WHEN DOES ISOMAP RECOVER THE NATURAL PARAMETERIZATION
OF FAMILIES OF ARTICULATED IMAGES?

by

David L. Donoho
Carrie Grimes

August 2002

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
WHEN DOES ISOMAP RECOVER THE NATURAL PARAMETERIZATION
OF FAMILIES OF ARTICULATED IMAGES?

by

David L. Donoho
Department of Statistics
Stanford University

Carrie Grimes
Department of Statistics
Stanford University

August 2002

This research was supported in part by National Science Foundation
grant DMS 00-72661

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
When does ISOMAP Recover the Natural Parameterization of Families of Articulated Images?

David L. Donoho and Carrie Grimes
Department of Statistics
Stanford University
{donoho, carrie}@stat.stanford.edu
August 3, 2002

Abstract

In science and engineering, there is a need to observe high dimensional data and "learn" a potential underlying nonlinear parametrization for the data. Recently, the ISOMAP procedure was proposed as a new way to recover hidden parametrizations of high-dimensional data. We consider a specific kind of data – families of images generated by the articulation of an object – in an idealization where the images are functions on the continuum plane. Using the ambient $L^2$-distance as a metric between articulated images makes the articulation family a nonlinear manifold. We introduce Continuum ISOMAP, an analog of the ISOMAP procedure where we obtain geodesic distance between points of the articulation manifold and attempt to realize that distance as a Euclidean metric on a Euclidean space. We study the question of when Continuum ISOMAP can truly recover the underlying parametrization of a family of images. We show that for images with edges, this idealization suffers from various infinities, but a natural renormalization of the notion of geodesic distance is well-defined. We exhibit a set of interesting image manifolds where the geodesic distance on the manifold is exactly proportional to the Euclidean distance in parameter space, and therefore, Continuum ISOMAP works perfectly to recover the natural parametrization up to a rigid motion. Examples of such successes include: translations of a disk, rotations of a closed figure, articulations of a horizon, independent non-occluding motions of 'fingers' of a cartoon 'hand', and gestures of a cartoon 'face', with articulated features. The theoretical predictions of the Continuum ISOMAP model are borne out by empirical experiments with published ISOMAP code. However, we demonstrate that in the case where several components of the image articulate independently and occlusion is possible, recovery of the original parameter space may be precluded by the characteristics of the data manifold. This methodology suggests new modifications to the existing procedures for recovering underlying image parametrization.

Keywords: ISOMAP. Multidimensional Scaling. Manifolds of Articulated Images.

Acknowledgements: CG would like to thank Vin de Silva, Josh Tenenbaum, Gunnar Carlsson and George Papanicolaou and DLD would like to thank Raphy Coifman, in both cases for discussions and references. This work has been partially supported by National Science Foundation grant DMS 00-72661.
1 Introduction

An object in the world, when rendered into an image, might be recorded at various locations, scales, orientations, and perspectives.

The collection of all such images of an object, as the parameters (locations, scale, etc.) vary, can be thought of as a manifold in the high-dimensional space of all conceivable images. The ability to recover the underlying parameters (location, scale, etc.) can obviously be very important for image understanding, image coding, and a variety of other purposes; for expressions of interest in this viewpoint, see Nayar et al. [7], Belhumeur and Kriegman [1].

In one scientifically important class of settings, one is given many digital images \((I_i : i = 1, \ldots, N)\) of the same object in a variety of articulations and poses, but where the underlying parametrization of the articulation are unknown. The images are thus thought to arise from a manifold \(M\), without the manifold or the associated parametrization being known. The issue is to "learn" the manifold and its intrinsic parametrization from the "samples". This can be important for understanding articulated vehicles in target recognition, and for understanding articulated faces in facial recognition.

The general problem of learning the shape of a manifold from scattered observations has been around for a long time; it has been the source of many multivariate techniques, including principal components analysis, independent components analysis, multidimensional scaling, self-organizing mappings, and other important methodological developments.

Recently, Tenenbaum et al. [9] 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. The general principle of ISOMAP is to measure distance between images, not using Euclidean distance (which is ignorant of the manifold structures), but using distance according to the shortest path in the nearest neighbor graph; and to use this graph distance as input to a classical "principal coordinates" multidimensional scaling procedure.

1.1 Analysis of ISOMAP

Tenenbaum et al. [9] published a few interesting examples, for example mapping out the parameters underlying a face seen from a variety of viewpoints. These empirical successes lead to the ...

**Obvious Question:** how "correct" is the ISOMAP procedure; does it really recover the "true" underlying parametrization of families of articulated images?

This leads immediately to the ...

**Obvious Approach:** Test ISOMAP for synthetic data where we know a priori "the natural" parametrizations and see if it can recover the parametrization of image manifolds.

In following the 'obvious' path, we would construct datasets of artificial images undergoing standard articulations – translations, rotations, etc. – and ask if the ISOMAP parametrization correctly discovers the underlying parametrizations.

While the question and the investigative approach seem clear at first glance, on closer inspection, it is not clear that one can push this line of investigation very far. There are several reasons a purely empirical investigation based on running ISOMAP on artificial examples may not be enlightening.
• Sampling issues. Whether ISOMAP works or not will depend on how many model images $I_i$ are in the database. In some vague sense, it will be important for these images to be well-distributed across the image manifold, but exactly what this means in a specific instance is unclear. Consequently, if ISOMAP "fails", this may be due to poor experimental design rather than any intrinsic property of ISOMAP.

• Digitization issues. In some sense, the fact that images are discretized into pixels makes them "noisy" / "blocky"; this again causes some difficulty in interpreting "failure" of ISOMAP -- is it due to the pixelization or is it intrinsic to the type of articulation?

• "Big Picture" issues. Empirical work with ISOMAP really doesn't give us any intellectual framework that we can leverage into other settings, at least not the kind of framework that might be possible by a more theoretical approach.

For these and other reasons, we suggest that the best way to understand when ISOMAP works, and how, is not to study ISOMAP as originally proposed. Instead we adopt an abstract "continuum" viewpoint where neither sampling nor digitization can cause problems, and where a clear intellectual framework exists naturally. For this viewpoint, we think of an image as a function $I(x)$ of a continuous variable $x \in \mathbb{R}^2$. We consider articulations of a base image $I_0$, producing to a family of images $(I_\theta : \theta \in \Theta)$, where $\theta$ is the parameter of the articulation and $\Theta$ is the parameter space. Examples we will consider include translation families, where $\Theta = \mathbb{R}^2$, $I_\theta(x) = I_0(x - \theta)$.

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

In this continuum setting, a natural analog of ISOMAP can be considered. One lets $G(\theta_1, \theta_2)$ be the geodesic distance defined by taking the shortest path lying in $M$ going from $I_{\theta_1}$ to $I_{\theta_2}$. Then one asks if $G(\theta_1, \theta_2)$ is proportional to Euclidean distance between $\theta_1$ and $\theta_2$, for all pairs $(\theta_1, \theta_2)$. If proportionality holds for a certain articulation manifold, then one says that (as a definition) "Continuum ISOMAP works" for that manifold.

This is indeed a continuum analog of ISOMAP, where the collection of images is continuous and the domain is continuous as well. It avoids the sampling and discretization issues associated with the 'obvious' empirical approach to studying ISOMAP. It also will turn out to provide a clear theoretical framework with crisply-stated problems that are easily answered.

1.2 Results

Our criterion for saying that "Continuum ISOMAP works" is quite stringent, and therefore it may seem unlikely that we could show that Continuum ISOMAP would succeed in general. However, in the image manifold setting that we describe, there are a number of interesting cases where Continuum ISOMAP actually does work. These cases include:

• Translation of simple black objects on a white background;

• Pivoting certain simple black objects on a white background around a fixed point;

• Morphing of boundaries of black objects on a white background;
• Articulation of ‘fingers’ of a digital ‘hand’;

• Articulation of a cartoon face by arranging its eyebrows, eyelids, and lips.

It also turns out that Continuum ISOMAP works for ‘movies’, i.e. for images articulating in time; an example being a cartoon face gesturing in time according to a sufficiently rich and complete inventory of gestures. Our theory predicts that from ‘watching’ sufficiently many movies of a cartoon face gesturing in time, ISOMAP will correctly learn the correct parametrization of the facial gesturing. Details will be provided and derivations will be sketched. Many more such results are possible, but we do not pursue them in this paper.

Our examples provide a theoretical framework in the setting of understanding families of images which validates the claim implicit in early uses of ISOMAP. Specifically, we show that geodesic distance can perfectly reveal the natural parametrization of certain image articulation manifolds.

However, we also point out 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 often involve occlusion – several articulating objects which collide and cover each other for some settings of the articulation.

The article is organized as follows: Section 1 introduces the problem area; Section 2 covers the mathematical foundations for image manifolds in the case of a simple object; Section 3 gives a few general ‘simple’ images for which the mathematical methodology applies; Section 4 tests the theoretical results against generated image data; Section 5 interprets the methodology for more complex example images (such as a horizon parametrized by a basis function expansion); Section 6 discusses composite articulations of more than one object; Section 7 has more exotic mathematical examples such as simple movies composed of images; Section 8 offers examples where the ISOMAP methodology fails to recover the natural parametrization perfectly; Section 9 discusses the case of occluding objects in an image; Sections 10 discusses effects related to non-convex parameter spaces; and further issues and discussion appear in Section 11.

2 Continuum ISOMAP

In this section, we develop the notion of a Continuum ISOMAP procedure. This involves a review of some basic geometric notions, such as arclength and geodesic, showing the ill-posedness of the ‘obvious’ approach to Continuum ISOMAP, and defining a special renormalized procedure which we then show is well-posed.

2.1 Some Simple Parametrized Manifolds

Consider a smooth function \( f_0 : \mathbb{R}^2 \to \mathbb{R}^2 \), which has \( L^2 \)-norm one, and consider the family

\[
f_\theta(x_1, x_2) = f_0(x_1 - \theta, x_2),
\]

so that the ‘image’ is translating right-to-left. A simple calculation shows that

\[
\|f_\theta - f_\theta_0\|_{L^2} = \left( \int |f_\theta(t) - f_\theta_0(t)|^2 \, dt \right)^{1/2} = \eta(||\theta - \theta_0||),
\]
where $\eta : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a nonlinear function. If we consider the case where $f_0$ is a reasonably localized function such that, for $|\theta_0 - \theta_1|$ large, $f_{\theta_0}$ and $f_{\theta_1}$ are almost entirely separated, we can see that $\eta(\cdot)$ is continuous and obeys $\eta(0) = 0$, $\eta(\infty) = \sqrt{2}$.

The significance of (2.2) is that, if we define the set $M = \{f_0 : \theta \in \Theta\}$, then equipping it with the metric $\mu(\theta_0, \theta_1) = \|f_{\theta_0} - f_{\theta_1}\|$ we have that $M$ is a smooth manifold (a curve actually) embedded in $L^2(\mathbb{R}^2)$. The tangent space to $M$ at $f_0$ is simply

$$\bar{f}_0 = \{f_0 + h f_0 : h \in \mathbb{R}\}$$

where

$$f_0 = \frac{\partial}{\partial \theta} f_0 = -f_0'(x - \theta);$$

The relation $\mu(\theta_0, \theta_1) = \eta(\|\theta_0 - \theta_1\|)$ means that, while the topology of $M$ is equivalent to the Euclidean topology, the embedding of the manifold is curved.

Consider now the two-dimensional case, starting with a smooth function $f_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$ and a two-parameter translation: $\theta = (\theta_0, \theta_1)$. For example, let $f_0$ be a radially symmetric function with effectively finite radius. Then we can calculate:

$$(T_{\theta} f)(x) = f(x - \theta);$$

then

$$f_\theta = T_{\theta} f_0, \quad \theta \in \Theta = \mathbb{R}^2$$

defines a family of translated objects in the plane, and the $L^2(\mathbb{R}^2)$ distance has the form

$$\|f_{\theta_0} - f_{\theta_1}\|_{L^2(\mathbb{R}^2)} = \eta(\|\theta_0 - \theta_1\|)$$

for an appropriate nonlinear function $\eta(\cdot)$. Again $M = \{f_\theta : \theta \in \Theta\}$ is a manifold, actually a 2-surface embedded in $L^2(\mathbb{R}^2)$. Due to (2.3), $M$ has a Euclidean topology, but the embedding is curved.

### 2.2 Geodesic Distance

Let $M$ be a smooth manifold embedded in $L^2$, and consider now a curve $\gamma : [0, 1] \rightarrow M$. Suppose that $\gamma$, viewed as a curve in $L^2$, is smooth, so that $\gamma(t)$ exists in $L^2$ for each $t \in [0, 1]$, and varies continuously in $t$. More precisely, $\gamma(t)$ is a function defined on the plane $(x_1, x_2)$ and has finite $L^2$ norm. Then the length of $\gamma$ is defined to be

$$L(\gamma) = \int_0^1 \|\gamma(t)\|_{L^2} \, dt.$$

The geodesic distance between points $\theta_0$ and $\theta_1$ in the smooth manifold $M$ is the length of the shortest curve joining the two points, or

$$G(\theta_0, \theta_1; M) = \inf\{L(\gamma) : \gamma(0) = f_{\theta_0}, \gamma(1) = f_{\theta_1}\}$$

Because $M \subset L^2$ and $L^2$-distance is the length of the shortest path in $L^2$ between end points $\theta_1$ and $\theta_0$,

$$G(\theta_1, \theta_0; M) \geq \mu(\theta_1, \theta_0; M)$$

As a simple example, suppose $f_0$ is a smooth radial function on $\mathbb{R}^2$, such as the Gaussian $f_0(x) = e^{-\frac{|x|^2}{2}}$, and $f_\theta(x) = f_0(x - \theta)$. Then

$$G(\theta_1, \theta_0; M) = c\|\theta_1 - \theta_0\|_2$$
where

\[ c^{-1} = \| \frac{\partial}{\partial x_1} f_0 \|_{L^2(\mathbb{R}^2)}. \]

In short, in this case, geodesic distance agrees globally with Euclidean distance on the natural parameter space.

2.3 Recovering the Natural Parametrization

When geodesic distance in \( M \) agrees with Euclidean distance in parameter space, we will say that “geodesic distance has recovered the natural parametrization.” To understand why, we look at the procedure for metric multidimensional scaling.

The usual procedure for metric multidimensional scaling [6, 2] for datasets of \( n \) objects is to first construct a matrix of pairwise distances between the objects. We then take the distance matrix \( (d_{i,j} : 1 \leq i, j \leq n) \) between pairs of objects, and define an auxiliary matrix \( A \) with entries \( a_{i,j} = -d_{i,j}^2 \). We then preprocess the \( A \) matrix, creating a matrix \( C \) with entries \( c_{i,j} = a_{i,j} - \bar{a}_{i} - \bar{a}_{j} + \bar{a} \). The matrix \( C \) resembles a covariance matrix in that if the original pairwise distances represent Euclidean distances in a \( k \) dimensional space, \( C \) will be symmetric and positive semidefinite of rank \( k \). One performs an eigenanalysis of \( C \); if it is of rank \( k \), then we can use the eigenanalysis to represent the objects as points in \( k \)-dimensional space, in such a way that the Euclidean distance between each pair of points \((i, j)\) is exactly \( d_{i,j} \).

To do so, we suppose there are \( k \) positive eigenvalues of \( C \), and we take the corresponding eigenvectors, call them \((v_j : 1 \leq j \leq k)\), and define data vectors \( y_j = \sqrt{\lambda_j} v_j \). If we form the \( n \)-by-\( k \) matrix \( Y \) with the \( y_j \) for the first \( k \) such columns, we obtain a representation of the \( n \) objects as follows: the \( i \)-th row of this matrix, \( x_i \), say, represents the \( i \)-th object. If \( C \) really is positive semidefinite, this representation is faithful to the original metric, so that

\[ \| x_i - x_j \| = d_{i,j}. \]

(Of course, in general \( C \) need not be positive semi-definite, which will not be true if there is in general no \( n \)-dimensional embedding representing the \( n \) objects with specified \( d_{i,j} \). In such cases the standard MDS procedure is to embed the data using only the ‘nice’ eigenvalues [6].)

In Continuum ISOMAP, we do not have a finite collection of objects; instead we have geodesic distance \( G(\theta_0, \theta_1) \) defined for all pairs with each \( \theta_i \in \Theta_i \), \( i = 0, 1 \). We take any finite collection of \( n \) points \((\theta_i, i = 1, \ldots, n)\), and set \( d_{i,j} = G(\theta_i, \theta_j) \) for \( 1 \leq i, j \leq n \). Then applying the metric multidimensional scaling to the distance matrix gives a faithful coordinate representation of the \( \theta_i \), in the sense that

\[ \| x_i - x_j \| = G(\theta_0, \theta_1). \]

Now if we have the situation of the last subsection, where for all \( \theta_0, \theta_1 \) in the underlying parameter space,

\[ G(\theta_0, \theta_1) = c \cdot \| \theta_0 - \theta_1 \| \]

it then follows that metric multidimensional scaling of \( G \) produces

\[ \| x_i - x_j \| = c \cdot \| \theta_i - \theta_j \|. \]

Now the \( x_i \) have been produced purely from knowledge of the \( d_{i,j} \), and not from any knowledge that there is a finite dimensional parameter space underlying those distances.
However, the $x_i$ have captured the essence of the situation: a $k$-dimensional Euclidean set of points.

Summarizing this discussion, we have

**Definition 2.1** Suppose that the geodesic distance $G(\theta_0, \theta_1; M)$ between points in the manifold $M$ is proportional to Euclidean distance in the parameter space $\Theta$. Then we say that Continuum ISOMAP works perfectly.

Of course, the recovery is correct only up to a scaling and rigid motion. Taking into account the previous subsections and the above discussion, we have:

**Corollary 2.1** Suppose we have a parametrized family of objects $f_\theta(x)$ defined by translation of a common prototype: $f_\theta(x) = f_0(x - \theta)$, and that $f_0$ is in $L^2$ and is differentiable in $L^2$. Then the geodesic distance between $f_{\theta_0}$ and $f_{\theta_1}$ has the form

$$G(\theta_0, \theta_1) = c \cdot \|\theta_0 - \theta_1\|.$$ 

Hence Continuum ISOMAP works perfectly.

As a nice side effect of this corollary, we will need only to be concerned with the geodesic distance $G(\theta_0, \theta_1)$ up to this constant of proportionality. For complete clarity, we also point out that, in these cases, using $L^2$ distance rather than geodesic distance would not recover the true parameter space. Indeed as we have seen $\mu(\theta_0, \theta_1) = \eta(\|\theta_0 - \theta_1\|)$ where $\eta(\cdot)$ is nonlinear. It then happens that if we apply multidimensional scaling to the interpoint distances, $C$ would not be of rank 2, but instead of rank $n$ in general, and multidimensional scaling would not recover an exact representation by a 2-dimensional parameter space. Therefore, the properties of geodesic distance make it uniquely successful in this setting.

### 2.4 Non-differentiable Image Manifolds

Real images are full of edges and so are non-differentiable.

Consider for simplicity the case of an indicator function $f(x) = 1_{\{|x|<1\}}$ on $\mathbb{R}^1$, and generate a family of functions by translating $f_0(x)$ to $f_\theta(x) = f_0(x - \theta)$ for $\theta \in \mathbb{R}^1$. Then let $M = \{f_\theta\}$.

Now we can write $\mu(\theta_1, \theta_0; M) = \|f_{\theta_1} - f_{\theta_0}\|_{L^2} = \eta(\|\theta_1 - \theta_0\|)$, for a new function $\eta(\cdot)$. However, we can calculate for a certain continuous function $\eta$ (different than before) that $\eta$ is no longer differentiable as a function of $\theta$. In particular, for a translation by $d$,

$$\eta(d) = c \cdot \min(d^{\frac{1}{2}}, \sqrt{2})$$

(2.5)

The fact that $\eta(d) \propto d^{1/2}$ as $d \to 0$ means that $M$ is not a smooth manifold. Similarly, for the case of an image with edges in $\mathbb{R}^2$ such as the indicator of the unit disk,

$$I_0(x) = 1_{\{|x|\leq 1\}},$$

and a family of translated images, $\theta = (\theta^{(1)}, \theta^{(2)}), \Theta = \mathbb{R}^2, I_\theta = I_0(x - \theta), \theta \in \Theta$. Then as before, we can calculate that $\eta$ is not differentiable and $\eta(d) \propto d^{1/2}$.

To illustrate a more general case of the phenomena caused by non-differentiability of images, we define a path $\theta(t)$, $t \in [0, 1]$, joining $\theta_0$ and $\theta_1$ in the parameter space $\Theta$ so that $\theta(0) = \theta_0$ and $\theta(1) = \theta_1$. In fact, we can see the following:

7
Theorem 2.1 If $M$ is the articulation manifold gotten by translating an image with edges, then $M$ is a continuous but not a differentiable manifold. Moreover shortest paths in $M$ do not have finite length.

To prove this last property, we argue as follows: Without loss of generality, suppose the existence of a shortest path $\gamma(t)$ from $I_{\theta_0}$ to $I_{\theta_1}$ in $M$, the geodesic $G(\theta_0, \theta_1; M)$. Let $\vartheta(t)$ be the corresponding path in $\Theta$, where $\vartheta(0) = \theta_0$ and $\vartheta(1) = \theta_1$. Mark points $\theta(i)$ along the path $\vartheta$ by this device: put a closed $L^2$ ball of radius $\epsilon$ around $\theta_0 = \theta(0)$, and let $\theta(1)$ be the last point on $\vartheta$ inside that ball, and continue in this manner until a ball of radius $\epsilon$ around $\theta(n)$ includes the curve's endpoint $\theta_1$. Now define $N(\epsilon) = n$ as the number of 'Euclidean steps' of length $\epsilon$ necessary to traverse the path from $\theta_0$ to $\theta_1$ for a given $\epsilon$ (here we will use $N(\epsilon)$, but the proof follows for the case of $N(\epsilon) = 1$, when the last $L^2$-ball needs a radius $\epsilon$ to encompass $\theta_1$).

Now, given that $G(\theta_1, \theta_0; M)$ is a geodesic, note that:

$$G(\theta_1, \theta_0; M) = \sum_{i=1}^{N(\epsilon)} G(\theta(i), \theta(i-1); M) \geq \sum_{i=1}^{N(\epsilon)} \mu(\theta(i), \theta(i-1); M) = \sum_{i=1}^{N(\epsilon)} \eta(\epsilon),$$

where $\eta(\epsilon)$ is the function associated with the distance measurement for this manifold $\mu(\theta_0, \theta_1; M) = \eta(||\theta_0 - \theta_1||)$. Now, $N(\epsilon) \cdot \eta(\epsilon) \geq \mu(\theta_0, \theta_1; M)$ from the above inequalities. In addition, we know that in $\Theta$,

$$\sum_{i=1}^{N(\epsilon)} ||\theta(i) - \theta(i-1)||_2 \geq ||\theta_1 - \theta_0||_2 = L,$$

where $L$ is a constant for a given parameter pair. Therefore, we have that $N(\epsilon) \geq L \cdot \frac{1}{\epsilon}$. Combining the two inequalities, and substituting the result from (2.5), we see that

$$G(\theta_1, \theta_0; M) \geq \frac{L}{\epsilon} \cdot \eta(\epsilon) \asymp \frac{\text{constant}}{\epsilon^{\frac{1}{2}}} \quad \text{for} \quad 0 < \epsilon < 1. \quad (2.6)$$

As $\epsilon \to 0$, the lower bound on $G(\theta_1, \theta_0; M)$ grows without bound, and thus we establish the unboundedness of $G$.

In short, geodesic distance is typically not well-defined on an image articulation manifold when the generating image has edges. Since any reasonably interesting image has edges, this seems at first glance to be fatal to the notion of geodesic distance in the continuum case.

2.5 Renormalized Geodesic Distance

It is possible to rescue the notion of geodesic distance in cases where the underlying image has edges. This requires a process of regularization and renormalization.

Consider again the case where $I_0 = 1_{|x| \leq 1}$ is the indicator of a disk, and define the regularized object $I_0^h$ by $I_0^h = I_0 * \phi_h$ where $\phi$ is a smooth radial function of width $h$ and $*$ stands for convolution. Each image $I_0(x) = I_0^h(x - \theta)$ is smooth, and if we let $M_h$ denote the image manifold generated by translations of $I_0^h$, with $L^2$ metric $\mu(\theta_1, \theta_0; M_h) = ||T_{\theta_0} I_0^h - T_{\theta_1} I_0^h||_{L^2}$, we get that $M_h$ is a smooth manifold. Obviously, $M_h$ tends to $M$ as $h \to 0$, and if we consider a curve $\vartheta(t)$ in the common parameter space $\Theta$ of all the $M_h$ and of $M$, then the lengths of the induced paths on the respective manifolds $M_h$ must converge
to the lengths of the limit manifold $M$. It follows that, if we let $G_h$ be the corresponding geodesic distance, we have

$$\lim_{h \to 0} G_h(\theta_1, \theta_0; M_h) = \infty.$$  

In short, regularization makes the geodesic distance finite, but this effect disappears as $h \to 0$.

To get a finite limiting result, we make the observation that, if $\tau_0$ and $\tau_1$ are two fixed 'landmarks', then the ratio

$$\frac{G(I_{\tau_0}^h, I_{\tau_1}^h; M^h)}{G(I_{\tau_0}, I_{\tau_1}; M^h)}$$

need not diverge as $h \to 0$. In effect, if we renormalize the distance so that $\tau_0$ and $\tau_1$ stay at unit distance on $M$ for all $h > 0$, we get a finite limiting ratio.

In our running example, let $I_0$ be the indicator of the unit disk in $\mathbb{R}^2$ and let $I_\theta(x) = I(x - \theta)$ be translation by $\theta$. For regularization, let $\phi_h$ denote the standard bivariate Gaussian density with covariance matrix $h^2 \cdot Id$. Let the path $\vartheta(t)$ be the line segment between $\tau_0 = (0,0)$ and $\tau_1 = (1,0)$. The length of the corresponding path (the 'landmark distance'), $\gamma^h_t$, on the smooth manifold $M_h$ generated by $\gamma^h_t(t) = \phi_h \ast I_{\vartheta(t)}$ is given by

$$L(\gamma^h_t) = \int_0^1 \| \frac{\partial}{\partial t} \gamma^h(t) \|_{L^2(\mathbb{R}^2)} dt = \int_0^1 \| \frac{\partial}{\partial t} I_{\vartheta(t)}^h \|_{L^2(\mathbb{R}^2)} dt.$$  

(2.7)

Now note that because of the translation invariance of the $L^2$ norm, for all $t$, we can write, for some $A_h$,

$$\| \frac{\partial}{\partial t} I_{\vartheta(t)}^h \|_{L^2(\mathbb{R}^2)} = \| \frac{\partial}{\partial t} I_{\vartheta(t)}^h \|_{L^2(\mathbb{R}^2)} |_{t=0} = A_h.$$
And therefore, \( L(\gamma_r^h) = A_n \). Now for any other chosen parameter points \( \theta_0 \) and \( \theta_1 \) in \( \mathbb{R}^2 \), connected by a line segment, consider the corresponding path \( \gamma_{01}^h \) in \( M_h \) between \( I_{\theta_0}^h \) and \( I_{\theta_1}^h \). We can perform the same calculation and see that

\[
L(\gamma_{01}^h) = A_n \|\theta_0 - \theta_1\|.
\]

Moreover \( \gamma_{01}^h \) is actually, as one might guess, the geodesic in \( M_h \) between \( I_{\theta_0}^h \) and \( I_{\theta_1}^h \). Hence

\[
\frac{G(I_{\theta_0}^h, I_{\theta_1}^h; M^h)}{L(\gamma_{01}^h; M^h)} = \frac{A_n \|\theta_0 - \theta_1\|}{A_n} = \|\theta_0 - \theta_1\|.
\]

In short, the dependence on \( h \) vanishes, and the result is a stable, well-defined (and interpretable!) quantity.

### 2.6 The Abstract Manifold \( \mathcal{M} \)

Inspired by the success of renormalization in the special case above, we propose the following shift of viewpoint. We define an abstract manifold \( \mathcal{M} \) which consists of a pair \( (\Theta, \delta) \), where \( \Theta \) is the parameter family underlying \( M \), and \( \delta \) is a notion of distance on \( \Theta \times \Theta \) which is derived from geodesic distance on the smoothed manifolds by the above limiting procedure. In this subsection, we set up the appropriate terminology.

**Definition 2.2** To define renormalized length of a path in the abstract manifold \( \mathcal{M} \), let \( \vartheta(t) \) denote a path in the parameter space \( \Theta \). Construct a family \( M_h \) of smooth manifolds, where \( h \) is the smoothing parameter, converging in in \( L^2 \) distance to \( M \) as \( h \to 0 \). Let \( \gamma^h \) denote the corresponding path in \( M_h \), and let \( \gamma^h_{\vartheta} \) be the path in \( M_h \) induced from the line segment \( \vartheta(t) \) in \( \Theta \). Then define the renormalized length as

\[
\lambda(\vartheta; \mathcal{M}) = \lim_{h \to 0} \frac{L(\gamma^h; M^h)}{L(\gamma^h_{\vartheta}; M^h)}.
\]

How does this work in our running example of translating the indicator of a disk? Fix landmarks \( \tau_0 = (0, 0) \) and \( \tau_1 = (1, 0) \); consider for our path the line segment \( \vartheta(t) = \theta_0 + t(\theta_1 - \theta_0) \) which runs from \( \theta_0 \) to \( \theta_1 \) according to a straight line. We find that

\[
\lambda(\vartheta; \mathcal{M}) = \|\theta_1 - \theta_0\|.
\]
Definition 2.3 By renormalized geodesic distance on $\mathcal{M}$ we mean: fix two landmarks $\tau_0$ and $\tau_1$, construct a sequence $M_h$ of smooth manifolds, where $h$ is the smoothing parameter, and which converge in $L^2$ distance to $\mathcal{M}$ as $h \to 0$. Then set

$$\delta(\theta_0, \theta_1; \mathcal{M}) = \min \{ \lambda(\phi(\cdot); M) : \phi(0) = \theta_0, \phi(1) = \theta_1 \}.$$ 

Again, in our running example of translating the indicator of a disk, the length of every line segment in the abstract manifold is the same as its Euclidean length. It follows that the geodesics for $\delta$ will just be the Euclidean geodesics. We also propose the following interpretation of $\delta$:

Definition 2.4 Suppose that $\delta$ obeys $\delta(\theta_1, \theta_2) = c\| \theta_1 - \theta_2 \|_2$, for all $\theta_1, \theta_2 \in \Theta$. Then we say that renormalized geodesic distance recovers the natural parametrization (up to a rigid motion).

Another way to put things: when the definition applies, Continuum ISOMAP works perfectly (up to a rigid motion of the parameter space).

We make two observations concerning the definition:

- The definition works in simple cases. Consider our running example of translating the indicator function of a disk in the plane. Set $\tau_0 = 0$ and $\tau_1 = (1, 0)$. For regularization, let $\phi$ denote the standard bivariate Gaussian density. Then, our calculation above shows that indeed

$$\delta(\theta_1, \theta_0; \mathcal{M}) = \| \theta_1 - \theta_0 \|_2.$$ 

In short, the definition leads to a computable quantity and the answer it gives makes sense. Incidentally, in this case we see that when the image manifold is generated by translation of a disk image, Continuum ISOMAP works correctly.

- The definition really matches the original purpose of our study: applications of ISOMAP to images. An object in a scene is rendered into a digital image by pixelization, and this process provides a limited-resolution image much as $\phi_h \ast I_0$ is a limited-resolution image. In some sense applying ISOMAP to sufficiently fine but limited-resolution imagery gives geodesic distances heuristically of the form $G(\cdot, \cdot; M_h)$. But those distances obey

$$G(\theta_1, \theta_0; M_h) \sim \delta(\theta_1, \theta_0; \mathcal{M})h^{-\frac{1}{2}}, \quad h \to 0.$$ 

Obviously, the factor $h^{-1/2}$ is completely transparent to all operations involved in metric multidimensional scaling, and will not change the reconstruction of pointsets except for a homothetic dilation that can easily be 'standardized away' by fixing the distance between landmark points. This suggests that what really matters for ISOMAP is the leading coefficient $\delta(\cdot)$.

There are several things that need to be checked about our definition of $\delta$: first, the choice of smoothing is not specified, and the definition might conceivably not be invariant to choice of regularization, so that different regularizations might conceivably lead to different answers; second, this might not define an actual distance; third, the manifold might not be continuous or smooth according to this distance. While there are satisfactory general answers to these questions, we prefer simply to say for now that in specific cases we have studied, these issues do not arise.
2.7 Calculation of $\lambda$, $\delta$

In order to show that renormalized length makes sense, we consider a somewhat general family of images and give a formula for the length of paths through the image manifold in such a family of articulated images. In particular, we will not assume a particular form (such as translation or rotation) for the articulations; merely that they involve smooth transformations of the object in question.

In this model, we consider black-and-white images where the black region $B_\theta$ is a kind of ‘blob’ – a compact set in the plane, with simple closed curve for boundary. The images themselves are $I_\theta(x) = 1_{B_\theta}(x)$, so that white is represented by 0 and black by 1. The boundary $\partial B_\theta$ can be parametrized by a curve $\beta(b; \theta)$, where the index $b \in [0, 1]$ and, because the beginning and ending points of $\beta$ are the same, $\beta(0; \theta) = \beta(1; \theta)$.

In defining the regularization procedure, we will use the 2-dimensional standard Gaussian kernel $\phi(x_1, x_2) = \exp\{-x_1^2 + x_2^2\}/2$, and, with $\phi_h(x) = h^{-2}\phi(h^{-1}x)$, define $I_\theta^h = \phi_h * I_\theta$.

We will make the quantitative assumption that $\beta$ is a $C^2$ function of both $\theta$ and $b$ jointly; hence, in particular, the radius of curvature of $\partial B_\theta$ is uniformly bounded over every compact subset of $\theta$. It follows from this that there is a tubular neighborhood of $\partial B_\theta$ with coordinates $(a, b)$ where any $x$ in that neighborhood has for its $b$-coordinate the closest point on $\partial B_\theta$ and for its $a$-coordinate the signed distance to $\partial B_\theta$.

The set $B_\theta$ has an outward-pointing normal at each point of the boundary, labelled $n(b; \theta)$. Each boundary point also undergoes a motion with changing $\theta$, so we use the notation $\nu_i(b; \theta)$ for the ‘motion vector’: the rate of change of a specific boundary points with change in specific parameter vector $\theta$. This is defined by

$$\nu_i(b; \theta) = \frac{\partial}{\partial \theta_i} \beta(b; \theta).$$

Finally, we need notation for the speed function of the parametrization of the boundary curve; set

$$s(b; \theta) = \|\frac{d}{db} \beta(b; \theta)\|.$$  

We have the following results.
Theorem 2.2 Let \((\vartheta(t) : t \in [0, 1])\) be a smooth curve in parameter space and, using the notation above, set
\[
v(t, b) = \sum_i \frac{d\theta_i}{dt} \nu_i(b; \theta).
\]
Then the length of the curve in \(\mathcal{M}\) is
\[
\lambda(\theta) = C_\tau \int_0^1 \left[ \int_0^1 \langle n(b), v(t, b) \rangle^2 \frac{db}{s(b; \theta)} \right]^{1/2} dt.
\]
(2.8)
Here \(C_\tau\) depends on the choice of landmark points, \(\tau_1\) and \(\tau_2\), used in defining the renormalized length.

Theorem 2.3 Let \((\vartheta(t) : t \in [0, 1])\) be a smooth curve in parameter space and, using the notation above, set
\[
g_{ij}(\theta) = \int_0^1 \langle n(b), \nu_i(b; \theta) \rangle \cdot \langle n(b), \nu_j(b; \theta) \rangle \frac{db}{s(b; \theta)}
\]
Then \(g_{ij}\) defines a Riemannian structure on \(\Theta\) and the renormalized length of the induced curve in \(\mathcal{M}\) is the same as the length of the curve in \(\Theta\) with respect to this Riemannian structure:
\[
\lambda(\theta) = C_\tau \int_0^1 \sqrt{\sum_{ij} g_{ij}(\theta) \frac{d\theta_i}{dt} \frac{d\theta_j}{dt}} dt.
\]
(2.9)
Here \(C\) depends on the choice of landmark points used in defining the renormalized length.

The proofs for these results are given in [3]. To see how the results can be applied, consider again our running example of translations of the disk. In this case,
\[
\beta(b; \theta) = \theta + (\cos(2\pi b), \sin(2\pi b)),
\]
while
\[
n(b; \theta) = (\cos(2\pi b), \sin(2\pi b)),
\]
and \(s(b; \theta) = 2\pi\), while
\[
\nu_i(b; \theta) = \begin{cases} (1, 0) & i = 1, \\ (0, 1) & i = 2. \end{cases}
\]
Again, if we translate from \(\theta_0\) to \(\theta_1\), the linear path \(\vartheta(t) = \theta_0 + t(\theta_1 - \theta_0)\), and so
\[
v(t, b) = (\theta_1 - \theta_0).
\]
Plugging all these formulas into the inner integral in (2.8) gives
\[
\int_0^1 \langle n(b), v(t, b) \rangle^2 \frac{db}{s(b; \theta)} = ||\vartheta_1 - \vartheta_0||^2 \int_0^1 \cos(2\pi b)^2 \frac{db}{2\pi} = \frac{1}{2} \cdot ||\vartheta_1 - \vartheta_0||^2.
\]
Hence we have, as expected:
\[
\lambda(\theta) = C_\tau \cdot ||\vartheta_1 - \vartheta_0||.
\]
As far as the Riemannian structure goes, we have
\[
g_{11}(\theta) = \int_0^1 \cos(2\pi b)^2 \frac{db}{2\pi} = 1
\]
and similarly $g_{12} = g_{21} = 0$, while $g_{22} = 1$. Hence the Riemannian structure is exactly the Euclidean structure and we see that the geodesic distance exactly recovers the natural structure of the parameter space.

It was perhaps not even necessary to do any computations in this case; one could see what the formulas would supply without calculation, just from understanding the terms going into the equations above. For example, formula (2.8) involves the integration, around the boundary, of the inner product between the motion vector $v(b; t)$ and the normal $n(b; \theta)$. But the motion vector is independent of $t$ and $b$, and the normal is independent of $\theta$, so clearly, the result is invariant under rotation of the coordinate system, and of the base vector $\theta_0$. Applying these observations to the formula (2.9) clearly gives that $(g_{ij}(\theta))$ does not depend on $\theta$, and is isotropic, so the Riemannian structure is therefore Euclidean.

2.8 Inferring Euclidean Structure

The preceding development gives us the picture of an abstract manifold $\mathcal{M}$ which is the parameter space $\Theta$ equipped with a distance $\delta(\cdot)$ (See Figure 2). They also give us the picture of the parameter space $\theta$ equipped with a Riemannian structure, $(g_{ij}(\theta))$. In particular, we will see several examples below where $g_{ij}$ is constant and proportional to the identity. This would seem to indicate that the abstract manifold is in fact isometric to a subset of Euclidean space, and that therefore ISOMAP works. In order to make such an inference from local characteristics to global ones, however, we need an extra element. In order to be globally Euclidean, the local structure must be Euclidean and the space $\Theta$ must be convex. If it is not convex, then shortest paths in $\Theta$ will not be dictated by the infinitesimal metric structure alone, but will in some cases be dictated by the nonconvexity of the space, eg. to avoid holes. Therefore, we will sometimes emphasize the issue of convexity.

In some cases, the demonstration that ISOMAP works can be done without computing either $\delta$ or $g$. In fact, all that is necessary is that $\Theta$ be convex, and that, for every pair of points $\theta_i \in \Theta$, the linear path in parameter space $\theta(t) = \theta_0 + t(\theta_1 - \theta_0)$ has a length proportional to the Euclidean distance $\|\theta_1 - \theta_0\|$, where the constant of proportionality does not depend on $\theta_i$. We will frequently check this condition, which implies both that $\lambda$ is proportional to Euclidean arclength and that $\delta$ is proportional to Euclidean distance.

3 Simple Examples

We now give a few simple examples of the preceding theory.

3.1 Pivoting an Object

Consider an object $B_0$ with smooth boundary, and suppose that 0 is a point on the boundary. Now consider a family of articulations obtained by pivoting the object around that fixed point, by making a planar rotation. More formally, let $R_\theta$ denote the rotation matrix

$$R_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix},$$

let $I_0(x) = 1_{B_0}(x)$, and let $I_\theta = I_0(R_\theta x)$.

Theorem 3.1 In the Pivot model, Continuum ISOMAP works perfectly; for $|\theta_0 - \theta_1| < \pi$ renormalized geodesic distance is proportional to the Euclidean distance in parameter space.
Proof: One has merely to notice that \( g_{11}(\theta) \) is independent of \( \theta \). This is easy to see by observing that the integrand in \( g_{11} \),

\[
\langle n(b), \nu_1(b; \theta) \rangle
\]

is independent of \( \theta \). Indeed, for any value of \( \theta \), \( n(b; \theta) \) points normal to \( \partial B \), while \( \nu_1(b; \theta) \) points normal to the line segment joining 0 to \( \beta(b; \theta) \). The angle between these two vectors is the same independent of \( \theta \) (See Figure 4).

3.2 Translating a 4-fold symmetric object

We consider a generalization of the problem of translating a disk, considering now a broader class of symmetric objects.

Definition 3.1 We will say that an object \( B_0 \) has 4-fold symmetry if it is invariant under rotations about the origin by 90, 180, and 270 degrees.

This class of objects includes \( \ell^p \) balls for \( 0 < p < \infty \):

\[
\ell^p = \{ x : |x_1|^p + |x_2|^p \leq 1 \},
\]

although for the cases \( p = 1, \infty \) the boundaries are not smooth.

Theorem 3.2 For translations of 4-fold symmetric figures with smooth boundary, Continuum ISOMAP works perfectly; the renormalized geodesic distance is proportional to the Euclidean distance in parameter space.

As \( \ell^p \) balls for \( p = 2 \) are just disks, this contains our earlier result about translating disks as a special case. However, this result is much more general than just to \( \ell^p \) balls; it works for a wide variety of nonconvex figures; see Figure 5.
Figure 5: Translation of a 4-fold symmetric object

**Proof.** Consider the integrand of (2.8):

\[ \int_0^1 \langle n(b), v(t, b) \rangle^2 \frac{db}{s(b; \theta)} \]

Now for a translation family, and a path \( \vartheta(t) = \theta_0 + t(\theta_1 - \theta_0) \), we have that \( v(t, b) = \theta_1 - \theta_0 \); this is constant, \( v \), say, independent of \( t \) and \( b \). Hence the integrand becomes:

\[ \int_0^1 \langle n(b), v \rangle^2 \frac{db}{s(b; \theta)} \]

Now for a 4-fold symmetric object, for every \( b \in [0, 1/4) \)

- the speed function \( s(b; 0) \) is the same at \( b, b + 1/4, b + 1/2, b + 3/4 \);
- we have antisymmetric pairs of normals at \( (b, b + 1/2) \), and \( (b + 1/4, b + 3/4) \), in the sense that the respective normals \( n(\cdot; 0) \) point in opposite directions at each member of the pair; and
- we have orthogonal pairs of normals at \( (b, b + 1/4) \), and \( (b + 1/2, b + 3/4) \) in the sense that associated pair of normals \( n(\cdot; 0) \) are mutually orthogonal.

Hence, using the equality of the speed function, we may write

\[ \int_0^1 \langle n(b), v \rangle^2 \frac{db}{s(b; \theta)} = \int_0^{1/4} \sum_{k=0}^3 \langle n(b + k/4), v \rangle^2 \frac{db}{s(b; \theta)} \]

On the other hand, letting \( n^\perp(b) \) be perpendicular to \( n(b) \), this is

\[ \sum_{k=0}^3 \langle n(b + k/4), v \rangle^2 = 2(\langle n(b), v \rangle^2 + \langle n^\perp(b), v \rangle^2) = 2\|v\|^2. \]
Hence,
\[
\int_0^1 \langle n(b), v \rangle^2 \frac{db}{s(b; \theta)} = \frac{\|v\|^2}{2} \int_0^1 \frac{db}{s(b; \theta)} = \frac{\|v\|^2}{2} L(\partial B_0)
\]
In short, recalling the definition of \( v = \theta_1 - \theta_0 \):
\[
\lambda(\theta) = \sqrt{L(\partial B_0)/2} \cdot \|\theta_1 - \theta_0\|
\]
This implies the proportionality of geodesic and Euclidean distance. \(\square\)

3.3 Extensions to Piecewise Smooth Boundaries

The results developed so far can be generalized to the case of piecewise smooth boundary curves.

On the one hand, as a limiting case of Theorem 3.2, consider \( p = \infty \). This is the unit square in the plane. Extrapolating the above result from \( p < \infty \) to \( p = \infty \) suggests that it might also be true that Continuum ISOMAP works perfectly for translations of the square, and indeed this is the case. To verify that, we need to know that (2.8) still works if \( \partial B_0 \) is merely piecewise smooth. The demonstration of this appears in [3]. It is clear that the same extension applies not only to squares, but to 4k-gons, for any \( k \) (squares are the case \( k = 1 \)). It is also clear that the same extension applies to the case of objects with piecewise smooth boundaries and 4-fold symmetry. In all cases, Continuum ISOMAP works perfectly.

On the other hand, as a limiting case of Theorem 3.1, consider a ‘pie slice’ \( B_0 \) outlined by two radial line segments, \( \{ (r \cos(\omega), r \sin(\omega)) : 0 \leq r < R \} \) and \( \{ (r \cos(-\omega), r \sin(-\omega)) : 0 \leq r < R \} \); and an arc connecting them, \( \{ (R \cos(\omega'), R \sin(\omega')) : -\omega \leq \omega' \leq \omega \} \).

Articulate the structure by a pivoting it according to \( R\theta \) as in Section 3.1 above. Hence, with \( I_0 = 1_{B_0} \), set \( I_{\theta}(x) = I_0(R\theta x) \).

This is a limiting case of Theorem 3.1 because the boundary is piecewise smooth rather than smooth. However, because (2.8) extends to the case of objects with piecewise smooth boundaries, the conclusion of Theorem 3.1 carries over to this case as well.

4 Empirical Results

In this section, we test our theoretical predictions about recovery of the parameter space for several example models against empirical results using the ISOMAP implementation.

4.1 Rigid translation of a single disk

Consider our standard example of an image \( I_0(x) = 1_{\{|x| \leq 1\}} \) that is the indicator of a unit disk. Translate the disk by a vector \( v \) in \( \mathbb{R}^2 \), and observe that we can then analytically evaluate (2.8):
\[
= \|v\|_2 \times \left( \int_{\mathbb{R}^2} \langle n(b), v(b) \rangle^2 \, db \right)^{1/3}
\]
\[
= \|v\|_2 \times \left( \int_0^{2\pi} \langle n(\phi), v(\phi) \rangle^2 \left\| \frac{d\beta}{d\phi} \right\|_2 \, d\phi \right)^{1/3}
\]
\[
= \|v\|_2 \times \pi^{1/3},
\]
which is exactly proportional to the Euclidean distance under the articulation \( v \). Therefore, we expect that an empirical trial (allowing for the vagaries of sampling) would show
Figure 6: Left panel: Original Disk Centers, colored by polar angle. Right panel: 2-dimensional embedding using ISOMAP, colored by original polar angle (optimally rotated and scaled).

essentially 'perfect' recovery of the original parameter space. Our test uses the publicly-available ISOMAP code ([9]), running with the nearest-neighbor parameter $k = 7$, and a set of 100 black-on-white images at 64-by-64 pixels. The translation of the prototype was performed by randomly selecting $x,y$-coordinates of the disk centers, restricting the centers so that the disk stayed inside the borders of the image. For the following examples, we will mimic this format, generally restricting the objects in the images to be at least 16 pixels wide in order to prevent pixelization from being a dominant embedding dimension. The results of the unit disk experiment are seen in Figure 6. ISOMAP almost perfectly recovers the original spacing and relative positioning of the disk centers. In fact, we note that the only noticeable deviations tend to occur where the sampling was particularly sparse, such as in the upper left corner of the image. The actual ISOMAP 2-dimensional embedding was optimally rotated and scaled using a Procrustes rotation [2] to match it to the original parameter space for graphical presentation. This accords with the theory that Continuum ISOMAP recovers the original parameter space up to a rigid motion.

4.2 Pivoting of a Single Disk

Consider pivoting a simple object, in particular, rotating a disk around a fixed boundary point. For simplicity, we choose the fixed point to lie at the origin of our coordinate system, and choose the disk to have radius 1. That is, the center of the disk lies on a unit circle around the origin. We choose the center of the disk as the candidate for our natural parametrization. Our theory states that the length of the induced curve is given by (2.8) after some calculation as

$$\lambda(\theta) \propto \pi^{\frac{1}{2}} \times |d\theta|$$

where the articulation is a pivoting by an angle $d\theta$. As in the previous experiment, we create a set of 100 images of a black-on-white disk at 64-by-64 pixels, and articulate the disk under pivoting operations of a randomly selected angle $\theta$.

The resulting experiment, optimally rotated and scaled, shows an almost exact match to the original parameter space in Figure 7. We plot the disk centers in two-dimensions for comparison to the image field, but the relative ordering is equally good in one dimension.
4.3 Translations of a Flower-shape

We now test our predictions about translations of a general four-fold symmetric object. In our experiment, we generated 100 images for a four-lobed flower, where the prototype is invariant under rotations by 90, 180, and 270 degrees. For a sample ‘flower’, see the third panel of Figure 5. The images are again 64-by-64 and we chose 100 examples. The articulation was formed by randomly selecting centers of the object in the image frame (restricted not to overlap the boundary).

The boundary of the ‘flower’ is formed by segments of four equally-sized disks. Therefore, we have a piecewise smooth and 4-fold symmetric boundary, and under the conditions give, we would expect Theorem 3.2 to hold. The theory is borne out by the empirical results. In Figure 8, the recovery of the original center locations is equally good as those in the single disk example. Similar empirical results hold for tests of other 4-fold symmetric polygons: squares, octagons, and so forth.
5 Variations

5.1 Horizon Articulation

We consider first a variation in both the image index set and the articulation model. The image $I(u, v)$ has indices $(u, v)$ spanning $0 \leq u \leq 1$ and $-\infty < v < \infty$, so forming a kind of vertical 'strip'. And the images of interest to us are black and white images with the boundary between the two colors being a horizon $v = \psi(u)$ -- hence

$$I(u, v; \psi) = 1_{\{v \leq \psi(u)\}}.$$

We are interested in parametric families $I_\theta$ of articulated images where the horizon is generated by a linear combination of basis elements $\psi_j$ (an example of an articulation appears in Figure 9)

$$\psi(u; \theta) = \sum_i \theta_i \psi_i(u).$$

Hence $\theta$ is a vector of expansion coefficients governing the shape of the horizon. We assume the basis functions are orthogonal for $L^2[0, 1]$:

$$\int_0^1 \psi_j(u)\psi_{j'}(u)du = 1_{\{j = j'\}}.$$

examples include orthonormal wavelets, and orthonormal sinusoids. We will assume that the basis functions are smooth: at least $C^2$.

We may think of the family $\{I_\theta\}$ of images defined by $I_\theta(u, v) = I(u, v; \psi(u; \theta))$ as generated by a kind of shearing. If we let $I_0(u, v)$ be the 'null' horizon image $I_0(u, v) = 1_{\{v \leq 0\}}$, and $S_\psi$ as the shearing operator on functions $f(u, v)$ according to

$$(S_\psi f)(u, v) = f(u, v - \psi(u)),$$

then of course $I_\theta = S_\psi(u, \theta)I_0$. Hence this family of images is the result of a 1-1 measure-preserving transformation of the image domain.
In defining the metric $\delta$ for this setting, we need to define image regularization. The natural choice in this case is to smooth ‘vertically’. So let $\varphi$ denote the 1-dimensional standard normal density, and the smoothing operation

$$(\varphi_h *_v f)(u, v) = \int_{-\infty}^{\infty} \varphi_h(w)f(u, v - w)dw;$$

define then the smoothed image $I^h$ by

$$I^h_\theta = \varphi_h *_v I_\theta.$$ 

For fixed $h > 0$, the collection of all $I^h_\theta$ for all $\theta \in \Theta$ defines a smooth image manifold $M_h$, and so we can define renormalized length in a fashion similar to (2.8). In this setting, the terms $\beta, n$, are no longer relevant; we define

$$v(t, u) = \sum_j \frac{d\theta_j}{dt}\psi_j(u);$$

and the formula for renormalized length is

$$\lambda(\theta) = C_\tau \cdot \int_0^1 \left[ \int_0^1 v(t, u)^2 du \right]^{1/2} dt. \quad (5.10)$$

When we do this, we find that, if we select a linear path in parameter space,

$$\theta(t) = \theta_0 + t(\theta_1 - \theta_0),$$

then

$$\lambda(\theta) = C \cdot \|\theta_1 - \theta_0\|.$$ 

Since the length of every line segment is proportional to the Euclidean distance, it follows that $\delta$ itself is proportional to Euclidean distance.

Also, the appropriate formula for the Riemannian metric is simply

$$g_{ij}(\theta) = C^2 \cdot \int_0^1 \phi_i(u)\phi_j(u) du = C^2 \cdot 1_{\{i=j\}},$$

which shows the same thing another way.

**Theorem 5.1** In the Horizon Articulation model where the parameter space $\Theta$ is convex, Continuous ISOMAP works perfectly.

An interesting feature here is that there is no limit to the dimensionality of $\theta$; in particular, we can consider quite complex articulations of a horizon; in each case, the natural parametrization will be recovered up to a rigid motion of parameter space.

### 5.2 Star-shaped Blobs, with Radial Articulation

We now consider the class of star-shaped ‘blobs’ in the plane. These ideal images are black-and-white where the black component $B$ is a set containing the origin, such that the line segment from $0$ to $\partial B$ lies completely in $B$. In our articulation model, $B_0$ will be the disk of radius one, and each $B_\theta$ will be a starshaped blob (for an illustration, see Figure 9).
In this example, however, to arrive at a satisfactory conclusion, we have to manipulate the ISOMAP problem using a weighted \(L^2\) distance, \(L^2(w(\|x\|)dx)\), and a special parametrization. To this end, we consider the image plane in polar coordinates \((r, \omega)\), \(r \geq 0\), \(\omega \in [0, 2\pi)\) and note that polar-domain objects which are star-shaped 'blobs' can be defined by

\[
J(r, \omega; \rho) = 1_{\{r \leq \rho(\omega)\}},
\]

(5.11)

where \(\rho(\omega) > 0\) is the distance from 0 to the boundary of the blob in direction \(\omega\). That is, if to each polar-domain object \(J\) defined by (5.11), we associated the cartesian-domain image \(I\) via

\[
I(r \cos(\omega), r \sin(\omega)) = J(r, \omega)
\]

then each such \(I\) is indeed a 'blob' image.

We will parametrize the boundary function \(\rho(\omega)\) by its expansion coefficients in an orthonormal basis. Let \((\psi_j)\) be a sequence of functions orthonormal for \(L^2[0, 2\pi)\), and, for a coefficient sequence \(\theta = (\theta_j)\), let

\[
\rho(\omega; \theta) = \exp\left\{\sum_j \theta_j \psi_j(\omega)\right\}.
\]

We assume that the basis functions are all smooth. We note that for every \(\theta\), \(\rho(\omega; \theta) > 0\), and so this defines a \(J\) which gives rise to an image \(I\) portraying a well-defined star-shaped object. We will compare polar-domain objects according to the metric

\[
\nu(\theta_0, \theta_1) = \|J(\cdot; \rho(\cdot; \theta_0)) - J(\cdot; \rho(\cdot; \theta_1))\|_{L^2(r^{-1}drd\theta)}.
\]

Note that this metric can also be written

\[
\nu(\theta_0, \theta_1) = \|I_{\theta_1} - I_{\theta_0}\|_{L^2([0,2\pi) \times [0,\infty))}
\]

so that we are working, as advertised earlier, in a weighted \(L^2\) space, with weight \(w(r) = r^{-2}\).

We now consider the manifold of such images \(I_\theta\) equipped with distance \(\nu\) and seek to define the renormalized geodesic distance. We smooth the images according to convolution in \(\log r\). That is, with \(\varphi_h\) the 1-D Gaussian as in the Horizon articulation case, we set

\[
J^h(r, \omega) = \int \varphi_h(u - \log(r))J(e^u, \omega)du.
\]

It follows that \(J^h\) is smooth, and that \(J^h \to J\) in \(L^2(r^{-1}drd\omega)\) metric as \(h \to 0\). We also note that from the definition of \(J\)

\[
J^h(r, \omega) = \Phi_h(\log(\rho(\omega)/r))
\]

where \(\Phi_h\) is the 1-dimensional normal distribution function. Now consider a path \(\vartheta(t)\) in \(\Theta\). Then

\[
\frac{d}{dt}J^h(r, \omega; \rho(\cdot; \vartheta(t))) = \Phi_h(\log \rho(\omega) - \log(r)) \cdot (-1) \cdot \frac{d\rho(\omega; \vartheta(t))}{dt}.
\]

Now note that

\[
\int_0^\infty (\phi_h(\log(r) - \log(\rho(\omega; \theta))))^2 r^{-1}dr = \int_0^\infty (\phi_h(u - \log(\rho(\omega; \theta))))^2du = \int_0^\infty \phi_h(u)^2du = C \cdot h^{-1}.
\]
Also, by orthogonality of the $\psi_j$,

$$
\int_0^{2\pi} \left( \sum_j \frac{d\theta_j}{dt} \psi_j(\omega) \right)^2 d\omega = \sum_j \left( \frac{d\theta_j}{dt} \right)^2
$$

Combining the last two displays, it follows that the speed of this curve is

$$
\| \frac{d}{dt} \mathbf{r}(r, \omega) \|_{L^2(r^{-1} dr d\omega)} = C \cdot h^{-1} \cdot \sum_j \left( \frac{d\theta_j}{dt} \right)^2
$$

Now in particular, if the path $\vartheta(t) = \vartheta_0 + t(\vartheta_1 - \vartheta_0)$ is a simple line segment in $\Theta$, then

$$
\sum_j \left( \frac{d\theta_j}{dt} \right)^2 = \| \vartheta_1 - \vartheta_0 \|^2.
$$

This implies that for all line segments in $\Theta$, renormalized geodesic distance is proportional to Euclidean distance. To conclude,

**Theorem 5.2** In the setting of collections of starshaped blobs with weighted $L^2(w(\|x\|)dx)$ distance, ($w(r) = r^{-2}$), and a convex parameter space $\Theta$, continuous ISOMAP works perfectly.

## 6 Composite Articulations

In this section, we discuss several restricted cases of multiple-object articulations. The restrictions are employed to demonstrate some of the successes of the Continuum ISOMAP algorithm. For detailed discussion of the necessary conditions and proofs of the results, see [3].

### 6.1 Bunny Ears

Consider a two-object ‘Bunny Ears’ composed of two wedges on a disk: Let $\theta = (\theta^1, \theta^2)$ denote a pair of angles, and let $B_0$ be the unit disk with two ‘ears’ attached, with the inclination of the ‘ears’ controlled by the components of $\theta$. Each ear is an annular wedge with angular opening $2 \omega$. Such a basic wedge is defined by boundary segments $\{(r \cos (\theta + \omega), r \sin (\theta + \omega)) : 1 \leq r < 2\}$ and $\{(r \cos (\theta - \omega), r \sin (\theta - \omega)) : 1 \leq r < 2\}$; and two arcs connecting them, along $r = 1$ and $r = 2$. The first ear is a wedge centered at $\theta = \theta^1$; the second ear is a wedge centered at $\theta = \theta^2$. The parameter space $\Theta$ is chosen so the ears don’t overlap:

$$
\Theta = \{ (\theta^1, \theta^2) : 2 \omega < \theta^1 - \theta^2 < 2\pi - 2 \omega \}.
$$

We can extend this to the Hands model, in which the unit disk has 5 ‘fingers’ attached, and each ‘fingert’ is again an annular wedge at specific angle $\theta^i$, $i = 1, \ldots, 5$.

$$
\Theta = \{ (\theta^1, \theta^2, \ldots, \theta^5) : 2 \omega < \theta^i - \theta^{i+1} < 2\pi - 2 \omega \}.
$$

As long as the condition that the mobile wedges cannot overlap is enforced, the demonstration can be given for any number of ‘fingers’.

**Theorem 6.1** Continuum ISOMAP works perfectly for the ‘Bunny Ears’ and ‘Hand’ Models.
6.2 Translating Several Ordered Nonoverlapping Disks

Consider now continuum images containing \( n \) disks, each of radius one, with centers \( \theta_i, i = 1, \ldots, n \). With \( I_0 \) the indicator of the unit disk, and \( \theta = (\theta_1, \ldots, \theta_n) \), put

\[
I_\theta(x) = \sum_{i=1}^{n} I_0(x - \theta_i).
\]

In this section we constrain

\[
\|\theta_i - \theta_j\| > 1
\]

so that no two terms in the sum overlap, and the image remains of the 'black-and-white' type in previous sections. We will go further than merely constrain the disks to be nonoverlapping; we do so in several ways, each defining a different parameter space \( \Theta \).

- **Single File.** Let \( \Theta_1 \) denote the set of \( \theta \) where the disk centers lie along the \( x \)-axis, so that

\[
\theta_{i,2} = 0 \quad i = 1, \ldots, n;
\]

(here \( \theta_{i,j} \) denotes the \( j \)-coordinate of \( \theta_i \)); moreover the centers are ordered along the \( x \)-axis:

\[
\theta_{1,1} < \theta_{2,1} < \cdots < \theta_{n,1}.
\]

- **Separated Columns.** Let \( \Theta_2 = \Theta_2((a_i)) \) denote the set of \( \theta \) arising where the disk centers can be distributed in the 2-dimensional plane relatively broadly, but are constrained to lie in zones defined by regions of the \( x \)-axis. We fix cutpoints \( a_0, a_1, \ldots, a_n \), with \( a_i > a_{i-1} + 2 \) and demand that

\[
a_{i-1} + 1 < \theta_{i,1} < a_i - 1 \quad i = 1, \ldots, n.
\]
• Northeast/Southwest ordering. We let \( \Theta_3 \) denote the parameter space where the disks are arranged so that the \( i \)-th one has all earlier disks lying to the southwest and all later disks lying in the northeast quadrant.

Each such constraint corresponds to a convex family \( \Theta \). Now we consider independent translations of the family of disks that maintain the assumed constraint, i.e. translations that preserve membership in \( \Theta \).

In this setting, we can adapt arguments from the case of a single disk to see that, if \( \theta_{(0)} = (\theta_0^1, \ldots, \theta_0^n) \) is a collection of disk centers and \( \theta_{(1)} = (\theta_1^1, \ldots, \theta_1^n) \) is another collection of disk centers, then for \( \vartheta(t) = \theta_{(0)} + t(\theta_{(1)} - \theta_{(0)}) \), we have the arclength

\[
\lambda(\vartheta; \mathcal{M}) = C_\tau \cdot \sqrt{\sum_{i=1}^n \left\| \theta_i^t - \theta_i^0 \right\|^2}
\]

and a Riemannian structure which is in fact Euclidean:

\[
g(\vartheta) = C_\tau \cdot Id, \quad \forall \vartheta \in \Theta.
\]

Thus, in this setting, the shortest distance between pairs of images indeed goes by morphing the centers along line segments.

**Theorem 6.2** Continuous ISOMAP works for the Ordered NonOverlapping Disks model with any of the parameter spaces \( \Theta_1, \Theta_2, \) or \( \Theta_3 \).

The extension from Disks to 4-fold symmetry seems likely to hold as well.

### 6.3 Two NonCrossing Horizons

Now consider a region which is white everywhere except between two horizons which are constrained never to cross. For example, suppose that, as in Figure 10, both horizons are touching at the endpoints \( u = 0, 1 \), and that elsewhere \( \psi_1(u) < \psi_2(u) \). Suppose that the two horizons are each parametrized as in Section 5.1, with respective parameter vectors \( \theta_i \), and that, with \( \theta = (\theta_1, \theta_2) \), we have

\[
I_\theta = 1_{\{\psi(u; \theta_1) \leq u \leq \psi(u; \theta_2)\}}.
\]

We let \( \Theta \) be the subset of \( \theta \) vectors where

\[
\psi(u; \theta_1) \leq \psi(u; \theta_2) \quad u \in [0, 1],
\]

so that the definition makes sense. Note that, for each fixed \( u \), \( \theta \mapsto \psi(u; \theta) \) is linear; therefore, \( \Theta \) is a convex parameter space.

It is not hard to see that if \( \theta_{(0)} = (\theta_0^1, \theta_0^2) \) is a collection of Horizon parameters in \( \Theta \) and \( \theta_{(1)} = (\theta_1^1, \theta_1^2) \) is another collection of horizon parameters in \( \Theta \), then for \( \vartheta(t) = \theta_{(0)} + t(\theta_{(1)} - \theta_{(0)}) \) a linear path we have the arclength

\[
\lambda(\vartheta; \mathcal{M}) = C_\tau \cdot \sqrt{\sum_{i=1}^2 \left\| \theta_i^t - \theta_i^0 \right\|^2}
\]

and a Riemannian structure which is in fact Euclidean:

\[
g(\vartheta) = Const \cdot Id, \quad \forall \vartheta \in \Theta.
\]

Thus, again in this setting, the shortest path between images defined by two horizons indeed goes by following a linear path in parameter space.
Theorem 6.3 Continuous ISOMAP works perfectly for the Two NonCrossing Horizons model.

6.4 Cartoon Faces

Now consider a very simple model of a cartoon face undergoing various articulations.

In the first, a fixed oval region in the plane is called the 'head', and five regions inside it, called 'brows', 'eyes' and 'mouth' are defined. Each of the brows, eyes and mouth is a black region defined by two articulating horizons. The eyes and mouth have horizons which, like in the previous section, are joined at the ends, creating an almond shape. The brows have two horizons ending in vertical line segments and having fixed widths. Examples of these faces appear in Figure 11. The parameter space under this model is convex for the same reasons as the parameter space in Section 6.3.

The underlying parameter vector $\theta = (\theta_1, ..., \theta_8)$; $\theta^1$ and $\theta^2$, give the upper and lower horizon of the mouth (i.e. the lips), $\theta^3$ and $\theta^4$ give the upper and lower horizon of the left eye, (i.e. the lids), etc. The parameter space $\Theta$ is constrained so that the upper and lower boundaries of an object such as the right eye never cross, and so that the upper boundary of the right eye never crosses the lower boundary of the right brow, etc.

We evaluate closeness of the images using $L^2(dx)$ as usual. Applying the same smoothing recipe as in the single horizon case, we get a renormalized distance.

It is not hard to see that if $\theta_{(0)} = (\theta^1_0, ..., \theta^8_0)$ is a collection of Horizon parameters in $\Theta$ and $\theta_{(1)} = (\theta^1_1, ..., \theta^8_1)$ is another collection of horizon parameters in $\Theta$, then for $\varphi(t) = \theta_{(0)} + t(\theta_{(1)} - \theta_{(0)})$ a linear path we have the arclength

$$\lambda(\varphi; M) = C_r \cdot \sqrt{\sum_{i=1}^{8} \| \theta^i_1 - \theta^i_0 \|^2}$$

and a Riemannian structure which is in fact Euclidean:

$$g_{ij}(\theta) = C_r \cdot Id, \quad \forall \ \theta \in \Theta.$$ 

Thus, again in this setting, the shortest path between images defined by two horizons indeed goes by following a linear path in parameter space.
Theorem 6.4 Continuous ISOMAP works perfectly for the Cartoon Face Model.

7 Movies

In this section only, we consider the case of 'movies', which are objects with an additional time index: \( I(x, \alpha) \) where \( x \in \mathbb{R}^2 \) and the time index \( \alpha \in [0, 1] \). Think of a (continuous) collection of 2-D images \( I(., \alpha) \); instead of looking at the parameter space for one movie, we consider the joint parameter space of an entire set of movies. We are able to study Continuum ISOMAP easily, as seen below. However, empirical tests of the theoretical predictions in these movie examples are beyond our computational resources because of the massive size of the underlying data matrix.

We will consider three kinds of articulations under the movie model.

7.1 Motions of a Disk

Let \( I_0 \) be a 2-dimensional image – the indicator function of a disk – and consider a movie showing motion of the disk. Let in fact \( \chi(\alpha) \) be the center of the disk as a function of time and put

\[
I(x, \alpha) = I_0(x - \chi(\alpha)).
\]

For a simple example, consider a movie which shows linear motion of the disk. Then

\[
\chi(\alpha) = v \cdot \alpha.
\]

The center of the disk is at \((0, 0)\) at the movie's beginning, at time \( \alpha = 0 \), and it is at \( v \) at the movie's end, at time \( \alpha = 1 \).

For a more interesting class of motions, we consider the parametric family:

\[
\chi(\alpha) = \chi(\alpha; \theta) = \sum_j \theta_j \psi_j(\alpha)
\]

where \( \psi_j \) are vector-valued and orthogonal in \( L^2[0,1] \).
Now consider a smooth path in $\Theta$. Because at each point of $\partial B_\theta$ the normal vector is either coincident with the motion vector or perpendicular to it, and because articulation in $\alpha_1$ alone has an impact corresponding to the vertical profile of the rectangle,

$$g_{11} = 2 \times \left[ \int_{x \in [-\alpha_2, \alpha_2]} dx \right]$$

In short,

$$\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} 2\alpha_2 & 0 \\ 0 & 2\alpha_1 \end{bmatrix}$$

As the metric tensor $g_{ij}$ depends in an essential way on $\theta$, it is not constant, and the abstract manifold $\mathcal{M}$ is not flat. Hence, the apparently 'natural' parameter space is not recovered by ISOMAP.

8.2 Ellipses

Another unsuccessful example occurs in the case of a family of ellipses with independent articulations of the semi-major and semi-minor axis lengths. Consider the indicator of an ellipse with boundary given by the curve $B(\omega; \alpha_1, \alpha_2) = (\alpha_1 \cos(\omega), \alpha_2 \sin(\omega))$ where $\omega \in [0, 2\pi)$. Consider 'morphing' the ellipse by changing the parameters $\alpha = (\alpha_1, \alpha_2) \rightarrow \alpha + \epsilon = (\alpha_1 + \epsilon_1, \alpha_2 + \epsilon_2)$. Let $\epsilon_2 = 0$. As in the top right panel of Figure 13, we can calculate the normal and motion vector at each point of the ellipse. The resulting metric tensor calculation for an articulation in $\alpha_1$ gives coefficient

$$g_{11} = \alpha_1 \left[ \int_0^{2\pi} \frac{\sin^2 \omega}{\sqrt{\alpha_1^2 \sin^2 \omega + \alpha_2^2 \cos^2 \omega}} d\omega \right]^{\frac{1}{2}}$$

Similarly, the coefficient in $\alpha_2$ is a function of both $\alpha_1$ and $\alpha_2$. As a result, we have curvature of the manifold as a function of the articulation parameter. Therefore, we note that continuous ISOMAP will not work perfectly on the case of ellipse morphing.

Both the rectangle- and the ellipse-morphing examples can be viewed as examples of the articulating 'blobs' in Section 5.2. In this case, we know that ISOMAP could be successful if the metric used were $\frac{dx}{\|x\|^2}$. However, in practice we wouldn't be able to adapt the choice of image metric to the parameter space very easily, so we consider this viewpoint an interesting curiosity rather than a realistic option.

8.3 Race Track

Consider the case where we allow two disks to move within a rectangular image space as in Figure 13. Both disks follow motions which are linear in parameter space. While their paths can cross, the disks are prevented from occupying the same space at the same time. However, because the disks cannot overlap, at any given position of the first disk, the parameter space of the second disk has a gap representing the parameters overlapping the first disk. As a result, the parameter space is not convex. This model can be generalized to the case of disks pursuing circular or elliptical paths.

8.4 Jointed Fingers

Imagine a model as in the fourth panel of Figure 13 where we define a single 'skeletal finger' composed of three connected line segments, thickened slightly into 'rods'. In this case, we
assume that the joints of the finger are negligible compared to the segment lengths (all segments are of length L). Parametrize each of the segments by an angular position: \( \theta_i, \) \( i = 1, 2, 3, \) where \( 0 \leq \theta_i \leq \pi/2. \) The entire skeleton of the finger is then, for \( l \in [0, L] \):

\[
S_1 = (l \cos(\theta_1), l \sin(\theta_1))
\]

\[
S_2 = (L \cos(\theta_1) + l \cos(\theta_2), L \sin(\theta_1) + l \sin(\theta_2))
\]

\[
S_3 = (L \cos(\theta_1) + L \cos(\theta_2) + l \cos(\theta_3), L \sin(\theta_1) + L \sin(\theta_2) + l \sin(\theta_3))
\]

As a result, the normal and motion vectors under a three-parameter articulation, \((\theta_1, \theta_2, \theta_3)\), depend on the values of the other variables. To achieve the minimal distance between two positions of \( \theta_1 \) requires that the other parameters \( \theta_i, \) \( i > 1 \) be in optimal position, and the definition of an ‘optimal’ position depends on the current value of \( \theta_1 \).

9 Occlusion

The composite articulations that we have considered up until now do not allow the objects in the image to occlude or overlap one another. Occlusion quite generally prevents Continuum ISOMAP from working perfectly.

9.1 Multi-disk examples

In Section 6.2 we defined several cases where multiple unit disks articulate in the image independently, but subject to certain restrictions. Consider a case of unrestricted motion; the image is the indicator of a single unit disk fixed with center at the origin, and a second disk which undergoes translations around the image; when the disks overlap, there will be occlusion. That is, \( I_{\theta}(x) = \max(0, \min(1_{\{||x||_2 \leq 1\}} + 1_{\{||x-\theta||_2 \leq 1\}}, 1)) \) for \( \theta \in \Theta \), where \( \Theta \) is the entire image space in \( \mathbb{R}^2 \). Where the two disks overlap, the image value is not additive; we maintain simple black-on-white structure throughout the image library.

In this case, the boundary of the black region has variable shape. The boundary, \( \partial B_{\theta} \), is the perimeter of two separate disks when \( ||\theta||_2 > 2 \) and the perimeter of the overlapping disks when the two disks overlap. The arclength of an infinitesimal curve element depends on whether \( ||\theta||_2 > 2 \), where it behaves in a Euclidean fashion, or \( ||\theta||_2 < 2 \), where it depends on a nonconstant function of \( \theta \). In effect, formula (2.8) says that the length and orientation of the boundary \( \partial B \) controls the arclength of the curve \( \theta \mapsto B_{\theta} \). In fact, the abstract manifold is curved as a function of the degree and angle of overlap. To see the curvature, consider calculating the coefficient \( g_{22} \) according to Theorem 2.3 (in effect, \( g_{22} \) measures the cost of infinitesimal translations of the mobile disk in the \( y \)-coordinate direction).

Let \( p \in [0, \frac{\pi}{2}] \) be the angle of the vector \( \theta = (c_1, c_2) \), i.e. arctan\((c_2/c_1)\). Let \( q \in [0, \frac{\pi}{2}] \) be half the angle subtending the overlapping arc, \( q = \arccos\left(\frac{1}{2} \times \sqrt{c_1^2 + c_2^2}\right) \), where \( v = (v_1, v_2) \) is the unit vector in the \( y \)-coordinate direction.

\[
g_{22} = \sqrt{\int_{b \in [0,1]} \langle v(b), n(b) \rangle^2 db}
\]

\[
= \left[ (v_1^2 + v_2^2)(\pi - q) - \cos(q)\sin(q) \right]
\times \left[ ((v_1^2 - v_2^2)\cos(2q) + 2v_1v_2\sin(2q))\right]^{\frac{1}{2}}
\]

The resulting function is ‘flat’ (independent of \( \theta \)) for \( ||\theta||_2 > 2 \), but is neither constant nor radially symmetric within the radius of occlusion. Figure 14 shows the coefficient \( g_{22} \)
7.4 Grimacing Faces

Suppose now that as in Section 6.4 we have a face made of a collage of pairs of nonoverlapping horizons. The horizons and their time articulations are parametrized as in the previous subsection. Combine all the parameters of all the horizons and their time-articulation in one vector, ranging through the convex parameter space $\Theta$. We have immediately from the previous analysis the following corollary of Theorem 7.3.

Corollary 7.1 Continuum ISOMAP works perfectly on movies of gesturing Faces.

8 Failures

The conditions established for Continuum ISOMAP to ‘work’ can fail in two different ways. First, the parameter space can be inherently non-convex. This failure is seen in the ‘Race Track’ example below. The more complicated failures occur in examples where, although the parameter space is convex, the manifold is not flat (i.e. the metric tensor is a nonconstant function of the parameters).

8.1 Rectangles

Consider the ‘world’ of rectangles of various volumes and aspect ratios and common centers. With $\alpha_1$ and $\alpha_2$ the halflengths in the axial directions, and $\theta = (\alpha_1, \alpha_2)$, the image $I_\theta$ is the the indicator of the region defined by $|x| \leq \alpha_1$, $|y| \leq \alpha_2$. Let $\Theta$ be the rectangle $\alpha_i \in (1/2, 2)$. 
Define now the family of articulated images $I_\theta(x) = I_0(x - \chi(\alpha; \theta))$, and for regularization, do 2-d smoothing within each frame:

$$I^h_\theta = \phi_h \ast_{(x, y)} I_\theta$$

here $\phi_h$ is a 2-d smoothing kernel in $(x, y)$ only (not in $\alpha$). For a smooth path $\vartheta(\cdot)$ in $\Theta$ we have the following formula for arclength:

$$\lambda(\vartheta) = C_r \int_0^1 \sqrt{\int (v, n)^2 \frac{db}{s(b; \alpha)}} d\alpha dt$$

where $b$ runs across the perimeter of the disk, $v = v(b, \alpha) = \frac{d}{dt} \beta(b, \alpha; t)$, $s(b; \alpha)$ is the determinant of the Jacobian $b \mapsto \beta(b; \alpha)$, and $n$ is the $(x, y)$-normal to the surface of the disk at $(b, \alpha)$. Then we notice that, just as in the case of translating a 2-d image,

$$\int (v, n)^2 \frac{db}{s(b; \alpha)} = C \cdot \| \frac{dx}{d\alpha} \|^2$$

We conclude that

$$\delta(\vartheta_0, \vartheta_1) = C \cdot \| \vartheta_1 - \vartheta_0 \|_2.$$ 

**Theorem 7.1** Continuous ISOMAP works perfectly for movies of a translating disk.

For similar reasons, Continuous ISOMAP works for movies of several translating disks which are ordered and nonoverlapping.

### 7.2 Gestures of a Hand

Consider the hand model of the previous section, only relabel the parameter $\theta$ in that model to be $\zeta$. Then $\zeta$ has several components, controlling the position of the various fingers. Now consider a movie showing motion of the hand components. Let in fact $\zeta(\alpha)$ be the parameters of the hand as a function of time and put

$$I(x, \alpha) = I_{\zeta(\alpha)}(x).$$

For a simple example, consider a movie which shows uniform rotation of the fingers around a common center. Then

$$\zeta(\alpha) = \zeta_0 + \alpha(\zeta_1 - \zeta_0).$$

The hand is in configuration $\zeta_0$ at the movie's beginning, at time $\alpha = 0$, and it is at $\zeta_1$ at the movie's end, at time $\alpha = 1$.

For a more interesting class of motions, we consider the parametric family:

$$\zeta(\alpha) = \zeta(\alpha; \theta) = \sum_j \theta_j \zeta_j(\alpha)$$

where $\zeta_j$ are vector-valued and orthogonal in $L^2[0, 1]$.

Define now the family of articulated images $I_\theta(x, \alpha) = I_{\zeta(\alpha)}(x)$. and for regularization, do 1-d smoothing angularly:

$$I^h_\theta = \phi_h \ast_\omega I_\theta$$
here $\phi_h$ is a 1-d smoothing kernel acting convolutionally in $r = \text{constant}$. For a smooth path $\theta(\cdot)$ in $\Theta$ we have the following formula for arclength:

$$\lambda(\theta; \mathcal{M}) = \int_0^1 \sqrt{\int \langle v, n \rangle^2 \frac{db}{s(b, \alpha)} d\alpha dt}$$

where $b$ runs across the perimeter of the hand within the two-dimensional image, $v = v(b, \alpha) = \frac{d}{dt} \beta(b, \alpha; t)$, $s(b, \alpha)$ is the speed of the curve, and $n$ is the $(x, y)$-normal to the surface of the disk at $(b, \alpha)$. Then we notice that, just as in the case of translating a 2-d image,

$$\int \langle v, n \rangle^2 \frac{db}{s(b, \alpha)} = C \cdot \left\| \frac{dC}{d\alpha} \right\|_2^2.$$

We conclude that

$$\delta(\theta_0, \theta_1) = C \cdot \left\| \theta_1 - \theta_0 \right\|_2.$$

**Theorem 7.2** Continuum ISOMAP works for movies of Hand Gestures.

### 7.3 Horizon Morphing

We now consider a movie of a horizon changing in time:

$$I(u, v, \alpha; \theta) = 1_{\{v \leq \psi(u, \alpha; \theta)\}}$$

where the horizon is parametrized as

$$\psi(u, \alpha; \theta) = \sum_{w, z} \theta_{w, z} \psi_z(u) \phi_w(\alpha)$$

with $(\psi_w)$ is orthonormal for $L^2([0, 1], du)$ and $(\phi_z)$ is orthonormal for $L^2([0, 1], d\alpha)$.

We use a 1-dimensional smoothing kernel

$$I^h_\theta = \phi_h *_v I_\theta$$

We get the following formula for arclength

$$\lambda(\theta) = \int_0^1 \sqrt{\int \left( \frac{d\psi(u, \alpha)}{dt} \right)^2 d\alpha dt}$$

where

$$\frac{d\psi(x, \alpha)}{dt} = \sum_{w, z} \frac{d\theta_{w, z}}{dt} \psi_z(u) \phi_w(\alpha)$$

If we consider a linear path $\theta(t) = \theta_0 + t(\theta_1 - \theta_0)$ we get

$$\int \left( \frac{d\psi(u, \alpha)}{dt} \right)^2 d\alpha dt = C \cdot \left\| \theta_1 - \theta_0 \right\|_2^2$$

We conclude that

$$\delta(\theta_0, \theta_1) = C \cdot \left\| \theta_1 - \theta_0 \right\|_2.$$

**Theorem 7.3** Continuous ISOMAP works perfectly on movies of articulating Horizons.
Figure 14: Image plot of square parameter space $\Theta$ in the Two-Disk model with fixed unit disk at $(0,0)$. $\theta$ represents the center of the mobile disk. Color indicates cost of infinitesimal motion of the disk center in the vertical direction.

plotted over the image space for the double-disk example. The behavior of the coefficient $g_{11}$ is identical, except that the image display is rotated by 90 degrees from the display of $g_{22}$. According to the previous definition, we note that we expect Continuum ISOMAP to fail under this occlusion example.

We can represent the manifold for a given quadrant of the image-plane using the matrix

$$
\begin{bmatrix}
g_{11} & g_{12} \\
g_{21} & g_{22}
\end{bmatrix} = \begin{bmatrix}
\left(\frac{\partial}{\partial u_1} \right)^2 & \frac{\partial}{\partial u_1 u_2} \\
\frac{\partial}{\partial u_1 u_2} & \left(\frac{\partial}{\partial u_2} \right)^2
\end{bmatrix}
$$


to generate an ellipse field on the image space. The resulting image is shown in Figure 15. Note that where the manifold is flat the ellipses become circles; where there is noticeable curvature (i.e. within the radius of occlusion) the ellipses have different dimension and orientation.

The implication of the resulting ellipse field for induced curves on the manifold is shown in Figure 16. The colored paths represent the projections of example ‘shortest paths’ on the image manifold (as defined by the $g_{ij}$ matrix) between the two parameter points. The paths were extracted iteratively using Floyd’s algorithm to find minimal cost paths on a grid. Once outside the radius of occlusion (drawn in blue), the paths are Euclidean. That is, we would expect that the shortest path between any two points on the parameter space would be a straight line. However, for points within or near the radius, the paths have significant deviation from the Euclidean path due to the reduced exposed boundary length within the occlusion region.

9.2 Empirical Verification

To test the theory, we do an empirical trial of 100 double-disk images of the type specified above, at the usual $64 \times 64$ resolution and apply the ISOMAP algorithm. The result
Figure 15: Domain is the first quadrant of the square parameter space $\Theta$ in the Two-Disk model. Each ellipse represents the Riemannian structure at the corresponding point of $\Theta$.

Figure 16: Projections of 'shortest paths' into the square parameter space, $\Theta$, plotted on the central region of $\Theta$. 
(to varying degrees, depending on the sample) is a two-dimensional embedding that reflects the deviations we see in the path projections in Figure 16. In the central panel of Figure 17, the resulting embedding is a 'spiny' structure, where each of the spines represents data points in a separate quadrant of the parameter space. In effect, the embedding says that the lowest cost paths from one quadrant to another always travel through the central fixed disk. Therefore, the embedded distances generated are in fact distances from and to occlusion, rather than distances reflecting the actual articulation. Furthermore, within each spine, lateral information about the relative locations of the parameter points is almost nonexistent.

10 Effects of Non-Convexity

The asymptotic consistency of the original ISOMAP algorithm, as demonstrated by Tenenbaum, de Silva, and Langford [9], relies on the data manifold being geodesically convex. However, it seems clear that where we are working not with a manifold but with a finite graph, datasets might not exhibit sufficiently straight point-to-point connectivity, and the resulting embedding would exhibit undesirable artifacts. A specific way in which this can occur is that the nearest neighbor graph has large holes in it, forcing the shortest path in the graph to curve significantly to get around the hole. This type of problem has been noted by Lee, Lendasse, and Verleysen [5] who refer to it as the 'Swiss Cheese' effect. The effect is characterized by an irregular 'soap bubble' appearance in the embedded neighborhood graph. Figure 18 shows the neighborhood graph of a 500-point sample from a three-dimensional Swiss roll manifold superimposed on the two-dimensional embedding produced by ISOMAP. While the original data manifold is geodesically convex, the actual sample dataset generates a nearest neighbor graph which exhibits gaps; the reconstructed geodesics step around the gaps in the sample.

For the case of manifolds of articulated images, we observe that the conditions for geodesic convexity hold automatically if the parameter space of the underlying articulation is convex. That is, for a prototype image $I_0$, where the manifold $M = \{I_\theta = T_\theta I_0 : \theta \in \Theta\}$, then the condition will be satisfied if $\Theta$ is convex and the local Riemannian structure is $g_{ij} = C \cdot I d$; that is, if any straight line between two parameter points $\theta_1$ and $\theta_2$ lies entirely in $\Theta$. As before, however, both low sampling and noise result in an effective non-convexity.
Figure 18: 2-dimensional embedding of Swiss Roll data; overlay of neighborhood graph of the sample nearest neighbor graph.

For example, in the case of the double disk images, the attempted solution to the occlusion problem would be to remove all images from the sample where the mobile disk overlaps the fixed central disk. However, the result of pruning the data sample in this way would be to remove all parameter points \( \theta \) where \( ||\theta||_2 < 2 \). But the parameter space \( \Theta \) would no longer be convex. The far right panel of Figure 19 shows results of embedding the test set of double-disk images without those where the two disks overlap. Coloring by radial angle reveals that the angular coordinate of the original parameters is ordered reasonably well around the origin. However, the point cloud of embedded centers, instead of being a plane with a missing center, is in fact attenuated to a ring. This effect is the result of the distortion of relative distances within the point cloud. Because there are no links across the middle region present in the data, those interpoint distances which need such a link are relatively extended, while those distances within the well-sampled regions are shortened by comparison. As a result, very little information about the true underlying radial distances survives.

In the case of the two-disk images, the problem with removing the occluded region is that the parameter space becomes non-convex. However, as we have already seen, we could construct reasonable examples where the remaining parameter space is actually convex. If we allow the image to be the indicator of a translated disk with a fixed 'bar' across the top of each image, then including occluded images results in a spike at the edge of the parameter space. However, by removing all occluded images (since the remaining parameter space is convex), we can produce an almost perfect embedding of the disk translations.

The success of the bar-and-disk example suggests a methodology to compensate for non-convex parameter sets. We could select overlapping subsets of the data that correspond to convex subsets of the parameter space. Then, embed the subsets separately and use Procrustes [2] analysis to scale and match the embeddings automatically. The fact
that the parameter space is Euclidean then allows the separate embeddings (matching the overlapping points) to be patched together into a ‘flat’ parameter space.

11 Discussion

11.1 Pixelization

Our theoretical approach may, or may not, describe the behavior of ISOMAP on actual datasets, as ISOMAP must cope with pixelization - as well as the chance of obtaining a truly terrible sample of the original parameter space. For example, several tests of the actual ISOMAP algorithm for objects with straight lines result in small but noticeable glitches in the second and third coordinates of the embedding. These examples have occurred in rotation of an object, when the important edges have reached nearly 45-degree angles with respect to the standard axes. Such a result could be explained by stair-step pixelization artifacts of diagonal lines. Also, our examples have been carefully scaled such that the objects in the image occupy at least 16 pixels to guarantee that the ‘object’ effects take precedence over the ‘pixel’ effects. For images with a wide variety of object sizes, the number of pixels on each object might have a strong relation to the quality of our continuum predictions.

11.2 Other Distance Metrics

The examples in this paper work with Euclidean ($L_2^2$) interpoint distances as the input to the ISOMAP processing. Two other common metrics have great appeal in the ISOMAP setting: $L_1^1$-distance and tangent distance. $L_1^1$-distance is intuitively reasonable for black and white images because it effectively measures the number of pixels different in two chosen
images. While similar mathematical analysis can be applied to $L^1$-distance, the generated manifold lacks the Riemannian properties of the manifold under $L^2$-distance.

Tangent distance has also been become very popular for image classification problems [4, 8], and is used for several of the successful ISOMAP embeddings [9]. By using distances through a tangent space of a particular dimension, tangent distance has the potential to ignore some of the additional dimensions introduced by occlusion. However, the problem needs more study and experimentation.

12 Conclusions

Our analysis of Continuum ISOMAP establishes that there is an underlying theoretical perspective for understanding the ISOMAP algorithm in the case of image manifolds. In effect we have shown for the first time that in some sense there is a set of natural examples where ISOMAP can be expected to work correctly in recovering the underlying parametrization of the image manifold. Although our analysis disregards several possible confounding factors such as pixelization, the comparison of theoretical predictions with empirical tests of ISOMAP suggests that the theory is accurate in predicting the actual empirical results. In addition, the theoretical perspective easily allows consideration of variations of the image case (movies and horizons, for example) that would not be easy to explore by empirical techniques.

References


