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

Romain Brégier, Frédéric Devernay, Laetitia Leyrit, James L. Crowley

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

Romain Brégier · Frédéric Devernay · Laetitia Leyrit · James L. Crowley

Abstract The pose of a rigid object is usually regarded as a rigid transformation, described by a translation and a rotation. However, equating the pose space with the space of rigid transformations is in general abusive, as it does not account for objects with proper symmetries – which are common among man-made objects. In this article, we define pose as a distinguishable static state of an object, and equate a pose with a set of rigid transformations. Based solely on geometric considerations, we propose a frame-invariant metric on the space of possible poses, valid for any physical rigid object, and requiring no arbitrary tuning. This distance can be evaluated efficiently using a representation of poses within a Euclidean space of at most 12 dimensions depending on the object's symmetries. This makes it possible to efficiently perform neighborhood queries such as *radius searches* or *k-nearest neighbor searches* within a large set of poses using off-the-shelf methods. Pose averaging considering this metric can similarly be performed easily, using a projection function from the Euclidean space onto the pose space. 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.

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

R. Brégier, F. Devernay, J. L. Crowley
Inria Grenoble Rhône-Alpes
UGA (Université Grenoble Alpes)

1 Introduction

Rigid body models play an important role in many technical and scientific fields, including physics, 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 $SE(3)$ makes the relative displacement of the object between two poses explicit and thus enables the definition of a distance between poses as the length of a shortest motion performing this displacement. This identification is 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). However while there already exists numerous works regarding $SE(3)$ metrics, choosing how to deal with poses of an object still remains challenging, as practitioners face questions such as how to tune the relative importance of position and orientation, even in the current deep learning era (Kendall et al 2015).

There are 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 based on a noisy set of votes is a good example of such a problem. While motion-based applications rely on local properties of the pose space which have been the subject of a large amount of research work, applications 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 theoretical interest. Consequently, similarity measures suffering from major flaws are still used in practical applications. Following the

work of Fanelli et al (2011), Tejani et al (2014) use a Mean Shift procedure based on the Euclidean distance between Euler angles as a representation of poses in their state-of-the-art object pose estimation method. Such a measure is fast to compute and enables the use of efficient tools developed for Euclidean spaces to perform the neighborhood queries and pose averaging required for Mean Shift, but it is not a distance. The parametrization of a rotation based on Euler angles notoriously suffers from border effects, singularities, and is dependent on the choice of frame. These 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 orientations. Such an example expresses the lack of tools for dealing efficiently with large sets of poses.

Lastly, there are cases where the pose of a rigid object cannot be identified as a single rigid transformation and therefore, for which existing results cannot be applied. Such cases occur when dealing with objects showing symmetry properties such as revolution objects or cuboids, and are, in fact, common among manufactured objects. The existing literature on object pose estimation does not usually discuss how such objects are handled, and the most widespread validation method used for symmetrical objects (Hinterstoisser et al 2012) consists in a relaxed similarity measure that cannot distinguish between poses such as a cylindrical can being flipped up or down.

Our goal in this paper is to address those issues by providing 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, equivalent to a set of rigid transformations (section 2). We then propose a physically meaningful distance over the pose space (section 4), and show how poses can be represented in a Euclidean space to enable fast distance computations and neighborhood queries (section 5). We show how the pose averaging problem can be solved quite efficiently (section 8) for this metric using a projection technique (section 7) and lastly we propose an example application for the problem of pose estimation of instances of a rigid object given a set of votes.

2 A definition for pose

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 rigid object is a distinguishable static state of this object.

We will refer to the set of possible poses as a *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 $SE(3)$

A 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 for this object.

A rigid transformation applied to 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 $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in SE(3)$ as 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, the rigid transformation corresponding to a given pose 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 displacements.

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 $\mathbf{T}$. For the *bunny* object figure 1d, M typically consists in the singleton $\{\mathbf{T}\}$. But M can also contain a continuum of poses in the case of a revolution object such as the *candlestick* figure 1a, or even be a discrete set, such as for the *rocket* object depicted figure 1e where the same pose can be represented by 3 different transformations.

By definition of M, $G \triangleq \{\mathbf{T}^{-1} \circ \mathbf{M}, \mathbf{M} \in M\}$ is the set of rigid transformations that have no effect on the static state of the object. This set therefore does not depend on the arbitrary transformation $\mathbf{T}$ considered. It is moreover a subgroup of $SE(3)$. Indeed, combinations and inversions of such transformations can be applied to the object while leaving 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 proper symmetry group

In the following, we propose a classification of the potential groups of proper symmetries for a physically meaningful bounded object. While models of infinite objects are commonly used e.g. for plane detection in 3D scene analysis, we do not consider those in this article as they do not correspond to actual physical objects and the definition of a suitable metric on the pose space of such objects is typically very dependent on the application. This classification will be helpful to derive the practical results associated with our proposed distance.

All proper symmetries of a bounded object necessarily have a common fixed point, thus we can consider the group of proper symmetries as a subgroup of the rotation group $SO(3)$ by choosing such a point as the origin of the object frame. 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, the potential groups of proper symmetries for a bounded object can actually be classified in a few categories.

In the 2D case, a bounded object will either show a circular symmetry – i.e. an invariance by any 2D rotation – or a cyclic symmetry of order $n \in \mathbb{N}^*$ – i.e. an invariance by rotation of $1/n$ turn. The special case $n = 1$ actually corresponds to a 2D object without any proper symmetry. Table 2 provides examples of such objects.

Similarly, we distinguish in the 3D case between five classes of proper symmetry groups, synthesized in table 1. A 3D bounded object can show a spherical symmetry – i.e. an invariance by any rotation – or a revolution symmetry – i.e. an invariance by rotation along a given axis of any angle. This latter class can actually be split into two, depending on whether the object is also invariant under reflection across a plane that is orthogonal to the revolution axis or not. We respectively refer to these classes as revolution symmetry with

or without rotoreflection invariance. In addition, we should also consider finite groups of proper symmetry, but there are an infinite number of them therefore they are considered in a general manner. We nevertheless distinguish the case of an object without proper symmetry (i.e. for which G contains only the identity transformation) from the other ones, because 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 rotational invariance of 180° along an arbitrary axis orthogonal to the revolution axis.

3 Prior work on metrics over the pose space

We propose in this section 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.

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

Infinite groups			Finite groups	
Revolution symmetry				
(a) Without rotoreflection invariance	(b) With rotoreflection invariance	(c) Spherical symmetry	(d) No proper symmetry	(e) 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

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 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.

For 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 $p \in [1, \infty]$:

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

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):

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

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, which can be evaluated quite easily e.g. from the following relations, using matrix or unit quaternion representations and respectively trace or inner product operators:

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

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

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:

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

Similar properties are obtained considering the Euclidean distance between representations of antipodal pairs of unit quaternions $\mathbf{q_1}$ and $\mathbf{q_2}$:

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

Merging position and orientation distances together requires setting the scaling factors a and b. The choice of those factors remains a heuristic issue, and the recent work of Kendall et al (2015) on camera pose regression using a deep neural network notably showed that this setting may have a great impact on performances. A reasonable choice in the case of object poses consists in 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), assuming 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.
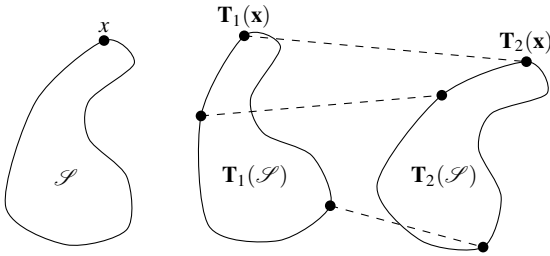
Fig. 1 Representation of corresponding points between instances at different poses of a rigid object without proper symmetries.

3.4 Geometric approaches

To avoid the need for arbitrary scaling factors, some distances are based only on geometric properties of the object. A particularly interesting possibility 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:

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

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 the use of the maximum displacement ($p = \infty$), and Hinterstoisser et al (2012) used the average displacement ($p = 1$) for a pose estimation evaluation. For the sake of tractability, those authors suggest to limit consideration to only some vertices of an object model since the integral has to be evaluated explicitly. Kazerounian and Rastegar (1992) on the other hand proposed the use of integral of square displacements ($p = 2$) on the whole object, and showed that it could 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 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 a 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 article 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, valid even for symmetric ones. This distance can be considered as an extension of the work of Kazerounian and Rastegar (1992) and Chirikjian and Zhou (1998) to arbitrary bounded objects. We also discuss some of its properties.

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 :

$$ d(\mathscr{P}_1, \mathscr{P}_2) \triangleq \min_{\mathbf{G}_1, \mathbf{G}_2 \in G} d_{\text{no\_sym}}(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2), $$

with

$$ d_{\text{no\_sym}}(\mathbf{T}_1, \mathbf{T}_2) \triangleq \sqrt{\frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \| \mathbf{T}_2(\mathbf{x}) - \mathbf{T}_1(\mathbf{x}) \|^2 ds} \qquad (10) $$

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 $d_{\text{no\_sym}}$. 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 d_{no_sym}, which is a direct consequence of Minkowski inequality. An equivalent formulation of this distance, involving a single minimization over G, is introduced in proposition 1.

In typical applications, one is particularly interested in the positioning of the surface of the object. Therefore in our experiments, we consider the surface of the object as set of points $\mathscr{S}$. 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 $\mu = 1$.

4.2 Objectiveness

The proposed distance is independent by construction of the choice of some arbitrary frames as it admits a purely geometric interpretation, a point we discuss in subsection 4.3.

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

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

Therefore, an arbitrary new reference pose $\mathscr{P}_3$ could be considered without any effect on the metric properties. Denoting $\mathbf{T}_3$ a rigid transformation identified to $\mathscr{P}_3$ relatively to the old reference pose $\mathscr{P}_0$, we verify the independence of d_{no_sym} from the choice of reference pose

$$d_{no\_sym}(\mathbf{T}_1, \mathbf{T}_2) = d_{no\_sym}(\mathbf{T}_3^{-1}\mathbf{T}_1, \mathbf{T}_3^{-1}\mathbf{T}_2), \qquad (12)$$

and hence the independence of the general distance:

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

4.3 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 these 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 d_{no_sym} as it enables efficient computations (see section 5):

$$d_{no\_sym}^2(\mathbf{T}_1, \mathbf{T}_2) = \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \|\mathbf{T}_2(\mathbf{x}) - \mathbf{T}_1(\mathbf{x})\|^2 ds. \qquad (14)$$

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):

$$d(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{G}_1, \mathbf{G}_2 \in G} d_{no\_sym}(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2). \qquad (15)$$

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 d_{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,*

$$d(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{G} \in G} d_{no\_sym}(\mathbf{T}_1, \mathbf{T}_2 \circ \mathbf{G}) \qquad (16)$$

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
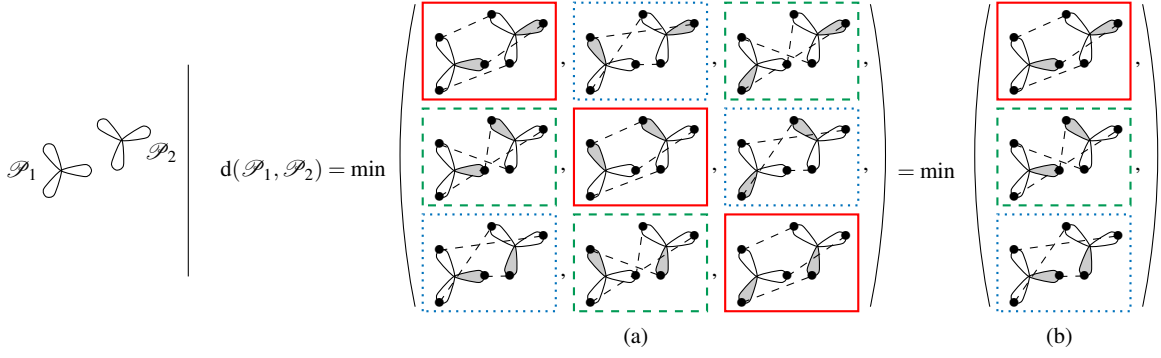
Fig. 2 Illustration of our proposed distance for a 2D object with a rotation symmetry of $2\pi/3$. (a) The distance between two poses consists in the minimum distance between two poses of an equivalent object without proper symmetry – here there are 3 possible poses of the equivalent object for each 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).

$x \leftarrow \mathbf{G}_1(x)$ and $\mathbf{G} \leftarrow \mathbf{G}_2 \circ \mathbf{G}_1^{-1}$ to write the following equality:

$$d_{\text{no\_sym}}^2(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2)$$

$$= \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \| \mathbf{T}_2 \circ \mathbf{G}_2(\mathbf{x}) - \mathbf{T}_1 \circ \mathbf{G}_1(\mathbf{x}) \|^2 ds \qquad (17)$$

$$= \frac{1}{S} \int_{\mathbf{G}_1(\mathscr{S})} \mu(\mathbf{G}_1^{-1}(\mathbf{x})) \| \mathbf{T}_2 \circ \mathbf{G}(\mathbf{x}) - \mathbf{T}_1(\mathbf{x}) \|^2 ds$$

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:

$$d_{\text{no\_sym}}^2(\mathbf{T}_1 \circ \mathbf{G}_1, \mathbf{T}_2 \circ \mathbf{G}_2)$$

$$= \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \| \mathbf{T}_2 \circ \mathbf{G}(\mathbf{x}) - \mathbf{T}_1(\mathbf{x}) \|^2 ds \qquad (18)$$

$$= d_{\text{no\_sym}}^2(\mathbf{T}_1, \mathbf{T}_2 \circ \mathbf{G}).$$

$\square$

4.4 Rotation anisotropy

In case of a pure rotational displacement around the center of mass of a non symmetric object, usual metrics are solely dependent on the angle of the relative rotation between the two poses. $d_{\text{no\_sym}}$ on the other hand – and the proposed distance as well – accounts for the object's geometry and as such also depends on the considered axis. More precisely, the distance between two poses linked by such a displacement depends on the angle θ and on the inertia moment $I_{\mathbf{k}}$ along the axis $\mathbf{k}$ of the relative rotation between the two poses, as follows:

$$d_{\text{no\_sym}}(\mathbf{T}_1, \mathbf{T}_2) = 2\sqrt{I_{\mathbf{k}}} \sin\left(\frac{\theta}{2}\right) \qquad (19)$$

where $I_{\mathbf{k}} = \frac{1}{S} \int \mu(\mathbf{x}) \| \mathbf{k} \times \mathbf{x} \|^2 ds$.

(a) d=9.2m   **(b)** d=18.9m
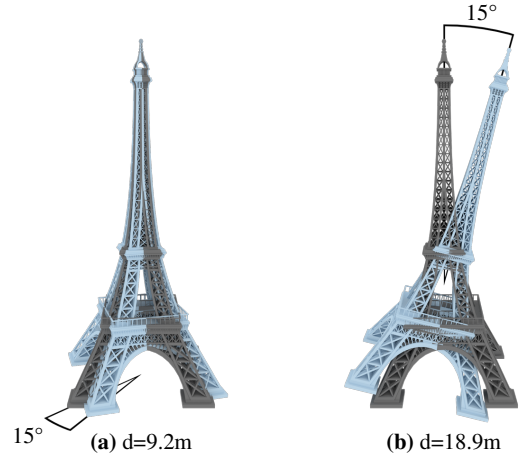
Fig. 3 Usual metrics would consider the distances between the two poses in cases (a) and in (b) equal – as in both cases the two poses are linked by a rotation of 15° around the center of mass of the object. Our distance accounts for the object geometry and discriminates these two configurations.

This result can be easily obtained by injecting Rodrigues' rotation formula in the expression of the proposed distance. We illustrate this property on figure 3 with an object consisting in a model of the Eiffel tower scaled to its actual dimensions, for two couples of poses linked by a smallest displacement consisting of a rotation of 15° around different axes. 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).

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 Euclidean space $\mathbb{R}^N$ of at most 12 dimensions, depending on the object's symmetries. 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,

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

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:

$$\forall \mathbf{p}_1 \in \mathscr{R}(\mathscr{P}_1), d(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{p}_2 \in \mathscr{R}(\mathscr{P}_2)} \|\mathbf{p}_2 - \mathbf{p}_1\|. \tag{21}$$

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.

The expression of pose representatives will be derived later in this section for the different classes of objects, and a synthesis is proposed in table 3 for 3D objects, and in table 4 for 2D ones.

5.1 Neighborhood query

The distance formulation (21) 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:

$$R = \bigcup_{\mathscr{P} \in S} \mathscr{R}(\mathscr{P}). \tag{22}$$

From (21), 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.

Such representatives can be retrieved 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 in an open ball of radius $T/2$, where T is a constant defined as follows:

Definition 3 (Min. distance between representatives)
We note T the minimum distance between different representatives of the same pose, with the convention $T = +\infty$ if a pose admits a single representative:

$$\forall \mathscr{P} \in \mathscr{C}, \forall \mathbf{p} \in \mathscr{R}(\mathscr{P}), T \triangleq \min_{\mathbf{q} \in \mathscr{R}(\mathscr{P}), \mathbf{q} \neq \mathbf{p}} \|\mathbf{q} - \mathbf{p}\|. \tag{23}$$

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 (see section 4.2), and considering an arbitrary representative $\mathbf{p} \in \mathscr{R}(\mathscr{P})$ because of the symmetry properties of representatives described in section 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. It is for example chosen at the center of the object for a spherical object, and on the revolution

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.		

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

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	$\{\mathbf{R}_z^\alpha \,\vert\, \alpha \in \mathbb{R}\}$	$(\lambda(\mathbf{Re}_z)^\top, \mathbf{t}^\top)^\top \in \mathbb{R}^6$
Revolution symmetry with rotoreflection invariance	$\{\mathbf{R}_x^\delta \mathbf{R}_z^\alpha \,\vert\, \delta \in \{0, \pi\}, \alpha \in \mathbb{R}\}$	$\{(\pm\lambda(\mathbf{Re}_z)^\top, \mathbf{t}^\top)^\top\} \subset \mathbb{R}^6$
No proper symmetry	$\{\mathbf{I}\}$	$(\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^\top, \mathbf{t}^\top)^\top \in \mathbb{R}^{12}$
Finite nontrivial	Finite	$\{(\mathrm{vec}(\mathbf{RG}\boldsymbol{\Lambda})^\top, \mathbf{t}^\top)^\top \,\vert\, \mathbf{G} \in G\} \subset \mathbb{R}^{12}$

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.		

$$\forall \alpha \in \mathbb{R}, e^{i\alpha} = (\cos(\alpha), \sin(\alpha)), \text{ and } \lambda \triangleq \left(\frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \|\mathbf{x}\|^2 ds \right)^{1/2}.$$

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 $n \in \mathbb{N}^*$)	$\{\mathbf{R}^{2k\pi/n} \,\vert\, k \in [\![0, n[\![\}$	$\{(\lambda e^{i(\theta + 2k\pi/n)}, \mathbf{t}^\top)^\top \,\vert\, k \in [\![0, n[\![\} \subset \mathbb{R}^4$

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:

$$\begin{aligned}
&\|(\mathbf{T}_2 \circ \mathbf{G}_2)(\mathbf{x}) - (\mathbf{T}_1 \circ \mathbf{G}_1)(\mathbf{x})\|^2 \\
&= \|\mathbf{R}_2 \mathbf{G}_2 \mathbf{x} + \mathbf{t}_2 - (\mathbf{R}_1 \mathbf{G}_1 \mathbf{x} + \mathbf{t}_1)\|^2 \\
&= \|\mathbf{R}_2 \mathbf{G}_2 \mathbf{x} - \mathbf{R}_1 \mathbf{G}_1 \mathbf{x}\|^2 + \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\
&\quad + 2(\mathbf{t}_2 - \mathbf{t}_1)^\top (\mathbf{R}_2 \mathbf{G}_2 - \mathbf{R}_1 \mathbf{G}_1) \mathbf{x}.
\end{aligned} \tag{24}$$

We add the further constraint that the origin of the object frame is the center of mass of the object's 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 (24) disappears during the integration, and the squared distance (10) can therefore be decomposed into a translation and a rota-

tion part:

$$\begin{aligned}
\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\
&+ \underbrace{\min_{\mathbf{G}_1, \mathbf{G}_2 \in G} \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \|\mathbf{R}_2 \mathbf{G}_2 \mathbf{x} - \mathbf{R}_1 \mathbf{G}_1 \mathbf{x}\|^2 ds}_{\mathrm{d}_{\mathrm{rot}}^2(\mathbf{R}_1, \mathbf{R}_2)}.
\end{aligned} \tag{25}$$

In the following subsections, 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 (25) can be expressed as follows:

$$\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) = \int_{\mathscr{S}} \mu(\mathbf{x}) \|\mathbf{R}_2 \mathbf{x} - \mathbf{R}_1 \mathbf{x}\|^2 ds \tag{26}$$

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:

$$\boldsymbol{\Lambda} \triangleq \left(\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\mathbf{x}\mathbf{x}^{\top}ds\right)^{1/2}. \tag{27}$$

$\boldsymbol{\Lambda}$ 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 (26) with a trace operator,

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

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

$$\begin{aligned}\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) &= \text{Tr}\left((\mathbf{R}_2 - \mathbf{R}_1)\boldsymbol{\Lambda}^2(\mathbf{R}_2 - \mathbf{R}_1)^{\top}\right) \\ &= \|\mathbf{R}_2\boldsymbol{\Lambda} - \mathbf{R}_1\boldsymbol{\Lambda}\|_F^2.\end{aligned} \tag{28}$$

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, such as

Object without proper symmetry:
$$\begin{aligned}\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{R}_2\boldsymbol{\Lambda} - \mathbf{R}_1\boldsymbol{\Lambda}\|_F^2 + \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\ &= \|\mathscr{R}(\mathscr{P}_2) - \mathscr{R}(\mathscr{P}_1)\|^2\end{aligned} \tag{29}$$
with $\mathscr{R}(\mathscr{P}) \triangleq \left(\text{vec}(\mathbf{R}\boldsymbol{\Lambda})^{\top}, \mathbf{t}^{\top}\right)^{\top} \in \mathbb{R}^{12}.$

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 the $\mathbf{e}_z$ axis, where ϕ is the angle of the considered rotation, and the proper symmetry group of the object consists in $G = \left\{\mathbf{R}_z^{\phi}\middle|\phi \in \mathbb{R}\right\}.$

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

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_r & 0 & 0 \\ 0 & \lambda_r & 0 \\ 0 & 0 & \lambda_z \end{pmatrix} \tag{30}$$

with $\lambda_r, \lambda_z \in \mathbb{R}^+$. This enables us to express the rotation part of the distance as a simple scaled distance between the revolution axes seen as 3D vectors:

$$\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) = \lambda^2\|\mathbf{R}_2\mathbf{e}_z - \mathbf{R}_1\mathbf{e}_z\|^2 \tag{31}$$

with $\lambda \triangleq \sqrt{\lambda_r^2 + \lambda_z^2}$. The reader is referred to 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, in order to efficiently evaluate distances:

Revolution object without rotoreflection invariance:
$$\begin{aligned}\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{t}_2 - \mathbf{t}_1\|^2 + \lambda^2\|\mathbf{R}_2\mathbf{e}_z - \mathbf{R}_1\mathbf{e}_z\|^2 \\ &= \|\mathscr{R}(\mathscr{P}_2) - \mathscr{R}(\mathscr{P}_1)\|^2\end{aligned} \tag{32}$$
with $\mathscr{R}(\mathscr{P}) \triangleq \left(\lambda(\mathbf{R}\mathbf{e}_z)^{\top}, \mathbf{t}^{\top}\right)^{\top} \in \mathbb{R}^6$
where $\lambda = \sqrt{\lambda_r^2 + \lambda_z^2}.$

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 (25) can thus be rewritten as follows:

$$\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{R}_1, \mathbf{R}_2}\left(\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\|\mathbf{R}_2\mathbf{x} - \mathbf{R}_1\mathbf{x}\|^2 ds\right). \tag{33}$$

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 also isometrically embedded into a $\mathbb{R}^3$ by representing a pose by the position of its center:

Spherical object:
$$\begin{aligned}\mathrm{d}^2(\mathscr{P}_1, \mathscr{P}_2) &= \|\mathbf{t}_2 - \mathbf{t}_1\|^2 \\ &= \|\mathscr{R}(\mathscr{P}_2) - \mathscr{R}(\mathscr{P}_1)\|^2\end{aligned} \tag{34}$$
With $\mathscr{R}(\mathscr{P}) = \mathbf{t} \in \mathbb{R}^3.$

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:

$$G = \left\{ \mathbf{R}_x^{\delta} \mathbf{R}_z^{\alpha} \mid \alpha \in \mathbb{R}, \delta \in \{0, \pi\} \right\}. \tag{35}$$

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

$$\min_{\delta_1, \delta_2, \phi_1, \phi_2} \mathrm{d}_{\mathrm{no\_sym}} \left((\mathbf{R}_1 \mathbf{R}_x^{\delta_1} \mathbf{R}_z^{\phi_1}, \mathbf{t}_1), (\mathbf{R}_2 \mathbf{R}_x^{\delta_2} \mathbf{R}_z^{\phi_2}, \mathbf{t}_2) \right). \tag{36}$$

We discussed in section 5.4 how to compute such an expression relatively to the symmetries along the revolution axis. Therefore using result (32), our distance can be rewritten as the minimum Euclidean distance between 6D points, two being assigned to each pose:

$$\mathrm{d}(\mathscr{P}_1, \mathscr{P}_2) = \min_{\delta_1, \delta_2 \in \{0, \pi\}} \| \mathbf{p}_2^{\delta_2} - \mathbf{p}_1^{\delta_1} \| \tag{37}$$

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 $\delta = 0, \pi$ and $i = 1, 2$.

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:

Revolution object with rotoreflection invariance:
$$\mathrm{d}(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{p}_1 \in \mathscr{R}(\mathscr{P}_1), \mathbf{p}_2 \in \mathscr{R}(\mathscr{P}_2)} \| \mathbf{p}_2 - \mathbf{p}_1 \| \tag{38}$$
With $\mathscr{R}(\mathscr{P}) \triangleq \left\{ \left(\pm \lambda (\mathbf{R} \mathbf{e}_z)^{\top}, \mathbf{t}^{\top} \right)^{\top} \right\} \subset \mathbb{R}^6.$

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, such as the object depicted in table 1e. The proposed distance between two poses of such an object can be written as:

$$\min_{\mathbf{G}_1, \mathbf{G}_2 \in G} \mathrm{d}_{\mathrm{no\_sym}} ((\mathbf{R}_1 \mathbf{G}_1, \mathbf{t}_1), (\mathbf{R}_2 \mathbf{G}_2, \mathbf{t}_2)). \tag{39}$$

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:

Object with a nontrivial finite proper symmetry group:
$$\mathrm{d}(\mathscr{P}_1, \mathscr{P}_2) = \min_{\mathbf{p}_1 \in \mathscr{R}(\mathscr{P}_1), \mathbf{p}_2 \in \mathscr{R}(\mathscr{P}_2)} \| \mathbf{p}_2 - \mathbf{p}_1 \|$$
With $\mathscr{R}(\mathscr{P}) \triangleq \left\{ \left(\mathrm{vec}(\mathbf{R} \mathbf{G} \boldsymbol{\Lambda})^{\top}, \mathbf{t}^{\top} \right)^{\top} \middle| \mathbf{G} \in G \right\} \subset \mathbb{R}^{12}.$
$$\tag{40}$$

5.8 2D object

The notion of pose representative can be applied 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 (25) and the expression of the rotation part as a Frobenius norm (28) are still valid in the 2D case, but they can be even further simplified. Indeed, a 2D rotation matrix can be parametrized by an angle θ as follows:

$$\mathbf{R}^{\theta} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{41}$$

Introducing the elements of the covariance matrix

$$\boldsymbol{\Lambda}^2 = \begin{pmatrix} \lambda_{xx}^2 & \lambda_{xy}^2 \\ \lambda_{xy}^2 & \lambda_{yy}^2 \end{pmatrix}, \tag{42}$$

the rotation part can be simplified into

$$\begin{aligned} \mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) &= \mathrm{Tr} \left((\mathbf{R}^{\theta_2} - \mathbf{R}^{\theta_1}) \boldsymbol{\Lambda}^2 (\mathbf{R}^{\theta_2} - \mathbf{R}^{\theta_1})^{\top} \right) \\ &= (\lambda_{xx}^2 + \lambda_{yy}^2) \| e^{i\theta_2} - e^{i\theta_1} \|^2 \end{aligned} \tag{43}$$

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:

$$
\boxed{
\begin{aligned}
&\text{2D object without proper symmetry:}\\
&d(\mathscr{P}_1,\mathscr{P}_2)=\min_{\mathbf{p}_1\in\mathscr{R}(\mathscr{P}_1),\mathbf{p}_2\in\mathscr{R}(\mathscr{P}_2)}\|\mathbf{p}_2-\mathbf{p}_1\|\\
&\text{with }\mathscr{R}(\mathscr{P})\triangleq\left(\lambda e^{i\theta},\mathbf{t}^{\top}\right)^{\top}\in\mathbb{R}^4\\
&\text{where }\lambda=\left(\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\|\mathbf{x}\|^2 ds\right)^{1/2}.
\end{aligned}}
\tag{44}
$$

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. Formally, for a given object, we define a finite group of symmetry operations $G_{\mathscr{R}}$ on the ambient space $\mathbb{R}^N$. This group consists in the identity singleton in the case of objects admitting a single representative per pose, and is defined in table 5 for the other objects classes. In this section we discuss some properties of this group. Those will be used in order to propose a method to properly average poses, in section 8.2.

First, we ensure that the proposed group is well defined:

Proposition 2 $G_{\mathscr{R}}$ *is a group for the composition operation.*

Proof This property derives directly from the group properties of $\mathbf{G}$, $\{1,-1\}$ and $\{e^{i2k\pi/n}|k\in[\![0,n]\!]\}$ for multiplication operations. $\qquad\square$

Then, we introduce the following lemma, which somehow expresses the fact that the geometry of the object is consistent with the object's symmetries:

Lemma 1 *For any proper symmetry* $\mathbf{G}\in G$, $\mathbf{G}$ *and* $\boldsymbol{\Lambda}$ *commute, i.e.* $\mathbf{G}\boldsymbol{\Lambda}=\boldsymbol{\Lambda}\mathbf{G}$.

Proof Let $\mathbf{G}$ be a proper symmetry in G. By definition of $\boldsymbol{\Lambda}^2$,

$$
\mathbf{G}\boldsymbol{\Lambda}^2=\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\mathbf{G}\mathbf{x}\mathbf{x}^{\top}ds.
\tag{45}
$$

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

$$
\frac{1}{S}\int_{\mathbf{G}(\mathscr{S})}\mu(\mathbf{G}^{-1}\mathbf{x})\mathbf{x}(\mathbf{G}^{-1}\mathbf{x})^{\top}ds.
\tag{46}
$$

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

$$
\begin{aligned}
\mathbf{G}\boldsymbol{\Lambda}^2&=\frac{1}{S}\int_{\mathscr{S}}\mu(\mathbf{x})\mathbf{x}\mathbf{x}^{\top}\mathbf{G}^{-\top}ds\\
&=\boldsymbol{\Lambda}^2\mathbf{G}^{-\top}.
\end{aligned}
\tag{47}
$$

$\mathbf{G}$ being a rotation, $\mathbf{G}^{-\top}=\mathbf{G}$, and therefore $\mathbf{G}$ and $\boldsymbol{\Lambda}^2$ commute, i.e.

$$
\mathbf{G}\boldsymbol{\Lambda}^2=\boldsymbol{\Lambda}^2\mathbf{G}.
\tag{48}
$$

Moreover, as a positive semi-definite symmetric matrix, $\boldsymbol{\Lambda}^2$ admits an eigenvalue decomposition $\boldsymbol{\Lambda}^2=\mathbf{U}\mathbf{D}\mathbf{U}^{\top}$, 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 48, we observe that $\mathbf{G}^{\top}\mathbf{U}$ is also an eigenbasis of $\boldsymbol{\Lambda}^2$:

$$
\boldsymbol{\Lambda}^2=(\mathbf{G}^{\top}\mathbf{U})\mathbf{D}(\mathbf{G}^{\top}\mathbf{U})^{\top}.
\tag{49}
$$

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

$$
\begin{cases}
\boldsymbol{\Lambda}=\mathbf{U}\mathbf{D}^{1/2}\mathbf{U}^{\top}\\
\boldsymbol{\Lambda}=(\mathbf{G}^{\top}\mathbf{U})\mathbf{D}^{1/2}(\mathbf{G}^{\top}\mathbf{U})^{\top}.
\end{cases}
\tag{50}
$$

Therefore, injecting the first equality into the second one, we proved that

$$
\boldsymbol{\Lambda}=\mathbf{G}^{\top}\boldsymbol{\Lambda}\mathbf{G}
\tag{51}
$$

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

Thanks to this lemma, it is now possible to exhibit the three following properties of those symmetries within the ambient space:

Proposition 3 $G_{\mathscr{R}}$ *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, i.e.*

$$
\{s(\mathbf{p})|s\in G_{\mathscr{R}}\}=\mathscr{R}(\mathscr{P}).
\tag{52}
$$

Proof This proposition is easily verified thanks to the expression of pose representatives tables 3 and 4. The only subtlety is the case of a 3D object with a finite proper symmetry group. In this case for any $\mathbf{G}\in G$, the symmetric by $s_{\mathbf{G}}$ of a pose representative $(\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^{\top},\mathbf{t}^{\top})^{\top}$, where $\mathbf{R}\in\mathscr{M}_{3,3}(\mathbb{R})$ and $\mathbf{t}\in\mathbb{R}^3$, can be expressed as

$$
s_{\mathbf{G}}\left((\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda})^{\top},\mathbf{t}^{\top})^{\top}\right)=(\mathrm{vec}(\mathbf{R}\boldsymbol{\Lambda}\mathbf{G})^{\top},\mathbf{t}^{\top})^{\top},
\tag{53}
$$

which according to lemma 1 is equal to $(\mathrm{vec}(\mathbf{R}\mathbf{G}\boldsymbol{\Lambda})^{\top},\mathbf{t}^{\top})^{\top}$. By definition of representatives for such object, it is therefore a representative of the same pose $\mathscr{P}$. $\qquad\square$

Table 5 Proposed symmetry operations on the ambient space for objects with multiple representatives per pose.

Notations
We decompose a point of the ambient space $\mathbb{R}^N$ into two parts as follows, depending on the dimension N of the space:
– $(\text{vec}(\mathbf{M})^\top, \mathbf{t}^\top)^\top$ for a 12D space, with $\mathbf{M} \in \mathcal{M}_{3,3}(\mathbb{R})$, and $\mathbf{t} \in \mathbb{R}^3$.
– $(\mathbf{a}^\top, \mathbf{t}^\top)^\top$ for a 6D space, with $\mathbf{a}, \mathbf{t} \in \mathbb{R}^3$.
– $(\mathbf{a}^\top, \mathbf{t}^\top)^\top$ for a 4D space, with $\mathbf{a}, \mathbf{t} \in \mathbb{R}^2$. **a**.
In the 4D case, we use the complex multiplication notation, assimilating **a** to a complex number.

Object type	Proper symmetry class	Symmetry group $G_\mathscr{R}$	Symmetry definition
3D	Finite	$\{s_\mathbf{G} \mid \mathbf{G} \in G\}$	$s_\mathbf{G} : \mathbb{R}^{12} \to \mathbb{R}^{12}$ $(\text{vec}(\mathbf{M})^\top, \mathbf{t}^\top)^\top \mapsto (\text{vec}(\mathbf{MG})^\top, \mathbf{t}^\top)^\top$
	Revolution with rotoreflection invariance	$\{s_{\text{rev},\delta} \mid \delta = \pm 1\}$	$s_{\text{rev},\delta} : \mathbb{R}^6 \to \mathbb{R}^6$ $(\mathbf{a}^\top, \mathbf{t}^\top)^\top \mapsto (\delta\mathbf{a}^\top, \mathbf{t}^\top)^\top$
2D	Cyclic (order $n \in \mathbb{N}^*$)	$\{s_{2\text{D},n,k} \mid k \in [\![0,n]\!]\}$	$s_{2\text{D},n,k} : \mathbb{R}^4 \to \mathbb{R}^4$ $(\mathbf{a}^\top, \mathbf{t}^\top)^\top \mapsto (e^{i2k\pi/n} \cdot \mathbf{a}, \mathbf{t}^\top)^\top.$

Proposition 4 *Elements of $G_\mathscr{R}$ are linear transformations of the ambient space, i.e. for $s \in G_\mathscr{R}$, and for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$, and $\alpha \in \mathbb{R}$,*

$$s(\mathbf{x}_1 + \alpha\mathbf{x}_2) = s(\mathbf{x}_1) + \alpha s(\mathbf{x}_2). \tag{54}$$

This proposition is a direct consequence of the definition of symmetries described table 5.

Proposition 5 *Elements of $G_\mathscr{R}$ are automorphisms of the ambient space: for any $s \in G_\mathscr{R}$, s is bijective, and for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$,*

$$\|s(\mathbf{x}_2) - s(\mathbf{x}_1)\| = \|\mathbf{x}_2 - \mathbf{x}_1\|. \tag{55}$$

Proof Bijectivity is straightforward since

– $(s_\mathbf{G})^{-1} = s_{\mathbf{G}^{-1}}$ for any $\mathbf{G} \in G$.
– $(s_{\text{rev},\delta})^{-1} = s_{\text{rev},\delta}$ for any $\delta \in \{-1, 1\}$.
– $(s_{2\text{D},n,k})^{-1} = s_{2\text{D},n,-k}$ for any $k \in \mathbb{N}$.

The morphism property comes from the linearity of those symmetry operations (proposition 4) and the fact they preserve the norm, since elements of $\mathbf{G}$, $\{1, -1\}$ and $\{e^{i2k\pi/n} \mid k \in [\![0,n]\!]\}$ are themselves of unit norm. $\qquad \square$

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 tables 3 and 4), 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 $\mathbf{x}$. The results of this section will be useful in section 8 to propose a method for pose averaging.

Definition 4 We define as projections of $\mathbf{x} \in \mathbb{R}^N$ the poses:

$$\text{proj}(\mathbf{x}) \triangleq \underset{\mathscr{P}}{\text{argmin}} \min_{\mathbf{p} \in \mathscr{R}(\mathscr{P})} \|\mathbf{p} - \mathbf{x}\|^2 \tag{56}$$

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 admits $\mathbf{x}$ for 3D coordinates.

7.2 Object of revolution

In the case of a revolution object 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:

$$\begin{aligned} \text{proj}(\mathbf{x}) &= \underset{\mathscr{P}}{\text{argmin}} \|\mathbf{x} - \mathscr{R}(\mathscr{P})\|^2 \\ &= \underset{\mathbf{a},\mathbf{t} \in \mathbb{R}^3/\|\mathbf{a}\|=1}{\text{argmin}} \left(\|\mathbf{x}_r - \lambda \mathbf{a}\|^2 + \|\mathbf{x}_t - \mathbf{t}\|^2 \right) \end{aligned} \quad (57)$$

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: $\mathbf{x} = (\text{vec}(\mathbf{X}_r)^\top, \mathbf{x}_t^\top)^\top$ 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:

$$\begin{aligned} \text{proj}(\mathbf{x}) &= \underset{\mathscr{P}}{\text{argmin}} \|\mathbf{x} - \mathscr{R}(\mathscr{P})\|^2 \\ &= \underset{\mathbf{R},\mathbf{t}}{\text{argmin}} \left(\|\mathbf{X}_r - \mathbf{R}\boldsymbol{\Lambda}\|_F^2 + \|\mathbf{x}_t - \mathbf{t}\|^2 \right) \end{aligned} \quad (58)$$

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

$$\mathbf{D} = \text{diag}(\alpha_1, \alpha_2, \alpha_3), \quad (59)$$

with $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq 0$ and

$$\mathbf{S} = \begin{cases} \mathbf{I} & \text{if } \det(\mathbf{U})\det(\mathbf{V}) > 0 \\ \text{diag}(1,1,-1) & \text{otherwise.} \end{cases} \quad (60)$$

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 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 great 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:

$$\text{mean}(S) \triangleq \underset{\mathscr{P} \in \mathscr{C}}{\text{argmin}} \Phi(\mathscr{P}), \quad (61)$$

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

$$\Phi(\mathscr{P}) \triangleq \sum_{i=1}^n w_i \mathrm{d}^2(\mathscr{P}_i, \mathscr{P}). \quad (62)$$

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 axes 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 admits closed form approximations which are "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:

$$\Phi(\mathscr{P}) = \sum_{i=1}^{n} w_i \min_{\mathbf{p}_i \in \mathscr{R}(\mathscr{P}_i), \mathbf{p} \in \mathscr{R}(\mathscr{P})} \|\mathbf{p}_i - \mathbf{p}\|^2. \quad (63)$$

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:

$$\sum_{i=1}^{n} w_i \|\mathbf{p}_i - \mathbf{p}\|^2 = \sum_i w_i \|\mathbf{p}_i - \mathbf{m}_P\|^2 \\ + \left(\sum_i w_i\right) \|\mathbf{p} - \mathbf{m}_P\|^2, \quad (64)$$

with the arithmetic mean

$$\mathbf{m}_P \triangleq \frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}. \quad (65)$$

The first term of (64) is independent of $\mathbf{p}$. Therefore, the problem of minimizing (64) 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:

$$\text{mean}(S) = \underset{\mathscr{P} \in \mathscr{A}}{\arg\min} \Phi(\mathscr{P}), \quad (66)$$

where

$$\mathscr{A} \triangleq \left\{ \text{proj}(\mathbf{m}_P) \,\middle|\, P \in \prod_i \mathscr{R}(\mathscr{P}_i) \right\}. \quad (67)$$

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 admit only one representative, a single tuple of representatives has to be considered. For those objects, the average pose exists and corresponds simply to the projection of the arithmetic average of their representatives:

$$\boxed{\text{mean}(S) = \text{proj}\left(\frac{\sum_i w_i \mathscr{R}(\mathscr{P}_i)}{\sum_i w_i}\right)} \quad (68)$$

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 4 of such a case, where three different choices for the initial pose lead to three different choices of "consistent" representatives combinations and therefore three different estimations of the average.

In this subsection, we propose an stricter definition of the *consistency* of a combination of representatives and prove that it enables an unambiguous estimation of the mean.

Definition 5 (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

$$\forall (i,j) \in [\![1,n]\!]^2, \forall \mathbf{q}_j \in \mathscr{R}(\mathscr{P}_j) \setminus \{\mathbf{p_j}\}, \\ \|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{q}_j - \mathbf{p}_i\|. \quad (69)$$

In other words, a consistent tuple is a set of pose representatives closer one another than to any other representatives.

Proposition 6 (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*

$$\{(s(\mathbf{p}_i))_{i=1\ldots n} | s \in G_{\mathscr{R}}\}. \quad (70)$$

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 $\mathbf{p}_j \neq \mathbf{q}_j$, and we know from the consistency definition of $(\mathbf{p}_i)_{i=1\ldots n}$ and $(\mathbf{q}_i)_{i=1\ldots n}$ that

$$\forall i \in [\![1,n]\!], \begin{cases} \|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{q}_j - \mathbf{p}_i\| \\ \|\mathbf{q}_j - \mathbf{q}_i\| < \|\mathbf{p}_j - \mathbf{q}_i\|. \end{cases} \quad (71)$$
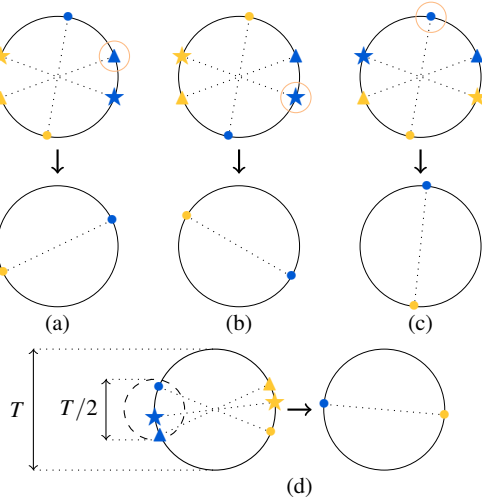
Fig. 4 Estimating the average of poses with multiple representatives: illustration with the orientation of a 2D object with a 180° rotation symmetry, that can be represented by a point on a circle, or the antipodal point. We consider three poses to average (triangle, star and disk shapes). (a, b, c) 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.

If there existed $i \in [\![1, n]\!]$ such as $\mathbf{p}_i = \mathbf{q}_i$, it would lead to the inequality $\|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{p}_j - \mathbf{p}_i\|$ which is a contradiction. The tuples $(\mathbf{p}_i)_{i=1...n}$ and $(\mathbf{q}_i)_{i=1...n}$ are thus pairwise disjoint, and there exists therefore at most $|\mathscr{R}(\bullet)|$ consistent tuples.

There are moreover exactly $|\mathscr{R}(\bullet)|$ different representative combinations symmetric to $(\mathbf{p}_i)_{i=1...n}$ – including itself (proposition 3):

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

Those combinations are consistent as much as $(\mathbf{p}_i)_{i=1...n}$, since symmetry operations are morphisms (proposition 5). Hence the uniqueness up to symmetry of a consistent tuple of representatives. $\qquad \square$

Proposition 7 (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 $s(\mathbf{x})$ correspond to the same pose:*

$$\text{proj}(s(\mathbf{x})) = \text{proj}(\mathbf{x}). \tag{73}$$

Proof This result can be easily verified in the case of a revolution object with rotoreflection symmetry or a 2D cyclic object. Therefore, 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

$$\mathbf{x} = (\text{vec}(\mathbf{M})^\top, \mathbf{t}^\top)^\top. \tag{74}$$

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

$$s_{\mathbf{G}}(\mathbf{x}) = (\text{vec}(\mathbf{MG})^\top, \mathbf{t}^\top)^\top. \tag{75}$$

The projection of $\mathbf{x}$ onto the pose space consists in the pose $[\hat{\mathbf{R}}, \mathbf{t}]$, with $\hat{\mathbf{R}} = \mathbf{USV}^\top$, considering a SVD decomposition $\mathbf{M\Lambda} = \mathbf{UDV}^\top$ 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 $\mathbf{MG\Lambda}$. We know from lemma 1 that this latter term can be rearranged into

$$\mathbf{MG\Lambda} = \mathbf{M\Lambda G}. \tag{76}$$

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

$$\begin{aligned} \mathbf{MG\Lambda} &= \mathbf{UDV}^\top \mathbf{G} \\ &= \mathbf{UD\tilde{V}}^\top \end{aligned} \tag{77}$$

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

$$\begin{aligned} \det(\tilde{\mathbf{V}}) &= \det(\mathbf{G})\det(\mathbf{V}) \\ &= \det(\mathbf{V}) \end{aligned} \tag{78}$$

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

$$\begin{aligned} \text{proj}(s_{\mathbf{G}}(\mathbf{x})) &= [\mathbf{US\tilde{V}}^\top, \mathbf{t}] \\ &= [\hat{\mathbf{R}}\mathbf{G}, \mathbf{t}]. \end{aligned} \tag{79}$$

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

$$[\hat{\mathbf{R}}\mathbf{G}, \mathbf{t}] = [\hat{\mathbf{R}}, \mathbf{t}] \tag{80}$$

which concludes the proof. $\qquad \square$

Based on those properties, it is possible to propose an unambiguous estimation of the mean, as follows:

Definition 6 (Mean estimation) Given a consistent tuple $(\mathbf{p}_i)_{i=1...n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ of representatives of the poses $S = \{\mathscr{P}_i\}_{i=1...n}$, we define as estimation of the mean of those poses

$$\widehat{\text{mean}}(S) \triangleq \text{proj}\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right). \tag{81}$$

Proof We show here that this expression is well-defined, i.e. that it does depends on the consistent tuple of representatives considered. Let $(\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ be a consistent tuple of representatives. The consistent tuples are the tuples symmetric to this one (proposition 6):

$$\{(s(\mathbf{p}_i))_{i=1\ldots n}|s \in G_{\mathscr{R}}\}. \tag{82}$$

Let us therefore consider an arbitrary consistent tuple $(s(\mathbf{p}_i))_{i=1\ldots n}$, 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 $(\mathbf{p}_i)_{i=1\ldots n}$.

By definition,

$$\mathscr{M} = \mathrm{proj}\left(\frac{\sum_i w_i s(\mathbf{p}_i)}{\sum_i w_i}\right). \tag{83}$$

Because of the linearity of symmetries (proposition 4), the arithmetic mean of $(s(\mathbf{p}_i))_{i=1\ldots n}$ corresponds to the symmetric of the arithmetic mean of $(\mathbf{p}_i)_{i=1\ldots n}$:

$$\frac{\sum_i w_i s(\mathbf{p}_i)}{\sum_i w_i} = s\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right), \tag{84}$$

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

$$\mathscr{M} = \mathrm{proj}\left(s\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right)\right). \tag{85}$$

Invariance of the projection under symmetry of representatives (proposition 7) enables to conclude this proof, since

$$\mathscr{M} = \mathrm{proj}\left(\frac{\sum_i w_i \mathbf{p}_i}{\sum_i w_i}\right). \tag{86}$$

$\square$

Such estimation corresponds most likely to the actual mean (61), unfortunately we do not have a proof of this conjecture.

8.3 Sufficient conditions of consistency

While the average of poses can be easily estimated given a consistent combination of representatives, there are however cases where no such combination exist, e.g. when trying to average poses spread out over the set of orientations such as illustrated on figure 4abc. In such a case, one might have to perform an exhaustive evaluation of the Fréchet variance for the different combinations in order to pick the one corresponding to the actual mean. Fortunately, this case is of limited practical interest, as the mean makes little sense.

Consistency is nonetheless not trivial to establish in the general case, and therefore in this section, we provide simple sufficient conditions for a combination of representatives to

be consistent. Consistency is in particular satisfied when the considered representatives are *close enough* one an other, relatively to the distance between their representatives:

Proposition 8 (Close-enough representatives) *Let* $(\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ *be a tuple of pose representatives. If its elements are closer one another than $T/2$, i.e. if*

$$\forall(i,j) \in [\![1,n]\!]^2, \|\mathbf{p}_i - \mathbf{p}_j\| < T/2, \tag{87}$$

then this tuple is consistent.

Proof Let us consider a tuple $(\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ that satisfies the condition (87). 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:

$$\begin{cases} \|\mathbf{q}_j - \mathbf{p}_j\| \le \|\mathbf{p}_j - \mathbf{p}_i\| + \|\mathbf{q}_j - \mathbf{p}_i\| & \text{(triangle inequality)} \\ \|\mathbf{q}_j - \mathbf{p}_j\| \ge T & \text{(definition 3)} \\ \|\mathbf{p}_j - \mathbf{p}_i\| < T/2. & \text{(condition (87))} \end{cases} \tag{88}$$

From those inequalities, we deduce that

$$\|\mathbf{p}_j - \mathbf{p}_i\| < \|\mathbf{q}_j - \mathbf{p}_i\|, \tag{89}$$

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

A special case of practical value of this criterion is obtained when the considered representatives are included in a ball small enough. It is illustrated on figure 4d, and is exploited in our application example section 10.

Proposition 9 (Representatives within a ball) *Let* $(\mathbf{p}_i)_{i=1\ldots n} \in \prod_i \mathscr{R}(\mathscr{P}_i)$ *be a tuple of pose representatives. If those representatives lie in a ball of radius $T/4$, i.e. if*

$$\exists \mathbf{c} \in \mathbb{R}^N / \forall i \in [\![1,n]\!]^2, \|\mathbf{p}_i - \mathbf{c}\| < T/4, \tag{90}$$

then this tuple is consistent.

Proof A tuple satisfying this condition also satisfies the one of proposition 8 because of the triangle inequality, since for any $(i,j) \in [\![1,n]\!]^2$,

$$\begin{aligned} \|\mathbf{p}_i - \mathbf{p}_j\| &\le \|\mathbf{p}_i - \mathbf{c}\| + \|\mathbf{p}_j - \mathbf{c}\| \\ &< T/4 + T/4. \end{aligned} \tag{91}$$

$\square$

These sufficient conditions can easily be generalized by considering that only the orientation parts of poses representatives have actually to be close enough one an other. This is a direct consequence to the fact that the pose space can be decomposed into a Cartesian product of a position and an orientation space, and that symmetry considerations only affect orientation for a bounded object.

9 Local properties

While we focus in this article on global metric properties, the proposed distance can be shown locally equivalent to a Riemannian metric over the pose space manifold. We therefore briefly discuss in this section those aspects.

Object with finite proper symmetry group In the case of a 3D object of finite proper symmetry group, the pose space can be seen as a manifold of dimension 6 (3 for translation, and 3 for rotation). Let us consider two poses of such an object, and $\mathbf{T}_1$ and $\mathbf{T}_2 \in SE(3)$ two associated rigid transformations, such as

$$\mathrm{d}([\mathbf{T}_1], [\mathbf{T}_2]) = \mathrm{d}_{\mathrm{no\_sym}}(\mathbf{T}_1, \mathbf{T}_2), \tag{92}$$

i.e. such as $\mathbf{T}_1^{-1} \circ \mathbf{T}_2$ to be a shortest displacement from pose $[\mathbf{T}_1]$ to pose $[\mathbf{T}_2]$. If the angle of the relative rotation between $\mathbf{T}_1$ and $\mathbf{T}_2$ is small, the corresponding displacement of a point $\mathbf{x} \in \mathbb{R}^3$ of the object between those two poses can be approximated by introducing the displacement vector $\mathbf{v} \in \mathbb{R}^3$ and the rotation vector $\omega \in \mathbb{R}^3$ between $\mathbf{T}_1$ and $\mathbf{T}_2$ as follows:

$$(\mathbf{T}_1^{-1} \circ \mathbf{T}_2)(\mathbf{x}) \underset{\theta \to 0}{\sim} \mathbf{x} + \omega \times \mathbf{x} + \mathbf{v}. \tag{93}$$

The distance between the two poses can then be approximated by

$$\mathrm{d}([\mathbf{T}_1], [\mathbf{T}_2]) \underset{\theta \to 0}{\sim} \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \|\omega \times \mathbf{x} + \mathbf{v}\|^2 ds. \tag{94}$$

When considering an infinitesimal displacement between $\mathbf{T}_1$ and $\mathbf{T}_2$, and assimilating $\mathbf{v}$ and ω to translational and angular velocities, this expression corresponds to the notion of kinetic energy (up to a factor $2/S$).

It is also a quadratic form, and therefore the proposed distance is locally equivalent to a Riemannian distance associated with the following metric tensor g, defined for any two tangent vectors $(\mathbf{v}_1^\top, \omega_1^\top)^\top$, $(\mathbf{v}_2^\top, \omega_2^\top)^\top \in \mathbb{R}^3 \times \mathbb{R}^3$ by

$$g\left((\mathbf{v}_1^\top, \omega_1^\top)^\top, (\mathbf{v}_2^\top, \omega_2^\top)^\top\right)$$
$$\triangleq \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x})(\omega_1 \times \mathbf{x} + \mathbf{v}_1)^\top (\omega_2 \times \mathbf{x} + \mathbf{v}_2) ds. \tag{95}$$

This Riemannian metric was already described in the litterature (Zefran and Kumar 1996; Lin and Burdick 2000), and Belta and Kumar (2002) notably suggested its use for interpolation on SE(3). If the covariance matrix considered for the object is isotropic – i.e. $\mathbf{\Lambda} = \lambda \mathbf{I}$, with $\lambda \in \mathbb{R}^{+*}$, which is the case e.g. for an object of spherical or cubic shape –, the proposed distance is moreover locally equivalent to the usual Riemannian distance (4) over $SE(3)$.

Other objects classes Similarly, the pose space of a 3D revolution object can be seen as a 5D manifold. The proposed distance for such an object is indeed locally equivalent to a Riemannian distance over $\mathbb{S}^2 \times \mathbb{R}^3$ induced by embedding $\mathbb{S}^2$ as a sphere in a 3D Euclidean space, and considering the usual Euclidean distance regarding the translation part.

Poses of a 2D object with a finite proper symmetry group lie likewise on a 3D manifold, and the proposed distance is locally equivalent to a Riemannian distance over $\mathbb{S}^1 \times \mathbb{R}^2$ induced by embedding $\mathbb{S}^1$ as a circle in a 2D Euclidean space, i.e. by considering as distance between two infinitesimally close orientations the angle of the relative rotation between those two (up to a scaling factor).

Finally, pose spaces of spherical 3D objects and circular 2D ones are respectively equivalent to $\mathbb{R}^3$ and $\mathbb{R}^2$, associated with the Euclidean distance.

Remarks Despite this local equivalence, the proposed pose spaces can be topologically different from the manifolds evoked above, because of the discrete symmetries of the object. As an example, the proposed pose space for a revolution object with rotoreflection invariance is actually homeomorphic with $\mathbb{RP}^2 \times \mathbb{R}^3$, where $\mathbb{RP}^2$ is the real projective plane – i.e. a sphere with antipodal points associated – instead of $\mathbb{S}^2 \times \mathbb{R}^3$.

Moreover, except for 3D spherical or 2D circular objects, the distance (10) is globally different from these Riemannian metrics. Compared to the proposed distance, those latter have several drawbacks.

They are indeed more expensive to estimate since they include trigonometric computations (e.g. for distance (4)). There is even no known closed-form expression for such distance in the case of an arbitrary 3D object of finite proper symmetry group. Moreover, they do not benefit from the same nice computational properties than the proposed distance regarding the problem of pose averaging, for which they may require iterative approaches (Pennec 1998). But more fundamentally, and as discussed in the introduction, our distance was proposed to quantify the *similarity* between poses at a global scale. The notion of *motion* between two poses, expressed in a Riemannian distance, is therefore irrelevant to the kind of applications we are interested in.

10 Application example

In this section, we illustrate the use of our metric on the problem of rigid object instances detection and pose estimation. Given an input depth image of a scene containing potentially multiple instances of a rigid object, our goal consists in recovering the poses of these instances. We perform experiments with three different objects of different symmetry classes among those shown in table 1 to illustrate the versatility of our approach:

- the Stanford bunny – an object without proper symmetry.
- a candlestick – considered as a revolution object without rotoreflection invariance.
- a cartoon-like space rocket – which is invariant by rotation of 120° along its axis.

For practical reasons, our example is based on synthetic 3D data, depicted in figure 5b. It was produced using an off-the-shelf stereo matching algorithm, on a pair of images of a virtual scene lit by a pseudo-random pattern projector (figure 5a). Those were synthesized with the rendering engine Blender Cycles (Blender Online Community 2016). The reader is referred to the work of Brégier et al (2017) for a more quantitative analysis of the interest of the proposed distance for pose estimation.

10.1 Mean Shift for pose recovery

Among the existing work adapted for a depth image input, 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. Tuzel et al (2005) and later Subbarao and Meer (2006) proposed versions of Mean Shift for Lie groups and Riemannian manifolds 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 generate a set of votes for object poses $\{\mathscr{P}_i\}_{i=1,...,n}$ thanks to our own implementation of the method of Drost et al (2010). It is based on a local aggregation scheme, and is performed by matching geometric features extracted from the input data with those extracted from a model of the object. We use a sampling rate of $\tau_d = 0.025$ and consider every samples as reference points – the interested reader is referred to the original description of Drost et al (2010) regarding the meaning of those parameters. This initial pose distribution is quite spread, as can be observed from the *blurred* effect of the representation figure 5c, column 1.

We then consider each of those votes as a starting pose for Mean Shift. 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 within the set of votes that are within a radius r of the current pose, compute their mean, shift the current pose to this mean and repeat until convergence.

We choose arbitrarily the radius r of our kernel to correspond to 1.5 times the smallest eigenvalue of the matrix $\boldsymbol{\Lambda}$ for the bunny and the candlestick – that is roughly 75% of 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 9 (see appendix B). A bigger radius value may also experimentally give good results, but does not provide the same theoretical guarantees.

Indeed, such choices of radii satisfy to the condition $r < T/4$, where T is the minimum distance between representatives of the same pose (see definition 3). 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 chosen 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 9 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, which can

(a) Stereo pair used for 3D reconstruction.
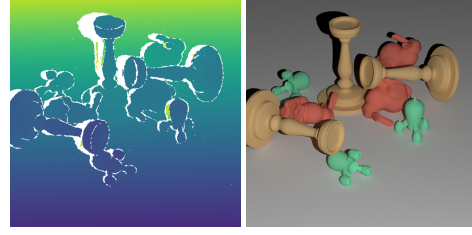
(b) Reconstructed 3D range data (RGB channel solely for visualization purposes).
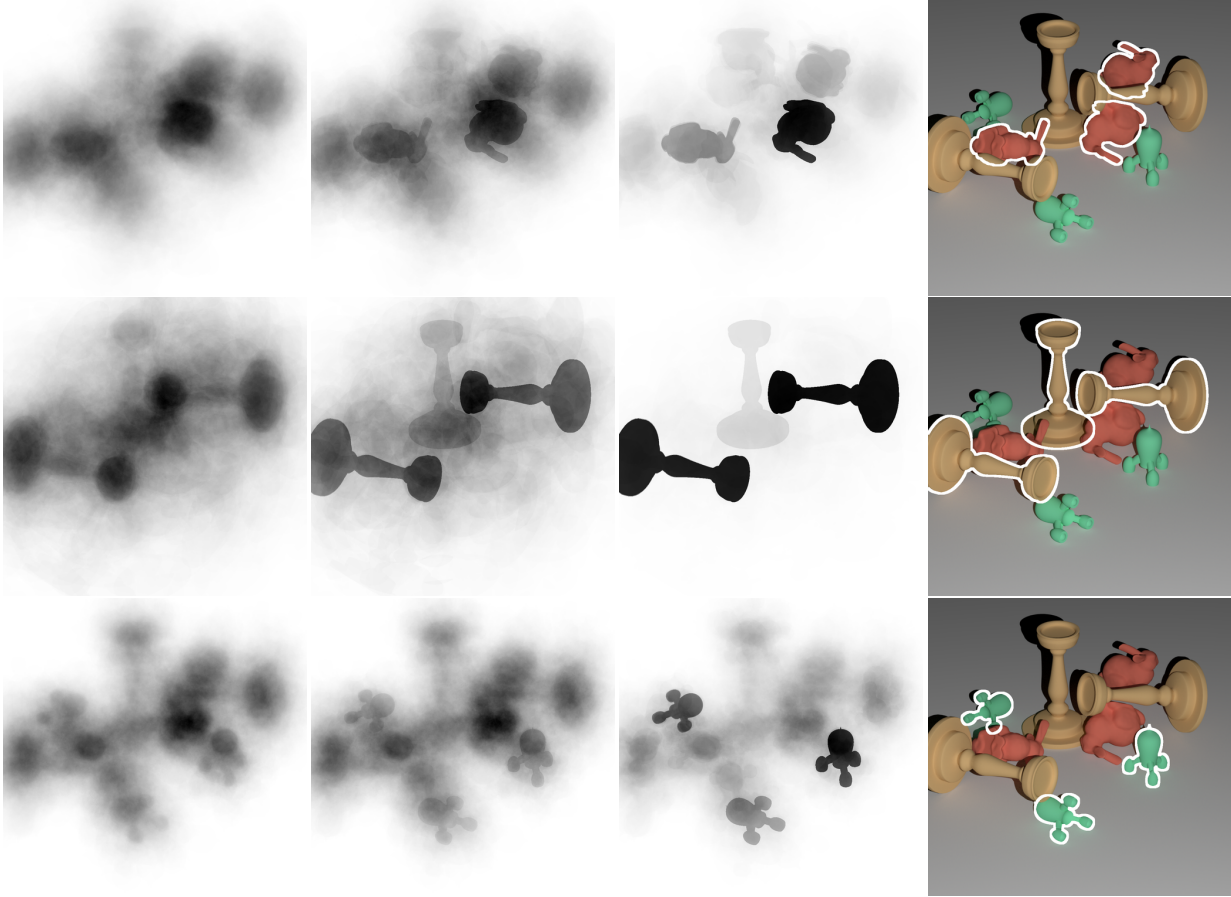
(c) From left to right: Pose distribution generated from the input 3D data using the method of Drost et al (2010). Shifted poses using Mean Shift. Shifted poses weighted by the density of the initial distribution. Recovered poses of object instances. A pose distribution is represented by accumulating the 2D silhouettes on the image plane of the object at its different poses (the more silhouettes a pixel belongs to, the darker it is). The process is performed independently for the three different objects.

Fig. 5 Application example: object instances detection and pose estimation via a Mean Shift procedure to extract the main modes of an initial pose distribution. Illustration with three different objects of various symmetry properties.

be performed as described in section 7. Pseudo-code of the adapted Mean Shift algorithm is proposed in algorithm 1.

The projection of the mean at each iteration is actually not required for the poses to shift towards meaningful modes in practice, and therefore we perform it only once after convergence. The pose distribution obtained after Mean Shift is *sharper* than the original one, as can be seen on figure 5c column 2 where the silhouettes of object instances emerge.

We then estimate the probability density (up to a scaling factor) at a mode $\mathscr{M}$ by kernel density estimation:

$$s(\mathscr{M}) = \sum_i H\left(\frac{\mathrm{d}(\mathscr{M}, \mathscr{P}_i)}{r}\right) \tag{96}$$

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

$$H(d) = \begin{cases} \frac{3}{4}(1 - d^2) & \text{if } |d| \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{97}$$

Algorithm 1 Mean Shift algorithm within the pose space

Input: $\{\mathscr{P}_i\}_{i=1,\dots,n}$ a set of poses,
　　$\mathbf{p}_{\text{in}}$ a representative of the pose to shift,
　　r the Mean Shift radius.
Output: A representative of the shifted pose.
　1: $R \leftarrow GetRepresentatives(\{\mathscr{P}_i\}_{i=1,\dots,n})$
　　# Preprocessing step independent of $\mathbf{p}_{\text{in}}$.
　　# R contains all poses representatives:
　　# $\forall i \in [\![1,n]\!], \{R[i,j]\}_{j=1,\dots,|\mathscr{R}(\bullet)|} = \mathscr{R}(\mathscr{P}_i)$.
　2: $\mathbf{p} \leftarrow \mathbf{p}_{in}$
　3: **repeat**
　4:　　$\mathbf{p}_{\text{old}} \leftarrow \mathbf{p}$
　5:　　$\mathscr{N} \leftarrow \mathsf{RadiusSearch}(R, \mathbf{p}, r)$
　6:　　**if** $\mathscr{N} \neq \emptyset$ **then**
　7:　　　$\mathbf{m} \leftarrow \left(\sum_{(i,j)\in\mathscr{N}} \mathbf{R}[i,j]\right)/|\mathscr{N}|$
　8:　　　$\mathbf{p} \leftarrow \mathsf{ClosestRepresentative}(\mathrm{proj}(\mathbf{m}))$
　9:　　**end if**
10: **until** $\mathbf{p} \neq \mathbf{p}_{\text{old}}$
11: **return** $\mathbf{p}$

The most significant modes based on this estimate can then be extracted. Those poses are assumed to be good pose hypotheses for the object instances of the scene, and typically stand up from the weighted distribution (figure 5c column 3). We refine them further through e.g. the ICP procedure (Besl and McKay 1992), and filter them by checking their consistency with the actual data in order to avoid false postives, to hopefully retrieve the poses of object instances in the 3D scene (figure 5c column 4).

Theoretical limitation The probabilistic interpretation 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 does not enter into such framework as we do not consider a Riemannian distance. Some theoretical results might nonetheless be obtained, since our metric is equivalent to a Riemannian metric for small Mean Shift radii (see section 9). Such considerations however, are out of the scope of this work, and $s(\mathscr{M})$ can simply be considered as a score for the pose $\mathscr{M}$.

10.2 Comparison with a $SE(3)$ metric

We compare these experimental results with those obtained using a more usual distance adapted to $SE(3)$

$$\mathrm{d}(\mathbf{T}_1, \mathbf{T}_2) = \sqrt{\|\mathbf{t}_2 - \mathbf{t}_1\|^2 + r^2 \|\mathbf{R}_2 - \mathbf{R}_1\|^2}. \tag{98}$$

We chose this particular distance because the Mean Shift approach depends on the ability to average multiple poses, and a Frobenius norm over the rotation space is quite suited for this task (Curtis et al 1993). To limit the comparison bias, we

choose the scaling factor r between the rotation and translation parts to be

$$r = \sqrt{\frac{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}{3}}, \tag{99}$$

where $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are the eigenvalues of $\mathbf{\Lambda}$. This choice is indeed consistent with our proposed metric, in that the rotational part of the distance between two poses of an object without proper symmetry corresponds respectively to $2\sqrt{2}r\sin(\theta/2)$ for distance (98), and $2\sqrt{I_\mathbf{k}}\sin(\theta/2)$ for the proposed one, where θ is the angle of the relative rotation between the two poses, and $I_\mathbf{k}$ is the inertia moment of the corresponding axis $\mathbf{k}$ (see section 4.4). Considering a typical value of $2/3(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)$ for $I_\mathbf{k}$ enables to identify those two terms.

As illustrated on figure 6, we do not observe much differences between the two approaches for the *bunny* object. This is actually not surprising because the two distances are in this case very similar, since the *bunny* is not symmetric, and has a limited anisotropy.

However, the benefit of our metric appears for the *candlestick* and the *rocket*, that both are symmetric. Initial votes for poses are indeed spread out over the space of rigid transformations $SE(3)$, and considering the $SE(3)$ distance (98) therefore leads to the detection of multiple modes corresponding to the same instance, because of the symmetries. One would have to filter out these duplicated poses hypotheses prior to any practical application, and because of these, it is required to check numerous modes to find every instance of the scene. In our example (figure 6b) we had to test up to respectively the 4th and 8th mode to recover the poses of the 3 rockets and candlesticks present in the scene.

On the other hand, the proposed distance enables to account for the proper symmetries of the object, and thus to better exploit the information contained in the initial set of votes than the $SE(3)$ distance. In our example, the 3 first modes extracted indeed correspond to the 3 actual instances for each object, without any duplicates. Moreover, these modes have more support from the initial set of votes and therefore stand out more clearly from the noise, which is important for the robustness of the method. The pose distribution obtained after Mean Shift is indeed visually sharper (figure 6, left), and spurious modes for the *rocket* and the *candlestick* have a score below respectively 59% and 66% of the ones of the modes corresponding to actual object instances, compared to 89% and 98% when using the distance (98).

11 Summary and discussion

In this paper, we address issues of the commonly used notion of pose of a rigid object, both in the 2D and 3D case.

(a) First modes of the pose distribution retrieved using the proposed metric.

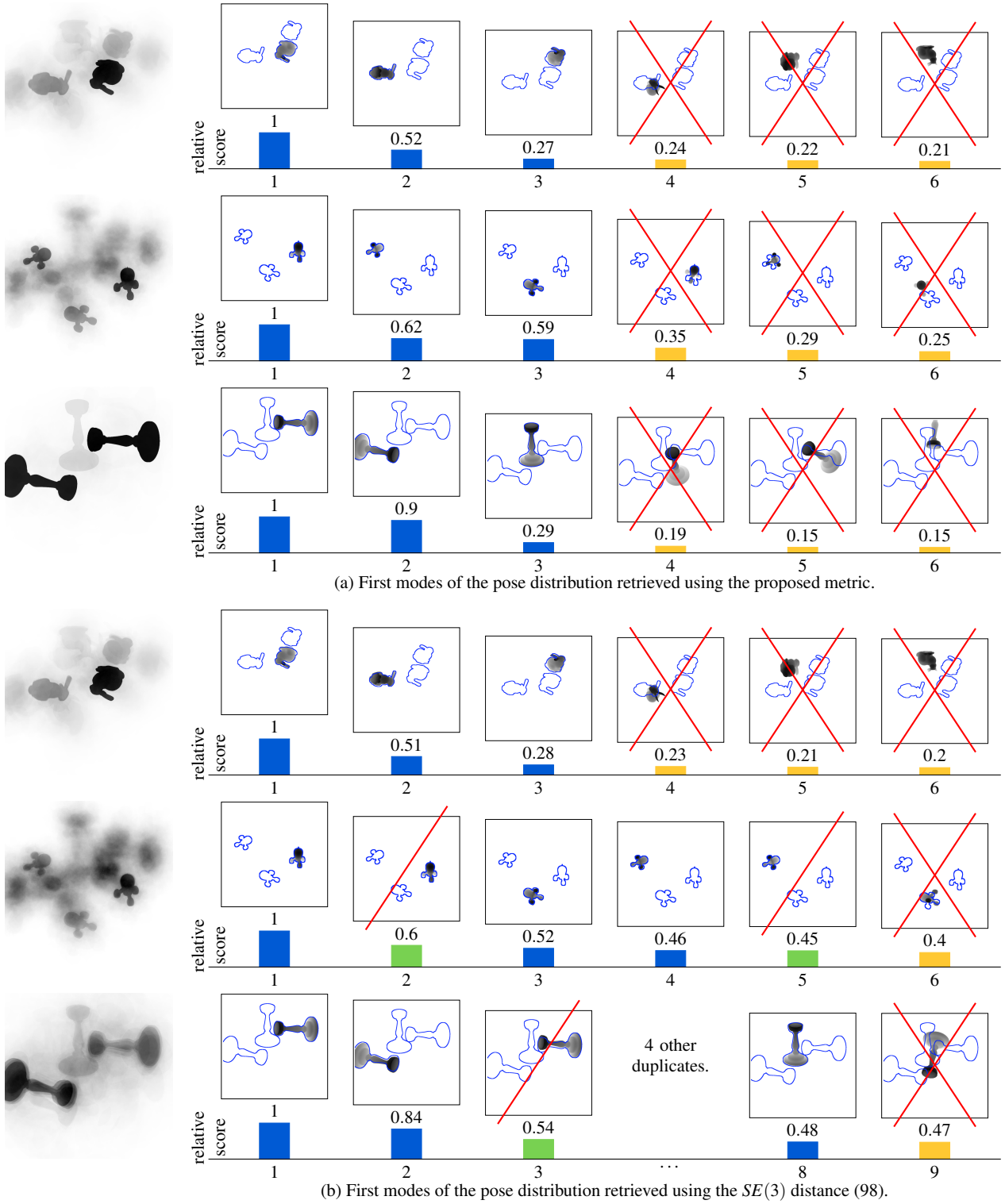(b) First modes of the pose distribution retrieved using the $SE(3)$ distance (98).

Fig. 6 Comparison of the proposed distance and a $SE(3)$ distance for pose estimation. Left: shifted poses weighted by the density of the initial pose distribution. Right: first modes retrieved from the pose distribution, sorted by descending score (with the contours of the actual poses of object instances superimposed). Modes are supposed to be good hypotheses regarding the poses of actual instances, and are classified as true positives (blue), duplicates (green, strikethrough) and false positives (yellow, double strikethrough). Both distances perform similarly well for the *bunny* object, which has no proper symmetry, and the first modes extracted correspond to the different object instances. However, the $SE(3)$ distance does not account for symmetries of the *rocket* and *candlestick* objects, and therefore leads to the generation of duplicated pose hypotheses, requiring to consider many of them to recover the pose of every instances. The proposed distance better exploit the information contained in the initial pose distribution, leading to a generation of poses hypotheses with no duplicates, and with a greater relative score gap between modes corresponding to actual instances and spurious ones.

While pose is usually assumed to be equivalent to a rigid transformation, this is not true in general due to potential symmetries of the object. We therefore propose a broader definition of the notion of pose, consisting in a distinguishable static state of the object. We show that with this definition, a pose can be considered as an equivalence class of the space of rigid transformations, thanks to the introduction of a proper symmetry group specific to the object. We believe this notion to be essential, as many of manufactured objects actually show some symmetry properties and could not be represented properly previously.

Based on this definition, we propose 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 propose a coherent framework to represent poses in a Euclidean space of at most 12 dimensions, 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 estimation based on a set of votes, where it appears to provide better results than a metric suited for $SE(3)$.

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 of the proposed distance for a revolution object without rotoreflection invariance can be rewritten in the following way:

$$
\begin{aligned}
&\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) \\
&= \min_{\phi_1, \phi_2} \frac{1}{S} \int_{\mathscr{S}} \mu(\mathbf{x}) \| \mathbf{R}_2 \mathbf{R}_z^{\phi_2} \mathbf{x} - \mathbf{R}_1 \mathbf{R}_z^{\phi_1} \mathbf{x} \|^2 ds \\
&= \min_{\phi_1, \phi_2} \| \mathbf{R}_2 \mathbf{R}_z^{\phi_2} \boldsymbol{\Lambda} - \mathbf{R}_1 \mathbf{R}_z^{\phi_1} \boldsymbol{\Lambda} \|_F^2 .
\end{aligned}
\tag{100}
$$

Frobenius norm being invariant under rotations, this expression can be rewritten with the relative rotation $\mathbf{R} \triangleq \mathbf{R}_1^{-1} \mathbf{R}_2$:

$$
\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) = \min_{\phi_1, \phi_2} \| \mathbf{R}_z^{-\phi_1} \mathbf{R} \mathbf{R}_z^{\phi_2} \boldsymbol{\Lambda} - \boldsymbol{\Lambda} \|_F^2 .
\tag{101}
$$

We parametrize $\mathbf{R}$ using Euler angles $(\tilde{\psi}, \theta, \tilde{\phi}) \in \mathbb{R}^3$ such as $\mathbf{R} = \mathbf{R}_z^{\tilde{\psi}} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\tilde{\phi}}$, considering the following elementary rotations:

$$
\mathbf{R}_z^{\alpha} \triangleq \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{R}_x^{\alpha} \triangleq \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}.
\tag{102}
$$

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:

$$
\begin{aligned}
\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) &= \min_{\phi_1, \phi_2} \| \mathbf{R}_z^{-\phi_1} \mathbf{R}_z^{\tilde{\psi}} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\tilde{\phi}} \mathbf{R}_z^{\phi_2} \boldsymbol{\Lambda} - \boldsymbol{\Lambda} \|_F^2 \\
&= \min_{\psi, \phi} \| \mathbf{R}_z^{\psi} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\phi} \boldsymbol{\Lambda} - \boldsymbol{\Lambda} \|_F^2 .
\end{aligned}
\tag{103}
$$

Because of the specific shape of $\boldsymbol{\Lambda}$ (equation 30), the term to minimize can be decomposed into two parts:

$$
\begin{aligned}
\| \mathbf{R}_z^{\psi} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\phi} \boldsymbol{\Lambda} - \boldsymbol{\Lambda} \|_F^2 = \lambda_z^2 \underbrace{\| \mathbf{R}_z^{\psi} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\phi} \mathbf{e}_z - \mathbf{e}_z \|^2}_{a_{\psi, \phi}} \\
+ \lambda_r^2 \underbrace{(\| \mathbf{R}_z^{\psi} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\phi} \mathbf{e}_x - \mathbf{e}_x \|^2 + \| \mathbf{R}_z^{\psi} \mathbf{R}_x^{\theta} \mathbf{R}_z^{\phi} \mathbf{e}_y - \mathbf{e}_y \|^2)}_{b_{\psi, \phi}} .
\end{aligned}
\tag{104}
$$

Developing this expression thanks to the definition of the elementary rotations (102), we evaluate those terms into:

$$
\begin{cases} a_{\psi, \phi} = 2(1 - \cos(\theta)) \\ b_{\psi, \phi} = 4 - 2\cos(\psi + \phi)(1 + \cos(\theta)). \end{cases}
\tag{105}
$$

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:

$$
\min_{\psi, \phi} b_{\psi, \phi} = 2(1 - \cos(\theta)).
\tag{106}
$$

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

$$
\begin{aligned}
2(1 - \cos(\theta)) &= \| \mathbf{R} \mathbf{e}_z - \mathbf{e}_z \|^2 \\
&= \| \mathbf{R}_2 \mathbf{e}_z - \mathbf{R}_1 \mathbf{e}_z \|^2
\end{aligned}
\tag{107}
$$

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:

$$
\mathrm{d}_{\mathrm{rot}}^2(\mathscr{P}_1, \mathscr{P}_2) = (\lambda_r^2 + \lambda_z^2) \| \mathbf{R}_2 \mathbf{e}_z - \mathbf{R}_1 \mathbf{e}_z \|^2 .
\tag{108}
$$

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 3) 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

$$
G = \left\{ \mathbf{I}, \mathbf{R}_z^{2\pi/3}, \mathbf{R}_z^{-2\pi/3} \right\}
\tag{109}
$$

and the square root of the covariance matrix as

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_r, \lambda_r, \lambda_z). \tag{110}$$

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

$$\left\{ \underbrace{\begin{pmatrix} \operatorname{vec}(\mathbf{\Lambda}) \\ \mathbf{0}_3 \end{pmatrix}}_{\mathbf{p}}, \begin{pmatrix} \operatorname{vec}(\mathbf{R}_z^{2\pi/3}\mathbf{\Lambda}) \\ \mathbf{0}_3 \end{pmatrix}, \begin{pmatrix} \operatorname{vec}(\mathbf{R}_z^{-2\pi/3}\mathbf{\Lambda}) \\ \mathbf{0}_3 \end{pmatrix} \right\}. \tag{111}$$

Thanks to those choices, we can evaluate T into:

$$\begin{aligned} T &= \min_{\mathbf{q} \in \mathscr{R}(\mathscr{P}_0), \mathbf{q} \neq \mathbf{p}} \|\mathbf{q} - \mathbf{p}\| \\ &= \min \|\mathbf{R}_z^{\pm 2\pi/3}\mathbf{\Lambda} - \mathbf{\Lambda}\|_F \\ &= \sqrt{6}\lambda_r. \end{aligned} \tag{112}$$

The threshold $\dfrac{T}{4}$ of proposition 9 therefore corresponds for the rocket to the value $\dfrac{\sqrt{3}}{2}\lambda_r$.

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:

$$S_{\mathbf{a},\mathbf{b},\mathbf{c}} = \frac{\|(\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})\|}{2}, \tag{113}$$

its center of mass through:

$$\mathbf{o}_{\mathbf{a},\mathbf{b},\mathbf{c}} = \frac{\mathbf{a} + \mathbf{b} + \mathbf{c}}{3}, \tag{114}$$

and its uncentered covariance matrix via:

$$\boldsymbol{\sigma}_{\mathbf{a},\mathbf{b},\mathbf{c}} = \frac{S_{\mathbf{a},\mathbf{b},\mathbf{c}}}{12}\left(9\mathbf{o}_{\mathbf{a},\mathbf{b},\mathbf{c}}\mathbf{o}_{\mathbf{a},\mathbf{b},\mathbf{c}}^\top + \mathbf{a}\mathbf{a}^\top + \mathbf{b}\mathbf{b}^\top + \mathbf{c}\mathbf{c}^\top\right). \tag{115}$$

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

$$S = \sum_i S_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}, \tag{116}$$

its center of mass:

$$\mathbf{o} = \sum_i S_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}\mathbf{o}_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}, \tag{117}$$

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

$$\mathbf{\Lambda}^2 = \frac{1}{S}\sum_i \boldsymbol{\sigma}_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}. \tag{118}$$

References

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

Belta C, Kumar V (2002) An SVD-based projection method for interpolation on SE (3). Robotics and Automation, IEEE Transactions on 18(3):334–345

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

Blender Online Community (2016) Blender - a 3D modelling and rendering package. Blender Foundation, Blender Institute, Amsterdam, URL http://www.blender.org

Brégier R, Devernay F, Leyrit L, Crowley JL (2017) Symmetry aware evaluation of 3d object detection and pose estimation in scenes of many parts in bulk. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2209–2218

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, Springer Netherlands, Dordrecht, pp 361–369. DOI 10.1007/978-1-4020-8600-7_38

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 (2012) Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In: Asian conference on computer vision, Springer, pp 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

Kendall A, Grimes M, Cipolla R (2015) PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2938–2946

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

Pennec X (1998) Computing the Mean of Geometric Features Application to the Mean Rotation. report, INRIA

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