IPSJ Transactions on Computer Vision and Applications Vol. 1 83–94 (Mar. 2009)

Invited Paper

A Survey of Manifold Learning for Images

Robert Pless[†1] and Richard Souvenir[†2]

Many natural image sets are samples of a low-dimensional manifold in the space of all possible images. Understanding this manifold is a key first step in understanding many sets of images, and manifold learning approaches have recently been used within many application domains, including face recognition, medical image segmentation, gait recognition and hand-written character recognition. This paper attempts to characterize the special features of manifold learning on image data sets, and to highlight the value and limitations of these approaches.

1. Introduction

Actions in video take many forms: people walk, cycle and run, birds fly, fish swim, kangaroos hop, water fountains gush, hearts beat, contrast agents permeate through tissue, and MRI tag lines slowly fade. These actions form the semantic content of the video. Recognizing or extracting representations of these actions is the key to semantic interpretation of video data. The same is true for large image data sets that are not temporal sequences. Images of specific objects may vary due to the pose, lighting, sensor noise, and distortions in the image capture process. Understanding or extracting representations of this variability is the key to effectively indexing the image data set.

What is an appropriate representation of actions in video or variability in image sets? In the context of naturally occurring patterns, D'Arcy Thompson proposes that a study of deformations is fundamental [63]:

> "In a very large part of morphology, our essential task lies in the comparison of related forms rather in the precise definition of each; and the deformation of a complicated figure may be a phenomenon easy of comprehension, though the figure itself may have to be left unanalyzed and undefined."

There are multiple possible goals of comparisons within a set of images, such as clustering — which partitions the data set into discrete groups of similar elements and regression — which discovers a continuous model from which the data set could be derived. For images, which can be considered as points in a high-dimensional space (where each pixel location is equivalent to a vector component), classical clustering and regression techniques often fail, as explained by the "curse of dimensionality" [5]. However, many image sets and videos vary due to a small number of degrees of freedom and the set of these points lie on or near some low-dimensional manifold embedded in the high dimensional space. This paper focuses on manifold learning tools in order to understand the behavior or variability in image data sets. To ground this exploration, we start in Section 2 by presenting a collection of image manifolds that occur in Computer Vision and Medical Imaging tasks, and highlight some of the features that these manifolds have.

Section 3 offers a brief description of current research in manifold learning algorithms, with a focus on those that have been used for image sets. Most commonly, these algorithms first compute some measure of the relationship (such as the distance) between all or some pairs of the original data points. Second, they find a set of low dimensional points that preserve, as best as possible, this relationship. In Section 4 we explore ways that manifold learning has been specialized to address special challenges when applied to large image sets.

Finally, Section 5 considers applications of manifold learning within a broader context where the end goal of the algorithm is not simply to create a perceptual organization of the image set, but to use the manifold parameters to support some larger goal, such as recognition or segmentation.

2. Example Image Manifolds

This section offers a collection of natural image sets that have a manifold structure. These examples highlight some ways that natural images tend to vary and provide a pictorial overview of some of the applications of manifold learning to images. For each, we describe the dimensionality and topology of the resulting

†1 Washington University, St. Louis
†2 University of North Carolina, Charlotte

Fig. 1 A data set of a small statue viewed from different viewpoints. The figure undergoes a rigid 3D transformation between different image.

image manifolds.

Figure 1, first described in Ref. 48), shows a montage of images of an object captured from different viewpoints. This set of images lies naturally on a 2D image manifold, parametrized by the object rotation angle and the camera elevation angle. Given an unordered collection of these images, manifold learning offers a tool for pose estimation — because the resulting manifold parameters should be related to the image viewpoint. As an example of the challenges that arise in manifold learning, we point out that images appear more similar as the object rotation changes (moving across the figure), and the images change more when the camera elevation angle changes (moving up and down the figure).

Another example of image variation is shown in **Fig. 2**. Within this data set of images of a bird flying across the sky [46)], there are two degrees of freedom: non-rigid deformations as the bird flaps its wings, and the rigid deformations as the bird flies past the camera. The second degree of freedom is shown in right column of this figure, with examples of the bird at the same part of its flying

Fig. 2 Examples of an image manifold created by cyclic deformation due to the flap cycle of the bird, and the rigid transformation as the bird passes the camera. This motion is highlighted in the right column, by selecting images at the same part of the bird flying cycle.

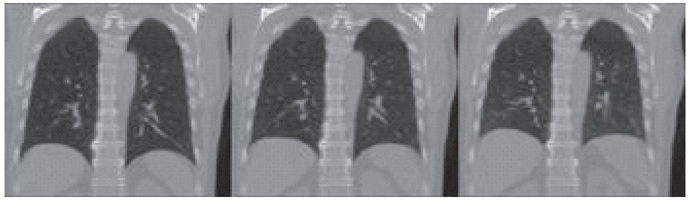


Fig. 3 Medical imagery has especially many compelling examples of images drawn from low-dimensional manifolds, as in this example of CT images of one patients lung as they breathe.

cycle, but viewed from different relative orientations.

The topology of this image manifold is cylindrical. That is, the images are most naturally parametrized by a cyclic parameter encoding what part of the flying cycle that image is captured during, and a linear parameter encoding how far it is from the camera. Many natural data sets of animals moving have cyclic topologies like this, and some of the efforts to make manifold learning successful for such topologies are detailed in Section 4.3.

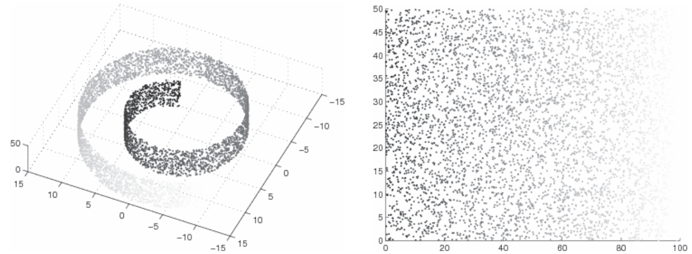
A final example of an image manifold is shown in **Fig. 3**. This shows examples of a slice of a CT image of a patient as they breathe. At this resolution, the lung has approximately the same appearance when the patient is inhaling as it does when the patient is exhaling, so these images are well described by the single parameter of lung-volume. However, at higher resolutions, images captured during inhalation are noticeably different than images captured during exhalation. In this case, the lung volume no longer uniquely defines the image, and the natural manifold parametrization must include a cyclic parameter.

These example manifolds are drawn from image sets with just a few degrees of freedom. In general, we have observed that manifold learning methods work best for data sets with few degrees of freedom. This is because most of the algorithms (described in the next section) require the underlying manifold to be densely sampled. Therefore, the data requirements (and corresponding computational requirements) become prohibitive for data sets with more than 3 or 4 degrees of freedom. Thus, applying manifold learning to data sets of one particular object under different viewpoints, or one particular patient as they breathe tends to work better than trying to use manifold learning to characterize, for example, the appearance variations between different people.

3. Manifold Learning Algorithms

Given the number of image sets with the natural low-dimensional, but nonlinear, structure shown in the previous section, it is natural to consider algorithms that automatically find this structure. Over the last decade, an astonishing number of these algorithms have been proposed and a recent summary of many of these highlights the broad interest in this area [11),41)]. In this section we present some of the more popular of these algorithms and show which choices are important for these algorithms to be effective on image data sets.

Manifold learning, or *nonlinear dimensionality reduction*, is the counterpart to PCA which aims to find a low dimensional parametrization for data sets which lie on *nonlinear* manifolds in a high-dimensional space. **Figure 4** shows a classic example of manifold learning on a synthetic data set. Figure 4 (a) depicts the so-called "Swiss Roll" data set which consists of 20,000 3-dimensional points. Intuitively, this data set can be visualized as points drawn off a rolled-up sheet of paper. While each of these points can be described by three coordinates (or more), there is an underlying two-dimensional representation. The goal of



(a) 3D Input Data (b) Low-dimensional Embedding

Fig. 4 Manifold learning example using the "Swiss Roll" data set which consists of 20,000 3-dimensional points. While each of these points can be described by three coordinates as shown in (a), there is an underlying two-dimensional representation as shown in (b). The goal of manifold learning is to automatically learn this representation. In this figure, the intensity of the points represents distance along the curved "axis."

manifold learning is to automatically learn this representation. The expected output of manifold learning algorithms is shown in Fig. 4 (b).

If the original data lies in a space R^D, then the manifold learning problem is to find coordinates for each point in a smaller dimensional space d. One formalization of this is:

> Given an input set $\mathcal{X}$, which is a finite subset of $\mathcal{R}^D$, for some dimension D, learn a parametrization which produces a mapping function $f : \mathcal{X} \longrightarrow \mathcal{R}^d$ which preserves *some properties* of the structure of $\mathcal{X}$.

This problem of nonlinear dimensionality reduction has been one of broad interest for many decades. However, for this survey, we will start the discussion in 2001 when renewed interest in the topic was sparked by two algorithms (Isometric Feature Mapping (Isomap)[62)] and Locally Linear Embedding (LLE)[52)]) which were presented simultaneously. At about the same time, typical desktop computers had the storage and processing power necessary for the analysis of large image sets and video, which opened up a fertile application area for these methods. Isomap and LLE are two methods which illustrate different ways to make the above definition more concrete. We first discuss these algorithms, then provide an overview of alternative approaches.

3.1 Isomap

Isomap is a manifold learning algorithm which preserves geometric features of the input set. Specifically, the goal of Isomap is to return an isometric mapping,

$$f : \mathcal{X} \longrightarrow \mathcal{Y} \tag{1}$$

for $\mathcal{X} \subset R^D$, $\mathcal{Y} \subset R^d$, and $d << D$ where, for all pair of points $X_i, X_j \in \mathcal{X}$,

$$|X_i - X_j|_{shortest\ path} = |Y_i - Y_j|_2 \tag{2}$$

That is, Isomap returns an embedding where the distances between points is approximately equal to the shortest path distance (on a graph defined on near-neighbors in the original space). Below, we describe the main steps of the algorithm.

Given: A set of points $\mathcal{X} \subset R^D$

(1) Compute the distance between all pairs of points (traditionally using the L_2 norm distance.)

(2) Define the set of points which comprise the neighborhood, $\mathcal{N}(X_i)$ for each point, $X_i \in \mathcal{X}$. This is typically done in one of two ways:
- *k-nearest neighbors.* Select the k closest points to X_i.
- *ϵ-ball.* Select all points $X_j \in \mathcal{X}$ such that $|X_j - X_i|_2 < \epsilon$.

(3) Define a graph with a node for each input point, X_i and weighted undirected edges connecting each node to the nodes corresponding to the points in $\mathcal{N}(X_i)$. For each edge, the weight equals the corresponding distance between the input points.

(4) Solve for the all-pairs shortest paths on this sparse graph to calculate a complete pair-wise distance matrix.

(5) Solve for the low-dimensional embedding $\mathcal{Y} \subset R^d$, using Multidimensional Scaling (MDS)[35] (described below). d is the dimension of the low-dimensional embedding and can be chosen as desired, but, ideally, is the number of degrees of freedom in the image set.

Output: $\mathcal{Y}$, the low-dimensional embedding of $\mathcal{X}$

3.1.1 Multidimensional Scaling (MDS)

The final step of Isomap requires embedding using MDS. MDS is the name of the eigenvalue problem to convert a matrix of pair-wise distances into absolute coordinates in some (typically low) dimensional space.

Given an $n \times n$ matrix D, such that $D(i, j)$ is the desired squared distance from point i to point j:

(1) Define $\tau = -HDH/2$, (H is the centering matrix: $H = I - \vec{e}\vec{e}^\top/n$, where $\vec{e} = [1, 1, \cdots, 1]^\top$).

(2) Let $s_1, s_2, \ldots$ be the (sorted in decreasing order) eigenvalues of τ, and let $v_1, v_2, \ldots$ be the corresponding (column) eigenvectors. The matrix $Y = [\sqrt{s_1}v_1|\sqrt{s_2}v_2|\ldots\sqrt{s_k}v_k]$ has row vectors which are the coordinates of the best k-dimensional embedding.

The matrix $YY^\top$ is the best rank k approximation to τ (with respect to the L_2 matrix norm). This process finds the k-dimensional coordinates that minimize:

$$\sum_{ij} \left(|Y_i - Y_j|_2^2 - D(i, j)\right)^2,$$

or, the low-dimensional coordinates that best fit the distance matrix D.

3.2 LLE

LLE is another manifold learning algorithm which makes different assumptions in order to learn a low-dimensional embedding. LLE attempts to represent the manifold locally by reconstructing each input point as weighted combination of its neighbors. The algorithm is:

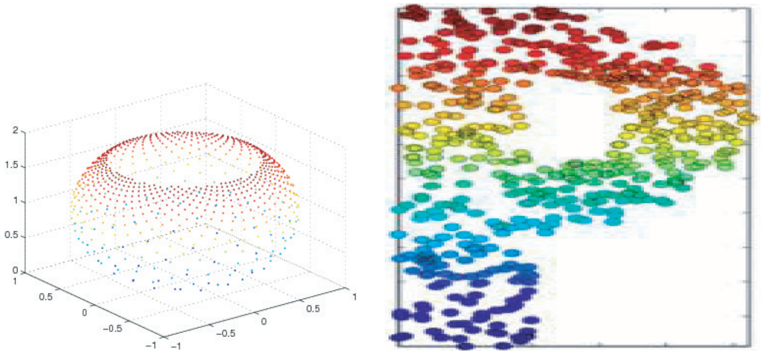
(1) Define $\mathcal{N}(X_i)$ for each point in $\mathcal{X}$. As in step 3 of Isomap (above), the k-nearest neighbors or ϵ-ball methods can be used.

(2) Solve for the reconstruction weights, W, used to reconstruct each X_i as a weighted sum of neighbors. $W(i, j)$ represents the weight of X_j used to reconstruct X_i; for $X_j \notin \mathcal{N}(X_i), W(i, j) = 0$. Normalize each row of W, such that $\sum_j W(i, j) = 1$, for each row i.

(3) Define the embedding coordinates $\mathcal{Y}$ using the weights W by solving an eigenproblem. Define $M = (I - W)' \times (I - W)$. Set $\mathcal{Y}$ to be the eigenvectors of M corresponding to the d smallest eigenvalues after discarding the smallest (with eigenvalue of zero.)

3.3 Other Methods

The original Isomap and LLE algorithms worked well for data sets such as the "Swiss Roll" in Fig. 4. However, data sets with high-curvature, self-intersections or near self-intersections, or non-convex sampling (within the manifold space) lead both Isomap and LLE to fail (see **Fig. 5**). **Table 1** list many recent varia-



(a) "Fishbowl" data set (b) Non-convexity example

Fig. 5 Examples of data sets for which early manifold learning algorithms fail. Table 1 lists a variety of manifold learning algorithms, some of which can accurately parametrize data sets such as these.

Table 1 Manifold learning algorithms.

Algorithm Class	Examples
Isomap variants	ST-Isomap [34] Continuum Isomap [20,74] Landmark Isomap [18] Conformal Isomap [18]
Charting	Manifold Charting [9] Nonlinear CCA & PCA [64]
Self-Organizing Maps	[1,8,65]
Graph Spectral Methods	Laplacian Eigenmaps [4,25] Kernel Eigenmaps [10] Hessian Eigenmaps [21] Locality Preserving Projections [31]
Supervised Methods	Local Fisher Embedding [17] Supervised LLE [16]
Other	Diffusion Maps [14] Manifold Tangent Learning [6] Proximity Graphs [12] Semidefinite Embedding [68] Stochastic Neighbor Embedding [32] Local Smoothing [44]

tions that help to address these and related shortcomings.

The variants of Isomap all follow the general steps described in Section 3.1. ST-Isomap can be applied to data with a temporal component, such as frames of a video, and works by modifying the local neighborhood structure and distance matrix to reduce the distance to both spatially and temporally adjacent points. Landmark Isomap trades off accuracy for speed by only using a subset of the points for the embedding step. Conformal Isomap describes how to faithfully embed data sets such as the "fishbowl" depicted in Fig. 5, given particular constraints on the sampling of the data points.

Another algorithm, Semidefinite Embedding (SDE), is related to Isomap in that the goal of the method is to provide an isometric embedding. In fact, the final step, embedding using MDS, is identical. The major difference is in the construction of the similarity matrix between all pairs of input points. SDE applies semi-definite programming to learn this kernel matrix. This method does not fail in the case of non-convexity like Isomap and can correctly parametrize the "P shape" in Fig. 5.

Self-Organizing Maps (SOMs)[51], also known as Kohonen Feature Maps, precede most of the algorithms in Table 1. Intended as a visualization tool for high-dimensional data, the use of SOMs for manifold learning was not discovered until later. SOMs follow the general framework of artificial neural networks trained using competitive learning[54] to discover a low-dimensional embedding for the input data.

The charting methods represent the high-dimensional manifold as a set of overlapping "charts" or "patches". The chart sizes may either be fixed or expandable until some assumption is violated, such as local planarity. In contrast to Isomap and LLE, most of these methods do not provide a globally consistent parametrization of the input data set, but rather parametrization within local regions.

All of the algorithms described so far are largely unsupervised, however, there exist a related class of semi- and fully supervised manifold learning algorithms which are variants of LLE. These algorithms generally retain the assumption that each point can be represented as a local combination of its neighbors, but introduce additional constraints on the solution (such as requiring particular points to have specific low-dimensional coordinates), in order to bias the final solution. The next section describes ways to specialize these general algorithms so that they are effective on images.

4. Specializing Manifold Learning for Image Applications

The previous section described algorithms which have been developed for the general problem of manifold learning. These can be directly applied to image data sets, because images can be thought of as points in a high-dimensional image space where each coordinate represents the intensity value of a single pixel. Since a set of images defines a set of points in this very high dimensional space, manifold learning can be directly applied to this point set; early papers provided results in this way on simple image sets [52),62)].

Three key challenges limit the usefulness of directly applying manifold learning algorithms to image data in this way. First, natural image manifolds are often very curved in this space. Second, many manifold learning algorithms do not give natural tools to project and re-project new images onto the manifold. That is, once nonlinear dimensionality reduction is applied to a set of images, there are no natural tools to find the embedded coordinate for a new image, nor are there natural tools to take parameters within the manifold and to generate an example image. Third, natural image data sets often arise from manifolds with cyclic parameters, which are challenging for many manifold learning algorithms

4.1 Image Distance Measures

Algorithms such as Isomap are based upon estimates of distances between images. Changing this distance metric is a natural approach to modifying the behavior of these manifold learning algorithms. The most common distance metric for manifold learning is the Euclidean distance. If images are treated as points in a high-dimensional space, where each dimension corresponds to the intensity at a pixel, then the Euclidean distance is equivalent to computing the square root of the sum-of-squared pixel intensity differences. However, often images vary due to motions or deformations of the object in view. In these cases, other distance measures may more faithfully capture how different a pair of images are.

There have been a large number of different image distance measures proposed, and many of them are summarized in the context of linear factorization methods [73)]. Image distance measures that have been explicitly used for manifold learning include the Earth Movers distance [53)] (which has recently been made dramatically faster [56)]). Methods to explicitly measure local deformations include using the phase shift of responses to a Gabor filter [60)].

Recently, there have been a number of algorithms that seek to learn the distance measure directly from the data set. One method for this takes the input data set and a collection of labels depicting that particular object pairs should be "similar" or "different" and learns a distance function as a weighted combination of image features [13)]. Alternatively, one can learn an embedding function that is invariant to particular transformations of the input [29)].

4.2 Projection and Re-projection

A second major limitations of manifold learning is that most methods define a mapping from the original data set to $\mathcal{R}^d$, where d is the low-dimension embedded space. That is, the result is a mapping

$$f : \mathcal{X} \longrightarrow \mathcal{R}^d$$

and not, as might be more convenient,

$$f : \mathcal{R}^D \longrightarrow \mathcal{R}^d.$$

This means that once the embedding of an data set $\mathcal{X}$ is computed, for $X' \notin \mathcal{X}$, the value of $f(X')$ is not well defined. Additionally, the inverse mapping is also problematic. For a point $Y \in \mathcal{R}^d$, if Y is not in the set of points defined by $f(\mathcal{X})$, then $f^{-1}(Y)$ is also not well defined.

Although a few generic approaches have been proposed to compute these "out of sample" projections, this remains, both theoretically and practically, a challenge for nonlinear dimensionality reduction techniques. Most commonly, these methods compute the embedding of a new point by finding near neighbors within the original data set and interpolating the embedded coordinates. Inversely, if given new coordinates within the embedding, an image for those coordinates is generated by linearly interpolating images which were embedded nearby [7)].

In some problem domains, there is additional information about the types of variation between images. For the cases of images which vary due to deformation, algorithms have been presented that explicitly solve for that deformation at each manifold coordinate using free form deformations [61]. More recent algorithms, seeking to characterize the changing shape of brains as they age, explicitly consider the infinite dimensional manifold of potential diffeomorphic transformations, and solve for a regression model characterizing one degree of freedom while ignoring others [15].

4.3 Interesting Topologies

The example image manifolds shown in Figs. 2 and 6 arise from underlying manifolds that have a cyclic structure (the flying cycle of the bird and the beating cycle of the heart). Given data points drawn from a manifold with an interesting topology, manifold learning algorithms have three potential behaviors. First, they may fail entirely to create an embedding that retains relationships between the original data points. Second, they may be able to output parameters in a Euclidean space that contains the appropriate manifold (e.g., embedding a one-D cyclic manifold as a circle on a plane). Third, they can directly provide coordinates correct for the type of the underlying manifold.

For simple cyclic manifolds, a 2-D Isomap embedding often succeeds at embedding points onto a circle, but in general Isomap requires that the distributions of the points on the underlying manifold be convex. Semi-Definite Embedding [68] relaxes this restriction and can, in certain cases, embed data samples with cyclic structures into a low-dimensional Euclidean space. Diffusion maps is another algorithm that is effective on cyclic manifolds. Furthermore, it is effective in cases where data is noisy enough that neighbors in the high-dimensional space are not always consistent with the intrinsic manifold structure [14].

A few algorithms seek to explicitly parametrize data sets with cyclic manifold structure. One approach for spherical manifolds is to estimate the curvature of the manifold in order to correct the geodesic distances computed by the Isomap algorithm, so that the new distances can be embedded in a Euclidean space of one higher dimension [47]. This approach has been extended to manifolds with negative curvature [3].

Learning manifolds with toroidal topological structure has been addressed by explicitly "cutting" the underlying neighborhood graph [19], or learning a kernel-based mapping from torus coordinates to the original data points [36]. Examining the connected components structure of the neighborhood graph allows for explicit reasoning about data sets drawn from multiple non-overlapping manifolds [71),72].

A harder problem arises when given data points from multiple, overlapping manifolds. These manifolds often occur in the context of human activity recognition, and may become increasingly important if manifold learning is to be applied to less structured data sets. The problem requires simultaneously clustering points and parametrize each point with respect to the correct manifold. This has been addressed through extensions to the LLE algorithm [26),49], Expectation Maximization approaches [30),59], and by applying minimum data encoding approaches [70].

5. Applications

To this point in the discussion, the manifold learning methods are simply used to find a low-dimensional embedding of the original image set. Typically, the underlying semantic content of the data is directly inferred from these parameters. This formed the basis of many of the earliest applications of image manifold learning where images were clustered or visualized with respect to their learned parameters. In this section, we summarize these methods and also present work that uses nonlinear dimensionality reduction as an initial step in order to facilitate common computer vision tasks such as video content analysis, pose estimation, image/video segmentation, and object tracking.

5.1 Content Analysis of Image Sets & Video

One of the first applications of image manifold learning was the visualization, or so-called *perceptual organization* of image sets. Most of the results depicted in the original Isomap and LLE papers were based on images, such as visualizing data sets of rendered faces and handwritten digits which varied due to a low number of degrees of freedom, pose and writing style, respectively. Additional work appearing shortly after these seminal papers used the algorithms to visualize large image sets [40], and classify images based on the embedded coordinates [66]. Directly using the learned parameters of manifold learning algorithms has also been used for face recognition [75),76] and estimating lighting directions [43),69] or head pose [50] from images. In Ref. 48), external domain knowledge was introduced to warp the output of nonlinear dimensionality reduction algorithms on images so that the learned parameters more closely matched the known causes of change

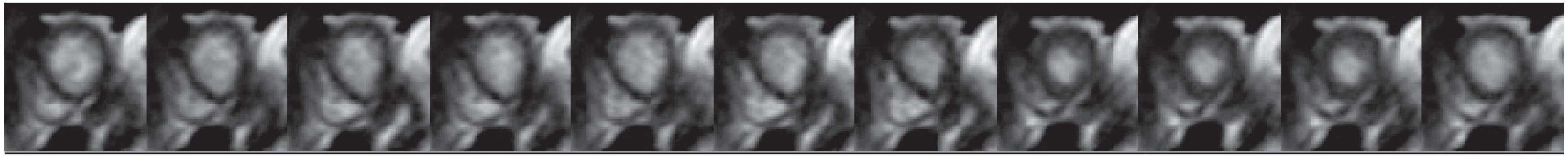


Fig. 6 Example frames from cine MRI of a beating heart.

in the image set.

The analysis of image sets evolved into video content analysis where the inherent temporal structure provided additional cues to the manifold structure of the set of frames. In Refs. 45) and 46) , image frames from videos of natural periodic motions (e.g., bird flying) were embedded using Isomap and certain patterns (e.g., cyclic, helical, knotted, and linear) emerged from the analysis of the distribution of points in the latent space. This work demonstrated that, for certain videos, the relationship between the learned parameters more accurately describes the relationship among frames in a video than the frame number and can be used for more intuitive video segmentation and temporal super-resolution. In Ref. 55), a similar approach was used for video compression where the frames were clustered based on their distances in latent space rather than temporal order.

5.2 Tracking and Pose Estimation

One area of computer vision where manifold learning has been successfully applied is in object tracking. The basic approach is supervised learning where latent variables are learned (generally using one of the manifold learning algorithms) and these latent variables are mapped to labeled (e.g., joint locations, silhouettes, contours) training images. For new test images, the latent variables are learned and the label (tracking output) is estimated.

In Refs. 27), 28), 39), human trackers are designed using dynamical models whose states live in the latent space of silhouettes, contours, or intensity images of humans walking. In Ref. 67), a similar approach is used to track other deformable objects, such as the contour of the lips while speaking and of fish swimming. In Refs. 38) and 42), face recognition and tracking systems are designed using a collection of sub-manifolds learned in training to detect and track faces in the presence of large rotations and occlusions.

There is also work which extends the idea of tracking silhouettes or contours in latent space to the problem of pose estimation for articulated objects. In Ref. 23), a style-independent pose manifold was learned for silhouettes of human walking and image sets of moving lips. In Ref. 22), *activity manifolds* are used to infer 3D body poses from image silhouettes. The forward and reverse mappings are learned from the ambient dimension of human silhouettes to the relevant latent space for a particular activity in addition to a mapping from latent space to 3D poses. The inference problem is then solved for new test silhouettes. In Ref. 37) and, similarly in Ref. 33), the ideas of the previous approach are extended to full tracking of 3D poses from video. In Ref. 57), *viewpoint manifolds* are modeled which represent the change in a high-dimensional action descriptor as a function of the camera viewpoint. The learned manifolds are used for action recognition and camera viewpoint estimation from human image silhouette sequences.

5.3 Medical Image Analysis

Medical image analysis, specifically of video data, is an area where manifold learning methods have shown promise and can have a significant impact. **Figure 6** depicts sample frames from cine MRI of a beating heart, rearranged to depict the image changes over the course of a single heartbeat cycle. In this example, the only cause of the image change is the non-rigid deformation due to the beating of the heart. Over the course of a single acquisition, many heartbeats take place and the resulting frames represent a sampling of a manifold which can be parametrized by the phase of the heartbeat. This scenario is typical of many medical image sets where the overall motion is periodic, there are only a small number of degrees of freedom, each (typically independent) cause of image change can lead to a complex image transformation (non-rigid deformation), and the latent space is close to uniformly-sampled. These properties make manifold learning an attractive choice for learning these latent parameters.

Image manifolds have been used to parametrize the complex non-rigid deformations commonly seen in medical data. In Refs. 58) and 61), cardiopulmonary image sets, similar to the one depicted in Fig. 6 are parametrized using Isomap and reordered based on these learned parameters. Under this ordering, there is minimal motion between neighboring images compared to the original temporal ordering, which simplifies the point correspondence problem and allows pairwise deformations to be estimated and extended into global deformation models. In Ref. 24), the learned parameters are used align CT volume slices of a lung in order to perform a 4D (3 spatial + time) reconstruction.

In Refs. 77) and 78), active contours for image segmentation are extended and applied to noisy cardiopulmonary images. The authors use Isomap with domain-specific distance metrics in order to learn a parametrization of the data with respect to the underlying degrees of freedom. The contours are then calculated over all of the images on the manifold simultaneously, using the learned parameters as additional shape constraints in a manner similar to how temporal constraints have been incorporated into active contours [2)].

6. Conclusions

In this article, we have tried to highlight the issues that most affect the special case of learning manifolds of images. Image data lies in an extremely high dimensional space, and often arises from naturally cyclic parameters; we believe that key future directions of research lie in more fully exploring manifold learning on curved surfaces and extending the topologies that can be effectively parametrized. Additionally, it is important to integrate the manifold learning process more tightly in specific image understanding applications, extending beyond what has already been demonstrated in image segmentation and visual tracking.

References

1) Agrafiotis, D.K. and Xu, H.: A self-organizing principle for learning nonlinear manifolds, *Proc. National Academy of Sciences*, Vol.99, pp.15869–15872 (2002).
2) Bascle, B. and Deriche, R.: Region Tracking Through Image Sequences, *Proc. IEEE International Conference on Computer Vision*, pp.302–307 (1995).
3) Begelfor, E. and Werman, M.: The world is not always flot or learning curved manifolds, Technical Report 65, School of Engineering and Computer Science, Hebrew University of Jerusalem (2005).
4) Belkin, M. and Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering, *Advances in Neural Information Processing Systems* (*NIPS*) (2002).
5) Bellman, R.: *Adaptive Control Processes: A Guided Tour*, Princeton University Press (1961).
6) Bengio, Y. and Monperrus, M.: Non-Local Manifold Tangent Learning, *Advances in Neural Information Processing Systems* (*NIPS*) (2005).
7) Bengio, Y., Paiement, J., Vincent, P., Delalleau, O., Roux, N.L. and Ouimet, M.: Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering, *Advances in Neural Information Processing Systems 16*, Thrun, S., Saul, L. and Schölkopf, B. (Eds.), MIT Press, Cambridge, MA (2004).
8) Bishop, C.M., Svensen, M. and Williams, C.K.I.: GTM: The Generative Topographic Mapping, *Neural Computation*, Vol.10, No.1, pp.215–234 (1998).
9) Brand, M.: Charting a Manifold, *Advances in Neural Information Processing Systems* (*NIPS*), S. Becker, S.T. and Obermayer, K. (Eds.), pp.961–968, MIT Press, Cambridge, MA, (2003).
10) Brand, M.: Continuous Nonlinear Dimensionality Reduction by Kernel Eigenmaps, *International Joint Conference on Artificial Intelligence* (*IJCAI*) (2003).
11) Brun, A.: Manifolds in Image Science and Visualization, PhD Thesis, Linköping University (2008).
12) Carreira-Perpiñán, M.A. and Zemel, R.S.: Proximity Graphs for Clustering and Manifold Learning, *Advances in Neural Information Processing Systems 17*, Saul, L.K., Weiss, Y. and Bottou, L. (Eds.), pp.225–232, MIT Press, Cambridge, MA, (2005).
13) Chen, H., Liu, T. and Fuh, C.: Learning Effective Image Metrics from Few Pairwise Examples, *Proc. IEEE International Conference on Computer Vision*, pp.II: 1371–1378 (2005).
14) Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F. and Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, *PNAS*, Vol.102, No.21, pp.7426–7431 (2005).
15) Davis, B., Fletcher, P., Bullitt, E. and Joshi, S.: Population Shape Regression from Random Design Data, *International Journal of Computer Vision*, Vol.Marr Prize Special Issue, p. to appear (2009). Marr Prize Special Issue.
16) de Ridder, D., Kouropteva, O., Okun, O., Pietikainen, M. and Duin, R.: Supervised locally linear embedding, *Artificial Neural Networks and Neural Information Processing, ICANN/ICONIP 2003 Proceedings, Lecture Notes in Computer Science 2714*, Springer, pp.333–341 (2003).

17) de Ridder, D., Loog, M. and Reinders, M.: Local Fisher Embedding, *Proc. International Conference on Pattern Recognition*, pp.295–298 (2004).
18) de Silva, V. and Tenenbaum, J.B.: Global Versus Local Methods in Nonlinear Dimensionality Reduction, *Advances in Neural Information Processing Systems* (*NIPS*), S. Becker, S.T. and Obermayer, K. (Eds.), MIT Press, Cambridge, MA, pp.705–712 (2003).
19) Dixon, M., Jacobs, N. and Pless, R.: Finding Minimal Parameterizations of Cylindrical Image Manifolds, *IEEE CVPR Workshop on Perceptual Organization in Computer Vision* (*POCV*), New York City, NY (2006).
20) Donoho, D. and Grimes, C.: When Does ISOMAP Recover the Natural Parameterization of Families of Articulated Images?, Technical report, Stanford University (2002).
21) Donoho, D.L. and Grimes, C.: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data, *PNAS*, Vol.100, No.10, pp.5591–5596 (2003).
22) Elgammal, A. and Lee, C.-S.: Inferring 3D body pose from silhouettes using activity manifold learning, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.681–688 (2004).
23) Elgammal, A. and su Lee, C.: Separating style and content on a nonlinear manifold, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.478–485 (2004).
24) Georg, M., Souvenir, R., Hope, A. and Pless, R.: Simultaneous data volume reconstruction and pose estimation from slice samples, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–6 (2008).
25) Gerber, S., Tasdizen, T. and Whitaker, R.: Robust non-linear dimensionality reduction using successive 1-dimensional Laplacian Eigenmaps, *ICML '07: Proc. 24th international conference on Machine learning*, New York, NY, USA, ACM, pp.281–288 (2007).
26) Goh, A. and Vidal, R.: Segmenting Motions of Different Types by Unsupervised Manifold Clustering, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–6 (2007).
27) Guo, F. and Qian, G.: 3D Human Motion Tracking using Manifold Learning, *Proc. IEEE International Conference on Image Processing*, Vol.1, pp.357–360 (2007).
28) Guo, F. and Qian, G.: Monocular 3D tracking of articulated human motion in silhouette and pose manifolds, *J. Image Video Process.*, Vol.2008, No.3, pp.1–18 (2008).
29) Hadsell, R., Chopra, S. and LeCun, Y.: Dimensionality Reduction by Learning an Invariant Mapping, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1735–1742 (2006).
30) Haro, G., Randall, G. and Sapiro, G.: Translated Poisson Mixture Model for Stratification Learning, *International Journal of Computer Vision*, Vol.80, No.3, pp.358–374 (2008).
31) He, X. and Niyogi, P.: Locality Preserving Projections, *Advances in Neural Information Processing Systems 16*, Thrun, S., Saul, L. and Schölkopf, B. (Eds.), MIT Press, Cambridge, MA (2004).
32) Hinton, G. and Roweis, S.: Stochastic Neighbor Embedding, *Advances in Neural Information Processing Systems* (*NIPS*) (2002).
33) Jaeggli, T., Koller-Meier, E. and Gool, L.V.: Multi-Activity Tracking in LLE Body Pose Space, *2nd Workshop on HUMAN MOTION Understanding, Modeling, Capture and Animation, ICCV* (2007).
34) Jenkins, O.C. and Mataric, M.J.: A spatio-temporal extension to Isomap nonlinear dimension reduction, *ICML '04: Twenty-first international conference on Machine learning*, New York, NY, USA, ACM Press (2004).
35) Kruskal, J.B. and Wish, M.: *Multidimensional Scaling* (*Quantitative Applications in the Social Sciences*), SAGE Publications (1978).
36) Lee, C.-S. and Elgammal, A.: Simultaneous Inference of View and Body Pose using Torus Manifolds, *Proc. International Conference on Pattern Recognition*, IEEE Computer Society, pp.489–494 (2006).
37) Lee, C.-S. and Elgammal, A.: Modeling View and Posture Manifolds for Tracking, *Proc. IEEE International Conference on Computer Vision*, pp.1–8 (2007).
38) Lee, K.-C. and Kriegman, D.: Online Learning of Probabilistic Appearance Manifolds for Video-Based Recognition and Tracking, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.852–859 (2005).
39) Lim, H., Camps, O.I., Sznaier, M. and Morariu, V.I.: Dynamic Appearance Modeling for Human Tracking, *CVPR '06: Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, IEEE Computer Society, pp.751–757 (2006).
40) Lim, I.S., Ciechomski, P. d.H., Sarni, S. and Thalmann, D.: Planar Arrangement of High-dimensional Biomedical Data Sets by Isomap Coordinates, *Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems* (*CBMS 2003*), pp.50–55 (2003). application.
41) Lin, T. and Zha, H.: Riemannian Manifold Learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.30, No.5, pp.796–809 (2008).
42) Liu, C.-B., Lin, R.-S., Yang, M.-H., Ahuja, N. and Levinson, S.: Object Tracking Using Globally Coordinated Nonlinear Manifolds, *Proc. International Conference on Pattern Recognition*, IEEE Computer Society, pp.844–847 (2006).
43) Murase, H. and Nayar, S.K.: Illumination Planning for Object Recognition Using Parametric Eigenspaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16, No.12, pp.1219–1227 (1994).
44) Park, J., Zhang, Z., Zha, H. and Kasturi, R.: Local Smoothing for Manifold Learning, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.452–459 (2004).
45) Pless, R.: Using Isomap to Explore Video Sequences, *Proc. IEEE International*

Conference on Computer Vision, pp.1433–1440 (2003).
46) Pless, R.: Images Spaces and Video Trajectories: Using Isomap to Explore Video Sequences, *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2003).
47) Pless, R. and Simon, I.: Embedding Images in non-Flat Spaces, *Proceedings of the International Conference on Imaging Science, Systems, and Technology* (2002).
48) Pless, R. and Simon, I.: Using thousands of images of an object, *International Conference on Computer Vision, Pattern Recognition and Image Processing* (2002).
49) Polito, M. and Perona, P.: Grouping and Dimensionality reduction by Locally Linear Embedding, *Advances in Neural Information Processing Systems* (*NIPS*), pp.1255–1262, MIT Press (2002).
50) Raytchev, B., Yoda, I. and Sakaue, K.: Head Pose Estimation by Nonlinear Manifold Learning, *Proc. International Conference on Pattern Recognition*, pp.462–466, IEEE Computer Society, (2004).
51) Ritter, H.J. and Kohonen, T.: Self-organizing semantic maps, *Biological Cybernetics*, Vol.61, pp.241–254 (1989).
52) Roweis, S.T. and Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, Vol.290, No.5500, pp.2323–2326 (2000).
53) Rubner, Y., Tomasi, C. and Guibas, L.J.: A Metric for Distributions with Applications to Image Databases, *Proc. IEEE International Conference on Computer Vision*, pp.59–66 (1998). distance metric.
54) Rumelhart, D.E. and Zipser, D.: *Feature discovery by competitive learning*, MIT Press, Cambridge, MA, USA (1986).
55) Schweitzer, H.: Computing Content-Plots for Video, *Proc. European Conference on Computer Vision*, pp.491–501 (2002).
56) Shirdhonkar, S. and Jacobs, D.W.: Approximate earth mover's distance in linear time, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8 (2008).
57) Souvenir, R. and Babbs, J.: Learning the Viewpoint Manifold for Action Recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2008).
58) Souvenir, R. and Pless, R.: Isomap and Non-Parametric Models of Image Deformation, *Proc. IEEE Workshop on Motion and Video Computing* (*WMVC*), Breckenridge, CO (2005).
59) Souvenir, R. and Pless, R.: Manifold Clustering, *Proc. IEEE International Conference on Computer Vision*, Beijing, China (2005).
60) Souvenir, R. and Pless, R.: Image distance functions for manifold learning, *Image Vision Comput.*, Vol.25, No.3, pp.365–373 (2007).
61) Souvenir, R., Zhang, Q. and Pless, R.: Image Manifold Interpolation using Free-Form Deformations, *Proc. IEEE International Conference on Image Processing*, pp. 1437–1440 (2006).
62) Tenenbaum, J.B., de Silva, V. and Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol.290, No.5500, pp.2319–2323 (2000).
63) Thompson, D.: *On Growth and Form*, Cambridge Universtiy Press (1917).
64) Verbeek, J.J., Roweis, S.T. and Vlassis, N.: Non-linear CCA and PCA by Alignment of Local Models, *Advances in Neural Information Processing Systems 16*, Thrun, S., Saul, L. and Schölkopf, B. (Eds.), MIT Press, Cambridge, MA (2004).
65) Verbeek, J., Vlassis, N. and Krose, B.: The Generative Self-Organizing Map: A Probabilistic Generalization of Kohonen's SOM, Technical Report IAS-UVA-02-03, University of Amsterdam, The Netherlands (2002).
66) Vlachos, M., Domeniconi, C., Gunopulos, D., Kollios, G. and Koudas, N.: Non-linear dimensionality reduction techniques for classification and visualization, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.645–651, ACM Press (2002).
67) Wang, Q., Xu, G. and Ai, H.: Learning Object Intrinsic Structure for Robust Visual Tracking, *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2003).
68) Weinberger, K.Q. and Saul, L.K.: Unsupervised Learning of Image Manifolds by Semidefinite Programming, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Washington D.C., pp.988–995 (2004).
69) Winnemöeller, H., Mohan, A., Tumblin, J. and Gooch, B.: Light Waving: Estimating Light Positions From Photographs Alone, *Computer Graphics Forum*, Vol.24, No.3, pp.433–438 (2005).
70) Wright, J., Tao, Y., Lin, Z., Ma, Y. and Shum, H.-Y.: Classification via Minimum Incremental Coding Length (MICL), *Advances in Neural Information Processing Systems 20*, Platt, J., Koller, D., Singer, Y. and Roweis, S. (Eds.), pp.1633–1640, MIT Press, Cambridge, MA, (2008).
71) Wu, Y. and Chan, K.L.: An Extended Isomap Algorithm for Learning Multi-Class Manifold, *Proc. IEEE International Conference on Machine Learning and Cybernetics* (*ICML*), Shanghai, China (2004).
72) Wu, Y., Chan, K.L. and Tsai, F.S.: F-Isomap: An Algorithm for Learning Fragmented Manifold, *Advances in Neural Information Processing Systems* (*NIPS*) (2004).
73) Xue, Y., Tong, C.S. and Zhang, W.: Survey of Distance Measures for NMF-Based Face Recognition, *International Conference on Computational Intelligence and Security* (*LNCS 4456*), pp.1039–1049 (2006).
74) Zha, H. and Zhang, Z.: Isometric Embedding and Continuum ISOMAP, *Proc. of the Twentieth International Conference on Machine Learning* (*ICML-2003*), Washington, DC (2003).
75) Zhang, J., Li, S.Z. and Wang, J.: Nearest manifold approach for face recognition, *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp.223–228 (2004).

76) Zhang, J., Li, S.Z. and Wang, J.: Manifold learning and applications in recognition, *in Intelligent Multimedia Processing with Soft Computing*, Springer-Verlag, pp.281–300 (2004).
77) Zhang, Q., Souvenir, R. and Pless, R.: Segmentation informed by manifold learning, *International workshop Energy Minimization Methods in Computer Vision and Pattern Recognition* (*EMMCVPR*) (2005).
78) Zhang, Q., Souvenir, R. and Pless, R.: On Manifold Structure of Cardiac MRI Data: Application to Segmentation, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1092–1098 (2006).

(Received November 11, 2008)
(Accepted November 30, 2008)
(Released March 31, 2009)

(Communicated by *Yoichi Sato*)

Robert Pless was born in 1972. He received his B.S. from Cornell University in 1994, and his Ph.D. from the University of Maryland in 2000. He has since been on the faculty at Washington University in St. Louis, and was promoted to Associate Professor in 2006. His current research interests are on mathemetical frameworks to understand visual motion for surveillance and medical imaging applications. He received a National Science Foundation Career Award and the Emerson Electric Excellence in Teaching Award in 2006. He has served as Chair of the IEEE Workshop on Omnidirectional Vision in 2003, and the MICCAI Workshop on Manifolds in Medical Imaging in 2008, and the IEEE Workshop on Motion and Video Computing in 2009.

Richard Souvenir was born in 1979. He received his D.Sc. degree from Washington University in St. Louis in 2006 where he was awarded the National Science Foundation Graduate Research Fellowship. Since then, he has been on the faculty at The University of North Carolina at Charlotte as an Assistant Professor in the Department of Computer Science. His research focuses on applying learning methods to problems in computer vision with applications to human motion analysis and medical imaging. He served as an Area Chair for the MICCAI Workshop on Manifolds in Medical Imaging in 2008 and is a member of the IEEE Computer Society.