ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 199-204

# When does geodesic distance recover the true hidden parametrization of families of articulated images?

David Donoho, Carrie Grimes
Department of Statistics, Stanford University

390 Serra Mall, Stanford, CA 94305

**Abstract**. Throughout the sciences and engineering, there is an urgent need to study observed high dimensional data and "learn" any nonlinear parametrization which might be underlying the data. Recently, the ISOMAP procedure was proposed as a new way to recover such hidden parametrizations of high-dimensional data [3]. But for which kinds of phenomena can ISOMAP truly recover the underlying structure?

We consider a specific kind of data – families of images generated by articulation of an object, in an idealization where the images are functions on the continuum plane. In these cases, using the ambient $L^2$ distance as a metric for articulated images makes the family a nonlinear manifold. An analog of the ISOMAP procedure is to obtain geodesic distance between points of the manifold and attempt to realize that distance as a euclidean metric on a euclidean space.

For a natural renormalization of geodesic distance, we give a set of example articulations where the geodesic distance on the manifold is exactly proportional to the euclidean distance in parameter space, and hence, the continuum analog of ISOMAP is successful. Examples include: translations of a disk or of a square, rotations of a square, rescalings of a circle, modifications of a horizon, and morphings of a disk. However, in the case where several components of the image articulate separately and occlusion is possible, characteristics of the data manifold may preclude recovery of the original parameter space.

## 1 Introduction

Recently, Tenenbaum et al. [3] proposed the ISOMAP procedure as a general tool for recovering the unknown parametrization underlying a set of digital images, $\{I_i\}$, of faces in various attitudes and articulations. Similar work on the underlying parameters of families of articulated images can be seen in Nayar et al. [2], Belhumeur and Kriegman [1]. The general principle of ISOMAP is to measure distance between images, not using euclidean distance (which is

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 199-204

ignorant of the manifold structures), but according to the shortest path in the nearest neighbor graph; and to use this graph distance as input to a classical "principal components" multidimensional scaling procedure.

These published examples naturally raise the question of how "accurate" or "correct" the ISOMAP procedure is; an obvious way to test this would be to attempt recovering the parametrization of data manifolds when we know what seems to be "the natural" parametrization because the data are generated from an analytical model.

Because of sampling issues (for example, how many model points $I_i$ do we have? Are they densely distributed?) and digitization issues (How finely spaced are the image pixels?), it is difficult to give simple and precise answers to this question as originally posed.

Instead, we adopt an abstract "continuum" viewpoint where neither sampling nor digitization can cause problems. We think of an image as a function $I(x)$ of a continuous variable $x \in \mathbb{R}^2$ and consider articulations of a base image $I_0$ according to a family of transformations $(T_\theta : \theta \in \Theta)$, where $\theta$ is the parameter of the transformation and $\Theta$ is the parameter space.

We consider the space of images as a subspace of $L^2(\mathbb{R}^2)$ and let $\mu(\theta_1, \theta_2) = ||I_{\theta_1} - I_{\theta_2}||_{L^2}$. With this metric, the space $M = \{I_\theta : \theta \in \Theta\}$ is a continuous manifold, and generally a nonlinear manifold. $M$ may loosely be called curved although the manifold is not necessarily differentiable and thus its curvature may not be well-defined.

We consider the following problem: when does the geodesic distance on the manifold $M$ reveal the true underlying parametrization? By this we mean: when is the geodesic distance on the manifold proportional to the euclidean distance between parameter points? While this may seem unlikely to happen for general datasets, in the image manifold setting, there are a number of cases where this is either exactly or approximately true. These cases include: translation of simple black objects on a white background; rotation of certain simple black objects on a white background; and morphing of boundaries of black objects on a white background. Combinations of these cases can also be successful, given that the articulations remain disjoint.

These examples provide a theoretical 'imaging' setting which supports the principle implicit in ISOMAP – that geodesic distance can reveal the underlying parametrization of the image manifold.

However, we have also found an important class of image manifolds where geodesic distance fails: i.e. where the natural euclidean parametrization is not recovered by the analysis of geodesic distance. Those examples involve *occlusion* – several articulating objects which collide and cover each other for some settings of the articulation.

Mathematical details and derivations are sketched elsewhere (as there is simply no space), along with many related results. This paper gives in Section 2 a methodology and simple examples where geodesic distance precisely recovers the natural parametrization, and in Section 3 more complex examples and exceptions.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 199-204

## 2 Basic Examples

In a longer treatment, we would show that for images with edges (i.e. practically all interesting ones) the image manifold generates "infinities" when geodesic distance is calculated. To remedy this, we consider smoothing the images with window size $h$, creating a smooth manifold $M_h$ and considering the limit $\delta(\theta_a, \theta_b)$ of geodesic distance $G_h$ renormalized for each $h$ so that for two distinct points $\theta_a$ and $\theta_b$, $G_h(\theta_a, \theta_b) = 1 \,\forall\, h$. We can prove, under simple regularity conditions on the image and the articulation, that this limiting process defines a $C^2$ Riemannian manifold.

Our fundamental tool will be the following: Suppose we are articulating an image according to $I_\theta = T_\theta I_0$, and we want to know the distance between $I_\theta$ and $I_{\theta+\eta}$, for $\eta$ a small displacement vector. Then if the image $I_0$ is the indicator of a region $R$, for small $\eta$ we have

$$\delta(\theta, \theta+\eta) \approx Const \cdot ||\eta||_2 \times \left( \int_{\partial R} \langle n(p), \eta \rangle^2 \, dp \right)^{1/2} \tag{1}$$

where $n(p) = \begin{bmatrix} n(p_1) \\ n(p_2) \end{bmatrix}$ is the unit vector normal to the boundary of the region at any boundary point $p \in \partial R$. This expression suggests a rough analogy between an object moving through the plane with a ship moving through water; the expression above is a measure of how much "water" the boundary has to "cross," in a mean-square sense.

If different parts of the boundary undergo different local motions during a given articulation, then the above formula can be changed to incorporate not a fixed motion vector $\eta$, but instead a variable motion vector $\eta(p)$.

### 2.1 Translation

Consider the situation where the prototype image $I_0 = 1_{\{||x||<1\}}$, a disk at the center of the image, is articulated in the following translation family: $\Theta = \mathbb{R}^2$, $(T_\theta I)(x) = I(x-\theta)$; $I_\theta = T_\theta I_0$. Then the geodesic distance between any two images parameterized by $\theta_1$ and $\theta_2$ is given by

$$||\theta_1 - \theta_2||_2 \times \left( \int_{\partial R} \langle n(p), \eta \rangle^2 \, dp \right)^{\frac{1}{2}} = ||\theta_1 - \theta_2||_2 \times \left( \int_0^{2\pi} (\cos\phi)^2 \, d\phi \right)^{\frac{1}{2}}$$

where $n(p)$ is again the unit vector normal to the circle at any point $p \in \partial R$, the disk boundary. Similarly, $\eta$ is the chosen direction of motion — in the case of a rigid translation the direction and magnitude of motion is constant around the boundary for a given articulation. Then evaluating the integral gives a constant value for any direction of motion $\eta$: $\sqrt{\pi} \times (\eta_1^2 + \eta_2^2)^{\frac{1}{2}} \times ||\theta_1 - \theta_2||_2$, and if we choose $\eta$ as a unit vector, we observe that the geodesic distance is, up to a constant factor, simply Euclidean distance in parameter space. In this case, *renormalized geodesic distance recovers the natural parametrization.* We can check the theoretical calculations against the empirical behavior of ISOMAP.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 199-204

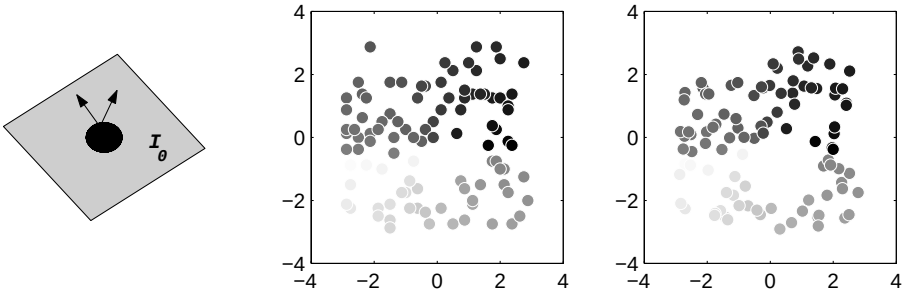


Figure 1: (left) Motion and Normal vectors for a disk image; (center) Original parameter space of disk centers sampled in a plane; (right) ISOMAP 2D embedding matched to original coordinates.

We applied the Tenenbaum, De Silva, and Langford algorithm [3] on a set of 100 images, each at resolution 64-by-64 pixels, using the k-nearest-neighbors option, k=7. As seen in Figure 1, the ISOMAP algorithm recovers the plane structure of the original parameter space — up to a rigid motion. To visually compare the "true" and "ISOMAP" solutions, we scale and rotate the ISOMAP embedding using a Procrustes rotation with optimal scaling. The result agrees with our theoretical calculations.

### 2.2 Rotation

As a second example, consider a radial 'pie wedge' rotated around its apex at the origin; this has the form of indicator of a region of points $(l\cos(\omega), l\sin(\omega))$ for $0 < l < L$ and $\omega_0 < \omega < \omega_1$, with $\omega_0 = a - \theta$ and $\omega_1 = b - \theta$. For points on the boundary of the wedge where $\omega \in \{\omega_1, \omega_2\}$, the normal $n(p)$ is collinear with the motion vector $\eta(p)$; while at the boundary points on the circular segment at the edge of the wedge, the motion vector is tangent to the boundary, hence $\langle n(p), \eta(p)\rangle = 0$. However, for a given articulation, the length of the path traveled by any point on the boundary is a function of both the articulation and of the distance between the point and the center of the rotation. We adapt the integral in (1) to include a varying path over the boundary of the wedge, where the distance traveled by a boundary point is a segment of a circle with radius $l$. Therefore, the distance between $\theta_1$ and $\theta_2$ is $\frac{1}{\sqrt{2}} \times L^3 \times |\theta_2 - \theta_1|$. In short, *geodesic distance exactly recovers the true parametrization.*

## 3 Composite Examples

### 3.1 Independent Articulations

Consider a "hand" made up of a unit circle with radial blade fingers of angular width $2\gamma$, each finger centered around a radial line at angle $\theta^{(i)}$ for $i = 1, ..., 5$.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 199-204

Then the image prototype is

$$I_0 = 1_{\{||x||_2 \leq 1\}} + \sum_{i=1}^{5} 1_{\{||x||_2 \leq L+1\}} 1_{\{<x,\, \theta^{(i)}>/||x||_2 \leq \cos^2 \gamma\}}$$

We can calculate the distance between any two positions for the $i$th finger by a constant function identical to that derived in the previous radial 'pie wedge' example. As long as the $\theta^{(i)}$ are restricted such that the fingers may not overlap, the full hand distances may be calculated by summing the resulting value over all fingers.

However, for examples where the movements are not independent, such as an articulated "finger" where the first joint may move the entire finger, but subsequent joints move sequentially smaller segments, the underlying parametrization may neither be convex (depending on the constraints) nor simple to recover.

## 3.2 Overlapping Articulations

Consider an image containing two independently articulated disks. If we fix the center of the first disk, and move the second around, the true underlying parametrization should be two-dimensional. Let $I_\theta = 1_{\{||x||_2 \leq 1\}} + 1_{\{||x-\theta||_2 \leq 1\}}$. If the disks are forcibly prevented from occluding, then the distance between two images, $I_{\theta_1}$ and $I_{\theta_2}$, is entirely described by the euclidean distance in parameter space, just as in the single disk images.

For empirical verification of this theoretical effect, consider the right panel of Figure 2, which shows the result of applying the ISOMAP algorithm to 100 images containing a fixed and a variable disk. ISOMAP recovers the ring structure of the non-occluding disk centers in approximately the correct angle groupings. The left panel shows the original parameter space sample with the boundary of the fixed second disk marked (points are colored by angle from 0).

Also in the continuum model, consider the case where the parameter space is not restricted to non-overlapping disks. In this case, the metric structure becomes radically more complex. Essentially, when the mobile disk is overlapping the fixed central disk, the cost for motion is reduced due to the smaller amount of exposed boundary. In effect, because the boundary is changing during articulation, and the different directions of motion cause different boundary changes, and other exotica crop up. In fact, the formal embedding dimension becomes infinite, although we do not have space here to fully document the effect.

In a discretized version of that situation, the ISOMAP algorithm result for a 2-dimensional embedding has an appearance very different from the 'natural parametrization'. It is an unexpected (but in agreement with the theoretical results) multiply-spined structure (middle panel of Figure 2). To travel from one quadrant of the parameter space to another, ISOMAP typically chooses a path that passes the mobile disk directly through the immobile one.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 199-204

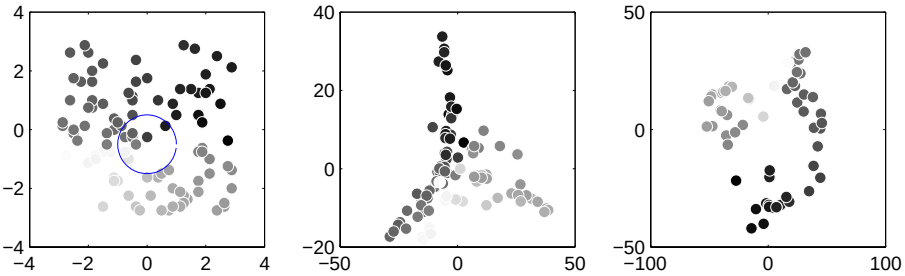


Figure 2: (left) Original parameter space sample for two disk images, with fixed disk outlined; (center) ISOMAP 2D embedding of all disk images given by original parameter space; (right) ISOMAP 2D embedding for non-occluding disks only.

# 4 Discussion

There are a number of additional issues relating to the use of an ISOMAP-type algorithm in image data. One example, illustrated by the right panel of Figure 2, is the convexity of the underlying parameter space. ISOMAP [3] requires that the image manifold be geodesically convex. However, in our case we note that it is sufficient for the parameter space to be convex. For the double disk images, removing centers where the two disks occlude effectively punches a hole in the parameter space, and results in a distortion of the relative distances within points on one side of the missing region compared to distances across the missing region. In cases where the parameter space is not sampled adequately, we could expect similar problems to occur at gaps in the data.

Furthermore, actual experimentation with the algorithm reveals discretization artifacts which depend on the relative size of objects in the image to the pixel resolution of the image. For example, rotating a straight line such as the fan wedge results in systematic deviations at the 45-degree angle which may be a result of the pixelization.

# References

[1] Peter N. Belhumeur and David J. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28(3):1–16, 1998.

[2] S.K. Nayar, S. Baker, and H. Murase. Parametric feature detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 471–477, San Francisco, California, 1996.

[3] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.