A-1 DEFINITIONS

For convenience of the reader, we also repeat various definitions which have been omitted or only stated informally. Introductory references for the differential geometric concepts used in this paper are [1, 2].

$\mathbb{R}^+_0$ is the set of positive reals. The generic $m$-sphere is $\mathbb{S}^m$. $\mathbb{S}^2$ is the 3D sphere, and $\mathbb{S}^1$ is the circle. The target manifold is $\mathcal{M}$ and $d$ indicates its geodesic distance. The set of unknown points to be reconstructed is $\mathcal{S} = \{s_1\}$. 

Definition A-1. The radius of a set $\mathcal{S} = \{s_1\}$ is defined as

$$\text{rad}(\mathcal{S}) \triangleq \min_i \max_j d(s_i, s_j).$$

The diameter is twice the radius.

Definition A-2. The informative radius $\text{infr}(f)$ of $f$ is the maximum $r$ such that $f$ is invertible in $[0, r]$.

Definition A-3. An isometry is a map $\varphi : \mathcal{M} \to \mathcal{M}$ that preserves distances: $d(\varphi(s_1), \varphi(s_2)) = d(s_1, s_2)$.

Definition A-4. A conformal map is a map that preserves the angles between geodesics.

A map is conformal if and only if its Jacobian is proportional to an orthogonal matrix.

Definition A-5. A generic warping is a map $\varphi_m : \mathcal{M} \to \mathcal{M}$ such that $d(\varphi(s_1), \varphi(s_2)) = m(d(s_1, s_2))$, for some monotonic function $m : \mathbb{R}^+_0 \to \mathbb{R}^+_0$.

Definition A-6. A linear warping is a map $\varphi_\alpha$ from $\mathcal{M}$ to itself such that $d(\varphi_\alpha(s_1), \varphi_\alpha(s_2)) = \alpha d(s_1, s_2)$ for some $\alpha > 0$.

Definition A-7. A wigging of a set $\{s_i\} \subset \mathcal{M}$ is a map $\varphi : \mathcal{M} \to \mathcal{M}$ that preserves the order of distances: for all $i, j, k, l$: $d(s_i, s_j) < d(s_k, s_l) \iff d(\varphi(s_i), \varphi(s_j)) < d(\varphi(s_k), \varphi(s_l))$.

Definition A-8. A geodesic curve $g(A, B, t)$ from point $A$ to point $B$, for $t \in [0, 1]$, is the curve on the manifold such that

$$d(g(A, B, t), A) = td(A, B),$$
$$d(g(A, B, t), B) = (1 - t)d(A, B).$$

In particular, $g(A, B, 0) = A$ and $g(A, B, 1) = B$.

A-2 PROOF OF PROPOSITION 2

We use some language (Haar measure) from group theory; for reference, see, e.g., [3, 4].

The luminance at pixel $s \in \mathbb{S}^2$ at time $t$ can be written as

$$y(s, t) = h(t(t), R(t) s),$$

where $t \in \mathbb{R}^m$ is the sensor position, $R \in \text{SO}(3)$ is the sensor orientation, and $h : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$ is a function that describes the environment. In the following, we drop the dependence on time.

We will show that the statistics depends only on the distance between two pixels, by showing that any pair of pixels with the same distance give the same value of the statistics. Consider two pairs of pixels $(s_i, s_j)$ and $(s_k, s_l)$ having the same distance:

$$d(s_i, s_j) = d(s_k, s_l).$$
Because \( d(s_i, s_j) = d(s_k, s_l) \), there exists an \( X \) such that

\[
s_k = Xs_i, \quad s_l = Xs_j.
\]

(2)

If the probability distribution of \( R \) is uniform on \( SO(3) \), that is, it is the Haar measure \( SO(3) \), then it is also invariant to a rotation (i.e., left/right actions): for all functions \( z \) and rotations \( X \), \( \mathbb{E}\{z(R)\} = \mathbb{E}\{z(RX)\} \).

If we choose the \( X \) such that (2), we obtain that

\[
\mathbb{E}\{g(y(s_i), y(s_j))\} = \mathbb{E}\{g(h(t, R s_i), h(t, R s_j))\} = \mathbb{E}\{g(h(t, R X s_i), h(t, R X s_j))\} = \mathbb{E}\{g(y(s_k), y(s_l))\}.
\]

(3)

\[
(4)
\]

A-3 PROOF OF PROPOSITION 8

A-3.1 Proof overview

The starting point is considering that the largest unobservable transformations are the set of wigglings (Definition 7), because they are exactly those that keep constant the order of the inter-points distances, which is the sufficient statistics for the estimation problem. All other symmetries—isometries (Definition A-3), linear warping (Definition 6), generic warping (Definition A-5) are a specialized version of wigglings. Moreover, an isometry is a linear warping with \( \alpha = 1 \), and a linear warping is a specialization of a generic warping. In summary, just by the definition of the various transformations, we have the following chain of inclusions:

\[
\text{isometries} \subset \text{linear warpings} \subset \text{generic warpings} \subset \text{wigglings}.
\]

Isometries and warpings are very structured transformations, but wigglings are in general discontinuous. The next step in the analysis is understanding in what cases the set of wigglings is more structured. Proposition A-9 shows that, as the number of points becomes large (in the limit, infinite), wigglings are constrained to be generic warpings. Thus, if \( \delta \) has an infinite number of points, we have the following:

\[
\text{isometries} \subset \text{linear warpings} \subset \text{generic warpings} \overset{n \rightarrow \infty}{\rightarrow} \text{wigglings}.
\]

With this assumption, we now can study a much more well-behaved set of transformations. Proposition A-10 gives the unexpected result that, in general, there exist no generic nonlinear warpings (Definition A-5), a result that does not depend on the manifold yet (i.e., we did not consider topology or curvature). Intuitively, there is no way to deform the distances in a nonlinear way that maintains the consistency of all constraints. The proof is based on an elementary argument based on the fact that any generic warping must preserve geodesics (Lemma A-11). Thus, only by assuming that the number of points is large, and with no assumptions on the manifold, we can conclude that:

\[
\text{isometries} \subset \text{linear warpings} = \text{generic warpings} = \text{wigglings}
\]

This means that the largest group of symmetries of the problem is composed by linear warping. At this point, we have to consider the property of the manifold. Proposition A-12 shows that, if the manifold has nonpositive curvature, then all linear warpings are necessarily isometries (the scaling factor is 1):

\[
\text{M curved} \overset{\text{linear warpings}}{\rightarrow} \text{generic warpings} = \text{wigglings}
\]

This means that for the sphere \( S^2 \) and the hyperbolic plane, isometries are the largest group of symmetries. This is surprising, because it means that we can recover the scale, even though the measurements available are completely non-metric. Instead, for Euclidean spaces, it is easy to see that a linear warping is always unobservable. Finally, Proposition A-13 discusses the special case of the circle. Because the topology is not simply connected, it is possible to establish additional constraints: intuitively, an arbitrary warping is not allowed, because if the distribution \( \delta \) is inflated too much, the tails will “crash” into each other and violate the problem constraints.

A-3.2 Proof details

Proposition A-9. If \( \delta \) is a connected open set, all wigglings are generic warpings.

Proof: The intuition is that a non-trivial wiggling is possible only if there are “gaps” between the points; as the points get denser, the gaps close and the wiggling degenerates to a warping.
Note that the definition of wiggling does not imply any particular property of the map \( \varphi \) such as continuity. It is a map defined only the subset \( S \) of \( \tilde{M} \). There is no information of how \( \varphi \) behaves outside of \( S \). However, if \( S \) is an open subset of \( \tilde{M} \), then necessarily \( \varphi \) must have certain regularities.

First of all, it should necessarily be a continuous map. This can be seen directly from the relation \( d(s_i, s_j) < d(s_k, s_l) \Leftrightarrow d(\varphi(s_i), \varphi(s_j)) < d(\varphi(s_k), \varphi(s_l)) \) if we let \( s_i = s_k \) and consider two sequences \( s_j^m \rightarrow \infty \) \( s_i \) and \( s_k^m \rightarrow \infty \) \( s_l \).

Consider two pairs of points \( s_i, s_j \) at distance \( \delta = d(s_i, s_j) \). Consider two other pairs of points \( s_k, s_l \) with the same relative distance \( \delta = d(s_k, s_l) \) — since the set is open, and the distance is continuous, \( s_k \) can be found in a neighborhood of \( s_i \) and \( s_l \) in a neighborhood of \( s_j \). Because \( d(s_i, s_j) = d(s_k, s_l) \), the wiggling direction constraint implies that \( d(\varphi(s_i), \varphi(s_j)) = d(\varphi(s_k), \varphi(s_l)) \). Because \( s_k, s_l \) have no other relation to \( s_i, s_j \) other than their distance, it follows that the distance of two points transformed by \( \varphi \) only depends on their initial distance: \( d(\varphi(s_i), \varphi(s_j)) = d(m(s_i, s_j)) \), for some possibly nonlinear function \( m \). Because \( \varphi \) is continuous, this holds for all points in \( S \), therefore \( \varphi \) is a generic warping.

**Proposition A-10.** All generic warpings are linear warpings.

*Proof:* The proof relies on Lemma A-11 below, which says that generic warpings preserve the geodesics. This means that, if the midpoint between \( A \) and \( B \) is a midpoint, then \( \varphi(C) \) is the midpoint between \( \varphi(A) \) and \( \varphi(B) \). Let \( d(A, C) = d(C, B) = \ell \). Then \( d(\varphi(A), \varphi(C)) = d(\varphi(C), \varphi(B)) = m(\ell) \). We can find two different expressions for \( d(\varphi(A), \varphi(B)) \):

\[
d(\varphi(A), \varphi(B)) = m(d(A, B)) = m(2\ell), \quad \text{and} \quad d(\varphi(A), \varphi(B)) = d(\varphi(A), \varphi(C)) + d(\varphi(C), \varphi(B)) = 2m(\ell).
\]

It follows that \( m(\ell) = \frac{1}{2} m(2\ell) \). Generalize this reasoning to an equal division of the geodesics in \( k \) parts, to derive \( m(x) = \frac{1}{2^m} m(kx) \) for all \( x > 0 \) and integers \( k \geq 1 \). Take the derivative of both sides with respect to \( x \) to obtain \( m'(y) = m(0) \), which implies that \( m \) is a linear function.

**Lemma A-11.** A generic warping preserves geodesics. More formally, for \( A, B, t \in [0, 1] \), let \( g(A, B, t) \) be the geodesic between \( A \) and \( B \). If \( \varphi : \tilde{M} \rightarrow M \) is a warping, then \( g(\varphi(A), \varphi(B), t) = \varphi(g(A, B, t)) \).

*Proof:* As a base case, we prove the statement for the midpoint. Suppose that there exists a geodesic between \( A \) and \( B \). Let \( C \) be the midpoint between \( A \) and \( B \), with \( d(A, C) = d(C, B) = \ell \). Let \( a = \varphi(A) \) and \( b = \varphi(B) \) be the transformed points. Let \( c = g(a, b, \frac{1}{2}) \) be the midpoint between \( a \) and \( b \), with \( d(a, c) = d(c, b) = \ell \). Using some elementary properties of geodesics, we shall derive that \( \varphi(C) = c \).

Because \( c \) is the midpoint, the shortest path between \( a \) and \( b \) goes through \( c \):

\[
d(a, c) + d(c, b) \leq d(a, x) + d(x, b), \quad \text{for all } x.
\]

Write this for \( x = \varphi(C) \):

\[
d(a, c) + d(c, b) \leq d(a, \varphi(C)) + d(\varphi(C), b)
\]

On the right-hand side, substitute \( d(a, \varphi(C)) = d(\varphi(A), \varphi(C)) = m(d(A, C)) \), using the definition of warping. Likewise \( d(\varphi(C), b) = d(\varphi(C), \varphi(B)) = m(d(C, B)) \), giving

\[
d(a, c) + d(c, b) \leq m(d(A, C)) + m(d(C, B)).
\]

The point \( c \) is the midpoint, so let \( \ell = d(a, c) = d(c, b) \), and \( L = d(A, C) = d(C, B) \). We obtain that \( \ell \leq m(L) \).

We can do the same computation with \( A \) and \( B \). Because \( C \) is the midpoint between \( A \) and \( B \), we have that \( d(A, C) + d(C, B) \leq d(A, x) + d(x, B) \), for all \( x \). Write it for \( x = \varphi^{-1}(c) \) and substitute \( A = \varphi^{-1}(a) \) and \( B = \varphi^{-1}(b) \) to obtain \( 2L \leq d(A, \varphi^{-1}(c)) + d(\varphi^{-1}(c), B) = d(\varphi^{-1}(a), \varphi^{-1}(c)) + d(\varphi^{-1}(c), \varphi^{-1}(b)) = m^{-1}(d(a, c)) + m^{-1}(d(c, b)) = 2m^{-1}(\ell) \), which gives us \( \ell \geq m(L) \). Together with \( \ell \leq m(L) \), we conclude that \( \ell = m(L) \). This means that \( d(a, \varphi(C)) = d(\varphi(C), b) = \ell \), and hence \( \varphi(C) \) is the midpoint between \( a \) and \( b \). Because the midpoint is unique, it follows that \( c = \varphi(C) \).

We have proved that \( g(\varphi(A), \varphi(B), \frac{1}{2}) = \varphi(g(A, B, \frac{1}{2})) \). By dividing the original geodesics, and applying the reasoning above recursively, one can show that \( g(\varphi(A), \varphi(B), \frac{a}{2^b}) = \varphi(g(A, B, \frac{a}{2^b})) \) for all integers \( b \geq 0 \) and \( a \leq 2^b \).

The set of dyadic rationals \( a/2^b \) is dense in \([0, 1]\), and the functions \( t \mapsto g(\varphi(A), \varphi(B), t) \) and \( t \mapsto \varphi(g(A, B, t)) \) are continuous, because they are compositions of continuous functions. If two continuous functions on the same domain \( X \) agree on a dense subset of \( X \), the agree on the whole domain. Therefore, it holds that \( g(\varphi(A), \varphi(B), t) = \varphi(g(A, B, t)) \) for all \( t \in [0, 1] \).

**Proposition A-12.** For \( S^m, m \geq 2 \) and the hyperbolic plane, all linear warpings are isometries.
Proof: This is true for all manifolds with nonzero curvature, but the \( m \)-sphere and the hyperbolic plane admit an elementary proof based on spherical/hyperbolic geometry. Firstly, note that a linear warping is a conformal map (Definition A-4) as the Jacobian is uniformly \( \alpha \) times an orthogonal matrix. Conformal maps preserve angles between geodesics.

Now it is time to recall high-school facts about spherical geometry: the sides of a spherical triangle are uniquely determined by its angles. The same is true for the hyperbolic plane [2].

Consider now three points in \( S \) and the induced spherical/hyperbolic triangle. Under a linear warping, its internal angles are preserved because a linear warping is conformal. Because the angles are preserved, the sides of the triangle are preserved as well, and therefore the distance between points is unchanged. Hence any linear warping is an isometry.

**Proposition A-13.** If \( M = S^1 \) and \( \text{rad}(S) + \text{infr}(f) < 2\pi \), a linear warping with \( \alpha \leq (2\pi - \text{rad}(S))/\text{infr}(f) \) is unobservable.

Proof: (sketch) This can be verified directly; the upper bound on \( \alpha \) ensures that the tails of \( S \) do not overlap in the informative range of \( f \). This result does not hold for \( S^2 \), where the geometry of the problem constrains linear warplings to be isometries (\( \alpha = 1 \)).

### A-4 ADDITIONAL EXPERIMENTS

#### A-4.1 Results for different similarity statistics

Proposition 2 ensures that any statistics is a function of the pixel distance, but this result is limited in three ways: 1) it is only an asymptotic result, valid as time tends to infinity; 2) it assumes a perfectly uniform attitude distribution; and 3) it does not ensure that the function \( f \) is invertible (monotonic). Therefore, it is still an engineering matter to find a statistics which is 1) robust to finite data size; 2) robust to a non-perfectly uniform trajectory; and 3) has a large invertible radius. An exhaustive treatment of this problem is outside the scope of this paper and delegated to future work. Here, we briefly show the results for three other statistics in addition to the luminance correlation. All statistics are defined as the correlation of an instantaneous function of the luminance and can be efficiently computed using streaming methods. The first variant consists in applying an instantaneous contrast transformation \( c: y \mapsto y^2 \) to the luminance before computing the correlation:

\[
Y_{ij} = \text{corr}( c(y_i(t)), c(y_j(t)) )
\]  
(5)

The second statistic is the correlation of the temporal derivative \( \dot{y} = \frac{d}{dt} y \) of the luminance:

\[
Y_{ij} = \text{corr}( \dot{y}_i(t), \dot{y}_j(t) )
\]  
(6)

This was inspired by recent developments in neuromorphic hardware [5]. Finally, we consider the correlation of the sign of the luminance change, as it is invariant to contrast transformations:

\[
Y_{ij} = \text{corr}( \text{sgn}(\dot{y}_i(t)), \text{sgn}(\dot{y}_j(t)) )
\]  
(7)

Table A-1 shows the Spearman score obtained by using these on the OMNI data (the most challenging dataset). We find, in this case, that the contrast-scaled luminance (5) is slightly better than the simple correlation; the solution found is qualitatively similar (Fig. B-7d). The two other similarity statistics (6) and (7) have much lower scores; for them, the monotonicity assumption is not well verified: their distributions are not informative for large distances (Fig. B-8, B-9). It is clear that there is a huge design space for similarity statistics. In the end, we did not find any statistic which was better than the simple correlation uniformly for all our three data sets. Therefore, we consider this an open research question.

<table>
<thead>
<tr>
<th>dataset</th>
<th>Spearman score</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMNI cor(y)</td>
<td>0.9173 0.9438</td>
</tr>
<tr>
<td>OMNI cor(c(y))</td>
<td>0.9212 0.9465</td>
</tr>
<tr>
<td>OMNI cor(\dot{y})</td>
<td>0.8550 0.9211</td>
</tr>
<tr>
<td>OMNI cor(sgn(\dot{y}))</td>
<td>0.8739 0.9077</td>
</tr>
</tbody>
</table>

(See complete results in Table B-3a)
REFERENCES


