representative, and leads to some symmetry properties within the set of representatives itself.

In this section we discuss some of those properties, that are required for the theoretical developments we propose in section 8.2.

Formally, we define for a given object a finite group of symmetry operations $G_{\mathscr{R}}$ on the ambient space $\mathbb{R}^N$ that notably satisfy the following properties:

Proposition 2 *$G_{\mathscr{R}}$ contains $|\mathscr{R}(\bullet)|$ elements, and given a pose $\mathscr{P}$ and one of its representative $\mathbf{p}\in\mathscr{R}(\mathscr{P})$, the set of elements symmetric to $\mathbf{p}$ (including itself) is the whole set of representatives of the pose:*

$$\{s(\mathbf{p})|s\in G_{\mathscr{R}}\} = \mathscr{R}(\mathscr{P}). \tag{43}$$

Proposition 3 *Elements of $G_{\mathscr{R}}$ are linear transformations of the ambient space, i.e. for $s\in G_{\mathscr{R}}$, and for any $\mathbf{x}_1,\mathbf{x}_2\in\mathbb{R}^N$, $\alpha\in\mathbb{R}$,*

$$s(\mathbf{x}_1+\alpha\mathbf{x}_2) = s(\mathbf{x}_1)+\alpha s(\mathbf{x}_2). \tag{44}$$

Proposition 4 *Elements of $G_{\mathscr{R}}$ are automorphisms of the ambient space: for any $s\in G_{\mathscr{R}}$, s is bijective, and for any $\mathbf{x}_1,\mathbf{x}_2\in\mathbb{R}^N$,*

$$\|s(\mathbf{x}_2)-s(\mathbf{x}_1)\| = \|\mathbf{x}_2-\mathbf{x}_1\|. \tag{45}$$

We detail the expression of those symmetries in the next subsections.

6.1 Object with a finite non trivial group

In the case of an object with a finite non trivial proper symmetry group G, we define the symmetry operations on representatives as follows:

$$G_{\mathscr{R}} = \{s_{\mathbf{G}}|G\in\mathbf{G}\}, \tag{46}$$

where for any proper symmetry $\mathbf{G}\in G$, splitting a 12D vector into two parts $\mathbf{M}\in\mathscr{M}_{3,3}(\mathbb{R})$ and $\mathbf{t}\in\mathbb{R}^3$,

$$\begin{aligned} s_{\mathbf{G}} &: \mathbb{R}^{12}\to\mathbb{R}^{12} \\ (\mathrm{vec}(\mathbf{M})^\top,\mathbf{t}^\top)^\top &\mapsto (\mathrm{vec}(\mathbf{MG})^\top,\mathbf{t}^\top)^\top. \end{aligned} \tag{47}$$

In particular, the symmetric of any point $(\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^\top,\mathbf{t}^\top)^\top$, where $\mathbf{R}\in\mathscr{M}_{3,3}(\mathbb{R})$ and $\mathbf{t}\in\mathbb{R}^3$, can be expressed

$$s_{\mathbf{G}}\left((\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^\top,\mathbf{t}^\top)^\top\right) = (\mathrm{vec}(\mathbf{RG}\boldsymbol{\Lambda})^\top,\mathbf{t}^\top)^\top \tag{48}$$

thanks to the lemma 1, and the proposition 2 is therefore satisfied because of our definition of the representatives of such object.

Those symmetry operations are moreover linear (i.e. satisfy proposition 3), and we verify that, for any given $\mathbf{G}\in G$, $s_{\mathbf{G}}$ is an automorphism (proposition 4). Bijectivity is indeed straightforward as we can exhibit the inverse $(s_G)^{-1} = s_{G^{-1}}$. Regarding the morphism property, let us consider two arbitrary 12D vectors $\left(\mathbf{x}_i = (\mathrm{vec}(\mathbf{M}_i)^\top,\mathbf{t}_i^\top)^\top\right)_{i=1,2}$. The square distance between their respective symmetric points can be expressed as follows:

$$\begin{aligned} &\|s_{\mathbf{G}}(\mathbf{x}_2)-s_{\mathbf{G}}(\mathbf{x}_1)\|^2 \\ &= \|\mathbf{t}_2-\mathbf{t}_1\|^2+\|(\mathbf{M}_2-\mathbf{M}_1)\mathbf{G}\|_F^2. \end{aligned} \tag{49}$$

The Frobenius norm being invariant by rotation, it simplifies into

$$\|\mathbf{t}_2-\mathbf{t}_1\|^2+\|\mathbf{M}_2-\mathbf{M}_1\|_F^2 \tag{50}$$

which corresponds to the square distance between the two points $\|\mathbf{x}_2-\mathbf{x}_1\|^2$, thus proving that $s_{\mathbf{G}}$ is indeed a morphism.

6.2 Revolution object with rotoreflection invariance

We define for a revolution object with rotoreflection invariance two symmetry operations on representatives as follows, to account for the invariance by reflection normal to the revolution axis:

$$G_{\mathscr{R}} = \left\{s_{\mathrm{rev},\delta}\middle|\delta=\pm1\right\}, \tag{51}$$

where, splitting a 6D vector into two parts $\mathbf{a},\mathbf{t}\in\mathbb{R}^3$,

$$s_{\mathrm{rev},\delta} : \mathbb{R}^6\to\mathbb{R}^6 \tag{52}$$

$$(\mathbf{a}^\top,\mathbf{t}^\top)^\top \mapsto (\delta\mathbf{a}^\top,\mathbf{t}^\top)^\top. \tag{53}$$

Property 2 is satisfied because of the definition of the pose representatives for such object. Such functions are also linear (property 3). Regarding property 4, bijectivity is again straightforward as $s_{\mathrm{rev},\delta}^{-1} = s_{\mathrm{rev},\delta}$ for any $\delta \in \{-1,1\}$. They are also morphisms since for any pair of points of the ambient space $(\mathbf{x}_i = (\mathbf{a}_i^\top, \mathbf{t}_i^\top)^\top))_{i=1,2}$,

$$\begin{aligned}
&\|s_{\mathrm{rev},\delta}(\mathbf{x}_2) - s_{\mathrm{rev},\delta}(\mathbf{x}_1)\|^2 \\
&= \delta^2 \|\mathbf{a}_2 - \mathbf{a}_1\|^2 + \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\
&= \|\mathbf{a}_2 - \mathbf{a}_1\|^2 + \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\
&= \|\mathbf{x}_2 - \mathbf{x}_1\|^2.
\end{aligned} \tag{54}$$

6.3 2D object with cyclic symmetry

Lastly, we define the following group of symmetry for a 2D object with a rotational symmetry of order $n \in \mathbb{N}^*$:

$$G_{\mathscr{R}} = \left\{ s_{\mathrm{2D},k} \middle| k \in [\![0,n[\![\right\}, \tag{55}$$

where, splitting a 4D vector into two parts $\mathbf{a},\mathbf{t} \in \mathbb{R}^2$, and using the complex multiplication,

$$s_{\mathrm{2D},k} : \mathbb{R}^4 \to \mathbb{R}^4 \tag{56}$$

$$(\mathbf{a}^\top, \mathbf{t}^\top)^\top \mapsto (e^{i2k\pi/n} \cdot \mathbf{a}, \mathbf{t}^\top)^\top. \tag{57}$$

This group satisfies the proposition 2 by definition of the pose representatives of such object, since for any $\theta \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^2$,

$$s_{\mathrm{2D},k}\left((\lambda e^{i\theta}, \mathbf{t}^\top)^\top \right) = (\lambda e^{i(\theta + 2k\pi/n)}, \mathbf{t}^\top)^\top. \tag{58}$$

It also consists of linear transformations (proposition 3) that preserve the norm (proposition 4), because of the properties of the complex multiplication and because $e^{i2k\pi/n}$ is of unit modulus, for any $k \in [\![0,n[\![$.

7 Projection onto the pose space

In section 5, we discussed how a pose $\mathscr{P}$ can be identified to a finite pointset $\mathscr{R}(\mathscr{P})$ of an Euclidean space of finite dimension $\mathbb{R}^N$ and how elements of $\mathscr{R}(\mathscr{P})$ can be computed easily from any rigid transformation $(\mathbf{R},\mathbf{t})$ associated to the pose. The backward mapping is possible, and for any element of $\mathscr{R}(\mathscr{P})$ we can compute a rigid transformation fully describing the pose $\mathscr{P}$. Hence we consider an element of $\mathscr{R}(\mathscr{P})$ as a *representative* of $\mathscr{P}$.

This computation is actually straightforward given the expressions of poses representatives (see table 3), thus we choose to discuss this assertion in the more general framework of projection onto the pose space: given an arbitrary N-D vector $\mathbf{x}$, find out what pose has the most similar representative to $\mathbf{x}$.

Definition 3 We define as projections of $\mathbf{x} \in \mathbb{R}^N$ the poses:

$$\mathrm{proj}(\mathbf{x}) \triangleq \underset{\mathscr{P}}{\mathrm{argmin}} \min_{\mathbf{p} \in \mathscr{R}(\mathscr{P})} \|\mathbf{p} - \mathbf{x}\|^2 \tag{59}$$

In nondegenerate cases, the projection is unique, and we propose in the next subsections its expression for the different classes of bounded objects, based on the computation of the closest pose representative to the query point.

7.1 Spherical object

Projection is trivial in the case of a spherical object, since all points of $\mathbb{R}^3$ are valid representatives of poses. A point $\mathbf{x} \in \mathbb{R}^3$ therefore projects onto the pose having $\mathbf{x}$ for representative, namely the pose in which the center of the object is $\mathbf{x}$.

7.2 Object of revolution

In the case of an object of revolution without rotoreflection invariance, the position of the center of mass and the oriented revolution axis of the object are well defined at any given pose. Reciprocally, a pose can be defined by the position of its center of mass $\mathbf{t}$ and its oriented revolution axis, that we represent by a normalized vector $\mathbf{a} \in \mathbb{R}^3$. The unique representative of such a pose is $(\lambda \mathbf{a}^\top, \mathbf{t}^\top)^\top$ as we defined in 5.4.

Let $\mathbf{x} \in \mathbb{R}^6$ be a point to project onto the pose space. Without loss of generality, $\mathbf{x}$ can be split into two parts: $\mathbf{x} = (\mathbf{x}_r^\top, \mathbf{x}_t^\top)^\top$ with $\mathbf{x}_r, \mathbf{x}_t \in \mathbb{R}^3$. The projection problem can therefore be reformulated into:

$$\begin{aligned}
\mathrm{proj}(\mathbf{x}) &= \underset{\mathscr{P}}{\mathrm{argmin}} \|\mathbf{x} - \mathscr{R}(\mathscr{P})\|^2 \\
&= \underset{\mathbf{a},\mathbf{t} \in \mathbb{R}^3 / \|\mathbf{a}\| = 1}{\mathrm{argmin}} \left(\|\mathbf{x}_r - \lambda \mathbf{a}\|^2 + \|\mathbf{x}_t - \mathbf{t}\|^2 \right)
\end{aligned} \tag{60}$$

This problem admits an unique solution as long as $\mathbf{x}_r \neq \mathbf{0}$, and in that case the projection of $\mathbf{x}$ is the pose of center of mass $\hat{\mathbf{t}} = \mathbf{x}_t$ and of axis $\hat{\mathbf{a}} = \mathbf{x}_r / \|\mathbf{x}_r\|$. This result holds true in the case of an object of revolution with rotoreflection invariance, since $(\lambda \hat{\mathbf{a}}^\top, \hat{\mathbf{t}}^\top)^\top$ is the closest pose representative to $\mathbf{x}$.

7.3 Object with a finite proper symmetry group

The representative of a pose of an object without proper symmetry is a 12D vector, the first 9 dimensions representing the orientation in the form of a vectorized matrix and the 3 others the position of the object. Therefore, and without loss of generality, we can split a point $\mathbf{x} \in \mathbb{R}^{12}$ to project in a similar fashion: $\mathbf{x} = (\mathrm{vec}(\mathbf{X}_r)^\top, \mathbf{x}_t^\top)^\top$ with $\mathbf{x}_t \in \mathbb{R}^3$ and

$\mathbf{X}_r \in \mathscr{M}_{3,3}(\mathbb{R})$. The projection problem for such a point $\mathbf{x}$ – in the case of an object without proper symmetry – thus consists in:

$$\begin{aligned}\text{proj}(\mathbf{x}) &= \underset{\mathscr{P}}{\text{argmin}} \|\mathbf{x} - \mathscr{R}(\mathscr{P})\|^2 \\ &= \underset{\mathbf{R},\mathbf{t}}{\text{argmin}} \left(\|\mathbf{X}_r - \mathbf{R}\boldsymbol{\Lambda}\|_F^2 + \|\mathbf{x}_t - \mathbf{t}\|^2\right)\end{aligned} \tag{61}$$

The two terms being independent, we conclude again that the position of the center of mass of the object for a projection of $\mathbf{x}$ is $\hat{\mathbf{t}} = \mathbf{x}_t$. The minimization problem regarding the orientation part is in the form of the so-called constrained orthogonal Procrustes problem (Schönemann 1966; Umeyama 1991) and admits the solution $\hat{\mathbf{R}} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, where $\mathbf{U}\mathbf{D}\mathbf{V}^\top$ is a singular value decomposition of $\mathbf{X}_r\boldsymbol{\Lambda}$ such as $\mathbf{D} = \text{diag}(\alpha_1, \alpha_2, \alpha_3), \alpha_1 \geq \alpha_2 \geq \alpha_3 \geq 0$, and

$$\mathbf{S} = \begin{cases} \mathbf{I} \text{ if } \det(\mathbf{U})\det(\mathbf{V}) > 0 \\ \text{diag}(1,1,-1) \text{ otherwise.} \end{cases} \tag{62}$$

The projection is unique if $\text{rank}(\mathbf{X}_r\boldsymbol{\Lambda}^\top) \geq 2$ (Umeyama 1991), a condition that is fulfilled in most practical cases. This result also holds true in the general case of an object with a finite proper symmetry group, as $(\text{vec}(\hat{\mathbf{R}}\boldsymbol{\Lambda})^\top, \hat{\mathbf{t}}^\top)^\top$ is the closest pose representative to $\mathbf{x}$.

7.4 2D object

The projection problem for a 2D object is similar to the 3D case.

In the case of a circular object, any point of $\mathbf{x} \in \mathbb{R}^2$ is a valid representative of a pose and therefore projects onto the pose of center $\mathbf{x}$.

Regarding an object with cyclic symmetry, we conclude by the same reasoning as in the case of a 3D revolution object (section 7.2) that a 4D vector $\mathbf{x} = (\mathbf{a}^\top, \mathbf{t}^\top)^\top$, where $\mathbf{a}, \mathbf{t} \in \mathbb{R}^2$, admits an unique projection as long as $\|\mathbf{a}\| \neq 0$. The projection admits a representative $(\lambda/\|\mathbf{a}\| \cdot \mathbf{a}^\top, \mathbf{t}^\top)^\top$, and consists of the pose defined by a translation $\mathbf{t}$ and a rotation of angle $\arg(\mathbf{a})$, where $\arg(\mathbf{a})$ is the argument of $\mathbf{a}$ seen as a complex number.

8 Averaging poses

Pose averaging is of greater use for applications such as denoising, modes detection or interpolation. Definition of the average is not obvious in non-vector spaces such as ours, and we therefore consider a generalization of the average to arbitrary metric spaces, known as the Fréchet mean.

Let us consider a finite set of poses $S = \{\mathscr{P}_i\}_{i=1..n}$ and a set of strictly positive weights $\{w_i\}_{i=1..n}$ assigned to each of those. The weighted mean of poses S is by definition the pose which minimizes the corresponding Fréchet variance:

$$\text{mean}(S) \triangleq \underset{\mathscr{P}\in\mathscr{C}}{\text{argmin}}\, \Phi(\mathscr{P}), \tag{63}$$

the Fréchet variance at a pose $\mathscr{P} \in \mathscr{C}$ being expressed as follows:

$$\Phi(\mathscr{P}) \triangleq \sum_{i=1}^{n} w_i \, \mathrm{d}^2(\mathscr{P}_i, \mathscr{P}). \tag{64}$$

This mean is not necessarily well defined, since the minimum of Fréchet variance is not necessarily reached at a unique pose. However, such cases typically occur in configurations where the average would actually be meaningless, e.g. when averaging two poses of opposite orientations for a revolution object without rotoreflection invariance.

The problem of pose averaging has already been studied for objects without proper symmetry with various metrics. Sharf et al (2010) notably compare different averaging techniques for the rotation part of a pose, using common metrics. While there is no known closed-form solution for the Riemannian metric (4), it can be computed iteratively, and it admits a closed form approximation which is "good enough" for practical applications (Gramkow 2001). A good approximation, when dealing with more than two poses, is based on computing the average of rotation matrices, and actually corresponds to the exact average when considering the distance (7) (Curtis et al 1993).

In the case of our proposed distance, the expression of the Fréchet variance can be developed into:

$$\Phi(\mathscr{P}) = \sum_{i=1}^{n} w_i \min_{\mathbf{p}_i \in \mathscr{R}(\mathscr{P}_i), \mathbf{p} \in \mathscr{R}(\mathscr{P})} \|\mathbf{p}_i - \mathbf{p}\|^2. \tag{65}$$

Considering a given tuple $P = (\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ of representatives of the poses to average, the weighted sum of square distances from a pose representative $\mathbf{p}$ to elements of P can be split into two terms, through the introduction of the arithmetic mean $\mathbf{m}_P$ of the elements of P:

$$\begin{aligned}\sum_{i=1}^{n} w_i \|\mathbf{p}_i - \mathbf{p}\|^2 &= \sum_i w_i \|\mathbf{p}_i - \mathbf{m}_P\|^2 \\ &\quad + \left(\sum_i w_i\right) \|\mathbf{p} - \mathbf{m}_P\|^2,\end{aligned} \tag{66}$$

with the arithmetic mean:

$$\mathbf{m}_P \triangleq \frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}. \tag{67}$$

The first term of (66) is independent of $\mathbf{p}$. Therefore, the problem of minimizing (66) for a given tuple P is reduced to the problem of finding a pose $\mathscr{P}$ which minimizes $\min_{\mathbf{p}\in\mathscr{R}(\mathscr{P})} \|\mathbf{p} - \mathbf{m}_P\|^2$. This corresponds to the projection problem we discussed and solved in section 7. The average

pose, if well defined, is thus the projection of the arithmetic average of a combination of representatives of the poses to average, or more formally:

$$\text{mean}(S) = \underset{\mathscr{P}\in\mathscr{A}}{\text{argmin}}\,\Phi(\mathscr{P}), \tag{68}$$

where

$$\mathscr{A} \triangleq \left\{ \text{proj}\,(\mathbf{m}_P) \,\middle|\, P \in \prod_i \mathscr{R}(\mathscr{P}_i) \right\} \tag{69}$$

8.1 Objects with a single representative per pose

Because the projection is unique in nondegenerate cases, the conclusion is straightforward for spherical objects, objects of revolution without rotoreflection invariance, and objects without proper symmetry. Since poses of such objects have only one representative, only a tuple of representatives is possible. For those objects, the average pose exists and corresponds simply to the projection of the arithmetic average of their representatives:

$$\boxed{\text{mean}(S) = \text{proj}\left(\frac{\sum_i w_i \mathscr{R}(\mathscr{P}_i)}{\sum_i w_i}\right)} \tag{70}$$

8.2 Objects with multiple representatives per pose

In the other cases, a pose admits several representatives and one should consider the different combinations of representatives to find the exact average – assuming its existence and uniqueness in nonpathological cases. This problem is not specific to our method and is similar to the issue encountered when averaging orientations of an object without proper symmetries through the arithmetic mean of their quaternions representatives, each orientation admitting two antipodal quaternions as representatives. Because the number of combinations of representatives is exponential in the number of considered poses, the exact computation of the average might easily become expensive.

A common practice to circumvent this issue when averaging orientations based on a quaternion representation is to compute the arithmetic mean of a "consistent" combination of representatives and consider its projection as the average. Such a combination is usually built by choosing a representative for an arbitrary initial pose, and then picking for each pose the nearest representative to the initial one (Gramkow 2001). This approach is simple, but is in the general case ill-defined. Indeed, the chosen combination – and hence the estimated average – are in general dependent of the initial choice. We depict an illustration on figure 3 of such a case, where three different choices for the initial pose

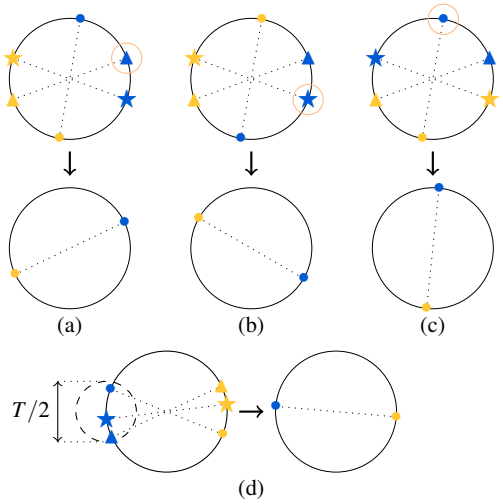


Fig. 3 Estimating the average of poses with multiple representatives: illustration with the orientation of a 2D object with a 180° rotation symmetry, that can be represented by a point on a circle, or the antipodal point. We consider three poses to average (triangle, star and disk shapes). (a, b, c) The choice of a "consistent" combination of poses representatives (blue clusters) in the sense of (Gramkow 2001) is dependent of the initial choice (circled, first row), resulting potentially in different estimations of the average pose (second row). (d) We propose a definition of consistency – which is in particular satisfied when the considered representatives are close enough one an other – ensuring an unambiguous estimation of the average pose.

lead to three different choices of "consistent" representatives combinations and therefore three different estimations of the average.

Such technique is nonetheless useful since it enables fast computations, and we show here that, under some conditions, it leads to an unambiguous estimation of the mean.

Definition 4 (Consistency) A tuple $(\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ is said *consistent* if and only if

$$\forall (i,j) \in [\![1,n]\!]^2, \forall \mathbf{q}_j \in \mathscr{R}(\mathscr{P}_j) \setminus \{\mathbf{p_j}\}, \\ \|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{q}_j - \mathbf{p}_i\|. \tag{71}$$

Proposition 5 (Uniqueness of a consistent tuple, up to symmetry) *If $(\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ is consistent, then the set of consistent tuples of $\prod_i \mathscr{R}(\mathscr{P}_i)$ is the set composed of $(\mathbf{p}_i)_{i=1\ldots n}$ and its symmetric tuples:*

$$\{(s(\mathbf{p}_i))_{i=1\ldots n} | s \in G_{\mathscr{R}}\}. \tag{72}$$

Proof Let $(\mathbf{p}_i)_{i=1\ldots n}, (\mathbf{q}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ be two different consistent tuples. There exists $j \in [\![1,n]\!]$ such as $\mathbf{p}_j \neq \mathbf{q}_j$. Because $(\mathbf{p}_i)_{i=1\ldots n}$ is consistent and $\mathbf{p}_j \neq \mathbf{q}_j$, we know that

$$\forall i \in [\![1,n]\!], \|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{q}_j - \mathbf{p}_i\|. \tag{73}$$

Therefore for any $i \in [\![1,n]\!]$, $\mathbf{p}_j \neq \mathbf{q}_j$. There are thus at most $|\mathscr{R}(\bullet)|$ consistent tuples.

Moreover, there are exactly $|\mathscr{R}(\bullet)|$ different representative combinations symmetric to $(\mathbf{p}_i)_{i=1\ldots n}$ – including itself (proposition 2):

$$\{(s(\mathbf{p}_i))_{i=1\ldots n} | s \in G_{\mathscr{R}}\}. \tag{74}$$

Those combinations are consistent as much as $(\mathbf{p}_i)_{i=1\ldots n}$, since symmetry operations are morphisms (proposition 4). Hence the uniqueness up to symmetry of a consistent tuple of representatives. □

Lemma 1 *For any* $\mathbf{G} \in G$, $\mathbf{G}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\mathbf{G}$.

Proof Let us consider $\mathbf{G} \in G$. By definition of $\boldsymbol{\Lambda}^2$,

$$\mathbf{G}\boldsymbol{\Lambda}^2 = \frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\mathbf{G}\mathbf{x}\mathbf{x}^\top ds. \tag{75}$$

Performing the change of variable $\mathbf{x} \leftarrow \mathbf{G}\mathbf{x}$ enables to rewrite this expression into:

$$\frac{1}{S}\int_{\mathbf{G}(\mathscr{S})} \mu(\mathbf{G}^{-1}\mathbf{x})\mathbf{x}(\mathbf{G}^{-1}\mathbf{x})^\top ds. \tag{76}$$

Thanks to the invariance of $\mathscr{S}$ and μ to the proper symmetries of the object, we exhibit back $\boldsymbol{\Lambda}^2$ as follows:

$$\begin{aligned}\mathbf{G}\boldsymbol{\Lambda}^2 &= \frac{1}{S}\int_{\mathscr{S}} \mu(\mathbf{x})\mathbf{x}\mathbf{x}^\top\mathbf{G}^{-\top} ds \\ &= \boldsymbol{\Lambda}^2\mathbf{G}^{-\top}.\end{aligned} \tag{77}$$

$\mathbf{G}$ being a rotation, $\mathbf{G}^{-\top} = \mathbf{G}$, and therefore we showed that $\mathbf{G}$ and $\boldsymbol{\Lambda}^2$ commute:

$$\mathbf{G}\boldsymbol{\Lambda}^2 = \boldsymbol{\Lambda}^2\mathbf{G}. \tag{78}$$

Moreover, because $\boldsymbol{\Lambda}^2$ is a positive semi-definite symmetric matrix, it admits an eigenvalue decomposition

$$\boldsymbol{\Lambda}^2 = \mathbf{U}\mathbf{D}\mathbf{U}^\top \tag{79}$$

where $\mathbf{U} \in SO(3)$ and $\mathbf{D}$ is a positive semi-definite diagonal matrix. Injecting this decomposition into the right hand side of equation 78, we observe that $\mathbf{G}^\top\mathbf{U}$ is also an eigenbasis of $\boldsymbol{\Lambda}^2$:

$$\boldsymbol{\Lambda}^2 = (\mathbf{G}^\top\mathbf{U})\mathbf{D}(\mathbf{G}^\top\mathbf{U})^\top. \tag{80}$$

$\boldsymbol{\Lambda}$ being the principal square root of $\boldsymbol{\Lambda}^2$, both share the same eigenspaces, thus:

$$\begin{cases} \boldsymbol{\Lambda} = \mathbf{U}\mathbf{D}^{1/2}\mathbf{U}^\top \\ \boldsymbol{\Lambda} = (\mathbf{G}^\top\mathbf{U})\mathbf{D}^{1/2}(\mathbf{G}^\top\mathbf{U})^\top. \end{cases} \tag{81}$$

Therefore, substituing $\mathbf{U}\mathbf{D}^{1/2}\mathbf{U}^\top$ within the second equality by $\boldsymbol{\Lambda}$, we proved that

$$\boldsymbol{\Lambda} = \mathbf{G}^\top\boldsymbol{\Lambda}\mathbf{G} \tag{82}$$

i.e., that $\mathbf{G}$ and $\boldsymbol{\Lambda}$ commute:

$$\mathbf{G}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\mathbf{G}. \tag{83}$$

□

Proposition 6 (Invariance of the projection under symmetry of representatives) *Let* $\mathbf{x} \in \mathbb{R}^N$ *be a point of the ambient space, and* $s \in G_{\mathscr{R}}$. *The projection of* $\mathbf{x}$ *and its symmetric* $s(\mathbf{x})$ *corresponds to the same pose:*

$$\text{proj}(s(\mathbf{x})) = \text{proj}(\mathbf{x}). \tag{84}$$

Proof This result can be easily verified in the case of a revolution object with rotoreflection symmetry or a 2D cyclic object. Therefore for the sake of conciseness, we only discuss the case of an object with a finite proper symmetry group. Let $\mathbf{x} \in \mathbb{R}^{12}$ be a point of the ambient space, and $s_{\mathbf{G}} \in G_{\mathscr{R}}$, where $\mathbf{G} \in G$. We split $\mathbf{x}$ into two parts $\mathbf{M} \in \mathscr{M}_{3,3}(\mathbb{R})$ and $\mathbf{t} \in \mathbb{R}^3$ such as

$$\mathbf{x} = (\text{vec}(\mathbf{M})^\top, \mathbf{t}^\top)^\top. \tag{85}$$

The symmetric of $\mathbf{x}$ can thus by definition be written as

$$s_{\mathbf{G}}(\mathbf{x}) = (\text{vec}(\mathbf{M}\mathbf{G})^\top, \mathbf{t}^\top)^\top. \tag{86}$$

The projection of $\mathbf{x}$ onto the pose space consists in the pose $[\hat{\mathbf{R}}, \mathbf{t}]$, with $\hat{\mathbf{R}} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, considering a SVD decomposition $\mathbf{M}\boldsymbol{\Lambda} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ and using the same conventions for $\mathbf{U}, \mathbf{V}, \mathbf{S}$ and $\mathbf{D}$ than in subsection 7.3 in which we detailed this result.

Similarly, the projection of $s_{\mathbf{G}}(\mathbf{x})$ can be deduced from a SVD decomposition of $\mathbf{M}\mathbf{G}\boldsymbol{\Lambda}$. We know from lemma 1 that this latter term can be rearranged into

$$\mathbf{M}\mathbf{G}\boldsymbol{\Lambda} = \mathbf{M}\boldsymbol{\Lambda}\mathbf{G}. \tag{87}$$

Thus, injecting the previous decomposition into this expression enables us to exhibit a SVD decomposition of $\mathbf{M}\mathbf{G}\boldsymbol{\Lambda}$:

$$\begin{aligned}\mathbf{M}\mathbf{G}\boldsymbol{\Lambda} &= \mathbf{U}\mathbf{D}\mathbf{V}^\top\mathbf{G} \\ &= \mathbf{U}\mathbf{D}\tilde{\mathbf{V}}^\top\end{aligned} \tag{88}$$

where $\tilde{\mathbf{V}} = \mathbf{G}^\top\mathbf{V}$. Because G is a rotation matrix,

$$\begin{aligned}\det(\tilde{\mathbf{V}}) &= \det(\mathbf{G})\det(\mathbf{V}) \\ &= \det(\mathbf{V})\end{aligned} \tag{89}$$

and the projection of $s_{\mathbf{G}}(\mathbf{x})$ is therefore

$$\begin{aligned}\text{proj}(s_{\mathbf{G}}(\mathbf{x})) &= [\mathbf{U}\mathbf{S}\tilde{\mathbf{V}}^\top, \mathbf{t}] \\ &= [\hat{\mathbf{R}}\mathbf{G}, \mathbf{t}].\end{aligned} \tag{90}$$

Since $\mathbf{G}$ is a proper symmetry of the object,

$$[\hat{\mathbf{R}}\mathbf{G}, \mathbf{t}] = [\hat{\mathbf{R}}, \mathbf{t}] \tag{91}$$

which concludes the proof. □

Definition 5 (Mean estimation) Given a consistent tuple $(\mathbf{p}_i)_{i=1\dots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ of representatives of the poses $S = \{\mathscr{P}_i\}_{i=1\dots n}$, we define as estimation of the mean of those poses

$$\widehat{\text{mean}}(S) \triangleq \text{proj}\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right). \tag{92}$$

Proof We show here that this expression is well-defined, i.e. that it does depends on the consistent tuple of representatives considered. Let $(\mathbf{p}_i)_{i=1\dots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ be a consistent tuple of representatives. The consistent tuples are the tuples symmetric to this one (proposition 5):

$$\{(s(\mathbf{p}_i))_{i=1\dots n} | s \in G_{\mathscr{R}}\}. \tag{93}$$

Let us therefore consider an arbitrary consistent tuple $(s(\mathbf{p}_i))_{i=1\dots n}$, with $s \in G_{\mathscr{R}}$ and show that it leads to the same estimation $\mathscr{M}$ of the average pose than an estimation performed with $(\mathbf{p}_i)_{i=1\dots n}$.

By definition,

$$\mathscr{M} = \text{proj}\left(\frac{\sum_i w_i s(\mathbf{p}_i)}{\sum_i w_i}\right). \tag{94}$$

Because of the linearity of symmetries (proposition 3), the arithmetic mean of $(s(\mathbf{p}_i))_{i=1\dots n}$ corresponds to the symmetric of the arithmetic mean of $(\mathbf{p}_i)_{i=1\dots n}$:

$$\frac{\sum_i w_i s(\mathbf{p}_i)}{\sum_i w_i} = s\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right), \tag{95}$$

hence this expression of $\mathscr{M}$:

$$\mathscr{M} = \text{proj}\left(s\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right)\right). \tag{96}$$

Invariance of the projection under symmetry of representatives (proposition 6) enables to conclude this proof:

$$\mathscr{M} = \text{proj}\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right). \tag{97}$$

□

8.3 Sufficient conditions of consistency

In the previous section we showed how the average of poses could be estimated easily as the projection onto the pose space of the arithmetic average of a consistent combination of representatives of the considered poses. However, the consistency of a combination it is not trivial to establish in the general case. In this section, we provide simple sufficient conditions for a combination of representatives to be consistent and therefore enable an efficient and unambiguous estimation of the average pose.

Definition 6 (Distance between representatives) Let T be the minimum distance between different representatives of the same pose, with the convention $T = +\infty$ if a pose admits a single representative:

$$\forall \mathscr{P} \in \mathscr{C}, \forall \mathbf{p} \in \mathscr{R}(\mathscr{P}), T \triangleq \min_{\mathbf{q} \in \mathscr{R}(\mathscr{P}), \mathbf{q} \neq \mathbf{p}} \|\mathbf{q} - \mathbf{p}\| \tag{98}$$

T can be computed considering an arbitrary pose $\mathscr{P}$ - because of the invariance of our underlying metric to the choice of a reference pose - and considering an arbitrary representative $\mathbf{p} \in \mathscr{R}(\mathscr{P})$ because of the symmetry between representatives (see section 6).

Proposition 7 (Close-enough representatives) *Let $(\mathbf{p}_i)_{i=1\dots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ be a tuple of pose representatives. If its elements are closer one another than $T/2$, i.e. if*

$$\forall (i,j) \in [\![1,n]\!]^2, \|\mathbf{p}_i - \mathbf{p}_j\| < T/2, \tag{99}$$

then this tuple is consistent.

Proof Let us consider a tuple $(\mathbf{p}_i)_{i=1\dots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ that satisfies the condition (99). For any $(i,j) \in [\![1,n]\!]^2$ and $\mathbf{q}_j \in \mathscr{R}(\mathscr{P}_j) \setminus \{\mathbf{p}_j\}$, the below properties hold true:

$$\begin{cases} \|\mathbf{q}_j - \mathbf{p}_j\| \leq \|\mathbf{p}_j - \mathbf{p}_i\| + \|\mathbf{q}_j - \mathbf{p}_i\| & \text{(triangle inequality)} \\ \|\mathbf{q}_j - \mathbf{p}_j\| \geq T & \text{(definition 6)} \\ \|\mathbf{p}_j - \mathbf{p}_i\| < T/2. & \text{(condition (99))} \end{cases} \tag{100}$$

From those inequalities, we deduce that

$$\|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{q}_j - \mathbf{p}_i\|, \tag{101}$$

hence the consistency of $(\mathbf{p}_i)_{i=1\dots n}$. □

A particular case of practical value is the following, illustrated on figure 3d:

Proposition 8 (Representatives within a ball) *Let $(\mathbf{p}_i)_{i=1\dots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ be a tuple of pose representatives. If those representatives lie in a ball of radius $T/4$, i.e. if*

$$\exists \mathbf{c} \in \mathbb{R}^N / \forall i \in [\![1,n]\!]^2, \|\mathbf{p}_i - \mathbf{c}\| < T/4, \tag{102}$$

then this tuple is consistent.

We exploit this property in our application example section 10.

Proof The condition of proposition 7 is satisfied because of the triangle inequality, since for any $(i,j) \in [\![1,n]\!]^2$,

$$\begin{aligned} \|\mathbf{p}_i - \mathbf{p}_j\| &\leq \|\mathbf{p}_i - \mathbf{c}\| + \|\mathbf{p}_j - \mathbf{c}\| \\ &< T/4 + T/4. \end{aligned} \tag{103}$$

□

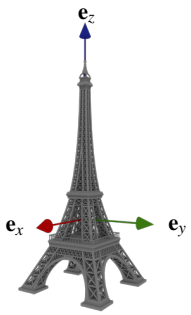


Fig. 4 Object frame centered on its centroid, here aligned with its principal axes.

9 Metric properties

We illustrate some of the properties of our metric with a toy example, using a 3D model of the Eiffel tower scaled to the actual tower dimensions.

We consider this object to be invariant under rotation of 90° about its central vertical axis, and we ensured this property for numerical computations by merging its triangular mesh model with its images by rotation of ±90° and 180° around the symmetry axis as it was slightly non symmetric.

The centroid and the covariance matrix of the surface of the object can be computed easily by summing the contributions of the different triangles of the mesh (see appendix C). We thus choose an object frame centered on the object centroid, and while it is not required in our formulation we consider axes aligned with the principal axes of the covariance matrix such as depicted on figure 4, as it makes the principal square root of the covariance matrix diagonal, and eases the interpretations:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & \lambda_z \end{pmatrix}, \tag{104}$$

with $\lambda_x = \lambda_y = 24.8\,\mathrm{m}$, and $\lambda_z = 68.0\,\mathrm{m}$ the standard deviations of the surface of the object along respectively the $\mathbf{e}_x$, $\mathbf{e}_y$ and $\mathbf{e}_z$ axes, depicted on figure 4.

The proper symmetry group of the object is finite of cardinal 4 and can be expressed in the chosen frame as

$$G = \left\{ \mathbf{I}, \mathbf{R}_z^{\pi/2}, \mathbf{R}_z^{\pi}, \mathbf{R}_z^{-\pi/2} \right\}. \tag{105}$$

Within our framework, a pose $\mathscr{P} = [\mathbf{R}, \mathbf{t}]$ can therefore be represented by four different 12D representatives:

$$\mathscr{R}(\mathscr{P}) = \left\{ (\mathrm{vec}(\mathbf{RG}\boldsymbol{\Lambda})^\top, \mathbf{t}^\top)^\top, \mathbf{G} \in G \right\}. \tag{106}$$

The distance between two poses $\mathscr{P}_1$ and $\mathscr{P}_2$ can thus be computed via four Euclidean distance computations in $\mathbb{R}^{12}$, given an arbitrary representative $\mathbf{p}_1 \in \mathscr{R}(\mathscr{P}_1)$ of pose $\mathscr{P}_1$:

$$\mathrm{d}(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{p}_2 \in \mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2 - \mathbf{p}_1\|$$

As we proved in 5.2, the square distance can be decomposed into a translation part and an orientation part. The translation part corresponds to the canonical Euclidean distance with which most people are familiar with. Therefore, we focus only on pure rotations around the center of mass of the object to illustrate two main properties of our metric.

9.1 Rotation anisotropy

As opposed to usual distances that are solely dependent on the angle of the relative rotation between two poses (when considering an object without proper symmetry), ours also depends on the considered axis. We illustrate this property on figure 5, for two couples of poses linked by a smallest displacement consisting of a rotation of 15° around different axes, respectively $\mathbf{e}_z$ and $\mathbf{e}_x$ (subfigures (a) and (b)). While the angle of the relative rotation is identical in both cases, displacements of surface points are quite different and we visually tend to consider poses in case (b) as being farther away one another than in case (a). Our framework formalizes this intuition, resulting in a distance between the poses in configuration (b) approximatively 2.1 times greater than the one between the poses in configuration (a). This factor actually corresponds to the square root of the moments of inertia ratio of the respective rotation axes: $\sqrt{I_x/I_z}$, with $I_z = \lambda_x^2 + \lambda_y^2$ and $I_x = \lambda_y^2 + \lambda_z^2$.

More precisely, the rotation part of the distance between two poses of an object without invariance depends on the angle θ and on the inertia moment $I_{\mathbf{k}}$ of the axis $\mathbf{k}$ of the relative rotation between the two poses. Injecting Rodrigues' rotation formula in our distance expression, we can indeed obtain the following equality:

$$\|\mathbf{R}_2\boldsymbol{\Lambda} - \mathbf{R}_1\boldsymbol{\Lambda}\|_F = 2\sin\left(\frac{\theta}{2}\right)\sqrt{I_{\mathbf{k}}} \quad \text{where } I_{\mathbf{k}} = \frac{1}{S}\int \mu(\mathbf{x})\|\mathbf{k}\times\mathbf{x}\|^2 ds. \tag{107}$$

9.2 Proper symmetry of the object

By construction, our distance takes into account the proper symmetries of the object. We illustrate this point on figure 6, where we consider the distance between a reference pose and its rotation about the symmetry axis for several angles. Since the two poses are identical for a rotation of 90°, we observe in this case a null distance between the two.

10 Mean Shift for pose recovery

We propose an example of application on the problem of pose estimation for instances of a rigid object. We perform

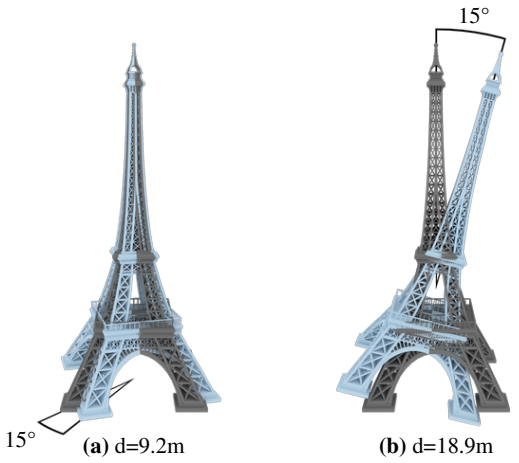


Fig. 5 With typical metrics, the distances between the two poses in cases (a) and in (b) are equal – as in both cases the two poses are linked by a rotation of 15° around the center of mass of the object. Our distance accounts for the object geometry and discriminates these two configurations.

experiments on three different objects among those shown in table 1, with different proper symmetries to illustrate the versatility of our approach:

- the Stanford bunny – an object without proper symmetry invariance.
- a candlestick – considered as a revolution object without rotoreflection invariance.
- a cartoon-like space rocket – whose proper symmetry group is the finite cyclic group of generator a rotation of 120° along the object axis.

We use as input data a synthetic depth map corresponding to a view of multiple instances of the considered object in different poses, as represented on figure 7a. Our goal consists in recovering the poses of these instances. Among the existing work adapted to such modality, some popular approaches (Drost et al 2010; Fanelli et al 2011; Tejani et al 2014) process in a bottom-up approach, generating votes for poses candidates in a Hough-like manner and, identifying those votes to the sampling of a pose distribution, look for its main modes which hopefully correspond to the actual poses of object instances. We place ourselves within such a modes-seeking framework.

Modes detection in a distribution on the pose space is not an easy problem. Grid-based accumulation techniques traditionally used in Hough-like methods are unpractical due to the high dimension of the pose space, except through the use of a sparse structure (Rodrigues et al 2012) or solely as a preprocessing technique used on a few dimensions (Drost et al 2010; Rodrigues et al 2012; Tejani et al 2014). A popular approach for modes detection more adapted to high dimensional problems is Mean-Shift, a local and non parametric iterative method based on a kernel density estimation of the probability distribution. Unfortunately, this method is designed for vector spaces, which the pose space is not. Fanelli et al (2011) and Tejani et al (2014) used nonetheless Mean Shift with a global parametrization of the pose space, but such approach suffers from the intrinsic drawbacks we evoked in the introduction, and illustrate in section 10.1. Tuzel et al (2005) and later Subbarao and Meer (2006) proposed versions of Mean Shift for Lie groups and Riemannian manifolds through the use of local parametrization that might circumvent those issues, but their approach is computationally expensive as it requires at each iteration to map the samples to the local tangent plane of the point to shift, compute the shift vector through the classical Mean Shift procedure and map it back to get the updated point.

In this example, we show how standard Mean Shift algorithm can be adapted to perform modes detection on the pose space quite efficiently through the use of our distance, even for objects showing proper symmetries properties.

Given an input depth map, we use our own implementation of the method of Drost et al (2010) – a simple method based on geometric features of pairs of 3D points and their associated normals – to generate a set of votes for object poses and associated weights $\{\mathscr{P}_i, w_i\}_{i=1,\dots,n}$. In our experiments, we use a sampling rate of $\tau_d = 0.025$ and consider every samples as reference points.

We consider a starting pose for Mean Shift at each of those votes. A usual practice to speed up drastically computations when seeking modes is to consider only a subset taken from the votes as starting points e.g. by random sampling, but we do not use such approach here to avoid the introduction of additional parameters.

For each of the poses to shift, we process iteratively following the usual Mean Shift procedure. Considering a flat Mean Shift kernel of radius r, we find the poses and associated weights within the set of votes that are within a radius r of the pose, compute their weighted average, update the pose to shift to this averaged pose and repeat until convergence.

In our experiments, we choose arbitrarily the radius r of our kernel to correspond to the smallest eigenvalue of the matrix $\boldsymbol{\Lambda}$ for the bunny and the candlestick – a value which roughly corresponds to half the smallest typical dimension of the object. For the rocket, we use a smaller radius of $\sqrt{3}/2$ times this eigenvalue, which is the greatest value that satisfies the condition of proposition 8 (see appendix B for details). Bigger radius values also experimentally give good results, but do not provide the same theoretical guarantees.

Indeed, such choice ensures that the condition $r < T/4$ is satisfied, where T is the minimum distance between representatives of the same pose (see definition 6). We know from section 5.1 that given a representative $\mathbf{p}$ of the pose $\mathscr{P}$ to shift, poses closer than r from $\mathscr{P}$ are the poses who have one representative within a ball of radius r around $\mathbf{p}$. These representatives can be retrieved efficiently through an off-the-shelf *radius search* method. Moreover, because r is cho-

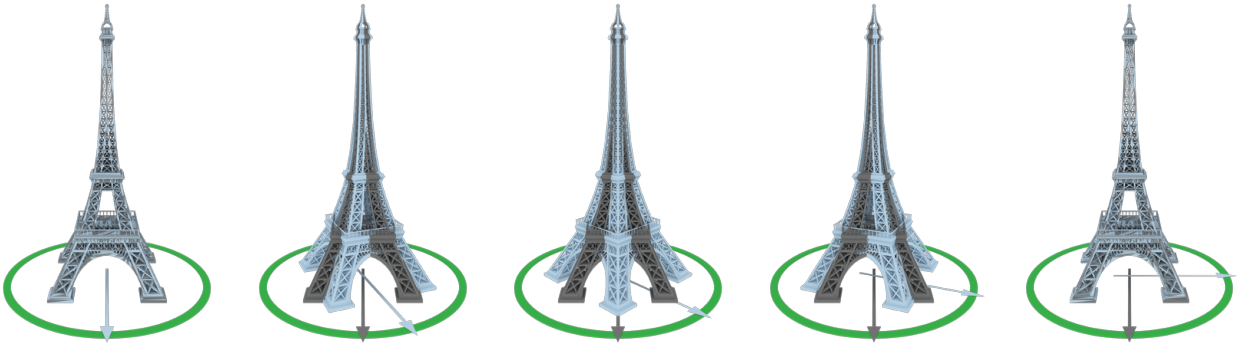


Fig. 6 Our distance takes into account the proper symmetries of the object: distance between a reference pose (dark gray) and a pose (light blue) generated by the rotation about the symmetry axis of the reference one, for several angles.

sen strictly smaller than $T/2$, the representatives retrieved by such query necessarily correspond to different poses and there are therefore no duplicates. Furthermore, these representatives lie in a ball of radius $T/4$, therefore we have the insurance from proposition 8 that we can unambiguously estimate the average of the corresponding poses as the projection on the pose space of the arithmetic mean of these representatives.

As a consequence, adapting the Mean Shift procedure to our pose space given the chosen radius only requires an additional step compared to the usual procedure in a vector space. This step consists in projecting the arithmetic mean of the retrieved representatives on the pose space and can be easily performed as we described in section 7. Pseudo-code of the adapted Mean Shift algorithm is proposed in algorithm 1.

Algorithm 1 Mean Shift algorithm within the pose space

Input: $(\mathscr{P}_i, w_i)_{i=1,\dots,n}$ a set of poses and associated weights,
$\mathbf{p}_{\text{in}}$ a representative of the pose to shift,
r the Mean Shift radius.
Output: A representative of the shifted pose

$R \leftarrow GetRepresentatives((\mathscr{P}_i)_{i=1,\dots,n})$
\# Preprocessing step independent of $\mathbf{p}_{\text{in}}$
\# R contains all poses representatives:
\# $\forall i \in [\![1,n]\!], \{R[i,j]\}_{j=1,\dots,|\mathscr{R}(\bullet)|} = \mathscr{R}(\mathscr{P}_i)$
$\mathbf{p} \leftarrow \mathbf{p}_{in}$
repeat
  $\mathbf{p}_{\text{old}} \leftarrow \mathbf{p}$
  $\mathscr{N} \leftarrow \text{RadiusSearch}(R, \mathbf{p}, r)$
  if $\mathscr{N} \neq \emptyset$ **then**
    $\mathbf{m} \leftarrow \left(\sum_{(i,j)\in\mathscr{N}} w_i \mathbf{R}[i,j]\right) / \left(\sum_{(i,j)\in\mathscr{N}} w_i\right)$
    $\mathbf{p} \leftarrow \text{ClosestRepresentative}(\text{proj}(\mathbf{m}))$
  end if
until $\mathbf{p} \neq \mathbf{p}_{\text{old}}$
return $\mathbf{p}$

In our experiments, we observed that the projection of the mean at each iteration is not required for the poses to shift to meaningful modes. Therefore we actually perform the projection onto the pose space only once after convergence for each representative to shift. Visual results of our experiments are shown on figure 7, where the intensity of a pixel corresponds to the sum of weights of poses whose silhouette contain this pixel, the higher the darker (rows b, c and d). The initial set of votes for poses is quite spread, leading to a blurry representation (row b). Those poses are shifted through the Mean Shift algorithm to modes of the original poses distribution, leading to a visually sharper representation where the silhouettes of object instances emerge (column c).

Shifted poses converge to modes of the probability density, and after fusing together those having converged to the same poses to avoid duplicates, we can estimate the probability density up to a scaling factor at a mode $\mathscr{M}$ by kernel density estimation:

$$s(\mathscr{M}) = \sum_i w_i H\left(\frac{\mathrm{d}(\mathscr{M}, \mathscr{P}_i)}{r}\right) \tag{108}$$

with H the Epanechnikov kernel to which is associated the flat Mean Shift kernel (Fukunaga and Hostetler 1975):

$$H(d) = \begin{cases} \frac{3}{4}(1-d^2) & \text{if } |d| \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{109}$$

Based on this estimate, the most significant modes of the distribution can be extracted (row d). Those poses are assumed to be good pose estimations for the object instances and could be further refined through e.g. the ICP procedure (Besl and McKay 1992), or filtered by checking their consistency with the actual data.

Theoretical limitations: The probabilistic interpretation we used here is abusive and should only be considered as a way to give the intuition of the Mean Shift approach. Kernel density estimation over a Riemannian manifold has been mathematically studied by Pelletier (2005), but our approach

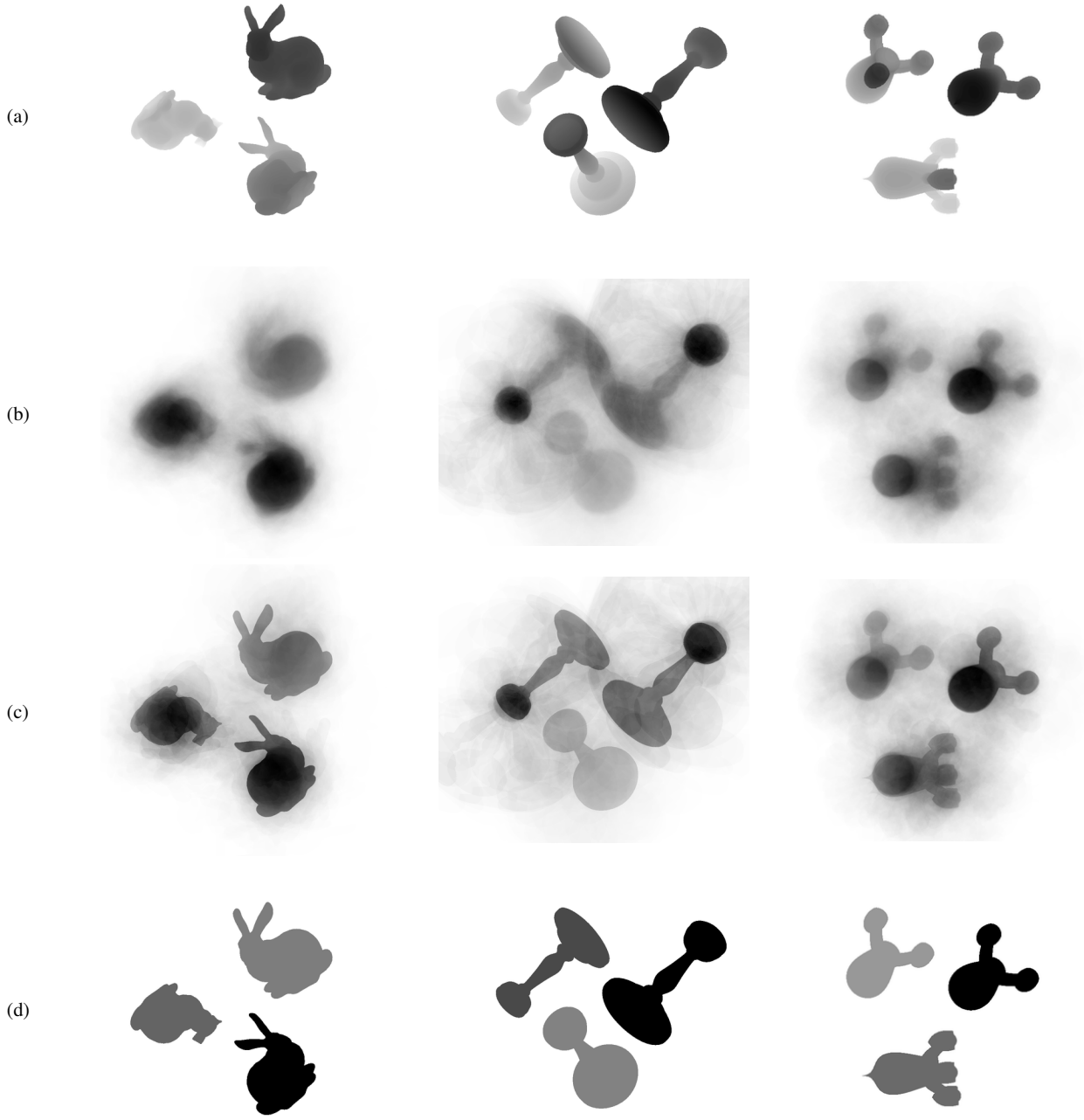


Fig. 7 Example of application: pose recovery by modes detection in a pose distribution. (a) Input depth map, (b) Pose distribution generated from the depth map with the method of Drost et al (2010), (c) Shifted votes using Mean Shift, (d) Main detected modes of the pose distribution.

does not enter into such framework. Some theoretical results might nonetheless be obtained, since as the number of pose samples increase, it becomes possible to consider a smaller Mean Shift radius and in the limit case of our distance tending towards 0, the proposed distance is most certainly equivalent to a Riemannian geodesic distance over the pose space – more precisely the one induced by the Riemannian tensor associated to the notion of kinetic energy (Zefran and Kumar 1996) in the case of an object without proper symmetry. Such considerations however, are out of the scope of this work, and we will only consider $s(\mathscr{M})$ as a score for the pose $\mathscr{M}$.

10.1 Limitations of a metric based on Euler angles

In the introduction, we mentioned the drawbacks of distances based on Euler angles representation to justify the need for a different metric. In this subsection, we illustrate those in the context of pose estimation of an object without proper symmetries.

An arbitrary rotation $\mathbf{R} \in SO(3)$ can be decomposed into a sequence of three consecutive rotations along three defined axis – a result attributed to Leonhard Euler. It can therefore be fully described by the three angles corresponding to those elementary rotations, usually referred to as Euler angles. In our experiments, we use a popular variant of this decomposition, known as Tait–Bryan angles, which describes a rotation $\mathbf{R}$ by three angles (α, β, γ) that satisfy

$$\mathbf{R} = \mathbf{R}_z^{\gamma} \mathbf{R}_y^{\beta} \mathbf{R}_x^{\alpha} \tag{110}$$

with

$$(\alpha, \beta, \gamma) \in]-\pi, \pi] \times [-\pi/2, \pi/2] \times]-\pi, \pi]. \tag{111}$$

Because of the concision of this description, rigid displacements – and therefore poses of rigid objects without proper symmetry (cf. section 2.2) – are commonly represented by a translation vector $\mathbf{t}$ and such three angles.

For the sake of consistency with our framework, we introduce a scaling factor r to balance between the rotation part and the translation part, and define a representative of a pose to be:

$$\mathscr{R}_{\text{Euler}}([\mathbf{R}, \mathbf{t}]) \triangleq (\mathbf{t}^\top, r(\alpha, \beta, \gamma))^\top. \tag{112}$$

It is tempting to define a distance between poses based on such representation as follows:

$$\mathrm{d}_{\text{Euler}}(\mathscr{P}_1, \mathscr{P}_2) \triangleq \|\mathscr{R}_{\text{Euler}}(\mathscr{P}_2) - \mathscr{R}_{\text{Euler}}(\mathscr{P}_1)\|. \tag{113}$$

Unfortunately, $\mathrm{d}_{\text{Euler}}$ is not well-defined – at least not on the whole pose space, because Euler angles parametrization is not unique. When $\beta = -\pi/2$ or $\pi/2$, $\mathbf{R}$ depends solely on respectively $\alpha + \gamma$ and $\alpha - \gamma$ and therefore the same pose can be described by a continuum of different Euler angles.

Moreover, this distance is not objective, in that it depends on the choice of inertial frame. We illustrate this phenomenon and its effects on figure 8 with the problem of pose estimation for a single instance of object, following the same pipeline as in section 10. Based on the same input pose distribution, we perform modes detection using our metric, and the pseudo-distance $\mathrm{d}_{\text{Euler}}$.

For our experiments, we choose the value of the scaling factor r to be:

$$r = \sqrt{\frac{2}{3}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}, \tag{114}$$

where $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are the eigenvalues of $\boldsymbol{\Lambda}$, in order to limit the bias due to this term. This choice is consistent with our proposed metric, in that the rotation part of the distance between poses near $\mathscr{P}_0$ is equivalent to respectively $r\theta$ for the distance (113) based on Euler angles, and $\sqrt{I_{\mathbf{k}}}\theta$ for ours, where θ is the angle of the relative rotation between the two poses, and $I_{\mathbf{k}}$ the inertia moment of the corresponding axis $\mathbf{k}$ (see equation (107)). Considering a typical value for the moment of inertia $2/3(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)$ enables to identify those two terms.

Using our metric (figure 8a), the main detected mode of the pose distribution is a good estimation of the actual pose of the object and stands out clearly, both in term of pose and score, of the other modes that are due to the noisy nature of the input pose distribution.

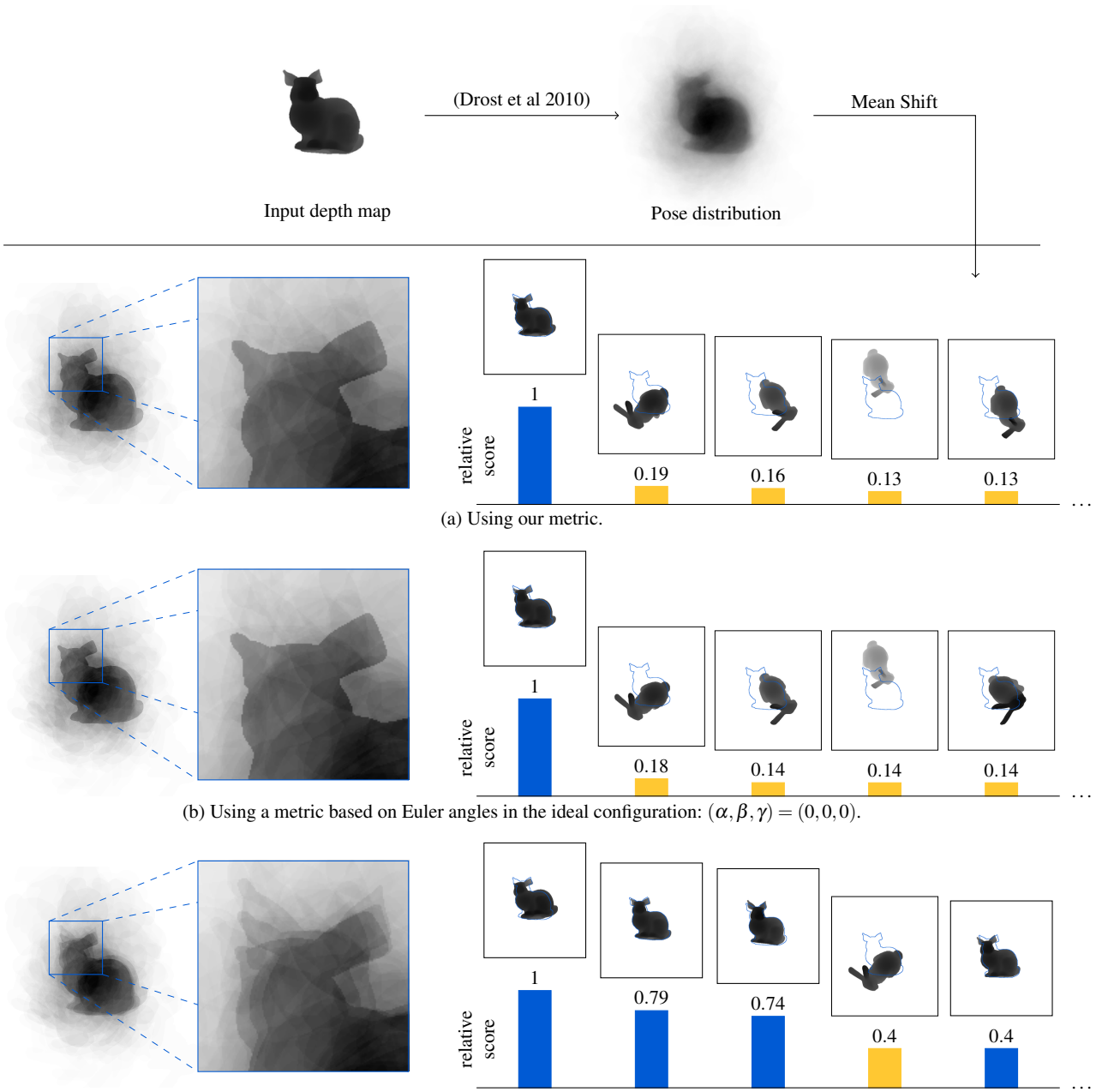
The distance based on Euler angles is dependent of the choice of orientation of the inertial and object frames, and this has an important impact on the result of the modes seeking algorithm. In a configuration where the inertial and object frames are chosen such as the actual orientation of the object is $(\alpha, \beta, \gamma) = (0, 0, 0)$ (figure 8b), we indeed obtain a good mode detection, with results similar to those obtained with our metric. This is due to the fact that the Euler angles of samples of the pose distribution corresponding to the main mode are around $(\alpha, \beta, \gamma) = (0, 0, 0)$, a configuration where the distance $\mathrm{d}_{\text{Euler}}$ is an approximation of the usual Riemannian metric (4).

However, in a configuration where the actual orientation of the object is $(\alpha, \beta, \gamma) = (\pi, 0, \pi)$ (figure 8c) votes are roughly split in four sets, depending on which side of the discontinuity $(\alpha, \beta, \gamma) = (\pm\pi, 0, \pm\pi)$ they are. This leads to the detection of four distinct modes corresponding approximately to the same pose. This is a source of potential duplicates that have to be filtered out in a post-processing step, but it also diminish the support of each modes from the votes of the pose distribution, making it harder to distinguish those from the noise and hence complicating the pose estimation.

11 Summary and discussion

In this paper, we addressed some issues of the existing notion of pose for a rigid object, both in the 2D and 3D case.

While poses of a rigid object are usually assumed to be equivalent to rigid transformations, we showed that this is not true in general. We therefore proposed a broader definition of the notion of pose, consisting in a distinguishable static state of the object. We showed that with this definition, a pose can be considered as an equivalence class of the space of rigid transformations, thanks to the introduction of the proper symmetry group of the object. We believe this notion to be essential, as lots of manufactured objects actually show some symmetry properties and could not be represented properly with the existing notion.



(a) Using our metric.

(b) Using a metric based on Euler angles in the ideal configuration: $(\alpha, \beta, \gamma) = (0, 0, 0)$.

(c) Using a metric based on Euler angles in a pathological configuration: $(\alpha, \beta, \gamma) = (\pm\pi, 0, \pm\pi)$.

Fig. 8 Comparison of the main modes retrieved using our metric and a metric based on an Euler angle parametrization, for pose estimation of an object with no invariance. Left: aggregation of silhouettes of the retrieved modes. Right: main modes retrieved (with the contour of the actual pose of the object superimposed in blue), sorted by descending score.

Based on this definition, we proposed a metric over the pose space as a measure of the smallest displacement between two poses, the length of a displacement consisting in the RMS displacement distance of surface points of the object. Besides being defined for any physical rigid object, such metric is interesting in that it does not depend on some arbitrary choice of frames or of scaling factors, while accounting for the geometry of the object.

With computation efficiency in mind, we proposed a coherent framework to represent poses in a low dimension Euclidean space, so as to enable efficient distance computations, neighborhood queries, and pose averaging, while providing theoretical proofs for those results.

Those developments enable the use of our metric for high level tasks such as pose recovery, where it appears to be a good replacement for a pseudo-metric currently used in some state-of-the-art methods, that is limited to objects

without proper symmetry and suffers from intrinsic drawbacks, as we illustrated experimentally.

Acknowledgements We based some of our illustrations on the following mesh models: "Stanford bunny", from the Stanford University Computer Graphics Laboratory; "Eiffel Tower" created by Pranav Panchal; and "Şamdan 2" (candlestick), from Metin N. Those were respectively available online at `http://graphics.stanford.edu/data/3Dscanrep`, and the GrabCAD and 3D Warehouse plateforms on May 2016.

A Distance simplification for a revolution object without rotoreflection invariance

Using the same definition of $\boldsymbol{\Lambda}$ as in section 5.3, the rotation part for a revolution object without rotoreflection invariance can be rewritten in the following way:

$$\begin{aligned}&\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1,\mathscr{P}_2)\\&=\min_{\phi_1,\phi_2}\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\|\mathbf{R}_2\mathbf{R}_z^{\phi_2}\mathbf{x}-\mathbf{R}_1\mathbf{R}_z^{\phi_1}\mathbf{x}\|^2 ds\\&=\min_{\phi_1,\phi_2}\|\mathbf{R}_2\mathbf{R}_z^{\phi_2}\boldsymbol{\Lambda}-\mathbf{R}_1\mathbf{R}_z^{\phi_1}\boldsymbol{\Lambda}\|_F^2.\end{aligned}\tag{115}$$

Frobenius norm being invariant under rotations, the expression (115) of the rotation part of the square distance for a revolution object without rotoreflection invariance can be rewritten with the relative rotation $\mathbf{R}\triangleq\mathbf{R}_1^{-1}\mathbf{R}_2$:

$$\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1,\mathscr{P}_2)=\min_{\phi_1,\phi_2}\|\mathbf{R}_z^{-\phi_1}\mathbf{R}\mathbf{R}_z^{\phi_2}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\|_F^2.\tag{116}$$

We consider a parametrization of $\mathbf{R}$ using Euler angles $(\tilde{\psi},\theta,\tilde{\phi})$ such as $\mathbf{R}=\mathbf{R}_z^{\tilde{\psi}}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\tilde{\phi}}$, considering the following elementary rotations:

$$\mathbf{R}_z^{\alpha}\triangleq\begin{pmatrix}\cos(\alpha)&-\sin(\alpha)&0\\\sin(\alpha)&\cos(\alpha)&0\\0&0&1\end{pmatrix},\mathbf{R}_x^{\alpha}\triangleq\begin{pmatrix}1&0&0\\0&\cos(\alpha)&-\sin(\alpha)\\0&\sin(\alpha)&\cos(\alpha)\end{pmatrix}.\tag{117}$$

Injecting this parametrization into the previous expression and performing the changes of variables $\psi\leftarrow\tilde{\psi}-\phi_1$ and $\phi\leftarrow\tilde{\phi}+\phi_2$ leads us to the following expression:

$$\begin{aligned}\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1,\mathscr{P}_2)&=\min_{\phi_1,\phi_2}\|\mathbf{R}_z^{-\phi_1}\mathbf{R}_z^{\tilde{\psi}}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\tilde{\phi}}\mathbf{R}_z^{\phi_2}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\|_F^2\\&=\min_{\psi,\phi}\|\mathbf{R}_z^{\psi}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\phi}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\|_F^2.\end{aligned}\tag{118}$$

Because of the specific shape of $\boldsymbol{\Lambda}$ (equation 28), the term to minimize can be decomposed into two parts:

$$\begin{aligned}\|\mathbf{R}_z^{\psi}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\phi}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\|_F^2&=\lambda_z^2\underbrace{\|\mathbf{R}_z^{\psi}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\phi}\mathbf{e}_z-\mathbf{e}_z\|^2}_{a_{\psi,\phi}}\\&+\lambda_r^2\underbrace{(\|\mathbf{R}_z^{\psi}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\phi}\mathbf{e}_x-\mathbf{e}_x\|^2+\|\mathbf{R}_z^{\psi}\mathbf{R}_x^{\theta}\mathbf{R}_z^{\phi}\mathbf{e}_y-\mathbf{e}_y\|^2)}_{b_{\psi,\phi}}.\end{aligned}\tag{119}$$

Developing this expression thanks to the definition of the elementary rotations (117), we evaluate those terms into:

$$\begin{cases}a_{\psi,\phi}=2(1-\cos(\theta))\\b_{\psi,\phi}=4-2\cos(\psi+\phi)(1+\cos(\theta))).\end{cases}\tag{120}$$

The first term is independent of ψ and ϕ. The second one can be minimized easily relatively to those two parameters, and admits a minimum that appears to be equal to the first term:

$$\min_{\psi,\phi}b_{\psi,\phi}=2(1-\cos(\theta)).\tag{121}$$

This result enables us to estimate the distance between the two poses in a closed form. However, having to refer to a relative rotation between the two poses and perform an Euler decomposition is cumbersome and would not enable to propose a representation of a pose efficient for neighborhood queries. We prefer instead to use the following property

$$\begin{aligned}2(1-\cos(\theta))&=\|\mathbf{R}\mathbf{e}_z-\mathbf{e}_z\|^2\\&=\|\mathbf{R}_2\mathbf{e}_z-\mathbf{R}_1\mathbf{e}_z\|^2\end{aligned}\tag{122}$$

in order to express the rotation part of the square distance as a function of the distance between the revolution axes of the object at the two poses:

$$\mathrm{d}^2_{\mathrm{rot}}(\mathscr{P}_1,\mathscr{P}_2)=(\lambda_r^2+\lambda_z^2)\|\mathbf{R}_2\mathbf{e}_z-\mathbf{R}_1\mathbf{e}_z\|^2.\tag{123}$$

B Minimum distance between representatives of the same pose

In this appendix, we show how to compute the minimum distance T between representatives of the same pose (see the definition 6) for the objects of our application example.

The bunny and the candlestick admit one representative per pose, hence $T=+\infty$ for those by convention.

The case of the rocket requires some calculus. For the sake of simplicity we consider an object frame whose z axis corresponds to the symmetry axis of the rocket. In this frame, the proper symmetry group of the rocket can be expressed as

$$G=\left\{\mathbf{I},\mathbf{R}_z^{2\pi/3},\mathbf{R}_z^{-2\pi/3}\right\}\tag{124}$$

and the square root of the covariance matrix as

$$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_r,\lambda_r,\lambda_z).\tag{125}$$

We choose to consider the reference pose $\mathscr{P}_0$ and one of its representatives $\mathbf{p}$ (underbraced below) for the computation of T as it makes the computation simpler. Representatives $\mathscr{R}(\mathscr{P}_0)$ of this pose are

$$\left\{\underbrace{\begin{pmatrix}\mathrm{vec}(\boldsymbol{\Lambda})\\\mathbf{0}_3\end{pmatrix}}_{\mathbf{p}},\begin{pmatrix}\mathrm{vec}(\mathbf{R}_z^{2\pi/3}\boldsymbol{\Lambda})\\\mathbf{0}_3\end{pmatrix},\begin{pmatrix}\mathrm{vec}(\mathbf{R}_z^{-2\pi/3}\boldsymbol{\Lambda})\\\mathbf{0}_3\end{pmatrix}\right\}.\tag{126}$$

Thanks to those choices, we can evaluate T into:

$$\begin{aligned}T&=\min_{\mathbf{q}\in\mathscr{R}(\mathscr{P}_0),\mathbf{q}\neq\mathbf{p}}\|\mathbf{q}-\mathbf{p}\|\\&=\min\|\mathbf{R}_z^{\pm2\pi/3}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\|_F\\&=\sqrt{6}\lambda_r.\end{aligned}\tag{127}$$

The threshold $\frac{T}{4}$ of proposition 8 therefore corresponds for the rocket to the value $\frac{\sqrt{3}}{2}\lambda_r$.

C Numerical recipes for a triangular mesh

Center of mass, area and covariance matrix of the surface of a triangular mesh $\mathscr{S} = \bigcup_i \mathscr{T}(\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i)$ – where $\mathscr{T}(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is a triangle defined by three vertices $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ – can be computed easily through the contributions of its triangles.

Let $\mathscr{T}(\mathbf{a}, \mathbf{b}, \mathbf{c})$ be a given triangle. Its area can be computed thanks to a cross product:

$$S_{\mathbf{a},\mathbf{b},\mathbf{c}} = \frac{\|(\mathbf{b}-\mathbf{a}) \times (\mathbf{c}-\mathbf{a})\|}{2}, \tag{128}$$

its center of mass through:

$$\mathbf{o}_{\mathbf{a},\mathbf{b},\mathbf{c}} = \frac{\mathbf{a}+\mathbf{b}+\mathbf{c}}{3}, \tag{129}$$

and its uncentered covariance matrix via:

$$\boldsymbol{\sigma}_{\mathbf{a},\mathbf{b},\mathbf{c}} = \frac{1}{12}\left(9\mathbf{o}_{\mathbf{a},\mathbf{b},\mathbf{c}}\mathbf{o}_{\mathbf{a},\mathbf{b},\mathbf{c}}^\top + \mathbf{a}\mathbf{a}^\top + \mathbf{b}\mathbf{b}^\top + \mathbf{c}\mathbf{c}^\top\right). \tag{130}$$

From those results, we deduce the expression of the surface area of the mesh:

$$S = \sum_i S_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}, \tag{131}$$

its center of mass:

$$\mathbf{o} = \sum_i S_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}\mathbf{o}_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i} \tag{132}$$

and its covariance matrix, if the center of mass of the mesh is chosen as origin of the object frame:

$$\boldsymbol{\Lambda}^2 = \frac{1}{S}\sum_i S_{\mathbf{a},\mathbf{b},\mathbf{c}}\boldsymbol{\sigma}_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}. \tag{133}$$

References

Angeles J (2006) Is there a characteristic length of a rigid-body displacement? Mechanism and Machine Theory 41(8):884–896

Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14(2):239–256, DOI 10.1109/34.121791

Chirikjian GS (2015) Partial Bi-Invariance of SE (3) metrics. Journal of Computing and Information Science in Engineering 15(1):011,008

Chirikjian GS, Zhou S (1998) Metrics on motion and deformation of solid models. Journal of Mechanical Design 120(2):252–261

Curtis W, Janin A, Zikan K (1993) A note on averaging rotations. In: , 1993 IEEE Virtual Reality Annual International Symposium, 1993, pp 377–385, DOI 10.1109/VRAIS.1993.380755

Di Gregorio R (2008) A novel point of view to define the distance between two rigid-body poses. In: Advances in Robot Kinematics: Analysis and Design, Springer, p 361–369

Drost B, Ulrich M, Navab N, Ilic S (2010) Model globally, match locally: Efficient and robust 3D object recognition. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, p 998–1005

Eberharter JK, Ravani B (2004) Local metrics for rigid body displacements. Journal of Mechanical Design 126(5):805–812, DOI 10.1115/1.1767816

Etzel KR, McCarthy JM (1996) A metric for spatial displacement using biquaternions on so (4). In: Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on, IEEE, vol 4, p 3185–3190

Fanelli G, Gall J, Van Gool L (2011) Real time head pose estimation with random regression forests. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 617–624, DOI 10.1109/CVPR.2011.5995458

Fukunaga K, Hostetler LD (1975) The estimation of the gradient of a density function, with applications in pattern recognition. Information Theory, IEEE Transactions on 21(1):32–40

Gramkow C (2001) On averaging rotations. Journal of Mathematical Imaging and Vision 15(1-2):7–16

Gupta KC (1997) Measures of positional error for a rigid body. Journal of mechanical design 119(3):346–348

Hinterstoisser S, Lepetit V, Ilic S, Holzer S, Bradski G, Konolige K, Navab N (2013) Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Computer Vision–ACCV 2012, Springer, p 548–562

Kazerounian K, Rastegar J (1992) Object norms: A class of coordinate and metric independent norms for displacements. Flexible Mechanisms, Dynamics, and Analysis ASME DE-Vol 47:271–275

Larochelle PM, Murray AP, Angeles J (2007) A distance metric for finite sets of rigid-body displacements via the polar decomposition. Journal of Mechanical Design 129(8):883–886

Lin Q, Burdick JW (2000) Objective and frame-invariant kinematic metric functions for rigid bodies. The International Journal of Robotics Research 19(6):612–625

Martinez JMR, Duffy J (1995) On the metrics of rigid body displacements for infinite and finite bodies. Journal of Mechanical Design 117(1):41–47

Muja M, Lowe DG (2009) Fast approximate nearest neighbors with automatic algorithm configuration. In: VISAPP (1), p 331–340

Park FC (1995) Distance metrics on the rigid-body motions with applications to mechanism design. Journal of Mechanical Design 117(1):48–54

Pelletier B (2005) Kernel density estimation on riemannian manifolds. Statistics & Probability Letters 73(3):297–304, DOI 10.1016/j.spl.2005.04.004

Purwar A, Ge QJ (2009) Reconciling distance metric methods for rigid body displacements. In: ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, p 1295–1304

Rodrigues JJ, Kim J, Furukawa M, Xavier J, Aguiar P, Kanade T (2012) 6D pose estimation of textureless shiny objects using random ferns for bin-picking. In: Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, IEEE, p 3334–3341

Schönemann PH (1966) A generalized solution of the orthogonal procrustes problem. Psychometrika 31(1):1–10

Sharf I, Wolf A, Rubin M (2010) Arithmetic and geometric solutions for average rigid-body rotation. Mechanism and Machine Theory 45(9):1239–1251, DOI 10.1016/j.mechmachtheory.2010.05.002

Subbarao R, Meer P (2006) Nonlinear mean shift for clustering over analytic manifolds. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, IEEE, vol 1, p 1168–1175

Sucan I, Moll M, Kavraki L (2012) The open motion planning library. IEEE Robotics Automation Magazine 19(4):72–82, DOI 10.1109/MRA.2012.2205651

Tejani A, Tang D, Kouskouridas R, Kim T (2014) Latent-Class hough forests for 3D object detection and pose estimation. In: Computer Vision–ECCV 2014, Springer, p 462–477

Tjaden H, Schwanecke U, Schömer E (2016) Real-Time monocular segmentation and pose tracking of multiple objects. In: European Conference on Computer Vision, Springer, p 423–438

Tuzel O, Subbarao R, Meer P (2005) Simultaneous multiple 3D motion estimation via mode finding on lie groups. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, IEEE, vol 1, p 18–25

Umeyama S (1991) Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on pattern analysis and machine intelligence 13(4):376–380

Vainsthein BK (1994) Fundamentals of Crystals. Springer Berlin Heidelberg, Berlin, Heidelberg

Zefran M, Kumar V (1996) Planning of smooth motions on SE (3). In: Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on, IEEE, vol 1, p 121–126