



**HAL**  
open science

# Randomness and Geometric Features in Computer Vision

Xavier Pennec, Nicholas Ayache

► **To cite this version:**

Xavier Pennec, Nicholas Ayache. Randomness and Geometric Features in Computer Vision. RR-2820, INRIA. 1996. inria-00073871

**HAL Id: inria-00073871**

**<https://inria.hal.science/inria-00073871>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Randomness and Geometric Features in  
Computer Vision*

Xavier PENNEC, Nicholas AYACHE

**N° 2820**

March 1996

PROGRAMME 4



*Rapport  
de recherche*



# Randomness and Geometric Features in Computer Vision

Xavier PENNEC, Nicholas AYACHE

Programme 4 — Robotique, image et vision  
Projet Epidaure

Rapport de recherche n° 2820 — March 1996 — 51 pages

**Abstract:** It is often necessary to handle randomness and geometry in computer vision, for instance to match and fuse together noisy geometric features such as points, lines or 3D frames, or to estimate a geometric transformation from a set of matched features. However, the proper handling of these geometric features is far more difficult than for points, and a number of paradoxes can arise. We try to establish in this article the basic mathematical framework required to avoid them and analyze more specifically three basic problems:

- what is a random distribution of features,
- how to define a distance between features,
- and what is the “mean feature” of a number of feature measurements ?

We insist on the importance of an invariance hypothesis for these definitions relative to a group of transformations. We develop general methods to solve these three problems and illustrate them with 3D frame features under rigid transformations.

The first problem has a direct application in the computation of the prior probability of false match in classical model-based object recognition algorithms, and we present experimental results of the two others for a data fusion problem: the statistical analysis of anatomical features (extremal points) automatically extracted on 24 three dimensional images of the head of a single patient. These experiments successfully confirm the importance of the rigorous requirements presented in this article.

**Key-words:** Geometric features, Transformation groups, Randomness, Invariant measure, Invariant distance, Expected features, Mean features.

*(Résumé : tsvp)*

# Primitives Géométriques et Probabilités en Vision par Ordinateur

**Résumé :** De nombreux algorithmes en vision par ordinateur, et en particulier en reconnaissance des formes, utilisent des primitives géométriques simples comme les points. D'un autre côté, les modèles géométriques du monde réel amènent souvent à considérer des primitives plus complexes, par exemple des droites, des plans, des points, des repères. . . L'utilisation de telles primitives plus sélectives est par ailleurs justifiée par la réduction de complexité qu'elle peut apporter pour la reconnaissance et le gain de précision au niveau du recalage.

La gestion de ces primitives soulève cependant un certain nombre de problèmes qui peuvent conduire à des paradoxes. Nous étudions dans cet article trois problèmes fondamentaux et nous présentons pour chacun d'entre eux un exemple de paradoxe potentiel :

- quelle est la probabilité à priori d'apparier deux primitives quelconques ou, en d'autres termes, quelle est la probabilité d'un faux positif dans un problème de reconnaissance ?
- comment définir une notion de distance cohérente, qui permettra entre autres de classifier des appariements de primitives et de généraliser les algorithmes de type plus proche voisin ?
- qu'est-ce que la moyenne de primitives géométriques, ou plus généralement la notion d'espérance mathématique dans une variété ?

S'il existe parfois des solutions ad-hoc pour chacun de ces problèmes, celles-ci ne sont en général valables que pour certains types de primitives dans certaines conditions. Nous présentons dans cet article des solutions génériques basées sur la notion d'invariance de l'opérateur considéré (distance, espérance, sélectivité ou inversement probabilité de faux positif) sous l'action d'un groupe de transformation et nous montrerons que ce cadre théorique est applicable, puisqu'il nous permettra de déterminer la sélectivité des repères pour la reconnaissance (un gain d'un facteur 10 à 100 par rapport aux points seuls), et nous présenterons une application en fusion de données : la détermination de la stabilité statistique des points extrémaux extraits sur 24 images IRM 3D de la tête d'un même patient.

**Mots-clé :** Primitives géométriques, Groupes de transformation, Probabilités, Mesure invariante, Distance invariante, Espérance, Moyenne

## Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Sets of geometric features and sets of transforms</b>	<b>6</b>
2.1	Geometric features: Manifolds and representations . . . . .	6
2.2	Transformations: Lie groups . . . . .	7
2.3	From transformation of the Euclidean space to feature transformation	7
2.4	Homogeneous features: identification with the transformation group	8
2.5	Representation and basic operations for 3D rotations, frames and points	9
<b>3</b>	<b>Classical Geometric probabilities</b>	<b>11</b>
3.1	Bertrand’s paradox . . . . .	11
3.2	Left and right invariant measures on a Lie group (Haar measures) . .	13
3.3	Invariant measure on homogeneous manifolds . . . . .	14
3.4	Practical use: probability of a false match . . . . .	15
3.A	Appendix A: Proofs for invariant measures . . . . .	16
3.B	Appendix B: Proofs for invariant measures on rigid transformations .	17
<b>4</b>	<b>Probability density functions</b>	<b>20</b>
4.1	Definition . . . . .	20
4.2	From the “invariant” pdf to the classical pdf on the representation . .	21
4.3	Propagation of one random feature or transform . . . . .	21
4.4	Canonical operations on random features and transformations . . . .	22
4.5	Expectation of a real valued function . . . . .	23
4.6	Conclusion . . . . .	24
4.A	Appendix A: Proofs for probability density functions . . . . .	24
<b>5</b>	<b>Metric properties</b>	<b>28</b>
5.1	The paradox of the closest line . . . . .	28
5.2	Invariant distance on a manifold . . . . .	29
5.3	Invariant distances on a Lie group . . . . .	30
5.4	Distance induced on the manifold by the group distance . . . . .	30
5.5	Practical use on rigid transformations . . . . .	31
5.6	Utility of invariant distances . . . . .	32
5.A	Appendix A: Metric induced on the manifold by the group . . . . .	33
5.B	Appendix B: Metric on rigid transformations . . . . .	33

<b>6</b>	<b>Expectation and mean of random features</b>	<b>35</b>
6.1	Standard expectation of a 2D random line . . . . .	35
6.2	Frechet expectation of a random feature . . . . .	36
6.3	Frechet expectation of a random transformation . . . . .	38
6.4	Invariance properties . . . . .	38
6.5	Discussion . . . . .	39
6.6	Application to data fusion: the mean frames . . . . .	39
6.A	Appendix A: Stability of the expected and mean elements . . . . .	40
6.B	Appendix B: Hints for the propagation through the other operations	41
<b>7</b>	<b>Experiments: a data fusion problem</b>	<b>42</b>
7.1	3D medical images . . . . .	42
7.2	Extremal points and frames . . . . .	43
7.3	Building a model . . . . .	43
7.4	Results . . . . .	43
<b>8</b>	<b>Conclusion</b>	<b>45</b>

## 1 Introduction

Many algorithms in computer vision and object recognition deal with simple geometric features like points, for example the Iterative Closest Point ([BM92, Zha94]), the geometric hashing ([LW88, RH93]), and the alignment ([AF86, HU87]) algorithms. On the other hand, models of the real world often lead to the consideration of more complex features: lines ([Gri92]), planes, oriented points, frames ([PA94, PT95b]), etc. . .

The handling of these features raises some problems and can lead to paradoxes such as Bertrand's paradox concerning geometric probabilities. We have previously shown ([PT95a]) that additive noise is not suited for describing the uncertainty of frames and should be replaced by a compositive model of noise. Other examples are presented in this article and demonstrate the need for a particular attention when dealing with geometric features.

We determined that the key point to analyze geometric feature is their behavior under a transformation group. We can see for instance in figure 1 that the object comparison only makes sense relatively to a transformation group.

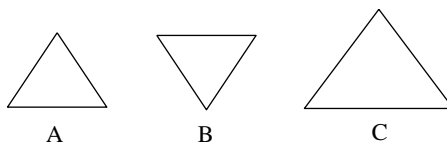


Figure 1: Comparison of geometric objects: we can say that all three triangles are similar (relative to similarity transformations); or that only **A** and **B** are congruent (relative to rigid transformations); or that they are all different (relative to translations).

The article is organized as follows. Section 2 focuses on the nature of geometric features, namely points on a manifold, and we investigate transformation groups (rigid, affine, . . .) that operate on this manifold. In Section 3, we investigate the standard geometric probabilities and in particular need to define an invariant measure on random features (under the considered transformation group) in order to obtain a meaningful result. This leads to the computation of the prior probability of a false match in recognition algorithms. In section 4, we extend the class of probabilities problems with the introduction of probability density functions (pdfs) on a manifold. Section 5 is devoted to metric properties and section 6 to their use in providing a definition of the expected and average features which is invariant of the representation and stable under the transformation group. In the fifth and final section, we present an experimental application of the theory to the data fusion problem.



## 2 Sets of geometric features and sets of transforms

### 2.1 Geometric features: Manifolds and representations

Geometric features are generally defined as sets of points in the plane or 3D space, and the set of all geometric features of a given type can be described by a parameter  $p$  and a function  $\beta(p, x)$  which associates the parameter  $p$  to the geometric feature (the set  $\{x \in \mathbb{R}^n \mid \beta(p, x) = 0\}$ ). The function  $\beta$  describe a particular type of geometric feature (lines, planes, curves, triangles, ...) with a specific *representation*  $p$ . 3D oriented planes can for instance be represented by  $p = (n, d)$  where  $n$  is a unit vector (the normal to the plane) and  $d$  the distance to the origin. The equation of “plane  $p$ ” is then:

$$\beta(p, x) = \langle n | x \rangle - d = 0$$

Usual sets of geometric features, such lines, curves, surfaces... are regular and constitute differential manifolds. This means that the set is locally diffeomorphic to a vector space  $\mathbb{R}^m$  (i.e there exists, at each point of the manifold  $\mathcal{M}$ , a locally differentiable one-to-one mapping from  $\mathcal{M}$  to  $\mathbb{R}^m$ );  $m$  being the dimension of the manifold. In the above example, we can see that the parameter  $p$  is four dimensional with a quadratic constraint (which is differentiable), and planes are then a 3D-manifold equivalent to  $S^2 \times \mathbb{R}^+$  ( $S^2$  is the unit sphere in 3D). Despite the rather complex mathematical formulation, this simply means that manifolds are not traditional vector spaces, but locally may be treated as if they were. Spheres or smooth surfaces are such manifolds, as is the set of rotation matrices which is equivalent to  $\mathcal{P}^3$  (the projective space of  $\mathbb{R}^4$ ) by means of unit quaternions (see [PT95b, Aya91]). Points trivially constitute a manifold since they already are a vector space. Another interesting type of features is oriented points, which are points associated with a vector. Such features can be extracted from a smooth surface, for instance, where the normal is attached at each point of the surface. A simple representation is given by  $u = (x, n)$  where  $x$  is the position and  $n$  a unit vector. The manifold of oriented points is thus equivalent to  $\mathbb{R}^3 \times S^2$  where  $S^2$  is the unit sphere of the 3D space.

There are often numerous ways to represent a given manifold, with different properties. We can for instance define a manifold as a subspace of  $\mathbb{R}^k$  with differentiable constraints and a one-to-one correspondence between features and parameters: this proves that the set of features is a differential manifold. For other purposes, in particular differentiation, it is necessary to have a *minimal representation* (where the dimension of the parameter is the dimension of the manifold), or more generally a set of charts forming an atlas of the manifold, exactly the same way we need several charts to represent the earth surface in a continuous way everywhere. Each chart is

defined by a one-to-one differentiable map  $\varphi_i(p)$  from the representation into the manifold and an open definition domain  $\mathcal{D}_i$ . The set of charts must cover the manifold and must overlap each other so that it is possible to move from one chart to another. A study of different representations for 2D and 3D lines, planes and rotations is presented in [Aya91]. We only assume for the moment that we have a one-to-one differentiable representation of the manifold, and we will identify this representation with the manifold.

## 2.2 Transformations: Lie groups

There are a number of familiar transformations: translations, rotations, similarities, affine transforms. . . More generally, a transformation of a set  $X$  is a one-to-one map of  $X$  onto  $X$ . If  $g$  is a transformation, we will denote by  $g \star x = g(x)$  the application of the transformation to an element  $x \in X$ , and by  $g^{(-1)}$  the inverse map. If  $g_1$  and  $g_2$  are two transformations, the map  $g(x) = g_2(g_1(x))$  is also a transformation: the composition of  $g_1$  and  $g_2$  ( $g = g_2 \circ g_1$ ). The set of all transformations with these operations is called the general transformation group of set  $X$ , and any subgroup  $\mathcal{G}$  is a transformation group of  $X$ .

The important class of Lie groups is obtained if  $\mathcal{G}$  has a separate topological structure (a Hausdorff space) and the composition and inversion maps are differentiable ( $\mathcal{G}$  is then a differentiable manifold). In fact, most usual transformation groups are Lie ones, as soon as the group is continuous (in the non-discrete sense) and has reasonable operations. In this article, we will use the 3D rigid motion group as an example application. An element of this group can be defined as the composition of a rotation with a translation. It can be represented by  $f = (R, t)$ , where the translation  $t$  belongs to  $\mathbb{R}^3$  and where  $R$  is a rotation matrix (a 3x3 matrix satisfying  $R.R^\top = R^\top.R = Id$  and  $\det(R) = +1$ ), and hence belongs to the special orthogonal group  $SO_3$ . The inverse and compose maps are easily written (“.” is the matrix multiplication):

$$f^{(-1)} = (R^\top, R^\top.t) \quad \text{and} \quad f = f_2 \circ f_1 = (R_2.R_1, R_2.t_1 + t_2)$$

## 2.3 From transformation of the Euclidean space to feature transformation

In the case of geometric objects, the transformation usually applies to the plane or the 3D space (or more generally  $\mathbb{R}^n$ ), but we wish to work directly on features and thus must take particular care that their nature is preserved during transformations.

Consider, for instance, that two orthonormal axes are no longer orthonormal after a general affine transformation. The first constraint is then for the manifold  $\mathcal{M}$  to be globally invariant under the considered transformation group  $\mathcal{G}$ . We can then define the image of the feature  $p$ , satisfying  $\beta(p, x) = 0$ , by a transformation  $g \in \mathcal{G}$  as being the feature  $p' \in \mathcal{M}$  realizing  $\beta(p', g \star x) = 0$ . We will write  $p' = g \star p$ . With this definition, the group  $\mathcal{G}$  is also a transformation group of the manifold  $\mathcal{M}$ . It can be very tricky to make the application operation explicit on some geometric features with some representations and this can lead to highly non-linear transformations. However, usual cases are generally simple: in the case of oriented planes presented above under rigid motion, we have

$$\beta(p, f \star x) = \langle n | R.x + t \rangle - d = \langle R^\top . n | x \rangle + \langle n | t \rangle - d = 0$$

which means that  $f \star p = (R^\top n, d - \langle n | t \rangle)$ . Similarly, the application of a rigid transform  $f$  to an oriented point  $u = (x, n)$  is:

$$u' = f \star u = (R.x + t, R.n)$$

We will be also interested in a third type of feature: frames. A frame is defined by a point with an associated orthonormal trihedron. We have already noted that we cannot use the affine group since orthonormal trihedra would not be conserved, but rigid motions are appropriate. A particularity of frames is that they are equivalent to rigid transformations. Indeed, any frame defines a basis for 3D space and so we can represent each frame by the rigid transformation which map the canonical basis to itself. It is then easy to verify that composition and application are equivalent.

## 2.4 Homogeneous features: identification with the transformation group

A special kind of relation between the manifold and the group will be very important: let  $\mathcal{O} \in \mathcal{M}$  be an element called the origin. The manifold  $\mathcal{M}$  is *transitive* or *homogeneous* for the group  $\mathcal{G}$  if any other element of the manifold can be obtained by a transformation of  $\mathcal{G}$ , i.e. if  $\mathcal{G} \star \mathcal{O} = \{q = g \star \mathcal{O} / g \in \mathcal{G}\} = \mathcal{M}$ . This means that the features we consider have no invariants. In fact, we assume that we can split the features into an invariant part (which we do not consider here) and a variable part under the group of transformation.

In the case of homogeneous features, we identify the manifold with equivalence classes of group elements in the following way. Let  $\mathcal{H}$  be the subset of transforms that leave  $\mathcal{O}$  invariant:

$$\mathcal{H} = \{h \in \mathcal{G} / h \star \mathcal{O} = \mathcal{O}\} \tag{1}$$

$\mathcal{H}$  is a group and is called the *isotropy* or *stability* group of  $\mathcal{G}$  at  $\mathcal{O}$ . The left cosets  $g \circ \mathcal{H}$  can be identified with elements of  $\mathcal{M}$ . Indeed, if  $g \star \mathcal{O} = x$ , then  $g \circ \mathcal{H}$  is the set of transformations which map  $\mathcal{O}$  to  $x$ . We will write

$$\hat{x} = \{g \in \mathcal{G} / g \star \mathcal{O} = x\} \tag{2}$$

If we consider for instance point features with origin  $\mathcal{O} = 0$  and rigid transformation, we have  $\mathcal{H} = \{(R, 0) / R \in SO_3\}$  and  $\hat{x} = \{(R, x) / R \in SO_3\}$  where  $R$  is any rotation. For frame, taking as the origin the canonical basis ( $\mathcal{O} = (I_d, 0)$ ), then  $\mathcal{H}$  and all its cosets are reduced to a single point:  $\hat{f} = \{f\}$ . This special case where the manifold is equivalent to the group leads to important simplifications in the theory. As an intermediate example, we consider oriented points: we set the origin to  $\mathcal{O} = (0, e_3)$ , which means that all rotations  $R_z$  around the third axis  $e_3$  will let it invariant:  $\mathcal{H} = \{R_z \in SO_3 / R.e_3 = e_3\}$ .

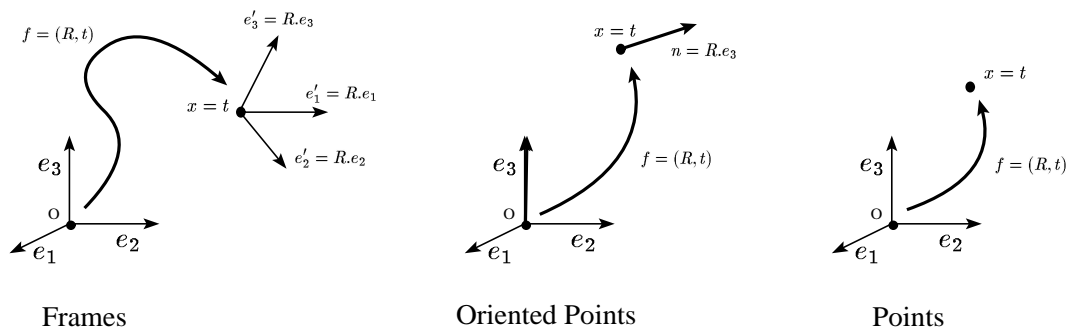


Figure 2: Chosen origins for frames, oriented points and points, and transformation of it by a rigid motion.

## 2.5 Representation and basic operations for 3D rotations, frames and points

It is well known that a rotation matrix can be characterized by an angle  $\theta$  around an axis  $n$  (unit vector), but since the axis is constrained, this couple is not minimal (the dimension of the representation is 4 instead of 3) and the axis is moreover not defined for the identity transformation. The rotation vector  $r = \theta.n$  is always defined (multiply, since  $\theta$  is modulo  $2\pi$ ) and differentiable( see [PT95a] for the equations). In order to define an atlas of rotations, we need in fact four charts.

**Chart 1:** Non reflection rotations are represented by rotations vectors  $r$  from the open ball  $\mathcal{B}^3(0, \pi)$ .

**Chart 2,3 and 4:** Non identity rotations with axis not orthogonal to the  $x$  axis (respectively  $y$ ,  $z$ ) are represented by rotation vectors  $r$  from the open half ball  $\mathcal{B}_{x^+}^3 = \{r \in \mathcal{B}^3(0, 2\pi) / r_x > 0\}$  (respectively  $\mathcal{B}_{y^+}^3$ ,  $\mathcal{B}_{z^+}^3$ ).

In theory we need to handle all four charts, but in practice only the principal chart (the first) is needed as long as it is remembered that, at the boundary of the domain,  $r = \pi.n$  and  $r' = -\pi.n$  are identical. Let  $\mathbf{R}(r)$  and  $\mathbf{r}(R)$  denote the mappings between rotation vectors and matrices, we can now write directly the composition and inversion laws on the representation:

$$r^{(-1)} = \mathbf{r}(\mathbf{R}(r)^\top) \quad \text{and} \quad r_2 \circ r_1 = \mathbf{r}(\mathbf{R}(r_2).\mathbf{R}(r_1))$$

Frames and motions are represented by a rotation vector and a translation: for convenient notation, we will write  $f = (r, t)$  and consider it as a column vector. In this framework, the representation of a point of the Euclidean space is denoted  $x$  (the standard coordinates). The standard operators can then be written:

- Composition:  $f = f_2 \circ f_1 = (r_2 \circ r_1, r_2 \star t_1 + t_2)$
- Inversion:  $f^{(-1)} = (r^{(-1)}, r^{(-1)} \star (-t))$
- Application to a point:  $x' = f \star x = r \star x + t$
- Cosets of points:  $\hat{x} = \{f_x = (r, x) / \mathbf{R}(r) \in SO_3\}$
- Application to a frame:  $f = f_2 \star f_1 = f_2 \circ f_1$
- Cosets of frames:  $\hat{f} = \{f\}$ .

### 3 Classical Geometric probabilities

The first class of problems in geometric probability are questions of how to measure the probability of occurrence of an event when some geometric elements are randomly distributed. Bertrand’s paradox illustrate the need to consider invariance by a transformation group in order to obtain a single and well defined result. In fact, the problem lie in the notion of a *uniform* distribution (or measure). Some more recent results in Lie group theory provide a means of computing the left invariant measure on the group  $\mathcal{G}$ , which induces the invariant measure on homogeneous manifolds. An application is presented with the generalization of the false positives analysis.

#### 3.1 Bertrand’s paradox

The problem raised by J. Bertrand in 1907 is to find the probability that a “random” chord of a circle has a length greater than the side of an inscribed equilateral triangle. Without loss of generality, we can fix the radius to 1 and the side length of the triangle is then  $\sqrt{3}$ . This problem can be tackled by at least three methods, which are illustrated in figure 3.

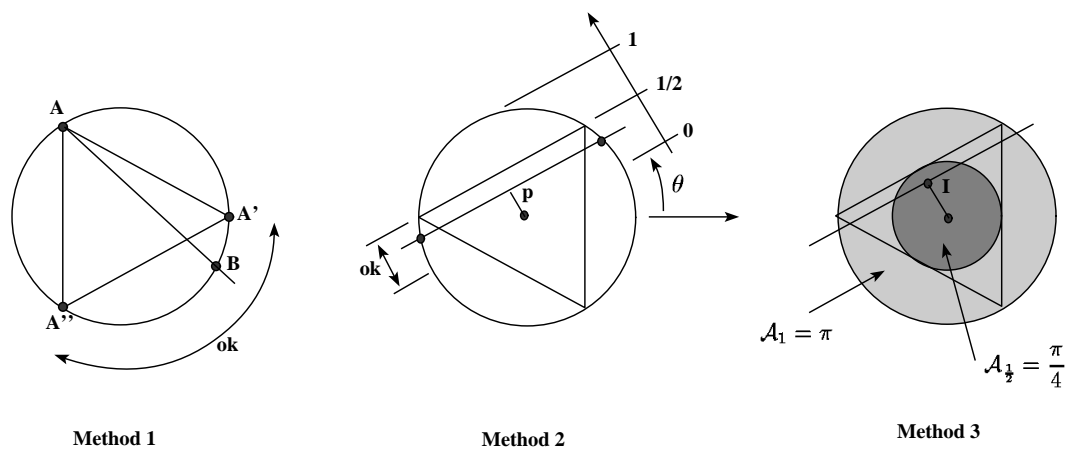


Figure 3: Three methods to compute the probability that a “random” chord of a circle has a length greater than the side of an inscribed equilateral triangle. From left to right, the methods give a probability of  $\frac{1}{3}$ ,  $\frac{1}{2}$  and  $\frac{1}{4}$ .

**Method 1:** By definition, a chord intersects the circle in two points, and we may assume that these two points are equally and independently distributed on the circle.

Assume that one of the points is  $A$  in figure 3. Then the second point has to lie between  $A'$  and  $A''$  in the circle for the chord to be greater than the triangle side. This is just  $\frac{1}{3}$  of the circumference and the searched probability is then  $\frac{1}{3}$ .

**Method 2:** A chord is characterized by its distance  $p$  to the center (between 0 and 1) and its orientation  $\theta$  w.r.t. a fixed line (between 0 and  $2\pi$ ). If we draw the equilateral triangle with a side parallel to the chord, we can see that the distance  $d$  has to be less than  $\frac{1}{2}$  in order to have a chord length greater than  $\sqrt{3}$ . Assuming a uniform orientation and distance to the origin, we find then a normalized probability of  $\frac{1}{2}$  (since  $p$  varies between 0 and 1).

**Method 3:** A chord is uniquely defined by the orthogonal projection  $I$  of the circle center onto it. It has to lie inside the disc of radius  $\frac{1}{2}$  in order to have a sufficient length. So, assuming  $I$  is uniformly distributed over the interior of the circle, the normalized probability is  $\frac{1}{4}$ .

The above three solutions are correct but they do not refer to the same notion of uniformity in the way we choose the chord. Using the  $(p, \theta)$  representation (described in the second method), we can compute (see [KM63]) that the probability measures are respectively

$$d\sigma_1 = \frac{dp \cdot d\theta}{2\pi\sqrt{1-p^2}} \quad d\sigma_2 = \frac{dp \cdot d\theta}{2\pi} \quad d\sigma_3 = p \cdot \frac{dp \cdot d\theta}{\pi}$$

and only the second one is invariant under the action of rotations, translations and reflections.

The solution to this problem is to impose an invariance constraint, or more precisely to *define* the notion of uniformity: for instance, uniform on  $\mathbb{R}$  means that the probability for a point to lie on an interval  $]x; x+d[$  is the same for all  $x$ . This is basically an invariance by translation. In the same way, and since we can only compare geometric objects with a transformation group, we define the uniform (or invariant) measure (the infinitesimal volume element) as the measure being invariant by the action of any fixed element  $f$  of the group. Let  $d\mathcal{M}(x)$  be such a measure ( $\mathcal{M}$  stands for manifold); this means that  $d\mathcal{M}(f \star x) = d\mathcal{M}(x)$  for any  $x$ . The invariant measure can be sought directly on a given representation (see [KM63]), but a general formalism is developed in [San76] to extract it from the left-invariant measure  $d_L\mathcal{G}$  of the group  $\mathcal{G}$ .

### 3.2 Left and right invariant measures on a Lie group (Haar measures)

We can require left invariance ( $d_L\mathcal{G}(g \circ f) = d_L\mathcal{G}(f)$ ) for any fixed  $g \in \mathcal{G}$ , or right invariance ( $d_R\mathcal{G}(f \circ g) = d_R\mathcal{G}(f)$ ). Since the group acts on the left (the application of transformation  $f$  to feature  $x$  is  $f \star x$ ), we are mainly interested in left-invariance and we will mean by invariant measure the left-invariant one.

To be mathematically correct, we require that for any continuous real function  $\alpha$  on  $\mathcal{G}$  with compact support, we have:

$$\forall g \in \mathcal{G} \quad \int_{f \in \mathcal{G}} \alpha(g \circ f) \cdot d_L\mathcal{G}(g \circ f) = \int_{f \in \mathcal{G}} \alpha(f) \cdot d_L\mathcal{G}(f) \quad (3)$$

If the group is locally compact, then it can be proven ([Hoc65]) that there exists only one left-invariant measure (up to a scale factor) that satisfies the above properties. This measure is called the (left) Haar measure of the group. In a symmetric way, there is also a unique right Haar measure.

It is interesting to note that the left and right invariant measures are generally different: their difference is quantified by the *module*  $\Delta\mathcal{G}$  defined by:  $d_L\mathcal{G}(f) = \Delta\mathcal{G}(f) \cdot d_R\mathcal{G}(f)$ . The group is called unimodular if the module is identity everywhere, which means that left and right measures are equal. A compact group is always unimodular, but only locally compact groups can have different left and right Haar measures. For instance, left and right Haar measures are identical on  $SO_3$ , since the 3D rotation group is compact, and the 3D rigid motion group is unimodular although the group is only locally compact due to the introduction of translations (see below).

The left and right invariant measures can be generally computed from the Maurer-Cartan equations ([San76]), but a very interesting theorem allows easy computation in the case of a *minimal* representation: assume that the definition domain of a chart almost covers the group<sup>1</sup> and that the Jacobian of the left translate of the identity  $J_L(f)$  exists and is continuous almost everywhere. Then the invariant measure can be written (see appendix 3.A.1):

$$d_L\mathcal{G}(f) = \frac{df}{|J_L(f)|} \quad \text{with} \quad J_L(f) = \left. \frac{\partial(f \circ e)}{\partial e} \right|_{e=I_d} \quad (4)$$

---

<sup>1</sup>Since we integrate function (and not distributions), we can “forget” a subset of the group that have a null measure.



where  $|J| = |\det(J)|$ . The right-invariant measure can be derived in the same way using the Jacobian  $J_R$  of the right translate of the identity, and the module is thus:

$$\Delta\mathcal{G}(f) = \frac{|J_R(f)|}{|J_L(f)|}$$

An interesting property of the module is that  $\Delta\mathcal{G}(f^{(-1)}) = \Delta\mathcal{G}(f)^{(-1)}$ . Using this scheme, we can show that the uniform measure for rotations using the rotation vector representation is

$$d_L\mathcal{G}(r) = d_R\mathcal{G}(r) = \frac{4 \sin^2(\theta/2)}{\theta^2} dr \quad (5)$$

where  $\theta = \|r\|$ . The invariant measure on rigid transformations is, with our representation  $f = (r, t)$ ,

$$d_L\mathcal{G}(f) = d_R\mathcal{G}(f) = \frac{4 \sin^2(\theta/2)}{\theta^2} dr.dt \quad (6)$$

Rotations and rigid transformations are thus unimodular.

### 3.3 Invariant measure on homogeneous manifolds

We saw in section 2.4 how to identify the homogeneous manifold  $\mathcal{M}$  with the quotient space  $\mathcal{G}/\mathcal{H}$ . We can find, thanks to the above section, the (left) invariant measures  $d_L\mathcal{G}$  and  $d_L\mathcal{H}$  on  $\mathcal{G}$  and  $\mathcal{H}$ , and write  $d_L\mathcal{G} = d\mathcal{M}.d_L\mathcal{H}$  where  $d\mathcal{M}$  is a measure on the manifold  $\mathcal{M}$  (or the quotient space  $\mathcal{G}/\mathcal{H}$ ). Santalo gives in [San76] several forms of a necessary and sufficient condition for  $d\mathcal{M}$  to be an invariant measure (i.e.  $d\mathcal{M}(g \star x) = d\mathcal{M}(x)$ ). One of them can be stated as follows ( $e$  is an element of the manifold  $\mathcal{M}$ ):

$$\forall h \in \mathcal{H}, \quad \left| \frac{\partial(h \star e)}{\partial e} \Big|_{e=\mathcal{O}} \right| = 1 \quad (7)$$

This means that the measure of the infinitesimal volume element at the origin remains unchanged under any transformation that keep the origin unchanged. If this condition is not satisfied, then there is no invariant measure on the manifold, otherwise we can compute it (with a minimal representation of the manifold) in a way similar to what we did for the group. Assume that the definition domain of a chart almost covers the manifold and that the Jacobian of the translation of the origin  $J(f)$  exists and is continuous almost everywhere. Then the invariant measure is

$$d\mathcal{M}(x) = \frac{dx}{|J(f_x)|} \quad \text{with} \quad J(f_x) = \frac{\partial(f_x \star e)}{\partial e} \Big|_{e=\mathcal{O}} \quad \text{and} \quad f_x \in \hat{x} \quad (8)$$

### 3.4 Practical use: probability of a false match

Assume that  $x$  is a uniform random feature in the first image (characterized by a set of possible features  $I_1$ ). What is the probability that this feature be accepted as a match with feature  $y$  in the second image, under a given global transformation  $f$  ?

If we characterize the possible matches for  $y$  by an “error volume”  $Z(y)$  around  $y$ , we can write this probability as the conditional probability

$$P(f \star x \leftrightarrow y) = P((f \star x) \in Z(y) | x \in I_1)$$

$$P(f \star x \leftrightarrow y) = \frac{\int_{(f \star I_1) \cap Z(y)} d\mathcal{M}(x)}{\int_{I_1} d\mathcal{M}(x)} = \frac{\mathcal{V}((f \star I_1) \cap Z(y))}{\mathcal{V}(I_1)}$$

where  $\mathcal{V}(X)$  is the “volume” of the set  $X$ . With the assumption that the volume  $\mathcal{V}(Z(y))$  is sufficiently small with respect to the volume of the image, we can consider that the transformed image  $I_1$  either contains the whole set  $Z(y)$  or does not intersect it at all. This allows us to approximate the above probability by

$$P(f \star x \leftrightarrow y) = \varepsilon \frac{\mathcal{V}(Z(y))}{\mathcal{V}(I_1)} \quad \text{with} \quad \begin{cases} \varepsilon = 1 & \text{if } f^{(-1)} \star y \in I_1 \\ \varepsilon = 0 & \text{otherwise} \end{cases}$$

A desirable property for our “error volume”  $Z(y)$  is that it be comparable at every point (as we usually fix the same bound for error on all the points): this means that, for any pair of points  $y$  and  $y'$  on the manifold, there exists a transformation  $f$  such that  $f \star y = y'$  and  $Z(y') = f \star Z(y)$  (the error volume is said to be homogeneous). A stronger hypothesis is that for every transformation  $f$ , the error volume on the transformed point is the transformation of the error volume:  $Z(f \star y) = f \star Z(y)$ . The volume is said to be isotropic in this case, and is completely determined by its shape around the origin. In both cases (homogeneity and isotropy), the volume of the error volume is invariant:  $\mathcal{V}(Z(y)) = \mathcal{V}(Z(\mathcal{O})) = \mathcal{V}_0$ . The basic probability of a false match we obtain can now be applied as usual in an analysis of the frequency of false positives (see [GH90, LW91, GHJ93, Pen93]).

As a practical example, we considered in [PA94] that two frames are matched if the distance between their point is less than a threshold  $d_0$  and if the rotation needed to adjust their trihedra has an angle less than a threshold  $\theta_0$ . This is in fact a bound on an invariant distance (see section 5.5), the volume is thus invariant and we can compute it at the origin. If  $f = (r, t)$  is a frame, it is in the error volume  $Z(I_d)$  if  $\theta = \|r\| < \theta_0$  and  $\|t\| < d_0$ . Using the invariant measure of equation 6, we can compute the volume of the error zone:

$$\mathcal{V}_0 = \int_{\theta < \theta_0} \int_{\|t\| < d_0} d\mathcal{M}(r, t) = \left( \int_{\theta < \theta_0} \frac{\sin^2(\theta/2)}{\theta^2} dr \right) \cdot \left( \int_{\|t\| < d_0} dt \right) = 2\pi(\theta_0 - \sin \theta_0) \cdot \left( \frac{4\pi}{3} d_0^3 \right)$$

If we assume a cubic image of side  $l$  (256 for instance) this gives a Euclidean volume  $V_I = l^3$  for points in which trihedrons are not constrained: the rotation volume is  $2\pi^2$ . We obtain finally the basic probability of false match:

$$P(f \star x \leftrightarrow y) = \left( \frac{\theta_0 - \sin \theta_0}{\pi} \right) \frac{4}{3} \left( \frac{d_0}{l} \right)^3 \varepsilon$$

We have isolated in the first term the probability of false match due to trihedra only, which reflects the gain in selectivity when using frames instead of points. This function is plotted in figure 4 and shows very interesting results: even for a bound of  $\theta_0 = \pi/2 = 90$  deg, more than 80% of the random matches are rejected. For a bound of  $\theta_0 = \pi/10 \simeq 20$  deg, the probability of a false match drops to 0.0016: we would have to divide the bound on the position by 10 to obtain an equivalent selectivity using points only.

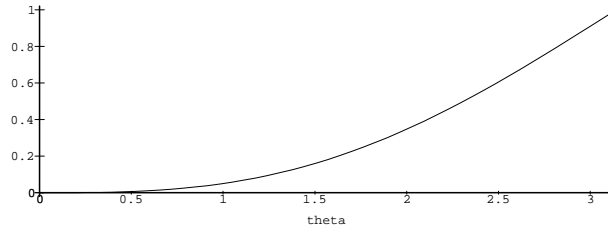


Figure 4: Basic probability of a false match for trihedrons with a bound on the angle for the adjustment rotation of theta.

### 3.A Appendix A: Proofs for invariant measures

#### 3.A.1 Invariant measures on a group

Let  $d_L \mathcal{G}(f)$  be the following measure:

$$d_L \mathcal{G}(f) = \frac{df}{|J_L(f)|} \quad \text{with} \quad J_L(f) = \left. \frac{\partial(f \circ e)}{\partial e} \right|_{e=I_d}$$

where  $|J| = |\det(J)|$ . We want to show that it is left invariant. We have:

$$d_L \mathcal{G}(g \circ f) = \frac{d(g \circ f)}{\left| \left. \frac{\partial((g \circ f) \circ e)}{\partial e} \right|_{e=I_d} \right|} \quad \text{and} \quad d(g \circ f) = \left| \det \left( \frac{\partial(g \circ f)}{\partial f} \right) \right| df$$

and we get by the chain rule:

$$\left. \frac{\partial((g \circ f) \circ e)}{\partial e} \right|_{e=I_d} = \frac{\partial(g \circ f)}{\partial f} \cdot \left. \frac{\partial(f \circ e)}{\partial e} \right|_{e=I_d}$$

Since  $\det(A.B) = \det(A) \cdot \det(B)$  for square matrices, we can conclude that  $d_L \mathcal{G}(g \circ f) = d_L \mathcal{G}(f)$ . The proof for the right invariance of the proposed measure  $d_R \mathcal{G}(f)$  is analogous.

### 3.A.2 Invariant measure on an homogeneous manifold

Let  $d\mathcal{M}(x)$  be the following measure:

$$d\mathcal{M}(x) = \frac{dx}{|J(f_x)|} \quad \text{with} \quad J(f_x) = \left. \frac{\partial(f_x \star e)}{\partial e} \right|_{e=\mathcal{O}} \quad \text{and} \quad f_x \in \hat{x}$$

$$\text{The condition (7):} \quad \forall h \in \mathcal{H}, \quad \left| \left. \frac{\partial(h \star e)}{\partial e} \right|_{e=\mathcal{O}} \right| = 1$$

is indeed required for  $d\mathcal{M}(x)$  to be invariant with respect to the choice of  $f_x \in \hat{x}$ . Let  $f_x$  and  $f'_x = f_x \circ h$  (with  $h \in \mathcal{H}$ ) be two transformations of  $\hat{x}$ . By the chain rule, we can write:

$$J(f'_x) = \left. \frac{\partial(f_x \star (h \star e))}{\partial e} \right|_{e=\mathcal{O}} = \left. \frac{\partial(f_x \star e')}{\partial e'} \right|_{e'=\mathcal{O}} \cdot \left. \frac{\partial(h \star e)}{\partial e} \right|_{e=\mathcal{O}} = J(f_x) \cdot \left. \frac{\partial(h \star e)}{\partial e} \right|_{e=\mathcal{O}}$$

and have  $|J(f_x)| = |J(f'_x)|$  if and only if  $\left| \left. \frac{\partial(h \star e)}{\partial e} \right|_{e=\mathcal{O}} \right| = 1$ . The proof of the invariance of the measure  $d\mathcal{M}(x)$  is then very similar to the proof for the group measure (and can be obtained by replacing  $(f \circ e)$  by  $(f \star e)$ ).

## 3.B Appendix B: Proofs for invariant measures on rigid transformations

The difficult part of the computations is the Jacobian of the composition of rotation vectors.

### 3.B.1 Composition of rotation vectors

The general operations on rotations (with rotation vectors and quaternions) and their Jacobians are detailed in [PT95a]. Since we are interested in a closed form for the Jacobian of the left and right translation of identity, we recall the results and expand the computations in this case.

Let  $r_1$  and  $r_2$  be two rotation vectors. To compute their composition, we compute the associated unit quaternions  $q_1$  and  $q_2$  and multiply them:  $q = q_2 * q_1$ . Then we come back to the rotation vector  $r(q)$ .

**Jacobian of the left translation of identity** From  $r_2 \circ r_1 = r(q(r_2) * q(r_1))$ , we derive with the chain rule:

$$\left. \frac{\partial r_2 \circ r_1}{\partial r_1} \right|_{r_1=0} = \left. \frac{\partial r(q)}{\partial q} \right|_{q=q_2} \cdot \left. \frac{\partial q_2 \star q_1}{\partial q_1} \right|_{q_2=q(r_2)} \cdot \left. \frac{\partial q(r_1)}{\partial r_1} \right|_{r_1=0}$$

**From rotation vector to unit quaternion** Let  $q = (\alpha, \beta)$  the unit quaternion associated with rotation vector  $r$ . Introducing the following functions of  $\theta = \|r\|$ :

$$b = \frac{\sin(\theta/2)}{\theta} = \frac{1}{2} - \frac{\theta^2}{48} + O(\theta^4) \quad c = \frac{\sin(\theta/2)}{\theta^3} - \frac{\cos(\theta/2)}{2\theta^2} = \frac{1}{24} \left( 1 - \frac{\theta^2}{40} \right) + O(\theta^4)$$

we obtain that  $\alpha = \cos(\theta/2)$ ,  $\beta = b.r$  and

$$\frac{\partial q}{\partial r} = \begin{bmatrix} \frac{\partial \alpha}{\partial r} \\ \frac{\partial \beta}{\partial r} \end{bmatrix} \quad \text{with} \quad \frac{\partial \alpha}{\partial r} = -\frac{\beta^\top}{2} \quad \text{and} \quad \frac{\partial \beta}{\partial r} = b I_d - c.r.r^\top$$

In our case, we have  $r_1 = 0$ , and thus  $q_1 = (1, 0)$  and  $\frac{\partial q_1}{\partial r_1} = \frac{1}{2} \begin{bmatrix} 0 \\ I_d \end{bmatrix}$ . On the other hand,  $r_2 = \theta.n$  gives  $q_2 = (\cos(\theta/2), \frac{\sin(\theta/2)}{\theta} r_2)$ .

**Composition of quaternions** This is simply the quaternion multiplication:

$$q = q_2 * q_1 = \begin{cases} \alpha_1 \alpha_2 - \langle \beta_1 | \beta_2 \rangle \\ \beta_2 \times \beta_1 + \alpha_2 \beta_1 + \alpha_1 \beta_2 \end{cases}$$

and since  $q_1 = (1, 0)$ , we obtain  $q = q_2$ . The Jacobian is thus easily obtained:

$$\frac{\partial q}{\partial q_1} = \frac{\partial q_2 * q_1}{\partial q_1} \Big|_{q_2=q(r_2)} = \frac{\cos(\theta/2)}{2} I_d + \frac{\sin(\theta/2)}{\theta} \begin{bmatrix} 0 & -r^\top \\ r & \tilde{r} \end{bmatrix}$$

**From unit quaternion to rotation vector** Let  $q = (\alpha, \beta)$  be a unit quaternion and  $a$  and  $c$  the following functions of  $\alpha$  and  $b = \|\beta\|$

$$a = 2 \operatorname{sign}(\alpha) \frac{\arcsin(b)}{b} = 2 \operatorname{sign}(\alpha) \left(1 + \frac{b^2}{6}\right) + O(b^4)$$

$$c = \frac{2\alpha - a}{b^2} = 2 \operatorname{sign}(\alpha) \frac{b\sqrt{1-b^2} - \arcsin(b)}{b^3} = -2 \operatorname{sign}(\alpha) \left(\frac{2}{3} + \frac{b^2}{5}\right) + O(b^4)$$

where  $\operatorname{sign}(0) = \pm 1$

$$r = a.\beta \quad \text{and} \quad \frac{\partial r}{\partial q} = \frac{\partial r}{\partial(\alpha, \beta)} = \begin{bmatrix} -2\beta; & a I_d + c.\beta.\beta^\top \end{bmatrix}$$

In our case, we obtain  $b = \|\beta\| = \sin \theta/2$   $a = \frac{\theta}{\sin(\theta/2)}$   $c = \frac{2 \cos(\theta/2) - \frac{\theta}{\sin(\theta/2)}}{\sin(\theta/2)^2}$  and

$$\frac{\partial r}{\partial q} = \begin{bmatrix} -2 \frac{\theta}{\sin(\theta/2)} r; & \frac{\theta}{\sin(\theta/2)} I_d + \left(2 \cos(\theta/2) - \frac{\theta}{\sin(\theta/2)}\right) \frac{r.r^\top}{\theta^2} \end{bmatrix}$$

**Composition of Jacobians** Since  $\frac{\partial q}{\partial q_1} \frac{\partial q_1}{\partial r_1} = \frac{\cos(\theta/2)}{2} \begin{bmatrix} 0 \\ I_d \end{bmatrix} + \frac{\sin(\theta/2)}{2\theta} \begin{bmatrix} -r^\top \\ \tilde{r} \end{bmatrix}$  we get

$$J_1(r) = \frac{\partial r \circ e}{\partial e} \Big|_{e=0} = \frac{\theta/2}{\tan(\theta/2)} I_d + \frac{1}{\theta^2} \left(1 - \frac{\theta/2}{\tan(\theta/2)}\right) r.r^\top + \frac{\tilde{r}}{2} \quad (9)$$

We note that this Jacobian is well define everywhere since  $\frac{\theta/2}{\tan(\theta/2)} = 1 - \frac{\theta^2}{12} + O(\theta^4)$ . The determinant of this Jacobian is

$$\det(J_1) = \frac{\theta^2}{4 \sin(\theta/2)^2}$$

**Jacobian of the right translation of identity** The derivation is similar to the left translation, except that

$$\frac{\partial q}{\partial q_2} = \frac{\partial q_2 * q_1}{\partial q_2} \Big|_{q_2=q(r_2)} = \frac{\cos(\theta/2)}{2} I_d + \frac{\sin(\theta/2)}{\theta} \begin{bmatrix} 0 & -r^\top \\ r & -\tilde{r} \end{bmatrix}$$

and we obtain

$$J_r(r) = \left. \frac{\partial r \circ e}{\partial e} \right|_{\epsilon=0} = \frac{\theta/2}{\tan(\theta/2)} I_d + \frac{1}{\theta^2} \left( 1 - \frac{\theta/2}{\tan(\theta/2)} \right) r \cdot r^\top - \frac{\tilde{r}}{2} \quad (10)$$

The determinant is this time:

$$\det(J_r) = \frac{-\theta^2}{4 \sin(\theta/2)^2}$$

### 3.B.2 Invariant measures on rigid transformations

The Jacobian of the left translation of identity is now easily derived: let  $f = (r, t)$  be a rigid transformation:

$$J_l(f) = \left. \frac{\partial f \circ e}{\partial e} \right|_{\epsilon=0} = \begin{bmatrix} \left. \frac{\partial r \circ r_1}{\partial r_1} \right|_{r_1=0} & \mathbf{0} \\ \mathbf{0} & R \end{bmatrix} = \begin{bmatrix} J_l(r) & \mathbf{0} \\ \mathbf{0} & R \end{bmatrix}$$

We have thus  $\det(J_l(f)) = \det(J_l(r)) \cdot \det(R) = \frac{\theta^2}{4 \sin(\theta/2)^2}$ . Similarly, the Jacobian of the right translation of identity is

$$J_r(f) = \left. \frac{\partial e \circ f}{\partial e} \right|_{\epsilon=0} = \begin{bmatrix} \left. \frac{\partial r_2 \circ r}{\partial r_2} \right|_{r_2=0} & \mathbf{0} \\ \left. \frac{\partial r_2 \circ t}{\partial r_2} \right|_{r_2=0} & I_d \end{bmatrix} = \begin{bmatrix} J_r(r) & \mathbf{0} \\ -\tilde{t} & I_d \end{bmatrix}$$

which determinant is  $\det(J_r(f)) = \det(J_r(r)) \cdot \det(I_d) = \frac{-\theta^2}{4 \sin(\theta/2)^2}$ . The left and right invariant measures are thus identical:

$$d_L \mathcal{G}(r) = d_R \mathcal{G}(r) = \frac{4 \sin^2(\theta/2)}{\theta^2} dr \cdot dt \quad (11)$$

## 4 Probability density functions

The invariant measure allows us to tackle geometric probability problems with *uniformly* distributed features. In the real world, however, we are often confronted with noisy measures of a specific feature where the probability distribution is not uniform but rather centered around the specific feature. The notion of probability density function (pdf) is especially well adapted to deal with this type of error in measurements, but rather difficult to use in a computational way. A well known approximation of the pdf is to use the first and second orders, which are the expected value and the covariance matrix. The problem we are confronted with is how to define these operations independently of the representation used, and in a way coherent with the action of the associated transformation group. We will see for instance in section 6.1 that the standard mean value fails to be “invariant” by a transformation, even with very simple features such as 2D vectorial lines.

In this section, we present a notion of probability density functions based on the invariant measure which is then independent of the chosen representation. From this pdf, we can easily derive the classical pdf in any representation. We will see in section 6 how to define expected values of these pdfs with the help of an invariant distance.

### 4.1 Definition

Let  $\mathcal{A}$  be the Borelian tribe of manifold  $\mathcal{M}$  (the tribe generated by the class of open sets of  $\mathcal{M}$ ). A probability  $P$  is a  $\sigma$ -additive function from  $\mathcal{A}$  to  $\mathbb{R}_+$  (a measure on  $\mathcal{A}$ ) such that  $P(\mathcal{M}) = 1$ . This is the function that maps to each set  $X$  of  $\mathcal{M}$  (each event) the probability of its occurrence  $P(X)$ . If there exists a continuous, real positive function  $p$  on  $\mathcal{M}$  with a compact support such that

$$\forall X \in \mathcal{A}, \quad P(\mathbf{x} \in X) = \int_X p(y) \cdot d\mathcal{M}(y) \quad \text{and} \quad P(\mathcal{M}) = \int_{\mathcal{M}} p(y) \cdot d\mathcal{M}(y) = 1$$

then  $p$  is called the probability density function (pdf) of probability  $P$ . Probability density functions are particularly well adapted to define the result of a measurement and are easy to manipulate. We used for instance in the previous section the uniform pdf over a bounded set  $X$  ( $\mathcal{V}(X)$  is the volume of the set  $X$  with respect to the invariant measure):

$$p_X(y) = \frac{1}{\int_X d\mathcal{M}} \text{Ind}_X(y) = \frac{\text{Ind}_X(y)}{\mathcal{V}(X)}$$

Pdfs can be defined in the same way for the transformation group  $\mathcal{G}$  using the left or right invariant measure. We note that if the group is not unimodular, left and right pdfs are different. From now on, we consider that the left-invariant measure is used.

## 4.2 From the “invariant” pdf to the classical pdf on the representation

Let  $p_{\mathbf{x}}$  be the pdf of a random feature  $\mathbf{x}$  (we identify here the point of the manifold with its representation  $x$ ). Since the representation is a vector space (limited by the definition domain), the classical pdf  $\rho_{\mathbf{x}}$  on the representation is defined using the Lebesgue measure  $dx$  instead of the invariant measure  $d\mathcal{M}(x)$ . Incorporating the definition of the invariant measure given in equation 8, the probability for  $\mathbf{x}$  to be in a set  $X$  is then:

$$P(\mathbf{x} \in X) = \int_X p_{\mathbf{x}}(y) \cdot d\mathcal{M}(y) = \int_X \frac{p_{\mathbf{x}}(y)}{|J(f_y)|} \cdot dy = \int_X \rho_{\mathbf{x}}(y) \cdot dy$$

where  $J(f_y)$  is the Jacobian of the application of  $f_y \in \hat{y}$  to the origin. We can then see that the classical pdf on the representation is given by

$$\rho_{\mathbf{x}}(y) = \frac{p_{\mathbf{x}}(y)}{|J(f_y)|} \tag{12}$$

Using the invariant measure to define the pdf has several advantages over other measures. The first one is that the pdf is invariant with respect to the representation: if  $x$  is a point of the manifold and  $\pi_1(x)$ ,  $\pi_2(x)$  two different representations of that feature, then  $p(\pi_1(x)) = p(\pi_2(x)) = p(x)$ , although in general  $\rho_{\pi_1}(\pi_1(x)) \neq \rho_{\pi_2}(\pi_2(x))$ .

## 4.3 Propagation of one random feature or transform

### 4.3.1 Application of a fixed transformation to a random feature

The second important property is the very simple transformation of a pdf under the action of a transformation (from the associated group  $\mathcal{G}$ ): let  $p_{\mathbf{x}}$  be the pdf of a random feature  $\mathbf{x}$ , and  $f_o \in \mathcal{G}$  a fixed transformation. Then the pdf of the random feature  $f_o \star \mathbf{x}$  is simply

$$p_{(f_o \star \mathbf{x})}(y) = p_{\mathbf{x}}(f_o^{-1} \star y) \tag{13}$$



### 4.3.2 Composition of fixed and random transformations

Unless the group is unimodular, an important difference occurs if the fixed transformation is on the left or on the right. Indeed, we consider here pdf defined with the left invariant measure. Let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$ ,  $f_o$  a fixed transformation, and  $\mathbf{g}_l = f_o \circ \mathbf{f}$  the left translation of  $\mathbf{f}$ . Following exactly the same proof as above, we obtain:

$$p_{(f_o \circ \mathbf{f})}(g) = p_{\mathbf{f}}(f_o^{(-1)} \circ g) \quad (14)$$

If we now consider the right composition  $\mathbf{g}_r = \mathbf{f} \circ f_o$ , we obtain

$$p_{(\mathbf{f} \circ f_o)}(g) = \frac{\Delta \mathcal{G}(g \circ f_o^{(-1)})}{\Delta \mathcal{G}(g)} \cdot p_{\mathbf{f}}(g \circ f_o^{(-1)}) \quad (15)$$

### 4.3.3 Application of a random transformation to a fixed feature

Let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$ . We determine first the pdf of the random feature  $\mathbf{y} = \mathbf{f} \star \mathcal{O}$  (see appendix 4.A):

$$p_{(\mathbf{f} \star \mathcal{O})}(y) = \int_{\mathcal{H}} p_{\mathbf{f}}(f_y \circ h) \cdot d\mathcal{H}(h) \quad (16)$$

If we want now to apply the random transformation  $\mathbf{f}$  to a fixed feature  $x$ , we can write:  $\mathbf{y} = \mathbf{f} \star x = (\mathbf{f} \circ f_x) \star \mathcal{O}$ , where  $f_x \in \hat{x}$ . We just have to replace the pdf of  $\mathbf{f}$  by its right translation with  $f_x$  in the above formula and we obtain:

$$p_{(\mathbf{f} \star x)}(y) = \int_{\mathcal{H}} p_{(\mathbf{f} \circ f_x)}(f_y \circ h) \cdot d\mathcal{H}(h) = \int_{\mathcal{H}} \frac{\Delta \mathcal{G}(f_y \circ h \circ f_x^{(-1)})}{\Delta \mathcal{G}(f_y \circ h)} \cdot p_{\mathbf{f}}(f_y \circ h \circ f_x^{(-1)}) \cdot d\mathcal{H}(h) \quad (17)$$

## 4.4 Canonical operations on random features and transformations

We just saw how the pdf propagates through the canonical operations if only one feature or transform is randomly distributed. The question is now how to propagate the pdf if all the geometric primitives are random ones. This will give us a coherent set of operations on probabilistic features and transformations, at least at a mathematical level.

#### 4.4.1 Composition of two random transformations

Let  $\mathbf{f}_1$  and  $\mathbf{f}_2$  be two random transformations of pdf  $p_{\mathbf{f}_1}$  and  $p_{\mathbf{f}_2}$ . The pdf of the random transformation  $\mathbf{f} = \mathbf{f}_1 \circ \mathbf{f}_2$  is (see appendix 4.A):

$$p_{(\mathbf{f}_1 \circ \mathbf{f}_2)}(f) = \int_{\mathcal{G}} p_{\mathbf{f}_1}(g) \cdot p_{\mathbf{f}_2}(g^{(-1)} \circ f) \cdot d_L \mathcal{G}(g) \quad (18)$$

We should note the very high similarity between this equation and the classical convolution product of pdf for the addition of random vectors in  $\mathbb{R}^n$ : let  $\mathbf{x}$  and  $\mathbf{y}$  be two random vectors of pdf  $p_{\mathbf{x}}$  and  $p_{\mathbf{y}}$ . The pdf of the random vector  $\mathbf{x} + \mathbf{y}$  is given by the convolution product of  $p_{\mathbf{x}}$  and  $p_{\mathbf{y}}$ :

$$p_{(\mathbf{x} + \mathbf{y})}(z) = p_{\mathbf{x}} \otimes p_{\mathbf{y}}(z) = \int_{\mathbb{R}^n} p_{\mathbf{x}}(t) \cdot p_{\mathbf{y}}(z - t) \cdot dt$$

This formula can be found as a particular case of equation (18): if we consider the translation group of  $\mathbb{R}^n$ , we find that  $d_L \mathcal{G}(f) = df$  and since  $g^{(-1)} = -g$  and  $f \circ g = f + g$ , we have  $p_{(\mathbf{f}_1 + \mathbf{f}_2)}(f) = \int_{\mathbb{R}^n} p_{\mathbf{f}_1}(g) \cdot p_{\mathbf{f}_2}(f - g) \cdot dg$ .

#### 4.4.2 Inversion of a random transformation

Let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$ . The pdf of its inverse is

$$p_{\mathbf{f}^{(-1)}}(g) = \Delta \mathcal{G}(g^{(-1)}) \cdot p_{\mathbf{f}}(g^{(-1)}) \quad (19)$$

#### 4.4.3 Application of a random transformation to a random feature

Let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$  and  $\mathbf{x}$  a random feature of pdf  $p_{\mathbf{x}}$ . The pdf of the random feature  $\mathbf{y} = \mathbf{f} \star \mathbf{x}$  is

$$p_{\mathbf{f} \star \mathbf{x}}(y) = \int_{\mathcal{G}} p_{\mathbf{f}}(g) \cdot p_{\mathbf{x}}(g^{(-1)} \star y) \cdot d_L \mathcal{G}(g) \quad (20)$$

Once again, this formula is very close to a convolution product.

#### 4.5 Expectation of a real valued function

Let  $\varphi(x)$  be a real valued function on the manifold  $\mathcal{M}$ . The expectation of this function for the random feature  $\mathbf{x}$  of pdf  $p_{\mathbf{x}}$  is defined by

$$E_{\mathbf{x}}(\varphi) = \int_{\mathcal{M}} \varphi(y) \cdot p_{\mathbf{x}}(y) \cdot d\mathcal{M}(y) \quad (21)$$

The expectation is a linear operator: if  $\lambda_1$  and  $\lambda_2$  are two scalars and  $\varphi_1$  and  $\varphi_2$  two real functions, we have:  $E_{\mathbf{x}}(\lambda_1\varphi_1 + \lambda_2\varphi_2) = \lambda_1 E_{\mathbf{x}}(\varphi_1) + \lambda_2 E_{\mathbf{x}}(\varphi_2)$ .

An interesting property is what becomes the expectation when we apply a fixed transformation  $f$  to the random feature  $\mathbf{x}$ :

$$\begin{aligned} E_{(f\star\mathbf{x})}(\varphi) &= \int_{\mathcal{M}} \varphi(y) \cdot p_{(f\star\mathbf{x})}(y) \cdot d\mathcal{M}(y) = \int_{\mathcal{M}} \varphi(y) \cdot p_{\mathbf{x}}((f^{(-1)} \star y)) \cdot d\mathcal{M}(y) \\ &= \int_{\mathcal{M}} \varphi(f \star z) \cdot p_{\mathbf{x}}(z) \cdot d\mathcal{M}(z) \end{aligned}$$

with the change in variable  $z = f_o^{(-1)} \star y$ . Let thus  $\varphi_f$  be the translation of  $\varphi$  by the (fixed) transformation  $f$ :  $\varphi_f(x) = \varphi(f \star x)$ , we obtain:

$$E_{(f\star\mathbf{x})}(\varphi) = E_{\mathbf{x}}(\varphi_f)$$

In particular, if  $\varphi$  is invariant ( $\varphi_f = \varphi$  for any transformation  $f$ ), then its expectation is also invariant.

The expectation of a real value function can be define in a very similar way for a random transformation and the last property holds for the left composition.

## 4.6 Conclusion

We developed in this section a probability theory on geometric primitives based on probability density functions defined with the (left) invariant measure. These pdfs are independent of the chosen representation for the group or the manifold, but can be easily related to standard pdfs on the representation space (defined with the Lebesgue measure). We derive quite simple equations to propagate the pdf through the canonical operations of composition, inversion of transformations and application to a feature. From a mathematical point of view, we have thus obtained a coherent set of operations to work on probabilistic geometric primitives.

From a computational point of view, however, we need to approximate the pdf by tractable elements to be able to use this theory: section 6 deals with a new definition of expectation with nice properties, with the help of an invariant distance (next section).

## 4.A Appendix A: Proofs for probability density functions

### 4.A.1 Application of a fixed transformation to a random feature

Let  $p_{\mathbf{x}}$  be the pdf of a random feature  $\mathbf{x}$ , and  $f_o \in \mathcal{G}$  a fixed transformation. The probability for  $\mathbf{y} = f_o \star \mathbf{x}$  to be in a set  $Y$  is

$$P(f_o \star \mathbf{x} \in Y) = P(\mathbf{x} \in f_o^{(-1)} \star Y) = \int_{(f_o^{(-1)} \star Y)} p_{\mathbf{x}}(z) \cdot d\mathcal{M}(z)$$

which gives, with the change in variables  $z = f_o^{(-1)} \star y$  and since  $d\mathcal{M}$  is the invariant measure:

$$P(f_o \star \mathbf{x} \in Y) = \int_Y p_{\mathbf{x}}(f_o^{(-1)} \star y).d\mathcal{M}(y)$$

From the definition, the pdf of the random variable  $\mathbf{y} = f_o \star \mathbf{x}$  is thus  $p_{(f_o \star \mathbf{x})}(y) = p_{\mathbf{x}}(f_o^{(-1)} \star y)$

#### 4.A.2 Right translation of a random transformation

Let  $\mathbf{f}$  be a random transformation, and  $f_o$  a fixed one. We want to determine the pdf of the right composition  $\mathbf{g}_r = \mathbf{f} \circ f_o$ . If  $Y \in \mathcal{G}$  is a set of transformations, the probability for  $\mathbf{g}$  to be in this set is

$$P((\mathbf{f} \circ f_o) \in Y) = P(\mathbf{f} \in (Y \circ f_o^{(-1)})) = \int_{(Y \circ f_o^{(-1)})} p_{\mathbf{f}}(f').d_L\mathcal{G}(f')$$

With the change in variable  $f' = g \circ f_o^{(-1)}$ , we obtain

$$P((\mathbf{f} \circ f_o) \in Y) = \int_Y p_{\mathbf{f}}(g \circ f_o^{(-1)}).d_L\mathcal{G}(g \circ f_o^{(-1)}) = \int_Y \Delta\mathcal{G}(g \circ f_o^{(-1)}).p_{\mathbf{f}}(g \circ f_o^{(-1)}).d_R\mathcal{G}(g \circ f_o^{(-1)})$$

Now since  $d_R\mathcal{G}$  is the right invariant measure, we can simplify and come back to the left invariant measure:

$$P((\mathbf{f} \circ f_o) \in Y) = \int_Y \Delta\mathcal{G}(g \circ f_o^{(-1)}).p_{\mathbf{f}}(g \circ f_o^{(-1)}).d_R\mathcal{G}(g) = \int_Y \frac{\Delta\mathcal{G}(g \circ f_o^{(-1)})}{\Delta\mathcal{G}(g)}.p_{\mathbf{f}}(g \circ f_o^{(-1)}).d_L\mathcal{G}(g)$$

From the definition of the pdf, we can then conclude that:

$$p_{(\mathbf{f} \circ f_o)}(g) = \frac{\Delta\mathcal{G}(g \circ f_o^{(-1)})}{\Delta\mathcal{G}(g)}.p_{\mathbf{f}}(g \circ f_o^{(-1)})$$

#### 4.A.3 Application of a random transformation to a fixed feature

Let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$ . We determine first the pdf of the random feature  $\mathbf{x} = \mathbf{f} \star \mathcal{O}$ . The probability that  $\mathbf{x} = \mathbf{f} \star \mathcal{O}$  be in a set  $X \in \mathcal{M}$  is given by

$$P(\mathbf{f} \star \mathcal{O} \in X) = \int_{\hat{X}} p_{\mathbf{f}}(f').d\mathcal{G}(f') \quad \text{with} \quad \hat{X} = \{f' \in \mathcal{G} / f' \star \mathcal{O} \in X\}$$

Let  $f_x$  be a representative of  $\hat{x}$  for every  $x$  (i.e. we choose one transformation  $f_x$  in  $\hat{x}$ ). Then the set of transformations  $\hat{X}$  is  $\hat{X} = \{f_x \circ h / x \in X, h \in \mathcal{H}\}$ . The above probability can thus be written:

$$P(\mathbf{x} \in X) = \int_{X, \mathcal{H}} p_{\mathbf{f}}(f_{x'} \circ h).d\mathcal{M}(x').d\mathcal{H}(h) = \int_X \left( \int_{\mathcal{H}} p_{\mathbf{f}}(f_{x'} \circ h).d\mathcal{H}(h) \right).d\mathcal{M}(x')$$

The pdf  $p_{\mathbf{x}}$  of  $\mathbf{x} = \mathbf{f} \star \mathcal{O}$  is then

$$p_{\mathbf{x}}(y) = \int_{\mathcal{H}} p_{\mathbf{f}}(f_y \circ h).d\mathcal{H}(h)$$

#### 4.A.4 Composition of two random transformations

Let  $\mathbf{f}_1$  and  $\mathbf{f}_2$  be two random transformations of pdf  $p_{\mathbf{f}_1}$  and  $p_{\mathbf{f}_2}$ ,  $\mathbf{f} = \mathbf{f}_1 \circ \mathbf{f}_2$  their composition and  $F \subset \mathcal{G}$  a set of transformations. The set  $A$  of couples  $(g_1, g_2)$  such that  $g_1 \circ g_2 \in F$  can be written in two ways:

$$(1) \quad A = \left\{ (g_1, g_2) / g_1 \in \mathcal{G}, g_2 \in g_1^{(-1)} \circ F \right\}$$

$$(2) \quad A = \left\{ (g_1, g_2) / g_2 \in \mathcal{G}, g_1 \in F \circ g_2^{(-1)} \right\}$$

Using the first formulation, the probability for  $\mathbf{f} = \mathbf{f}_1 \circ \mathbf{f}_2$  to be in the set  $F$  is

$$\begin{aligned} P(\mathbf{f}_1 \circ \mathbf{f}_2 \in F) &= \int_{\mathcal{G}} \left( \int_{g_1^{(-1)} \circ F} p_{\mathbf{f}_2}(g_2) \cdot d_L \mathcal{G}(g_2) \right) p_{\mathbf{f}_1}(g_1) \cdot d_L \mathcal{G}(g_1) \\ &= \int_{\mathcal{G}} \left( \int_F p_{\mathbf{f}_2}(g_1^{(-1)} \circ g_2) \cdot d_L \mathcal{G}(g_2) \right) p_{\mathbf{f}_1}(g_1) \cdot d_L \mathcal{G}(g_1) \\ &= \int_F \left( \int_{\mathcal{G}} p_{\mathbf{f}_2}(g_1^{(-1)} \circ g_2) \cdot p_{\mathbf{f}_1}(g_1) \cdot d_L \mathcal{G}(g_1) \right) d_L \mathcal{G}(g_2) \end{aligned}$$

We obtain thus the requested pdf:

$$p_{(\mathbf{f}_1 \circ \mathbf{f}_2)}(f) = \int_{\mathcal{G}} p_{\mathbf{f}_1}(g) \cdot p_{\mathbf{f}_2}(g^{(-1)} \circ f) \cdot d_L \mathcal{G}(g)$$

Using the second formulation of the set  $A$ , we have:

$$\begin{aligned} P(\mathbf{f}_1 \circ \mathbf{f}_2 \in F) &= \int_{\mathcal{G}} \left( \int_{F \circ g_2^{(-1)}} p_{\mathbf{f}_1}(g_1) \cdot d_L \mathcal{G}(g_1) \right) p_{\mathbf{f}_2}(g_2) \cdot d_L \mathcal{G}(g_2) \\ &= \int_{\mathcal{G}} \left( \int_F p_{\mathbf{f}_1}(g \circ g_2^{(-1)}) \cdot d_L \mathcal{G}(g \circ g_2^{(-1)}) \right) p_{\mathbf{f}_2}(g_2) \cdot d_L \mathcal{G}(g_2) \\ &= \int_{\mathcal{G}} \left( \int_F \frac{\Delta \mathcal{G}(f \circ g^{(-1)})}{\Delta \mathcal{G}(f)} p_{\mathbf{f}_1}(g \circ g_2^{(-1)}) \cdot d_L \mathcal{G}(g) \right) p_{\mathbf{f}_2}(g_2) \cdot d_L \mathcal{G}(g_2) \end{aligned}$$

This gives the pdf

$$p_{(\mathbf{f}_1 \circ \mathbf{f}_2)}(f) = \int_{\mathcal{G}} \frac{\Delta \mathcal{G}(f \circ g^{(-1)})}{\Delta \mathcal{G}(f)} \cdot p_{\mathbf{f}_1}(f \circ g^{(-1)}) \cdot p_{\mathbf{f}_2}(g) \cdot d_L \mathcal{G}(g)$$

However, we note that the change in variables  $g_1 = g \circ g_2^{(-1)}$  bring back to the first form of the pdf.

#### 4.A.5 Inversion of a random transformation

The first step is to determine the relationship between  $d_L \mathcal{G}(g^{(-1)})$  and  $d_L \mathcal{G}(g)$ : since  $d(g^{(-1)}) = \left| \frac{\partial g^{(-1)}}{\partial g} \right| \cdot dg$ , we obtain:

$$d_L \mathcal{G}(g^{(-1)}) = \left| \frac{\partial g^{(-1)}}{\partial g} \right| \cdot \frac{|J_L(g)|}{|J_L(g^{(-1)})|} \cdot d_L \mathcal{G}(g)$$

On the other hand, we have:

$$J_R(g^{(-1)}) = \frac{\partial(e \circ g^{(-1)})}{\partial e} \Big|_{e=I_d} = \frac{\partial(e \circ g^{(-1)})}{\partial(g \circ e^{(-1)})} \Big|_{e=I_d} \cdot \frac{\partial(g \circ e^{(-1)})}{\partial e^{(-1)}} \Big|_{e^{(-1)}=I_d} \cdot \frac{\partial(e^{(-1)})}{\partial e} \Big|_{e=I_d}$$

and thus  $J_R(g^{(-1)}) = \frac{\partial g^{(-1)}}{\partial g} \cdot J_L(g) \cdot \frac{\partial(\epsilon^{(-1)})}{\partial \epsilon} \Big|_{\epsilon=I_d}$ . But since  $\frac{\partial(\epsilon^{(-1)})}{\partial \epsilon} \Big|_{\epsilon=I_d} = \frac{\partial(\epsilon')}{\partial \epsilon'^{(-1)}} \Big|_{\epsilon'=I_d} = \left( \frac{\partial(\epsilon^{(-1)})}{\partial \epsilon} \Big|_{\epsilon=I_d} \right)^{(-1)}$ , this term is the identity matrix, and the relationship between the measures simplifies to

$$d_L \mathcal{G}(g^{(-1)}) = \frac{|J_R(g^{(-1)})|}{|J_L(g^{(-1)})|} \cdot d_L \mathcal{G}(g) = \Delta \mathcal{G}(g^{(-1)}) \cdot d_L \mathcal{G}(g)$$

Now, let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$ , and  $F$  a set of transformations:

$$\begin{aligned} P(\mathbf{f}^{(-1)} \in F) &= P(\mathbf{f} \in F^{(-1)}) = \int_{F^{(-1)}} p_{\mathbf{f}}(g) \cdot d_L \mathcal{G}(g) \\ &= \int_F p_{\mathbf{f}}(g^{(-1)}) \cdot d_L \mathcal{G}(g^{(-1)}) = \int_F \Delta \mathcal{G}(g^{(-1)}) p_{\mathbf{f}}(g^{(-1)}) \cdot d_L \mathcal{G}(g) \end{aligned}$$

Which gives the requested pdf:

$$p_{\mathbf{f}^{(-1)}}(g) = \Delta \mathcal{G}(g^{(-1)}) \cdot p_{\mathbf{f}}(g^{(-1)})$$

#### 4.A.6 Application of a random transformation to a random feature

Let  $\mathbf{f}$  be a random transformation of pdf  $p_{\mathbf{f}}$ ,  $\mathbf{x}$  a random feature of pdf  $p_{\mathbf{x}}$  and  $X$  a set of features. The set  $A$  of couples  $(f, x)$  such that  $f \star x \in X$  can be written

$$A = \left\{ (f, x) / f \in \mathcal{G}, x \in f^{(-1)} \star X \right\}$$

and the probability that  $\mathbf{y} = \mathbf{f} \star \mathbf{x}$  be in  $X$  is thus

$$\begin{aligned} P(\mathbf{f} \star \mathbf{x} \in X) &= \int_{\mathcal{G}} \left( \int_{g^{(-1)} \star X} p_{\mathbf{x}}(y) \cdot d\mathcal{M}(y) \right) p_{\mathbf{f}}(g) \cdot d_L \mathcal{G}(g) = \int_{\mathcal{G}} \left( \int_X p_{\mathbf{x}}(g^{(-1)} \star z) \cdot d\mathcal{M}(z) \right) p_{\mathbf{f}}(g) \cdot d_L \mathcal{G}(g) \\ &= \int_X \left( \int_{\mathcal{G}} p_{\mathbf{x}}(g^{(-1)} \star z) p_{\mathbf{f}}(g) \cdot d_L \mathcal{G}(g) \right) \cdot d\mathcal{M}(z) \end{aligned}$$

The pdf of the random feature  $\mathbf{y} = \mathbf{f} \star \mathbf{x}$  is thus

$$p_{\mathbf{f} \star \mathbf{x}}(y) = \int_{\mathcal{G}} p_{\mathbf{f}}(g) \cdot p_{\mathbf{x}}(g^{(-1)} \star y) \cdot d_L \mathcal{G}(g)$$

## 5 Metric properties

A distance between points is often used to classify, quantify and minimize. This is even the core of some algorithms (Iterative Closest Point for instance). In order to be able to compare features (and generalize a number of methods on points), we also need a distance function on features. However, the distance has to be carefully chosen: a highly desirable property for this distance is the invariance under the transformation group and the invariance under the chosen representation (see example below).

A general method is given to generate an invariant distance on the manifold from a metric on the transformation space, and this allows the generalization of several algorithms based on distance to homogeneous features. The invariance property automatically guarantees the stability of the result with respect to the considered group of transformations.

### 5.1 The paradox of the closest line

We saw with the paradox of Bertrand different representations of 2D line. We use in this section another minimal representation based on the line equation ([Aya91]): the equation of a 2D line is  $a.x + b.y + c = 0$ . In order to obtain a minimal representation, we need to eliminate one parameter:

- **Chart 1:** lines that are not parallel to the  $X$  axis are represented by  $d = (a, p) \in \mathbb{R}^2$  and have equation:  $a.x + y + p = 0$ .
- **Chart 2:** lines that are not parallel to the  $Y$  axis are represented by  $d' = (a', p') \in \mathbb{R}^2$  and have equation:  $x + a'.y + p' = 0$ .

In the first chart, the line  $d = (a, p)$  cuts the  $Y$  axes at the point  $(0, -p)$  and has a director vector  $(1, -a)$ . This is symmetric in the second chart: the line  $d' = (a', p')$  cuts the  $X$  axes at the point  $(-p', 0)$  and has a director vector  $(-a', 1)$ .

We draw in figure 5 three lines. The problem is to choose which line  $d_2$  or  $d_3$  is the closest one to  $d_1$ . A definition of the distance that seems to be reasonable is  $\text{dist}(d_1, d_2) = \sqrt{(a_1 - a_2)^2 + (p_1 - p_2)^2}$ . The coordinates of the three lines in the first chart are  $d_1 = (-1/2; -1)$ ,  $d_2 = (-1; -1)$ ,  $d_3 = (-1/4; -1)$ , and the distance between line  $d_1$  and  $d_2$  turns out to be equal to the distance between line  $d_1$  and  $d_3$ :

$$\text{dist}(d_1, d_2) = \text{dist}(d_1, d_3) = 1/2$$

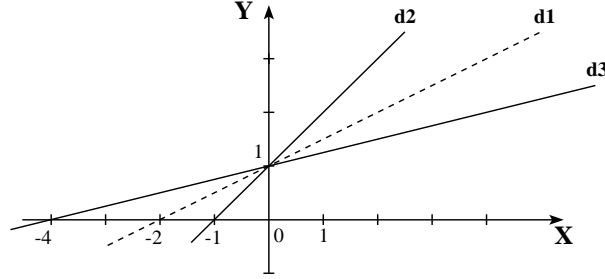


Figure 5: Three lines in the plane.

The line  $d_1$  is thus the mid-line between  $d_2$  and  $d_3$ . If we now consider the lines in the second chart, their coordinates are  $d_1 = (-2; 2)$ ,  $d_2 = (-1; 1)$ ,  $d_3 = (-4; 4)$  and the distances are now

$$\text{dist}(d_1, d_2) = \sqrt{2} \quad \text{and} \quad \text{dist}(d_1, d_3) = 2\sqrt{2}$$

We have then the following paradox:  $d_1$  is the mid-lines of  $d_2$  and  $d_3$  in one chart, and  $d_2$  is the closest line to  $d_1$  in the second chart. Which chart gives the most reasonable result? In fact, visually, the line  $d_3$  is the closest one: there is an angle of  $12.5^\circ$  between  $d_1$  and  $d_3$ , versus an angle of  $18.5^\circ$  between  $d_1$  and  $d_2$  (these angles are obviously invariant by rigid transformation).

## 5.2 Invariant distance on a manifold

Let  $x, y \in \mathcal{M}$  and  $g \in \mathcal{G}$ , then we require that  $\text{dist}(x, y) = \text{dist}(g \star x, g \star y)$ . This means in particular that this distance is completely defined by the distance  $N(x)$  of a feature  $x$  with the origin: if we use transformation  $f_y^{(-1)}$  or  $f_x^{(-1)}$ , we have

$$\text{dist}(x, y) = \text{dist}(f_y^{(-1)} \star x, \mathcal{O}) = N(f_y^{(-1)} \star x) = N(f_x^{(-1)} \star y) \quad (22)$$

The symmetry ( $\text{dist}(x, y) = \text{dist}(y, x)$ ), positivity ( $\text{dist}(x, y) \geq 0$ ), definite character ( $(\text{dist}(x, y) = 0) \Leftrightarrow (x = y)$ ) and triangular inequality ( $\text{dist}(x, y) + \text{dist}(y, z) \geq \text{dist}(x, z)$ ) for the distance are translated, under the invariance assumption, into the three following properties:

- $N(x) \geq 0$  and  $(N(x) = 0) \Leftrightarrow (x = \mathcal{O})$ .
- $N(f_x^{(-1)} \star \mathcal{O}) = N(x)$  for  $f_x \in \hat{x}$  and thus  $N(h \star x) = N(x)$  for any  $h \in \mathcal{H}$ .
- $N(x) + N(y) \geq N(f_y^{(-1)} \star x) = N(f_x^{(-1)} \star y)$  for any  $f_x \in \hat{x}$  and  $f_y \in \hat{y}$



These properties are very close to those required in order to define a norm on a vector space (without the positive linearity). To distinguish the function  $N$  from the associated invariant distance, we call  $N$  the “norm” of the manifold. We note that we have so far defined the metric on the manifold and not in a particular chart. In practice, we use a “principal chart”, centered around the origin and covering almost the manifold. The norm  $N$  is defined in this chart, and when we have to use the distance  $\text{dist}(x, y)$ , we compute  $N(f_y^{(-1)} \star x)$ .

### 5.3 Invariant distances on a Lie group

Assume that we work now on the transformation group  $\mathcal{G}$ . We can require the distance to be either left or right invariant. As above, a left invariant distance is determined by a norm  $N_g$  on the group, satisfying the following properties:

- $N_g(f^{(-1)}) = N_g(f)$ .
- $N_g(f) \geq 0$  and  $(N_g(f) = 0) \Leftrightarrow (f = I_d)$ .
- $N_g(f) + N_g(g) \geq N_g(g^{(-1)} \circ f) = N_g(f^{(-1)} \circ g)$ .

The triangular inequality becomes  $N_g(f) + N_g(g) \geq N_g(g \circ f^{(-1)}) = N_g(f \circ g^{(-1)})$  for a right invariant distance. The corresponding left and right invariant distances are

$$\begin{aligned} \text{dist}_L(f, g) &= N_g(g^{(-1)} \circ f) = N_g(f^{(-1)} \circ g) \\ \text{dist}_R(f, g) &= N_g(g \circ f^{(-1)}) = N_g(f \circ g^{(-1)}) \end{aligned}$$

We are only interested here in the left-invariant distance since it induces an invariant distance on an homogeneous manifold.

### 5.4 Distance induced on the manifold by the group distance

Let  $N_g$  be a norm on the group  $\mathcal{G}$ . We define the induced semi-norm on the homogeneous manifold  $\mathcal{M}$  as

$$N(x) = \inf_{(h \in \mathcal{H}, f_x \in \hat{x})} (N_g(h \circ f_x)) = \inf_{(h_1, h_2) \in \mathcal{H}^2} (N_g(h_1 \circ f_x \circ h_2)) \quad (23)$$

If the infimum of  $N_g(h_1 \circ f \circ h_2)$  is reached for every transformation  $f$  by  $(h_1, h_2) \in \mathcal{H}^2$ , then the semi-metric is separable and is thus a metric (see proofs in appendix 5.A). This property is always true if the isotropy group  $\mathcal{H}$  is compact but is not

automatically verified otherwise (for instance, there is no metric induced on points by the similarities or affine transformations).

Assuming that we have a norm, the distance associated with this metric is automatically invariant and satisfy

$$d(x, y) = \inf_{\{f_x \in \hat{x}; f_y \in \hat{y}\}} (\text{dist}_L(f_x, f_y)) = \inf_{\{(h_1, h_2) \in \mathcal{H}^2\}} \left( N_g(h_1 \circ f_x^{(-1)} \circ f_y \circ h_2) \right) \quad (24)$$

When a metric is chosen on the transformation space, we then automatically have an invariant distance on the manifold. Since features are usually objects abstracted from an Euclidean space, a reasonable requirement to make is that the distance induced on points of the original space is the canonical distance of the space (possibly up to a scale factor). We are then guaranteed reasonable invariant distances on all the features we consider.

### 5.5 Practical use on rigid transformations

The Euclidean distance on  $\mathbb{R}^3$  is induced by the  $L_2$  norm:  $d_t(x, y) = \|x - y\|$ . On the other hand, it can be shown ([Alt86]) that the angle  $\theta$  of a rotation is a norm that induce a left and right invariant distance on  $SO_3$ , the rotation group. With the rotation vector representation, we have then  $d_\theta(I_d, r) = \|r\| = \theta$  and thus  $d_\theta(r_1, r_2) = \|r_2^{(-1)} \circ r_1\| = \|r_1 \circ r_2^{(-1)}\|$  (the last term of the equality comes from the right invariance).

We define the norm on the rigid motion group as (see appendix 5.B):

$$N_\lambda(f) = N_\lambda((r, t)) = \|f\| = \sqrt{\lambda^2 \|r\|^2 + \|t\|^2}$$

where  $\lambda$  is a fixed parameter that allow to tune the importance of the trihedron (rotation part) with respect to the position (or translation part). Indeed, the angle of rotation  $\theta$  is in radian (or degrees, . . .) and the translation in millimeters, kilometers, or inches, . . . We usually scale each of the two terms by the inverse of their variation domain ( $\pi$  for  $\theta$  and the diameter  $l_0$  of the image or the interest object for the translation:  $\lambda = l_0/\pi$ ). When we have a an information about the noise level (i.e. standard deviations  $\sigma_\theta$  and  $\sigma_t$ ), we can also use  $\lambda = \sigma_t/\sigma_\theta$ .

We can check that the distance induced by this norm on the original space is the Euclidean distance (see appendix 5.B.1). The left-invariant distance is thus

$$\text{dist}_L(f_1, f_2)^2 = \|f_2^{(-1)} \circ f_1\|^2 = \lambda^2 \|r_2^{(-1)} \circ r_1\|^2 + \|t_1 - t_2\|^2$$

whereas the right invariant distance is

$$\text{dist}_R(f_1, f_2)^2 = \|f_1 \circ f_2^{(-1)}\|^2 = \lambda^2 \|r_1 \circ r_2^{(-1)}\|^2 + \|t_1 - (r_1 \circ r_2^{(-1)}) \star t_2\|^2$$

Although the rigid motion group is unimodular (and thus left and right invariant measures are identical), the left and right invariant distances are obviously different. This reflects the fact that  $N_\lambda(f_1 \circ f_2) \neq N_\lambda(f_2 \circ f_1)$ .

## 5.6 Utility of invariant distances

The distance between two points is often used in geometric algorithms: the Iterative Closest Point algorithm, developed in [BM92, Zha94] is the best example. Another example is the classic method of computing the transformation that maps a set of points  $x_i$  in one image, to a set of points  $y_i$  in another image: it is the transformation which minimizes the sum of squared distances  $C(f) = \sum_i \text{dist}^2(f \star x_i, y_i)$ .

All these algorithms can be extended to homogeneous features in a straightforward way using the invariant distance previously defined. The nice property that we have in using this distance is the ‘‘conservation’’ of the result under a change in the reference frame (the application of a transformation). Let  $F$  be the transformation minimizing the above criterion  $C(f)$  (we can assume for simplicity that it is unique, but the same results hold for a set  $F$  of minimums). Assuming that the features  $x_i$  are transformed by a transformation  $g$  ( $x'_i = g \star x_i$ ). The criterion is now

$$C'(f) = \sum_i \text{dist}^2(f \star x'_i, y_i) = \sum_i \text{dist}^2((f \circ g) \star x_i, y_i) = C(f \circ g)$$

With or without the invariance constraint, the new result is  $F' = F \circ g$ . If we assume now that the features of the second image are globally transformed:  $y'_i = g \star y_i$ , then the criterion is

$$C''(f) = \sum_i \text{dist}^2(f \star x_i, y'_i) = \sum_i \text{dist}^2(f \star x_i, g \star y_i)$$

We need here the invariance property of the distance to conclude

$$C''(f) = \sum_i \text{dist}^2((g^{(-1)} \circ f) \star x_i, y_i) = C(g^{(-1)} \circ f)$$

This means that the new minimum is  $F'' = g^{(-1)} \circ F$ , which gives the expected result  $F' = g \circ F''$ . The same experiment can be done if both images are transformed by

the *same* transformation  $g$  (this means a global change of the reference frame), and the invariance of the distance is required to prove that the transformation found is  $F''' = g^{(-1)} \circ F \circ g$ , i.e. only the change of the reference frame.

However, we note that the property of separability along each axis (equivalence with the  $L_\infty$  norm), used for instance in kd-trees to find the closest point, is not generally true for our invariant distance.

## 5.A Appendix A: Metric induced on the manifold by the group

Let  $N_g$  be a norm on the group  $\mathcal{G}$ . We define the induced semi-norm on the homogeneous manifold  $\mathcal{M}$  as

$$N(x) = \inf_{(h \in \mathcal{H}, f_x \in \hat{x})} (N_g(h \circ f_x)) = \inf_{(h_1, h_2) \in \mathcal{H}^2} (N_g(h_1 \circ f_x \circ h_2)) \quad (25)$$

The positivity of  $N$  follows immediately from the positivity of  $N_g$ . The symmetry comes from the symmetry of the metric on the transformation space and the symmetry of the metric definition:

$$N(f_x^{(-1)} \circ \mathcal{O}) = \inf_{(h_1, h_2) \in \mathcal{H}^2} (N_g(h_1 \circ f_x^{(-1)} \circ h_2)) = \inf_{(h'_1, h'_2) \in \mathcal{H}^2} (N_g(h'_1 \circ f_x^{(-1)} \circ h'_2)) = N(x)$$

The triangular inequality is preserved by the infimum:

$$\begin{aligned} \inf_{(h_1, h_2) \in \mathcal{H}^2} (N_g(h_1^{(-1)} \circ f_x^{(-1)} \circ f_y \circ h_2)) &\leq \inf_{h_1 \in \mathcal{H}} \{N_g(f_x \circ h_1)\} + \inf_{h_1 \in \mathcal{H}} \{N_g(f_y \circ h_2)\} \\ &\leq \inf_{h_1, h_2} (N_g(h_1 \circ f_x \circ h_2)) + \inf_{h_1, h_2} (N_g(h_1 \circ f_y \circ h_2)) \end{aligned}$$

and we eventually get  $N(f_x^{(-1)} \star y) \leq N(x) + N(y)$ . Now the definiteness: if the infimum of  $N_g(h_1 \circ f \circ h_2)$  is reached for every transformation  $f$  by  $(h_1, h_2) \in \mathcal{H}^2$ , then the semi-metric is separable and is thus a metric (this is in particular always true if the isotropy group  $\mathcal{H}$  is compact):

$$\begin{aligned} N(x) = 0 &\Leftrightarrow \exists (h_1, h_2) \in \mathcal{H}^2 / N_g(h_1 \circ f_x \circ h_2) = 0 \\ &\Leftrightarrow \exists (h_1, h_2) \in \mathcal{H}^2 / f_x = h_1 \circ h_2 \\ &\Leftrightarrow f_x \in \mathcal{H} \Leftrightarrow \hat{x} = \mathcal{H} = \hat{\mathcal{O}} \\ &\Leftrightarrow x = \mathcal{O} \end{aligned}$$

## 5.B Appendix B: Metric on rigid transformations

The norm definition is ( $\lambda$  is a fixed parameter):

$$N_\lambda(f)^2 = N_\lambda((r, t))^2 = \|f\|^2 = \lambda^2 \|r\|^2 + \|t\|^2$$

This norm is positive and null only for  $\|r\| = \|t\| = 0$ , that is for identity. If  $f = (r, t)$ , we have  $f^{(-1)} = (r^{(-1)}, r^{(-1)} \star (-t))$  (see [PT95a] for equations on rotations and frames) and since  $\theta = \|r\|$  is a norm on rotations:

$$N_\lambda(f^{(-1)})^2 = \lambda^2 \|r^{(-1)}\|^2 + \|-R^\top \cdot t\|^2 = N_\lambda(f)^2$$

The triangular inequality follows from triangular inequalities on the rotation and translation metrics: let  $f_1 = (r_1, t_1)$  and  $f_2 = (r_2, t_2)$  be two frames. We have

$$f_1^{(-1)} \circ f_2 = \left( r_1^{(-1)} \circ r_2 ; r_1^{(-1)} \star (t_2 - t_1) \right)$$

and thus  $N_\lambda(f_1^{(-1)} \circ f_2)^2 = \lambda^2 \|r_1^{(-1)} \circ r_2\|^2 + \|R_1^\top \cdot (t_2 - t_1)\|^2$ . The triangular inequality on rotations ensures that  $\theta(r_1^{(-1)} \circ r_2) \leq \theta_1 + \theta_2$  (where  $\theta_i = \|r_i\|$ ) and we have on vectors  $\|t_2 - t_1\|^2 = \|t_1\|^2 + \|t_2\|^2 - 2 \langle t_1, t_2 \rangle$ . Hence:

$$N_\lambda(f_1^{(-1)} \circ f_2)^2 \leq \theta_1^2 + \theta_2^2 + 2\theta_1\theta_2 + \|t_1\|^2 + \|t_2\|^2 + 2\|t_1\|\|t_2\|$$

Since  $(\theta_1\theta_2 + \|t_1\|\|t_2\|)^2 = (\theta_1^2 + \|t_1\|^2)(\theta_2^2 + \|t_2\|^2) - (\theta_1\|t_2\| - \theta_2\|t_1\|)^2$ , we obtain that

$$\begin{aligned} N_\lambda(f_1^{(-1)} \circ f_2)^2 &\leq (\theta_1^2 + \|t_1\|^2) + 2\sqrt{(\theta_1^2 + \|t_1\|^2)(\theta_2^2 + \|t_2\|^2)} + (\theta_2^2 + \|t_2\|^2) \\ &\leq \left(\sqrt{\theta_1^2 + \|t_1\|^2} + \sqrt{\theta_2^2 + \|t_2\|^2}\right)^2 \end{aligned}$$

taking the root-square, we obtain the requested inequality

$$N_\lambda(f_1^{(-1)} \circ f_2) \leq N_\lambda(f_1) + N_\lambda(f_2)$$

### 5.B.1 Metric induced on points

The norm induced on points is defined by

$$N(x) = \inf_{(h \in \mathcal{H}, f_x \in \hat{x})} (N_g(h \circ f_x)) = \inf_{(h_1, h_2) \in \mathcal{H}^2} (N_g(h_1 \circ f_x \circ h_2))$$

Let  $f_x = (0, x) \in \hat{x}$  and  $h_1 = (r_1, 0)$ ,  $h_2 = (r_2, 0) \in \mathcal{H}$ . We have thus  $h_1 \circ f_x \circ h_2 = (r_1 \circ r_2; r_1 \star x)$  and the (squared) norm of this transformation is simply

$$N_\lambda^2(h_1 \circ f_x \circ h_2) = \lambda \|r_1 \circ r_2\|^2 + \|x\|^2$$

The infimum is reached for  $r_1 = r_2^{(-1)}$  and we have  $N(x) = \|x\|$ . We note that in this case the isotropy group  $\mathcal{H}$  is reduced to the rotation group  $SO_3$ , which is compact.

## 6 Expectation and mean of random features

Uncertainty on geometric features (and more generally on measurements) is usually characterized by a probability density function whose expected value corresponds to the exact value. From a computational point of view, however, we need to keep only a few number of parameters characterizing this pdf. The usual way ([DW88, Aya91, ZF92]) is to consider the representation of the random feature as a random vector and, assuming that the pdf is quasi-Gaussian, approximate it up to the second order by its expectation and covariance matrix. We focus in the sequel on the expectation  $\bar{x}$  and its statistical measurement (or its discrete version): the empirical mean. The classical definition is, for a pdf  $p_x$  (in the parameter space) and a set of measured features  $\{x_i\}$ :

$$\bar{x} = \mathbf{E}(x) = \int_{\mathcal{D}} y \cdot p_x(y) \cdot dy \quad \text{and} \quad \mathbf{E}(\{x_i\}) = \frac{1}{n} \sum_i x_i$$

We claim that these operators are not properly defined. In particular, the result of the integral or the sum is *not* ensured to be in the definition domain and defines not necessarily a feature: for instance, the arithmetic mean of several rotation matrices is generally not an orthogonal matrix and is therefore not a rotation itself (particularly for large deviations). The second reason, is that the expectation does not commute in general with the application of a fixed transformation (see example below). This means that the mean value of a pdf depends on the chosen reference frame, which is unacceptable.

### 6.1 Standard expectation of a 2D random line

We consider for this example 2D oriented vector lines, which can be represented by a point on the unit circle, and therefore an angle  $\theta$  with a given axis. We fix the domain of  $\theta$  to be  $\mathcal{D} = ] - \pi, \pi]$ . The action of a rotation of angle  $\lambda$  is simply the addition (modulo  $2\pi$ ). We can define an uncertain line by its probability density function  $\rho(\theta)$ , and the classical way to obtain the expected value is to integrate in the parameter space:

$$\bar{\theta} = \mathbf{E}(\theta) = \int_{\mathcal{D}} \theta \cdot \rho(\theta) \frac{d\theta}{2\pi}$$

where the term  $2\pi$  is a normalization factor. We note that  $d\theta$  is the uniform measure for lines under rotation. Let  $\rho_0(\theta) = 2 \cos(\theta/2)^2$  be such a pdf (drawn in figure 6). It is normalized and its expectation is  $\bar{\theta}_0 = 0$ . If we now change the reference frame,

i.e. apply a rotation of angle  $\lambda$ , we obtain the new pdf  $\rho_\lambda(\theta) = \rho_0(\theta - \lambda)$ , and the expected value is now

$$\bar{\theta}(\lambda) = \int_{\mathcal{D}} \theta \cdot \rho_0(\theta - \lambda) \frac{d\theta}{2\pi} = 2 \sin(\lambda/2) \cos(\lambda/2)$$

which is different from the rotated expectation  $\underline{\theta}(\lambda) = \lambda$ . In particular, for a rotation of  $\lambda = \pi$ , we find that  $\underline{\theta}_\pi = \pi$  and  $\bar{\theta}_\pi = 0$ .

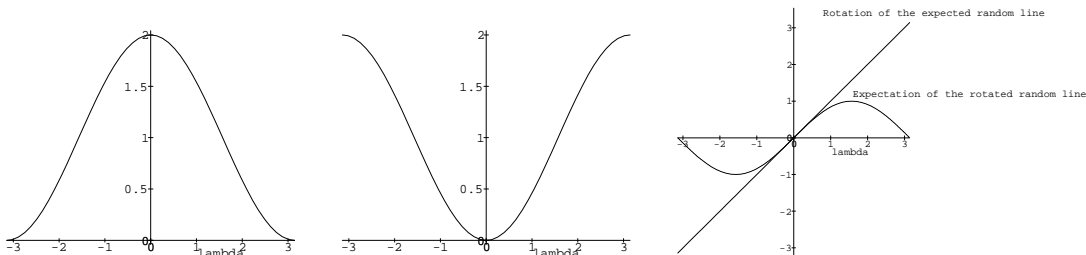


Figure 6: Left: Original pdf  $\rho_0(\theta) = 2 \cos(\theta/2)^2$ . Expectation is  $\bar{\theta}(0) = 0$ . Middle: pdf after a rotation of angle  $\lambda = \pi$ . The expectation  $\bar{\theta}(\pi)$  is also 0, whereas the rotated expectation is  $\underline{\theta}_\pi = \pi$ . Right: Comparison of the expectation of the rotated line and the rotation of the expected line.

The first idea to avoid this type of problems is to “center” the definition domain of the chart around the expected feature. If this can be easy in the case of a circle, as above, it may be more problematical for some other features such as 3D lines, or frames, where the manifold is far more complex. The question is similar for the mean value, especially with scattered measurements. Last, but not least, if the chart is centered around the expected (or the mean) feature, then the problem is already solved. The Frechet expectation is a well-posed formalism to implement this idea: the “centrality” of a feature is based on its distance with other measurements, and the mean or expected values are the features that optimize the “centrality”.

## 6.2 Frechet expectation of a random feature

Let  $\mathbf{v}$  be a random vector in  $\mathbb{R}^n$ . Frechet ([Fre48]) observed that the variance  $\sigma_{\mathbf{v}}^2(x) = \mathbf{E}(\text{dist}(\mathbf{v}, x)^2)$  is minimized at the expected value  $\bar{v}$ . The second point is that if the expectancy of the representation of a feature (a vectorial integral) is not well defined (because features are not vectors), the expectation of a real mapping (and in particular the distance to a fixed feature) is properly defined.

Let  $dist$  be an invariant distance on the manifold  $\mathcal{M}$  under the group  $\mathcal{G}$ . This distance is independent of the representation. In order to get rid of the representation

problem and be sure to integrate correctly on the manifold, we use from now on the “invariant” pdf previously introduced to describe random features. Let then  $\mathbf{x}$  be a random feature of pdf  $p_{\mathbf{x}}$ . The expected square distance of a deterministic feature  $y$  with the random feature  $\mathbf{x}$  is defined by

$$\sigma_{\mathbf{x}}^2(y) = \mathbf{E}_{\mathbf{x}}(\text{dist}(y, \mathbf{x})^2) = \int_{\mathcal{M}} \text{dist}(y, z)^2 \cdot p_{\mathbf{x}}(z) \cdot d\mathcal{M}(z) \quad (26)$$

If  $\sigma_{\mathbf{x}}^2(y)$  is finite for all  $y$  (and this is the case since we work with pdf that have a compact support, see definition in section 4.1), we call every feature  $\bar{x}$  minimizing  $\sigma_{\mathbf{x}}^2$  an expected feature, and we denote by  $\mathbb{E}(\mathbf{x})$  the set of all expected features of the random feature  $\mathbf{x}$ . We have thus

$$\mathbb{E}(\mathbf{x}) = \arg \min_{y \in \mathcal{M}} (\mathbf{E}_{\mathbf{x}}(\text{dist}(y, \mathbf{x})^2)) \quad (27)$$

If there is at least one expected feature  $\bar{x}$ , then  $\sigma_{\mathbf{x}} = \sigma_{\mathbf{x}}(\bar{x})$  is called the standard deviation of  $\mathbf{x}$  and  $\sigma_{\mathbf{x}}^2$  the variance. In general, there can be several expected values. However, Emery and Mokobodzki show in [EM91], under some strong conditions, the existence and uniqueness of the expected value.

In a very similar way, we can define the set of empirical means of features  $x_1, \dots, x_n$  with the empirical (or discrete) version of the expectation:

$$\mathbb{E}(\{x_i\}) = \arg \min_{y \in \mathcal{M}} (\mathbf{E}(\{\text{dist}(y, x_i)^2\})) = \arg \min_{y \in \mathcal{M}} \left( \frac{1}{n} \sum_i \text{dist}(y, x_i)^2 \right) \quad (28)$$

and if there is at least one mean  $\bar{y}$ , we call  $s = \sqrt{\frac{1}{n} \sum_i \text{dist}(y, x_i)^2}$  the empirical standard deviation and  $s^2$  the empirical variance.

Other types of central values can be defined in this framework: we define more generally the “mean deviation” at order  $\alpha$  by

$$\sigma_{\mathbf{x}, \alpha}(y) = (\mathbf{E}_{\mathbf{x}}(\text{dist}(y, \mathbf{x})^\alpha))^{1/\alpha} = \left( \int_{\mathcal{M}} \text{dist}(y, z)^\alpha \cdot p_{\mathbf{x}}(z) \cdot d\mathcal{M}(z) \right)^{1/\alpha} \quad (29)$$

If this function is finite over  $\mathcal{M}$ , the features  $\bar{x}_\alpha$  minimizing it are called central features of order  $\alpha$ . To be more practical, we obtain the modes of the pdf for  $\alpha = 0$ , the median features for  $\alpha = 1$ , of course the mean or expected features for  $\alpha = 2$ , and the “barycenter” of the support of the pdf (which is a compact set) if the mean deviation is finite for  $\alpha = \infty$ . This can also be applied to define the empirical central features at any order.



### 6.3 Frechet expectation of a random transformation

All the above definitions can be applied to the transformation group in order to define the expected and mean mean transformations of a random transformation: if  $dist$  is an invariant distance on the group  $\mathcal{G}$  and  $\mathbf{f}$  a random transformation of pdf  $p_{\mathbf{f}}$ , we have:

$$\sigma_{\mathbf{f}}^2(g) = \mathbf{E}_{\mathbf{f}}(\text{dist}(\mathbf{f}, g)^2) = \int_{\mathcal{G}} \text{dist}(g, f')^2 \cdot p_{\mathbf{f}}(f') \cdot d_L \mathcal{G}(f')$$

and the set of expected transformations is

$$\mathbb{E}(\mathbf{f}) = \arg \min_{g \in \mathcal{G}} (\mathbf{E}_{\mathbf{f}}(\text{dist}(g, \mathbf{f})^2)) \quad (30)$$

The set of mean transformations is similarly defined.

### 6.4 Invariance properties

#### 6.4.1 Application of a fixed transformation to a random feature

The nice properties of the Frechet expectation and mean features are, in our case, due to the invariant distance (see appendix 6.A for proofs); these sets are stable under the transformation group:

$$\mathbb{E}(g \star \mathbf{x}) = g \star \mathbb{E}(\mathbf{x}) \quad \text{and} \quad \mathbb{E}(\{g \star x_i\}) = g \star \mathbb{E}(\{x_i\}) \quad (31)$$

Since the distance we use does not depends on the representation, the results of all minimization are ensured to be also independent of the representation. We have thus obtained a stable definition (and a mean of computation via optimization) for the expected features of a pdf and for the empirical mean features.

#### 6.4.2 Left composition of a random transformation by a fixed one

Let  $\mathbf{f}$  be a random transformation and  $\mathbf{g} = f_o \circ \mathbf{f}$  its left composition. Then, with the same proof as above, we have

$$\mathbb{E}(f_o \star \mathbf{f}) = f_o \star \mathbb{E}(\mathbf{f}) \quad \text{and} \quad \mathbb{E}(\{f_o \star f_i\}) = f_o \star \mathbb{E}(\{f_i\}) \quad (32)$$

### 6.5 Discussion

In this definition of expectation, there two major components that influence the behavior of the result with respect to the different kind of operations we consider (application, inversion, composition with fixed or random elements). The first one is the measure, in our case the invariant measure. However, the right invariant measure has to be used for some derivations (right composition, inversion) and the module of the group (the relationship between left and right measure) is necessary for some equation of propagation of the pdf.

The second important component of the expectation is the distance function. In appendix (6.B), we were not able to conclude for the propagation of the expectation through all the canonical operations using only the (left) invariance of the distance. On the other hand, the two above invariance properties can be obtained with a weaker constraint than (left) invariance: if

$$\forall(x, y) \in \mathcal{M}^2 \quad \text{dist}(f \star x, f \star y) = \alpha(f) \cdot \text{dist}(x, y)$$

then the values of the minima are not conserved, but so are the elements realizing these minima. Similarly, we could think of an equivalent property for the group distance, with two different functions  $\alpha_L$  and  $\alpha_R$  for left and right composition. Examining such distances could lead to an important improvement in the properties of our expectation.

### 6.6 Application to data fusion: the mean frames

Assume that we have a set of frame measurements  $f_i = (r_i, x_i)$ . We are looking for the mean frame  $f = (r, x)$  in the Frechet sense. Since the distance from  $f$  to  $f_i$  is

$$\text{dist}(f, f_i) = N_\lambda(f^{(-1)} \circ f_i) = N_\lambda(f_i^{(-1)} \circ f)$$

with

$$f_i^{(-1)} \circ f = \left( r_i^{(-1)} \circ r ; r_i^{(-1)} \star (x - x_i) \right)$$

the least squares criterion reduces to

$$C_\lambda(f) = \lambda^2 \sum_i \|r_i^{(-1)} \circ r\|^2 + \sum_i \|x - x_i\|^2$$

We can thus minimize independently for the position and the orientation. Moreover, the solution is independent of the parameter  $\lambda$ .

The position is given by the barycenter of the frame positions, and the orientation is obtained by an iterative gradient descent (equations for the derivation can be found in [PT95b]). The gradient descent can be repeated for several starting points to verify that the global minimum is obtained and to test its uniqueness.

Another method following the same principles but incorporating second order informations (covariance matrices) was proposed in [PT95b].

## 6.A Appendix A: Stability of the expected and mean elements

### 6.A.1 Application of a fixed transformation

Assume that  $\mathbf{z} = g \star \mathbf{x}$  is the random feature obtained by the transformation of the random feature  $\mathbf{x}$  by the fixed transformation  $g$ :

$$\begin{aligned}\sigma_{\mathbf{z}}^2(y) &= \mathbf{E}_{\mathbf{z}} (\text{dist}(g \star \mathbf{x}, y)^2) = \int_{\mathcal{M}} \text{dist}(y, z)^2 \cdot p_{\mathbf{z}}(z) \cdot d\mathcal{M}(z) \\ &= \int_{\mathcal{M}} \text{dist}(y, z)^2 \cdot p_{\mathbf{x}}(g^{(-1)} \star z) \cdot d\mathcal{M}(z) = \int_{\mathcal{M}} \text{dist}(y, g \star x)^2 \cdot p_{\mathbf{x}}(x) \cdot d\mathcal{M}(x)\end{aligned}$$

With the invariance of the distance, we get

$$\sigma_{\mathbf{z}}^2(y) = \int_{\mathcal{M}} \text{dist}(g^{(-1)} \star y, x)^2 \cdot p_{\mathbf{x}}(x) \cdot d\mathcal{M}(x) = \sigma_{\mathbf{x}}^2(g^{(-1)} \star y)$$

$\bar{x}$  minimizes then  $\sigma_{\mathbf{x}}^2(x)$  is and only if  $\bar{z} = g \star \bar{x}$  minimizes  $\sigma_{\mathbf{z}}^2(z)$ :

$$\mathbb{E}(\mathbf{z}) = g \star \mathbb{E}(\mathbf{x}) \quad \text{and} \quad \sigma_{\mathbf{z}} = \sigma_{\mathbf{x}}$$

The same argument holds for the stability of central features of any order and also the mean features of a set  $x_i$ . If  $z_i = g \star x_i$ , we have

$$\mathbb{E}(\{z_i\}) = g \star \mathbb{E}(\{x_i\}) \quad \text{and} \quad s_z = s_x$$

### 6.A.2 Composition by a fixed transformation

Let  $\mathbf{g} = f_o \circ \mathbf{f}$  a random transformation. Exactly as above, we have:

$$\begin{aligned}\sigma_{\mathbf{g}}^2(d) &= \mathbf{E}_{\mathbf{g}} (\text{dist}(f_o \circ \mathbf{f}, d)^2) = \int_{\mathcal{G}} \text{dist}(d, g)^2 \cdot p_{\mathbf{g}}(g) \cdot d_L\mathcal{G}(g) \\ &= \int_{\mathcal{G}} \text{dist}(d, g)^2 \cdot p_{\mathbf{f}}(f_o^{(-1)} \circ g) \cdot d_L\mathcal{G}(g) = \int_{\mathcal{G}} \text{dist}(d, f_o \circ f)^2 \cdot p_{\mathbf{f}}(f) \cdot d_L\mathcal{G}(f)\end{aligned}$$

With the invariance of the distance, we get

$$\sigma_{\mathbf{g}}^2(d) = \int_{\mathcal{G}} \text{dist}(f_o^{(-1)} \circ d, f)^2 \cdot p_{\mathbf{f}}(f) \cdot d_L\mathcal{G}(f) = \sigma_{\mathbf{f}}^2(f_o^{(-1)} \circ d)$$

$\bar{f}$  minimizes then  $\sigma_{\mathbf{f}}^2(f)$  is and only if  $\bar{g} = f_o \circ \bar{f}$  minimizes  $\sigma_{\mathbf{g}}^2(g)$ :

$$\mathbb{E}(\mathbf{g}) = f_o \circ \mathbb{E}(\mathbf{f}) \quad \text{and} \quad \sigma_{\mathbf{g}} = \sigma_{\mathbf{f}}$$

The stability of the mean transformation by the left composition is also similar.

## 6.B Appendix B: Hints for the propagation through the other operations

### 6.B.1 Right composition by a fixed transformation

Let  $\mathbf{g} = \mathbf{f} \circ f_o$  a random transformation. Using the pdf of  $\mathbf{g}$  (equation 15), we have:

$$\begin{aligned} \sigma_{\mathbf{g}}^2(d) &= \mathbf{E}_{\mathbf{g}} (\text{dist}(\mathbf{f} \circ f_o, d)^2) = \int_{\mathcal{G}} \text{dist}(d, g)^2 \cdot p_{\mathbf{g}}(g) \cdot d_L \mathcal{G}(g) \\ &= \int_{\mathcal{G}} \text{dist}(d, g)^2 \cdot \Delta \mathcal{G}(g \circ f_o^{(-1)}) \cdot p_{\mathbf{f}}(g \circ f_o^{(-1)}) \cdot d_R \mathcal{G}(g) = \int_{\mathcal{G}} \text{dist}(d, f \circ f_o)^2 \cdot p_{\mathbf{f}}(f) \cdot d_L \mathcal{G}(f) \end{aligned}$$

This time, the distance is not supposed to be right invariant, and we can just conclude that

$$\sigma_{\mathbf{g}}^2(d) = \int_{\mathcal{G}} \text{dist}(d, f \circ f_o)^2 \cdot p_{\mathbf{f}}(f) \cdot d_L \mathcal{G}(f) = \mathbf{E}_{\mathbf{f}} (\text{dist}(\mathbf{f} \circ f_o, d)^2)$$

The question is: can we impose a weaker constraint than invariance on the distance that guarantee the stability of the expected transformation on both sides?

### 6.B.2 Application of a random transformation

Following the same type of derivation, we can write if the feature  $x$  is fixed:

$$\sigma_{(\mathbf{f} \star x)}^2(z) = \int_{\mathcal{G}} \text{dist}(f \star x, z)^2 \cdot p_{\mathbf{f}}(f) \cdot d_L \mathcal{G}(f)$$

and if both the transformation and the feature are random:

$$\sigma_{\mathbf{f} \star \mathbf{x}}^2(z) = \int_{\mathcal{G}} \sigma_{\mathbf{x}}^2(f^{(-1)} \star z) \cdot p_{\mathbf{f}}(f) \cdot d_L \mathcal{G}(f) = \int_{\mathcal{M}} \sigma_{(\mathbf{f} \star x)}^2 \cdot p_{\mathbf{x}}(x) \cdot d_{\mathcal{M}}(x)$$

Once again, what are the constraints needed on the distance in order to prove that  $\mathbb{E}(\mathbf{f} \star x) = \mathbb{E}(\mathbf{f}) \star x$  and something like  $\mathbb{E}(\mathbf{f} \star \mathbf{x}) \subset \mathbb{E}(\mathbf{f}) \star \mathbb{E}(\mathbf{x})$  ?

## 7 Experiments: a data fusion problem

We presented in [PT95b] an algorithm for the registration based on 3D frames which also quantifies the uncertainty on both the data and the transformation. We used it to register medical images and showed that the accuracy of the registration is far below both the voxel size and the uncertainty of the individual features. In this method, only the most stable features are used to compute the registration, and a lot of matches are discarded due to their large uncertainty.

The aim of the present experiments is to fuse together several registered images of a single patient in order to construct an average model based on extremal points. We are mainly interested here in the topological stability of the features, i.e. determine which features can be reliably matched in most images, even with large geometric deviations. The selectivity of the features (section 3.4) therefore very important. Such a study on several patients will eventually lead to identify the most stable anatomical features (landmarks), and will allow us to better reduce the complexity while increasing the robustness of the registration task.

The key point is to be able to compute the “mean feature” even in the presence of large deviations. We saw for instance in section 3.4 that the selectivity of a trihedron match remains high even for a large error bound: it is most interesting to keep in our model the mean frames and not only the mean points. In such a case, the Frechet expectation we defined in section 6.2 is particularly well suited. We use more precisely the mean feature (equation 28), defined with the invariant distance on rigid transformations (see 5.5 and section 6.6 for the algorithm).

### 7.1 3D medical images

We present results from an experiment performed using 3D Magnetic Resonance images (MRI). The images were provided by Pr. Ron Kikinis of the Brigham and Woman’s Hospital as part of an extensive study of the evolution of the Multiple Sclerosis (MS) disease. The same patient get a complete 3D MR examination several times during one year (typically 24 different 3D acquisitions). The aim is to register precisely in 3D all those images, in order to segment the lesions and evaluate very accurately their evolution (MS lesions are white spots in the images in figure 7). The images are first echo,  $256 \times 256 \times 54$  voxels, the voxel size is  $1mm \times 1mm \times 3mm$ .

## 7.2 Extremal points and frames

Our registration algorithm relies on the extraction of feature points in 3D medical images, defined with differential geometry criteria: the *Extremal Points* (defined in [Thi93]). These are points on the object surface for which both principal curvatures are extremal. The interesting thing is that not only do we get some invariant measurements associated with those points (the principal curvatures), which are used to reduce the complexity of the matching process, but we get also the principal directions, which form, with the normal to the surface and the extremal point itself, an orthonormal basis, that is, a frame.

Typically, we extract about 3000 extremal points from a 3.5 million voxels image. Our matching algorithm produces about 600 pairs of associated extremal points between two images with a residual mean square error (RMS) of about  $1mm$ , and about 1000 additional matches with a RMS around  $5mm$ .

## 7.3 Building a model

Among the 24 images of a patient, one is considered as the reference for registration: the algorithm presented in [PT95b] registers the 23 other ones using the 600 most geometrically stable frames. The accuracy obtained for the estimated transformation is sufficiently good (compared to the large deviations on frames that we want to handle) to consider that the transformations are exact.

We regroup then the features that we can match without ambiguity in several images, and determine the mean frame. To build a interesting model, we add to this mean frame its variance (more precisely a variance  $\sigma_\theta$  on the orientation, and a variance  $\sigma_x$  on the position), and its probability of observation (the number of images where it is observed divided by the total number of images, here 24). We consider that two frames  $f_1 = (r_1, x_1)$  and  $f_2 = (r_2, x_2)$  are matched without ambiguity if the distance between them is less than a given threshold (more exactly a threshold  $\theta_0$  on  $\|r_1^{(-1)} \circ r_2\|$  for the orientation and a threshold  $d_0$  on  $\|x_1 - x_2\|$  for the position), and if there is no other frame that can be matched with any of the two frames within these bounds.

## 7.4 Results

We present in figure 8 the surface of the brain and the crest lines extracted from the first image and the most stable frames. We observed that about 30 frames are extremely well preserved, both geometrically and topologically, and 70 to 200 others

are observed in more than 80% of the images, depending on the error bound we consider (see also figure 9). We plan to adapt automatically the error bound for each feature in order to be able to maximize for each model feature the number of non ambiguous matches. This should allow to compute more robustly the stable features. An interesting result is that the most stable extremal points are located on the surface of the brain and not on the skin or on the skull. Since the images comes from the magnetic resonance modality, the skull is indeed not very visible. This points out the fact that the registration is mainly done on the surface of the brain, which was expected.

In a next step, we plan to compute the models of several patients in order to compare them and characterize the anatomically stable extremal points. The results could then be incorporated into an anatomical atlas (see for instance [STA95]).

## 8 Conclusion

We showed in this paper that a wide range of paradoxes arise when we try to generalize to geometric features the classical algorithms used for points, and we demonstrate, in the case of homogeneous features, that they can be avoided by the careful definition of basic operators which respect the following rules:

- Invariance with respect to the representation.
- Invariance or commutativity with respect to the action of the associated transformation group.

We developed here general methods which allow to define three basic operators following these rules: the invariant measure, invariant distances and stable expectation and mean features. These methods were illustrated with frame features under the rigid transformation group, and an application to the data fusion problem is presented.

We believe that the application of these two basic rules to a great number of geometrical problems can lead to a proper mathematical framework that will give reasonable and robust results in any situation without the need to design ad-hoc heuristics. We are currently working on a theory of uncertainty on geometric features completing the formalism introduced in this paper.

## Acknowledgments

Part of this work was supported by a fellowship from D.R.E.T./D.C.N. (France). Images were provided by Pr. Ron Kikinis of the Brigham and Woman's Hospital. We want to thank the whole Epidaure Research team for the help provided during this work.



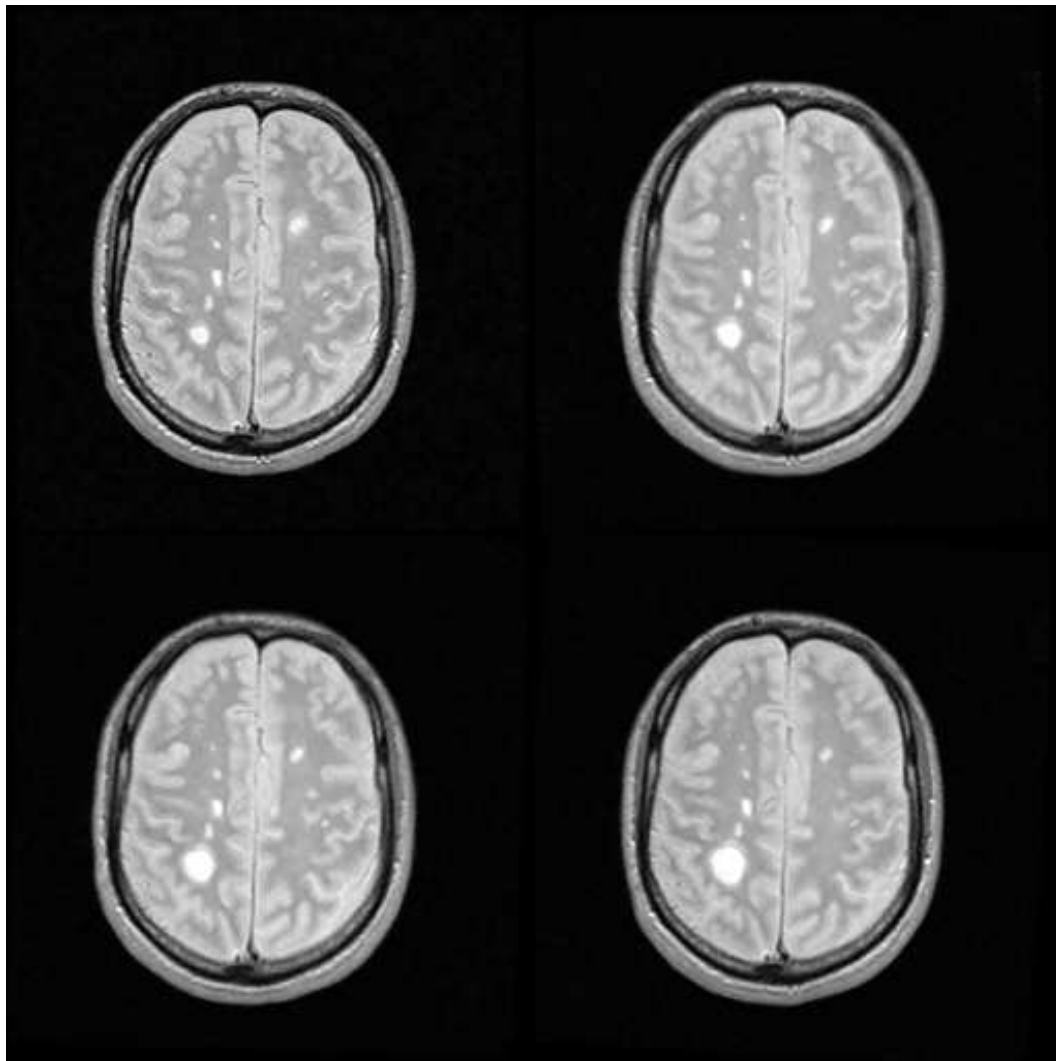


Figure 7: The same slices from four registered 3D MR images of the Multiple Sclerosis study. Note the evolution of two MS lesions (white spots). One is growing in the anterior left hemisphere, one is shrinking in the right posterior hemisphere. There is two weeks between each acquisition.

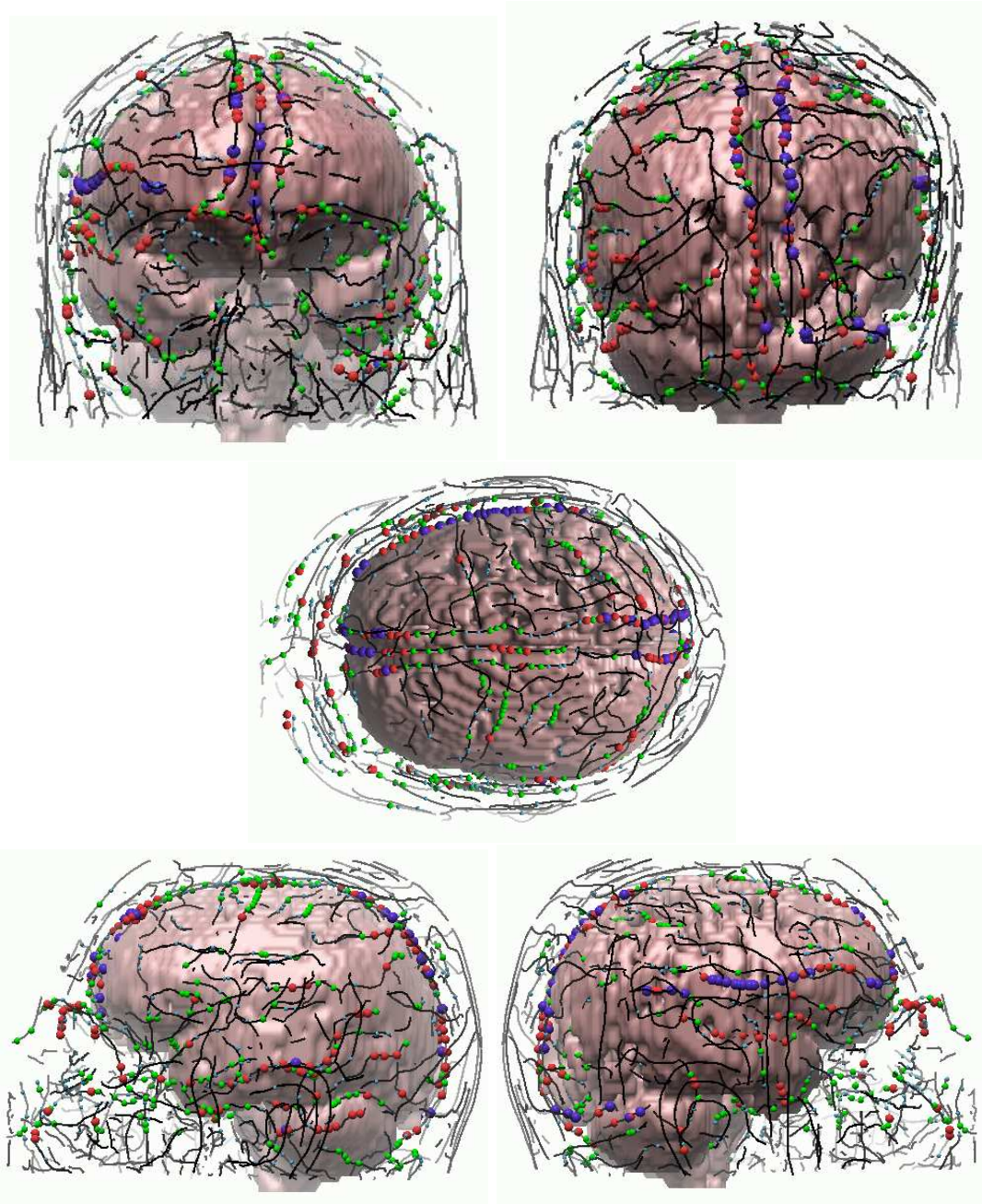


Figure 8: The surface of the brain is displayed along with the crest lines extracted from the first image. The most stable frames of the model are represented by spheres colored from decreasing stability order (and size) in purple, red, green and blue. Top line from left to right: front and rear views of the head. Middle: top view. Bottom line: left and right views.

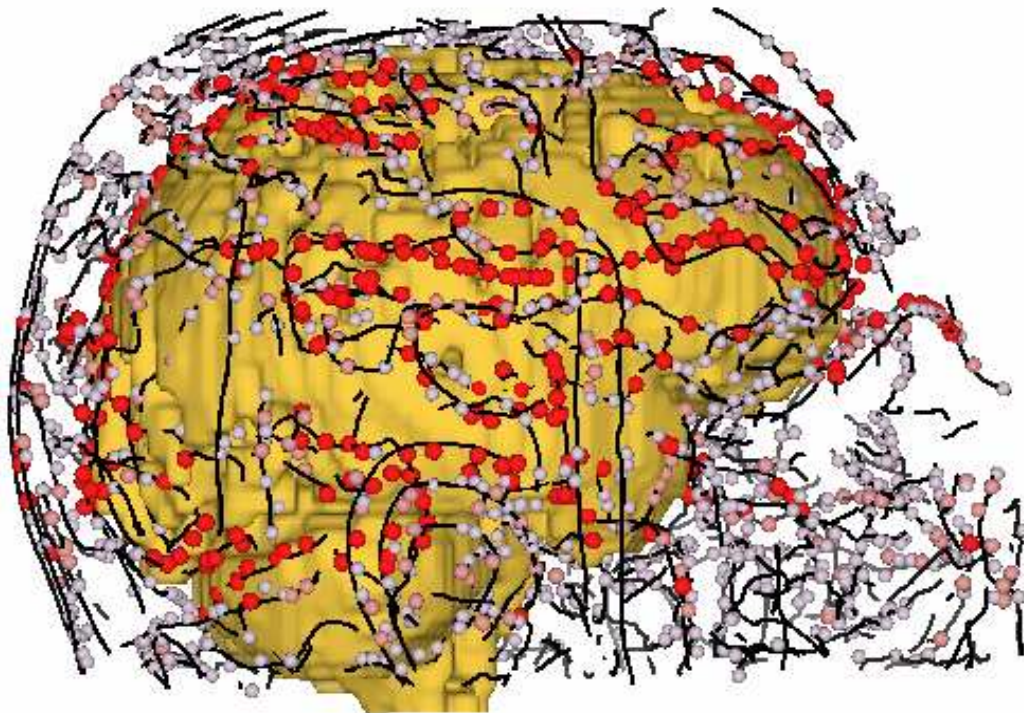


Figure 9: Extremal points matched in more than 80% of the 24 images with a larger error bound.

## References

- [AF86] N. Ayache and O.D. Faugeras. Hyper : A new approach for the recognition and positioning of two-dimensional objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):44–54, 1986.
- [Alt86] S. L. Altmann. *Rotations, Quaternions, and Double Groups*. Clarendon Press - Oxford, 1986.
- [Aya91] N. Ayache. *Artificial Vision for Mobile robots - Stereo-vision and Multisensor Perception*. MIT-Press, 1991.
- [BM92] Paul Besl and Neil McKay. A method for registration of 3–D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992.
- [DW88] H.F. Durrant-Whyte. *Integration, Coordination and Control of Multi-Sensor Robot Systems*. Kluwer Academic Publishers, 1988.
- [EM91] M. Emery and G. Mokobodzki. Sur le barycentre d’une probabilité dans une variété. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXV*, pages 220–233, Springer-Verlag, 1991.
- [Fre48] M. Frechet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310, 1948.
- [GH90] W.E.L. Grimson and D.P. Huttenlocher. On the sensitivity of geometric hashing. In *Proc. third ICCV*, pages 334–338, 1990.
- [GHJ93] W.E.L. Grimson, D.P. Huttenlocher, and D.W. Jacobs. A study of affine matching with bounded sensor error. 1993. to appear in *Int. Journ. of Comput. Vision*.
- [Gri92] W.E.L. Grimson. *Object Recognition by Computer - The role of Geometric Constraints*. MIT Press, 1992.
- [Hoc65] G. Hochschild. *The structure of Lie groups*. Holden-Day, 1965.
- [HU87] D.P. Huttenlocher and S. Ullman. Object recognition using alignment. In *Proc. of ICCV*, pages 72–78, 1987.
- [KM63] M.G. Kendall and P.A.P. Moran. *Geometrical probability. Griffin’s statistical monographs and courses*, Charlkes Griffin & Co. Ltd., 1963.

- [LW88] Y. Lamdan and H.J. Wolfson. Geometric hashing : A general and efficient model-based recognition scheme. In *Proc. of Second ICCV*, pages 238–289, 1988.
- [LW91] Y Lamdan and H.J. Wolfson. On the error analysis of geometric hashing. In *IEEE Int. Conf. on Comput. Vis. and Patt. Recog.*, pages 22–27, 1991.
- [PA94] X. Pennec and N. Ayache. An  $\mathcal{O}(n^2)$  algorithm for 3D substructure matching of proteins. In a. Califano, I. Rigoutsos, and H.J. Wolson, editors, *Shape and Pattern Matching in Computational Biology*, Plenum Publishing, 1994.
- [Pen93] X. Pennec. *Correctness and Robustness of 3D Rigid Matching with Bounded Sensor Error*. Research report 2111, INRIA, November 1993.
- [PT95a] X. Pennec and J.P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. 1995. Submitted to IJCV.
- [PT95b] X. Pennec and J.P. Thirion. Validation of 3D registration methods based on points and frames. In *Proceedings of the 5th Int. Conf on Comp. Vision (ICCV'95)*, pages 557–562, Cambridge, Ma, June 1995.
- [RH93] I. Rigoutsos and R. Hummel. Distributed bayesian object recognition. In *Proceedings of Int. Conf on Comput. Vis. and Pat. Recog*, pages 180–186, IEEE Computer Society Press, June 1993.
- [San76] L.A. Santalo. *Integral Geometry and Geometric Probability*. Volume 1 of *Encyclopedia of Mathematics and its Applications*, Addison-Wesley Publishing Company, 1976.
- [STA95] G. Subsol, J.Ph. Thirion, and N. Ayache. A General Scheme for Automatically Building 3D Morphometric Anatomical Atlases: application to a Skull Atlas. In *MRCAS'95*, November 1995. \*\*\*.
- [Thi93] J.P. Thirion. *New Feature Points based on geometric invariants for 3D Image Registration*. Research Report 1901, INRIA, Mai 1993. accepted for publication in IJCV.
- [ZF92] Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis: a stereo based approach*, chapter 2: Uncertainty Manipulation and Parameter Estimation,

pages 9–27. Volume 27 of *Springer series in information science*, Springer Verlag, 1992.

- [Zha94] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994. Also Research Report No.1658, INRIA Sophia-Antipolis, 1992.



---

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,  
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY  
Unité de recherche INRIA Rennes, Irista, Campus universitaire de Beaulieu, 35042 RENNES Cedex  
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1  
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex  
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

---

Éditeur  
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)  
ISSN 0249-6399