Joint Manifolds for Data Fusion

Mark A. Davenport, *Student Member, IEEE*, Chinmay Hegde, *Student Member, IEEE*
Marco F. Duarte, *Member, IEEE,* and Richard G. Baraniuk, *Fellow, IEEE*

Abstract

The emergence of low-cost sensing architectures for diverse modalities has made it possible to deploy sensor networks that capture a single event from a large number of vantage points and using multiple modalities. In many scenarios, these networks acquire large amounts of very high-dimensional data. For example, even a relatively small network of cameras can generate massive amounts of high-dimensional image and video data. One way to cope with such a data deluge is to develop low-dimensional data models. Manifold models provide a particularly powerful theoretical and algorithmic framework for capturing the structure of data governed by a low-dimensional set of parameters, as is often the case in a sensor network. However, these models do not typically take into account dependencies among multiple sensors. We thus propose a new *joint manifold* framework for data ensembles that exploits such dependencies. We show that joint manifold structure can lead to improved performance for a variety of signal processing algorithms for applications including classification and manifold learning. Additionally, recent results concerning random projections of manifolds enable us to formulate a universal, network-scalable dimensionality reduction scheme that efficiently fuses the data from all sensors.

I. Introduction

The emergence of low-cost sensing devices has made it possible to deploy sensor networks that capture a single event from a large number of vantage points and using multiple modalities. This can lead to a veritable data deluge, fueling the need for efficient algorithms for processing and efficient protocols for transmitting the data generated by such networks. In order to address these challenges, there is a clear need for a theoretical framework for modeling the complex interdependencies among signals acquired by these networks. This framework should support the development of efficient algorithms that can exploit this structure and efficient protocols that can cope with the massive data volume.

MAD, CH, and RGB are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX. MFD is with the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ. This work was supported by grants NSF CCF-0431150 and CCF-0728867, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936 and N00014-08-1-1112, AFOSR FA9550-07-1-0301, ARO MURI W911NF-07-1-0185, ARO MURI W911NF-09-1-0383, and the TI Leadership Program. Thanks to J.P. Slavinsky for his help in acquiring the data for the experimental results presented in this paper.

Consider, for example, a camera network consisting of J video acquisition devices each acquiring N-pixel images of a scene simultaneously. Ideally, all cameras would send their raw recorded images to a central processing unit, which could then holistically analyze all the data produced by the network. This naïve approach would in general provide the best performance, since it exploits complete access to all of the data. However, the amount of raw data generated by a camera network, on the order of JN, becomes untenably large even for fairly small networks operating at moderate resolutions and frame rates. In such settings, the amount of data can (and often does) overwhelm network resources such as power and communication bandwidth. While the naïve approach could easily be improved by requiring each camera to first compress the images using a compression algorithm such as JPEG or MPEG, this modification still fails to exploit any interdependencies between the cameras. Hence, the total power and bandwidth requirements of the network will still grow linearly with J.

Alternatively, exploiting the fact that in many cases the end goal is to solve some kind of inference problem, each camera could independently reach a decision or extract some relevant features, and then relay the result to the central processing unit which would then combine the results to provide the solution. Unfortunately, this approach also has disadvantages. First, the cameras must be "smart" in that they must possess some degree of sophistication so that they can execute nonlinear inference tasks. Such technology is expensive and can place severe demands on the available power resources. Perhaps more importantly, the total power and bandwidth requirement will still scale linearly with J.

In order to cope with such high-dimensional data, a common strategy is to develop appropriate models for the acquired images. A powerful model is the geometric notion of a low-dimensional *manifold*. Informally, manifold models arise in cases where *(i)* a K-dimensional parameter θ can be identified that carries the relevant information about a signal and *(ii)* the signal $f(\theta) \in \mathbb{R}^N$ changes as a continuous (typically nonlinear) function of these parameters. Typical examples include a one-dimensional (1-D) signal translated by an unknown time delay (parameterized by the translation variable), a recording of a speech signal (parameterized by the underlying phonemes spoken by the speaker), and an image of a 3-D object at an unknown location captured from an unknown viewing angle (parameterized by the three spatial coordinates of the object as well as its roll, pitch, and yaw). In these and many other cases, the geometry of the signal class forms a nonlinear K-dimensional manifold in $\mathbb{R}^N$,

$$\mathcal{M} = \{f(\theta) : \theta \in \Theta\}, \tag{1}$$

where Θ is the K-dimensional parameter space. In recent years, researchers in image processing have

become increasingly interested in manifold models due to the observation that a collection of images obtained from different target locations/poses/illuminations and sensor viewpoints form such a manifold [1]–[3]. As a result, manifold-based methods for image processing have attracted considerable attention, particularly in the machine learning community and can be applied to diverse applications as data visualization, classification, estimation, detection, control, clustering, and learning [3]–[5]. Low-dimensional manifolds have also been proposed as approximate models for a number of nonparametric signal classes such as images of human faces and handwritten digits [6]–[8].

In sensor networks, multiple observations of the same event are often acquired simultaneously, resulting in the acquisition of interdependent signals that share a common parameterization. Specifically, a camera network might observe a single event from a variety of vantage points, where the underlying event is described by a set of common global parameters (such as the location and orientation of an object of interest). Similarly, when sensing a single phenomenon using multiple modalities, such as video and audio, the underlying phenomenon may again be described by a single parameterization that spans all modalities (such as when analyzing a video and audio recording of a person speaking, where both are parameterized by the phonemes being spoken). In both examples, all of the acquired signals are functions of the same set of parameters, i.e., we can write each signal as $f_j(\theta)$ where $\theta \in \Theta$ is the same for all j.

Our contention in this paper is that we can obtain a simple model that captures the correlation between the sensor observations by matching the parameter values for the different manifolds observed by the sensors. More precisely, we observe that by simply concatenating points that are indexed by the same parameter value θ from the different component manifolds, i.e., by forming $\boldsymbol{f}(\theta) = [f_1(\theta), f_2(\theta), \ldots, f_J(\theta)]$, we obtain a new manifold, which we dub the *joint manifold*, that encompasses all of the component manifolds and shares the same parameterization. This structure captures the interdependencies between the signals in a straightforward manner. We can then apply the same manifold-based processing techniques that have been proposed for individual manifolds to the entire ensemble of component manifolds.

In this paper we conduct a careful examination of the topological and geometrical properties of joint manifolds; in particular, we compare joint manifolds to their component manifolds to see how properties like geodesic distances, curvature, branch separation, and condition number are affected. We then observe that these properties lead to improved performance and noise-tolerance for a variety of signal processing algorithms when they exploit the joint manifold structure. As a key advantage of our proposed model, we illustrate how the joint manifold structure can be exploited via a simple and efficient data fusion algorithm based on *random projections*. For the case of J cameras jointly acquiring N-pixel images of a

common scene characterized by K parameters, we demonstrate that the total power and communication bandwidth required by our scheme is linear in the dimension K and only *logarithmic* in J and N. Recent developments in the field of compressive sensing has made this data acquisition model practical in many interesting applications [9]–[11].

Related prior work has studied *manifold alignment*, where the goal is to discover maps between datasets that are governed by the same underlying low-dimensional structure. Lafon et al. proposed an algorithm to obtain a one-to-one matching between data points from several manifold-modeled classes [12]. The algorithm first applies dimensionality reduction using diffusion maps to obtain data representations that encode the intrinsic geometry of the class. Then, an affine function that matches a set of landmark points is computed and applied to the remainder of the datasets. This concept was extended by Wang and Mahadevan, who applied Procrustes analysis on the dimensionality-reduced datasets to obtain an alignment function between a pair of manifolds [13]. Since an alignment function is provided instead of a data point matching, the mapping obtained is applicable for the entire manifold rather than for the set of sampled points. In our setting, we assume that either *(i)* the manifold alignment is implicitly present, for example, via synchronization between the different sensors, or *(ii)* the manifolds have been aligned using one of these approaches. Our main focus is an analysis of the benefits provided by analyzing the joint manifold versus solving the task of interest separately on each of the manifolds. For concreteness, but without loss of generality, we couch our analysis in the language of camera networks, although much of our theory is sufficiently generic so as to apply to a variety of other scenarios.

This paper is organized as follows. Section II introduces and establishes some basic properties of joint manifolds. Section III provides discussion of practical examples of joint manifolds in the camera network setting and describes how to use random projections to exploit the joint manifold structure in such a setting. Sections IV and V then consider the application of joint manifolds to the tasks of classification and manifold learning, providing both a theoretical analysis as well as extensive simulations. Section VI concludes with a brief discussion.

II. Joint Manifolds: Theory

In this section we develop a theoretical framework for ensembles of manifolds that are *jointly* parameterized by a small number of *common* degrees of freedom. Informally, we propose a data structure for jointly modeling such ensembles; this is obtained simply by concatenating points from different ensembles that are indexed by the same articulation parameter to obtain a single point in a higher-dimensional space.

We begin by defining the joint manifold for the setting of general topological manifolds.[1] In order to simplify our notation, we will let $\boldsymbol{\mathcal{M}} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_J$ denote the *product manifold*. Furthermore, we will use the notation $\boldsymbol{p} = (p_1, p_2, \ldots, p_J)$ to denote a J-tuple of points, or concatenation of J points, which lies in the Cartesian product of J sets (e.g., $\boldsymbol{\mathcal{M}}$).

Definition 1. *Let $\{\mathcal{M}_j\}_{j=1}^J$ be an ensemble of J topological manifolds of equal dimension K. Suppose that the manifolds are homeomorphic to each other, in which case there exists a homeomorphism ψ_j between $\mathcal{M}_1$ and $\mathcal{M}_j$ for each j. For a particular set of $\{\psi_j\}_{j=2}^J$, we define the* **joint manifold** *as*

$$\mathcal{M}^* = \{\boldsymbol{p} \in \boldsymbol{\mathcal{M}} : p_j = \psi_j(p_1), 2 \leq j \leq J\}.$$

Furthermore, we say that $\{\mathcal{M}_j\}_{j=1}^J$ are the corresponding **component manifolds***.*

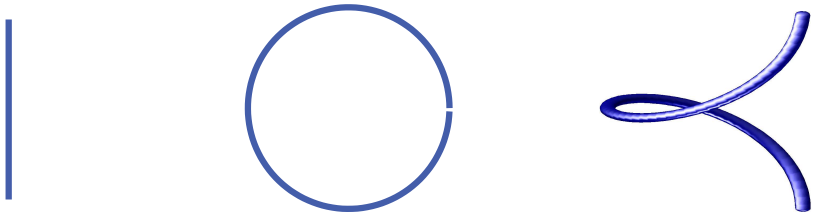
Note that $\mathcal{M}_1$ serves as a common *parameter space* for all the component manifolds. Since the component manifolds are homeomorphic to each other, this choice is ultimately arbitrary. In practice it may be more natural to think of each component manifold as being homeomorphic to some fixed K-dimensional parameter space Θ. However, in this case one could still define $\mathcal{M}^*$ as is done above by defining ψ_j as the composition of the homeomorphic mappings from $\mathcal{M}_1$ to Θ and from Θ to $\mathcal{M}_j$.

As an example, consider the one-dimensional manifolds in Fig. 1. Figures 1(a) and (b) show two isomorphic manifolds, where $\mathcal{M}_1 = (0, 2\pi)$ is an open interval, and $\mathcal{M}_2 = \{\psi_2(\theta) : \theta \in \mathcal{M}_1\}$ where $\psi_2(\theta) = (\cos(\theta), \sin(\theta))$, i.e., $\mathcal{M}_2 = S^1 \backslash (1, 0)$ is a circle with one point removed (so that it remains isomorphic to a line segment). In this case the joint manifold $\mathcal{M}^* = \{(\theta, \cos(\theta), \sin(\theta)) : \theta \in (0, 2\pi)\}$, illustrated in Fig. 1(c), is a helix. Notice that there exist other possible homeomorphic mappings from $\mathcal{M}_1$ to $\mathcal{M}_2$, and that the precise structure of the joint manifold as a submanifold of $\mathbb{R}^3$ is heavily dependent on the choice of this mapping.

Returning to the definition of $\mathcal{M}^*$, observe that although we have called $\mathcal{M}^*$ the joint manifold, we have not shown that it actually forms a topological manifold. To prove that $\mathcal{M}^*$ is indeed a manifold, we will make use of the fact that the joint manifold is a subset of the product manifold $\boldsymbol{\mathcal{M}}$. One can show that $\boldsymbol{\mathcal{M}}$ forms a JK-dimensional manifold using the product topology [14]. By comparison, we now show that $\mathcal{M}^*$ has dimension only K.

Proposition 1. *$\mathcal{M}^*$ is a K-dimensional submanifold of $\boldsymbol{\mathcal{M}}$.*

[1] We refer the reader to [14] for a comprehensive introduction to manifolds.



(a) $\mathcal{M}_1 \subseteq \mathbb{R}$: line segment (b) $\mathcal{M}_2 \subseteq \mathbb{R}^2$: circle segment (c) $\mathcal{M}^* \subseteq \mathbb{R}^3$: helix segment

Fig. 1. *A pair of isomorphic manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$, and the resulting joint manifold $\mathcal{M}^*$.*

Proof: We first observe that since $\mathcal{M}^* \subset \boldsymbol{\mathcal{M}}$, we automatically have that $\mathcal{M}^*$ is a second countable Hausdorff topological space. Thus, all that remains is to show that $\mathcal{M}^*$ is locally homeomorphic to $\mathbb{R}^K$. Suppose $\boldsymbol{p} \in \mathcal{M}^*$. Since $p_1 \in \mathcal{M}_1$, we have a pair (U_1, ϕ_1) such that $U_1 \subset \mathcal{M}_1$ is an open set containing p_1 and $\phi_1 : U_1 \to V$ is a homeomorphism where V is an open set in $\mathbb{R}^K$. We now define for $2 \leq j \leq J$ $U_j = \psi_j(U_1)$ and $\phi_j = \phi_1 \circ \psi_j^{-1} : U_j \to V$. Note that for each j, U_j is an open set and ϕ_j is a homeomorphism (since ψ_j is a homeomorphism).

Now set $\boldsymbol{U} = U_1 \times U_2 \times \cdots \times U_J$ and define $U^* = \boldsymbol{U} \cap \mathcal{M}^*$. Observe that U^* is an open set and that $\boldsymbol{p} \in U^*$. Furthermore, let $\boldsymbol{q}$ be any element of U^*. Then $\phi_j(q_j) = \phi_1 \circ \psi_j^{-1}(q_j) = \phi_1(q_1)$ for each $2 \leq j \leq J$. Thus, since the image of each $q_j \in U_j$ in V under their corresponding ϕ_j is the same, we can form a single homeomorphism $\phi^* : U^* \to V$ by assigning $\phi^*(q) = \phi_1(q_1)$. This shows that $\mathcal{M}^*$ is locally homeomorphic to $\mathbb{R}^K$ as desired. ■

Since $\mathcal{M}^*$ is a submanifold of $\boldsymbol{\mathcal{M}}$, it also inherits some desirable properties from $\{\mathcal{M}_j\}_{j=1}^J$.

Proposition 2. *Suppose that $\{\mathcal{M}_j\}_{j=1}^J$ are isomorphic topological manifolds and $\mathcal{M}^*$ is defined as above.*

1) *If $\{\mathcal{M}_j\}_{j=1}^J$ are Riemannian, then $\mathcal{M}^*$ is Riemannian.*
2) *If $\{\mathcal{M}_j\}_{j=1}^J$ are compact, then $\mathcal{M}^*$ is compact.*

Proof: The proofs of these facts are straightforward and follow from the fact that if the component manifolds are Riemannian or compact, then $\boldsymbol{\mathcal{M}}$ will be as well. $\mathcal{M}^*$ then inherits these properties as a submanifold of $\boldsymbol{\mathcal{M}}$ [14]. ■

Up to this point we have considered general topological manifolds. In particular, we have *not* assumed that the component manifolds are embedded in any particular space. If each component manifold $\mathcal{M}_j$ is embedded in $\mathbb{R}^{N_j}$, the joint manifold is naturally embedded in $\mathbb{R}^{N^*}$ where $N^* = \sum_{j=1}^J N_j$. Hence, the joint manifold can be viewed as a model for sets of data with *varying ambient dimension* linked

by a common parametrization. In the sequel, we assume that each manifold $\mathcal{M}_j$ is embedded in $\mathbb{R}^N$, which implies that $\mathcal{M}^* \subset \mathbb{R}^{JN}$. Observe that while the intrinsic dimension of the joint manifold remains constant at K, the ambient dimension increases by a factor of J. We now examine how a number of geometric properties of the joint manifold compare to those of the component manifolds.

We begin with the following simple observation that Euclidean distances[2] between points on the joint manifold are larger than distances on the component manifolds. The result follows directly from the definition of the Euclidean norm, so we omit the proof.

Proposition 3. *Let $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}^*$ be given. Then*

$$\|\boldsymbol{p} - \boldsymbol{q}\| = \sqrt{\sum_{j=1}^{J} \|p_j - q_j\|^2}.$$

While Euclidean distances are important (especially when noise is introduced), the natural measure of distance between a pair of points on a Riemannian manifold is not Euclidean distance, but rather the *geodesic distance*. The geodesic distance between points $p, q \in \mathcal{M}$ is defined as

$$d_{\mathcal{M}}(p, q) = \inf\{L(\gamma) : \gamma(0) = p, \gamma(1) = q\}, \tag{2}$$

where $\gamma : [0, 1] \to \mathcal{M}$ is a C^1-smooth curve joining p and q, and $L(\gamma)$ is the length of γ as measured by

$$L(\gamma) = \int_0^1 \|\dot{\gamma}(t)\| dt. \tag{3}$$

In order to see how geodesic distances on $\mathcal{M}^*$ compare to geodesic distances on the component manifolds, we will make use of the following lemma.

Lemma 1. *Suppose that $\{\mathcal{M}_j\}_{j=1}^J$ are Riemannian manifolds, and let $\boldsymbol{\gamma} : [0, 1] \to \mathcal{M}^*$ be a C^1-smooth curve on the joint manifold. Denote by γ_j the restriction of $\boldsymbol{\gamma}$ to the ambient dimensions of $\mathcal{M}^*$ corresponding to $\mathcal{M}_j$. Then each $\gamma_j : [0, 1] \to \mathcal{M}_j$ is a C^1-smooth curve on $\mathcal{M}_j$, and*

$$\frac{1}{\sqrt{J}} \sum_{j=1}^{J} L(\gamma_j) \leq L(\boldsymbol{\gamma}) \leq \sum_{j=1}^{J} L(\gamma_j).$$

Proof: We begin by observing that

$$L(\boldsymbol{\gamma}) = \int_0^1 \|\dot{\gamma}(t)\| dt = \int_0^1 \sqrt{\sum_{j=1}^{J} \|\dot{\gamma}_j(t)\|^2}\ dt. \tag{4}$$

[2]In the remainder of this paper, whenever we use the notation $\|\cdot\|$ we mean $\|\cdot\|_{\ell_2}$, i.e., the ℓ_2 (Euclidean) norm on $\mathbb{R}^N$. When we wish to differentiate this from other ℓ_p norms, we will be explicit.

For a fixed t, let $x_j = \|\dot{\gamma}_j(t)\|$, and observe that $\boldsymbol{x} = (x_1, x_2, \ldots, x_J)$ is a vector in $\mathbb{R}^J$. Thus we may apply the standard norm inequalities

$$\frac{1}{\sqrt{J}}\|\boldsymbol{x}\|_{\ell_1} \leq \|\boldsymbol{x}\|_{\ell_2} \leq \|\boldsymbol{x}\|_{\ell_1} \tag{5}$$

to obtain

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\|\dot{\gamma}_j(t)\| \leq \sqrt{\sum_{j=1}^{J}\|\dot{\gamma}_j(t)\|^2} \leq \sum_{j=1}^{J}\|\dot{\gamma}_j(t)\|. \tag{6}$$

Combining the right-hand side of (6) with (4) we obtain

$$L(\boldsymbol{\gamma}) \leq \int_0^1 \sum_{j=1}^{J}\|\dot{\gamma}_j(t)\|dt = \sum_{j=1}^{J}\int_0^1 \|\dot{\gamma}_j(t)\|dt = \sum_{j=1}^{J} L(\gamma_j).$$

Similarly, from the left-hand side of (6) we obtain

$$L(\boldsymbol{\gamma}) \geq \int_0^1 \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\|\dot{\gamma}_j(t)\|dt = \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\int_0^1 \|\dot{\gamma}_j(t)\|dt = \frac{1}{\sqrt{J}}\sum_{j=1}^{J} L(\gamma_j).$$

■

We are now in a position to compare geodesic distances on $\mathcal{M}^*$ to those on the component manifold.

Theorem 1. *Suppose that $\{\mathcal{M}_j\}_{j=1}^J$ are Riemannian manifolds. Let $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}^*$ be given. Then*

$$d_{\mathcal{M}^*}(\boldsymbol{p}, \boldsymbol{q}) \geq \frac{1}{\sqrt{J}}\sum_{j=1}^{J} d_{\mathcal{M}_j}(p_j, q_j). \tag{7}$$

If the mappings $\psi_2, \psi_3, \ldots, \psi_J$ are isometries, *i.e., $d_{\mathcal{M}_1}(p_1, q_1) = d_{\mathcal{M}_j}(\psi_j(p_1), \psi_j(q_1))$ for any j and for any pair of points $(\boldsymbol{p}, \boldsymbol{q})$, then*

$$d_{\mathcal{M}^*}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{\sqrt{J}}\sum_{j=1}^{J} d_{\mathcal{M}_j}(p_j, q_j) = \sqrt{J}\cdot d_{\mathcal{M}_1}(p_1, q_1). \tag{8}$$

Proof: If γ is a geodesic path between $\boldsymbol{p}$ and $\boldsymbol{q}$, then from Lemma 1,

$$d_{\mathcal{M}^*}(\boldsymbol{p}, \boldsymbol{q}) = L(\boldsymbol{\gamma}) \geq \frac{1}{\sqrt{J}}\sum_{j=1}^{J} L(\gamma_j).$$

By definition $L(\gamma_j) \geq d_{\mathcal{M}_j}(p_j, q_j)$; hence, this establishes (7).

Now observe that lower bound in Lemma 1 is derived from the lower inequality of (5). This inequality is attained with equality if and only if each term in the sum is equal, i.e., $L(\gamma_j) = L(\gamma_k)$ for all j and k. This is precisely the case when $\psi_2, \psi_3, \ldots, \psi_J$ are isometries. Thus we obtain

$$d_{\mathcal{M}^*}(\boldsymbol{p}, \boldsymbol{q}) = L(\boldsymbol{\gamma}) = \frac{1}{\sqrt{J}}\sum_{j=1}^{J} L(\gamma_j) = \sqrt{J}L(\gamma_1).$$

We now conclude that $L(\gamma_1) = d_{\mathcal{M}_1}(p_1, q_1)$ since if we could obtain a shorter path $\tilde{\gamma_1}$ from p_1 to q_1 this would contradict the assumption that $\boldsymbol{\gamma}$ is a geodesic on $\mathcal{M}^*$, which establishes (8). ■

Next, we study local smoothness and global self avoidance properties of the joint manifold.

Definition 2. [15] *Let $\mathcal{M}$ be a Riemannian submanifold of $\mathbb{R}^N$. The* **condition number** *is defined as $1/\tau$, where τ is the largest number satisfying the following: the open normal bundle about $\mathcal{M}$ of radius r is embedded in $\mathbb{R}^N$ for all $r < \tau$.*

The condition number controls both local smoothness properties and global properties of the manifold; as $1/\tau$ becomes smaller, the manifold becomes smoother and more self-avoiding, as observed in [15]. We will informally refer to manifolds with large τ as "good" manifolds.

Lemma 2. [15] *Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two distinct points on $\mathcal{M}$, and let $\gamma(t)$ denote a unit speed parameterization of the geodesic path joining p and q. Then*

$$\max_t \|\ddot{\gamma}(t)\| \leq \frac{1}{\tau}.$$

Lemma 3. [15] *Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points on $\mathcal{M}$ such that $\|p - q\| = d$. If $d \leq \tau/2$, then the geodesic distance $d_{\mathcal{M}}(p, q)$ is bounded by*

$$d_{\mathcal{M}}(p, q) \leq \tau(1 - \sqrt{1 - 2d/\tau}).$$

We wish to show that if the component manifolds are smooth and self avoiding, the joint manifold is as well. It is not easy to prove this in the most general case, where the only assumption is that there exists a homeomorphism (i.e., a continuous bijective map ψ) between every pair of manifolds. However, suppose the manifolds are *diffeomorphic*, i.e., there exists a continuous bijective map between tangent spaces at corresponding points on every pair of manifolds. In that case, we make the following assertion.

Theorem 2. *Suppose that $\{\mathcal{M}_j\}_{j=1}^J$ are Riemannian submanifolds of $\mathbb{R}^N$, and let $1/\tau_j$ denote the condition number of $\mathcal{M}_j$. Suppose also that the $\{\psi_j\}_{j=2}^J$ that define the corresponding joint manifold $\mathcal{M}^*$ are diffeomorphisms. If $1/\tau^*$ is the condition number of $\mathcal{M}^*$, then*

$$\frac{1}{\tau^*} \leq \max_{1 \leq j \leq J} \frac{1}{\tau_j}.$$

Proof: Let $\boldsymbol{p} \in \mathcal{M}^*$. Since the $\{\psi_j\}_{j=2}^J$ are diffeomorphisms, we may view $\mathcal{M}^*$ as being diffeomorphic to $\mathcal{M}_1$; i.e., we can build a diffeomorphic map from $\mathcal{M}_1$ to $\mathcal{M}^*$ as

$$\boldsymbol{p} = \psi^*(p_1) := (p_1, \psi_2(p_2), \ldots, \psi_J(p_J)).$$

We also know that given any two manifolds linked by a diffeomorphism $\psi_j : \mathcal{M}_1 \to \mathcal{M}_j$, each vector v_1 in the tangent space $T_1(p_1)$ of the manifold $\mathcal{M}_1$ at the point p_1 is *uniquely* mapped to a tangent vector $v_j := \phi_j(v_1)$ in the tangent space $T_j(p_j)$ of the manifold $\mathcal{M}_j$ at the point $p_j = \psi_j(p_1)$ through the map $\phi_j := \mathcal{J} \circ \psi_j(p_1)$, where $\mathcal{J}$ denotes the Jacobian operator.

Consider the application of this property to the diffeomorphic manifolds $\mathcal{M}_1$ and $\mathcal{M}^*$. In this case, the tangent vector $v_1 \in T_1(p_1)$ to the manifold $\mathcal{M}_1$ can be uniquely identified with a tangent vector $\boldsymbol{v} = \phi^*(v_1) \in T^*(p)$ to the manifold $\mathcal{M}^*$. This mapping is expressed as

$$\phi^*(v_1) = \mathcal{J} \circ \psi^*(p_1) = (v_1, \mathcal{J} \circ \psi_2(p_1), \ldots, \mathcal{J} \circ \psi_J(p_1)),$$

since the Jacobian operates componentwise. Therefore, the tangent vector $\boldsymbol{v}$ can be written as

$$\boldsymbol{v} = \phi^*(v_1) = (v_1, \phi_2(v_1), \ldots, \phi_J(p_1)).$$

In other words, a tangent vector to the joint manifold can be decomposed into J component vectors, each of which are tangent to the corresponding component manifolds.

Using this fact, we now show that a vector $\boldsymbol{\eta}$ that is normal to $\mathcal{M}^*$ can also be broken down into sub-vectors that are normal to the component manifolds. Consider $\boldsymbol{p} \in \mathcal{M}^*$, and denote $T^*(\boldsymbol{p})^\perp$ as the normal space at $\boldsymbol{p}$. Suppose $\boldsymbol{\eta} \in T^*(p)^\perp$. Decompose each η_j as a projection onto the component tangent and normal spaces, i.e., for $j = 1, \ldots, J$,

$$\eta_j = x_j + y_j, \qquad x_j \in T_j(p_j),\ y_j \in T_j(p_j)^\perp.$$

such that $\langle x_j, y_j \rangle = 0$ for each j. Then $\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{y}$, and since $\boldsymbol{x}$ is tangent to the joint manifold $\mathcal{M}^*$, we have $\langle \boldsymbol{\eta}, \boldsymbol{x} \rangle = \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} \rangle = 0$, and thus $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = -\|\boldsymbol{x}\|^2$. But, $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \sum_{j=1}^{J} \langle y_j, x_j \rangle = 0$. Hence $\boldsymbol{x} = 0$, i.e., each η_j is normal to $\mathcal{M}_j$.

Armed with this last fact, our goal now is to show that if $r < \min_{1 \le j \le J} \tau_j$ then the normal bundle of radius r is embedded in $\mathbb{R}^N$, or equivalently, for any $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}^*$, that $\boldsymbol{p} + \boldsymbol{\eta} \neq \boldsymbol{q} + \boldsymbol{\nu}$ provided that $\|\boldsymbol{\eta}\|, \|\boldsymbol{\nu}\| \le r$. Indeed, suppose $\|\boldsymbol{\eta}\|, \|\boldsymbol{\nu}\| \le r < \min_{1 \le j \le J} \tau_j$. Since $\|\eta_j\| \le \|\boldsymbol{\eta}\|$ and $\|\nu_j\| \le \|\boldsymbol{\nu}\|$ for all $1 \le j \le J$, we have that $\|\eta_j\|, \|\nu_j\| < \min_{1 \le i \le J} \tau_i \le \tau_j$. Since we have proved that η_j, ν_j are vectors in the normal bundle of $\mathcal{M}_j$ and their magnitudes are less than τ_j, then $p_j + \eta_j \neq q_j + \nu_j$ by the definition of condition number. Thus $\boldsymbol{p} + \boldsymbol{\eta} \neq \boldsymbol{q} + \boldsymbol{\nu}$ and the result follows. ■

This result states that for general manifolds, the most we can say is that the condition number of the joint manifold is guaranteed to be less than that of the *worst* manifold. However, in practice this is not likely to happen. As an example, Fig. 2 illustrates the point at which the normal bundle intersects itself

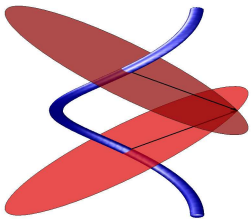


Fig. 2. *Point at which the normal bundle for the helix manifold from Fig. 1(c) intersects itself. Note that the helix has been slightly rotated.*

for the case of the joint manifold from Fig. 1(c). In this case we obtain $\tau^* = \sqrt{\pi^2/2 + 1} > 1$. Note that the condition numbers for the manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ generating $\mathcal{M}^*$ are given by $\tau_1 = \infty$ and $\tau_2 = 1$. Thus, while the condition number of the joint manifold is not as high as the best manifold, it is notably larger than that of the worst manifold. In general, even this example may be somewhat pessimistic and it is possible that the joint manifold may be better conditioned than even the best manifold.

III. Joint Manifolds: Practice

As noted in the Introduction, a growing number of algorithms exploit manifold models for tasks such as pattern classification, estimation, detection, control, clustering, and learning [3]–[5]. The performance of these algorithms often depends on the geometric properties of the manifold model, such as its condition number or geodesic distances along its surface. The theory developed in Section II suggests that the joint manifold preserves or improves such properties. In Sections IV and V we consider two illustrative applications and observe that when noise is added to the underlying signals, it can be extremely beneficial to use algorithms specifically designed to exploit the joint manifold structure. However, before we address these particular applications, we must first address some key practical concerns.

A. Acceptable deviations from theory

While manifolds are a natural way to model the structure of a set of images governed by a small number of parameters, the results in Section II make a number of assumptions concerning the structure of the component manifolds. In the most general case, we assume that the component manifolds are homeomorphic to each other. This means that between any pair of component manifolds there should exist a bijective mapping ϕ such that both ϕ and ϕ^{-1} are continuous. Such an assumption assures that the joint manifold is indeed a topological manifold. Unfortunately, this excludes some scenarios that can occur in practice. For example this assumption might not be applicable in a camera network featuring

non-overlapping fields of view. In such camera networks, there are cases in which only some cameras are sensitive to small changes in the parameter values. Strictly speaking, our theory may not apply in these cases, since the joint "manifold" as we have defined it is not necessarily even a topological manifold. As a result, one might expect to see significant performance degradation when exploiting techniques that heavily rely on this joint manifold structure. We provide additional discussion of this issues in Section V-B below, but for now we simply note that in Sections IV and V we conduct extensive experiments using both synthetic and real-world datasets and observe that in practice joint manifold-based processing techniques still exhibit significantly better performance than techniques that operate on each component manifold separately. While non-overlapping fields of view do pose a challenge (especially in the context of manifold learning), the fact that this results in non-homeomorphic manifolds seems to be more of a technical violation of our theory than a practical one.

In the context of manifold learning, we must actually assume that the component manifolds are isometric to each other. This is certainly not the case in a camera network with non-overlapping fields of view. Even with the restriction of a common field of view, this may seem an undue burden. In fact, this requirement is fulfilled by manifolds that are isometric to the parameter space that governs them—a class of manifolds that has been studied in [2]. Many examples from this class correspond to common image articulations that occur in vision applications, including:

- articulations of radially symmetric images, which are parameterized by a 2-D offset;
- articulations of four-fold symmetric images with smooth boundaries, such as discs, ℓ_p balls, etc.;
- pivoting of images containing smooth boundaries, which are parameterized by the pivoting angle;
- articulations of $\frac{K}{2}$ discs over distinct non-overlapping regions, with $\frac{K}{2} > 1$, producing a K-dimensional manifold.

These examples can be extended to objects with piecewise smooth boundaries as well as to video sequences of such articulations. In Section V we describe heuristics for dealing with the problem of non-overlapping fields of view and provide a number of experiments that suggest that these heuristics can overcome violations of the isometry assumption in practice.

In our theoretical results concerning condition number, we assume that the component manifolds are smooth, but the manifolds induced by the motion of an object where there are sharp edges or occlusions are nowhere differentiable. This problem can be addressed by applying a smoothing kernel to each captured image, inducing a smooth manifold [3]. More generally, we note that if the cameras have moderate computational capabilities, then it may be possible to perform simple preprocessing tasks such

as segmentation, background subtraction, and illumination compensation that will make the manifold assumption more rigorously supported in practice. This may be necessary in scenarios such as those involving multiple objects or challenging imaging conditions.

B. Efficient data fusion via joint manifolds using linear projections

Observe that when the number J and ambient dimension N of the manifolds become large, the ambient dimension of the joint manifold—JN—may be so large that it becomes impossible to perform any meaningful computations. Furthermore, it might appear that in order to exploit the joint manifold structure, we must collect all the data at a central location, which we earlier claimed was potentially impossible. In order to address this problem, we must exploit the joint manifold structure to develop a more efficient fusion scheme.

Specifically, given a network of J cameras, let $x_j \in \mathbb{R}^N$, denote the image acquired by camera j, which is assumed to belong in a manifold $\mathcal{M}_j$, and let $\boldsymbol{x}$ denote the corresponding point in the joint manifold $\mathcal{M}^*$. Rather than forming the vector $\boldsymbol{x}$, one could potentially estimate a K-dimensional parameter vector $\widehat{\theta}_j$ via the nonlinear mapping of x_j corresponding to the manifold $\mathcal{M}_j$. By collecting the $\widehat{\theta}_j$ at a central location, we would obtain a data representation of dimension JK. By simply concatenating each $\widehat{\theta}_j$, this approach essentially ignores the joint manifold structure present in the data, which is evident due to the fact that in an ideal setting the same K parameters will be obtained from each of the J cameras. Moreover, given noisy estimates for $\widehat{\theta}_j$, it is not obvious how to most effectively integrate the $\widehat{\theta}_j$ to obtain a single joint K-dimensional representation. Finally, while this approach eliminates the dependence on N, it still suffers from a linear dependence on J.

To address this challenge, we observe that if we had access to the vector $\boldsymbol{x}$, then we could exploit the joint manifold structure to map it to a parameter vector $\widehat{\theta}$ of length only K rather than JK. Unfortunately, this mapping will generally be nonlinear, and each element of $\widehat{\theta}$ could potentially depend on the entire vector $\boldsymbol{x}$, preventing us from operating individually on each x_j. Thus, rather than directly extract the features, we will instead restrict our focus to *linear* dimensionality reduction methods that, while acting on the concatenated data $\boldsymbol{x}$, can be implemented in a distributed fashion.

Specifically, we will aim to compute a dimensionally reduced representation of $\boldsymbol{x}$ denoted $\boldsymbol{y} = \boldsymbol{\Phi x}$, where $\boldsymbol{\Phi}$ is a standard linear projection operator. Since the operator is linear, we can take *local* projections of the images acquired by each camera, and still calculate the *global* projections of $\boldsymbol{x}$ in a distributed fashion. Let each camera calculate $y_j = \Phi_j x_j$, with the matrices $\Phi_j \in \mathbb{R}^{M \times N}, 1 \leq j \leq J$. Then, by

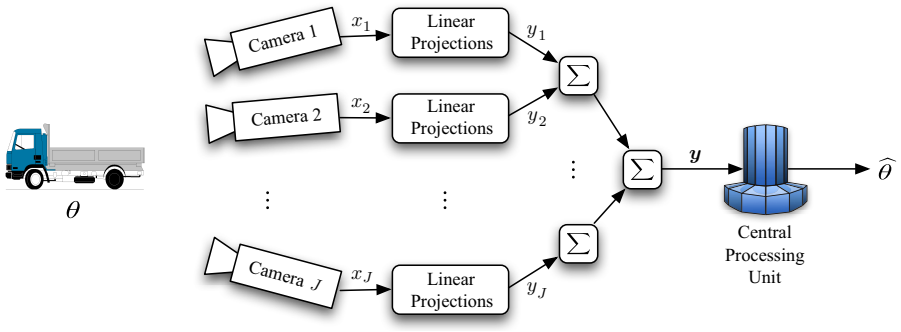


Fig. 3. *Distributed data fusion using linear projections in a camera network.*

defining the $M \times JN$ matrix $\mathbf{\Phi} = [\Phi_1 \ \Phi_2 \ \cdots \ \Phi_J]$, the global projections $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}$ can be obtained by

$$\begin{aligned}\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x} &= [\Phi_1 \quad \Phi_2 \quad \cdots \quad \Phi_J][x_1^T \quad x_2^T \quad \cdots \quad x_J^T]^T \\ &= \Phi_1 x_1 + \Phi_2 x_2 + \cdots + \Phi_J x_J.\end{aligned}$$

Thus, the final measurement vector can be obtained by simply *adding independent projections* of the images acquired by the individual cameras. This gives rise to the *compressive data fusion* protocol illustrated in Fig. 3. Suppose the individual cameras are associated with the nodes of a binary tree of size J, where the edges represent communication links between nodes. Let the root of the tree denote the final destination of the fused data (the central processing unit). Then the fusion process can be represented by the flow of data from the leaves to the root, with a binary addition occurring at every parent node. Recalling that the dimensionality of the data is M and the depth of the tree is $R = O(\log J)$, we observe that the *total* communication bandwidth requirement is given by $R \times M = O(M \log J)$, i.e., the communication burden grows only logarithmically in J.

The main challenge in designing such a scheme is the choice of a suitable matrix $\mathbf{\Phi}$. Given a specific joint manifold $\mathcal{M}^*$, there may be an optimal $\mathbf{\Phi}$ that preserves the Euclidean and the geodesic structures of $\mathcal{M}^*$ while ensuring that M is comparable to the dimension K of the joint manifold (and hence much less than the ambient dimension JN). Unfortunately, the general problem of computing an optimal linear projection of a manifold remains unsolved and, in our context, finding this projection matrix would also require full knowledge of the objects to be classified as well as the position/orientation of each camera in the network. Such information would typically not be available within the network.

Fortunately, we can exploit recent results concerning *random projections* to solve this problem without any prior knowledge of the structure of the network or the objects to be captured. Specifically, it

has been shown that the essential structure of a K-dimensional manifold with condition number $1/\tau$ residing in $\mathbb{R}^N$ is approximately preserved under an orthogonal projection into a random subspace of dimension $O(K\log(N/\tau)) \ll N$ [16]. This result has been leveraged in the design of efficient algorithms for inference applications, such as classification using multiscale navigation [17], intrinsic dimension estimation [18], and manifold learning [18]. In our context, this result implies that if the joint manifold has bounded condition number as given by Theorem 2, then we can project the joint data into a random subspace of dimension that is only logarithmic in J and N and still approximately preserve the manifold structure. This is formalized in the following theorem, which follows directly from [16].

Theorem 3. *Let $\mathcal{M}^*$ be a compact, smooth, Riemannian joint manifold in a JN-dimensional space with condition number $1/\tau^*$. Let $\mathbf{\Phi}$ denote an orthogonal linear mapping from $\mathcal{M}^*$ into a random M-dimensional subspace of $\mathbb{R}^{JN}$. Let $M = O(K\log(JN/\tau^*)/\epsilon^2)$. Then, with high probability, the geodesic and Euclidean distances between any pair of points on $\mathcal{M}^*$ are preserved up to distortion ϵ under $\mathbf{\Phi}$.*

Thus, we obtain a faithful embedding of the manifold via a representation of dimension only $O(K\log JN)$. This represents a massive improvement over the original JN-dimensional representation, and for large values of J it can be a significant improvement over the JK-dimensional representation obtained by performing separate (nonlinear) dimensionality reduction on each component manifold.[3]

Recalling that the total communication bandwidth required for our compressive data fusion scheme is $O(M\log J)$, we obtain that when using random projections the dependence of the required bandwidth on J is $O(\log^2 J)$; this offers a significant improvement from previous data fusion methods that necessarily require the communication bandwidth to scale linearly with the number of cameras. Joint manifold fusion via linear projections integrates the network transmission and data fusion steps in a fashion similar to the protocols discussed in randomized gossiping [20] and compressive wireless sensing [21].

Joint manifold fusion via random projections, like compressive sensing [9]–[11], is *universal* in that the measurement process is not dependent on the specific structure of the manifold. Thus, our sensing techniques need not be replaced for these extensions; only our underlying models (hypotheses) are updated. We have not discussed complicating factors such as clutter, varying or unknown backgrounds, etc. Promising progress in [22] suggests that such difficulties can potentially be handled even after acquiring

[3]Note that if we were interested in compressing only a fixed number of images P, we could apply the Johnson-Lindenstrauss lemma [19] to obtain that $M = O(\log P)$ would be sufficient to obtain the result in Theorem 3. However, the value of M required in Theorem 3 is independent of the number of images P, and therefore provides scalability to extremely large datasets.

the random projections. Furthermore, recent years have also seen the development of devices capable of *directly* acquiring arbitrary linear projections of the images [11], [23]. Our fusion scheme can directly operate on the measurements provided by such devices.

IV. Joint Manifold Classification

We now examine the application of joint manifold-based techniques to some common inference problems. In this section we will consider the problem of binary classification when the two classes corresponds to different manifolds. As an example, we will consider the scenario where a camera network acquires images of an unknown vehicle, and we wish to classify between two vehicle types. Since the location of the vehicle is unknown, each class forms a distinct low-dimensional manifold in the image space. The performance of a classifier in this setting will depend partially on topological quantities of the joint manifold described in Section II, which in particular provide the basis for the random projection-based version of our algorithms. However, the most significant factor determining the performance of the joint manifold-based classifier is of a slightly different flavor. Specifically, the probability of error is determined by the distance between the manifolds. Thus, we also provide additional theoretical analysis of how distances between the joint manifolds compare to those between the component manifolds.

A. Theory

The problem of manifold-based classification is defined as follows: given manifolds $\mathcal{M}$ and $\mathcal{N}$, suppose we observe a signal $y = x + n \in \mathbb{R}^N$ where either $x \in \mathcal{M}$ or $x \in \mathcal{N}$ and n is a noise vector, and we wish to find a function $f : \mathbb{R}^N \to \{\mathcal{M}, \mathcal{N}\}$ that attempts to determine which manifold "generated" y.

This problem has been explored in the context of automatic target recognition (ATR) using high-dimensional noisy data arising from a single manifold [24], [25]; the authors of this body of work adopt a fully Bayesian approach in deriving fundamental bounds on the performance of certain types of classification methods. A full analysis of our proposed framework for joint manifolds using this approach is beyond the scope of this paper; we defer that to future research. Instead, we consider a classification algorithm based on the *generalized maximum likelihood* framework described in [26]. The approach is to classify by computing the distance from the observed signal y to each manifold, and then classify based on which of these distances is smallest; i.e., our classifier is

$$f(y) = \arg\min \left[d(y, \mathcal{M}), d(y, \mathcal{N}) \right], \tag{9}$$

where $d(y, \mathcal{M}) = \inf_{z \in \mathcal{M}} \|y - x\|$. We will measure the performance of this algorithm by considering the probability of misclassifying a point from $\mathcal{M}$ as belonging to $\mathcal{N}$, which we denote $P_{\mathcal{MN}}$.

To analyze this problem, we employ three common notions of separation in metric spaces:

- The *minimum separation* distance between two manifolds $\mathcal{M}$ and $\mathcal{N}$, defined as

$$\delta(\mathcal{M},\mathcal{N}) = \inf_{p\in\mathcal{M}} d(p,\mathcal{N}).$$

- The *maximum separation* distance between manifolds $\mathcal{M}$ and $\mathcal{N}$, defined as

$$\Delta(\mathcal{M},\mathcal{N}) = \sup_{x\in\mathcal{M}} \sup_{y\in\mathcal{N}} \|x-y\|.$$

- The *Hausdorff distance* from $\mathcal{M}$ to $\mathcal{N}$, defined as

$$D(\mathcal{M},\mathcal{N}) = \sup_{p\in\mathcal{M}} d(p,\mathcal{N}).$$

Note that while $\delta(\mathcal{M},\mathcal{N}) = \delta(\mathcal{N},\mathcal{M})$ and $\Delta(\mathcal{M},\mathcal{N}) = \Delta(\mathcal{N},\mathcal{M})$, in general $D(\mathcal{M},\mathcal{N}) \neq D(\mathcal{N},\mathcal{M})$. As one might expect, $P_{\mathcal{MN}}$ is controlled by the separation distances. For example, suppose that $x \in \mathcal{M}$; if the noise vector n is bounded and satisfies $\|n\| < \delta(\mathcal{M},\mathcal{N})/2$, then we have that $d(y,\mathcal{M}) \leq \|n\| < \delta(\mathcal{M},\mathcal{N})/2$ and hence

$$\begin{aligned}\delta(\mathcal{M},\mathcal{N}) &= \inf_{p\in\mathcal{M},q\in\mathcal{N}} \|p-q\| = \inf_{p\in\mathcal{M},q\in\mathcal{N}} \|p-y+y-q\| \\ &\leq \inf_{p\in\mathcal{M},q\in\mathcal{N}} \|p-y\| + \|y-q\| = d(y,\mathcal{M}) + d(y,\mathcal{N}) < \delta(\mathcal{M},\mathcal{N})/2 + d(y,\mathcal{N}).\end{aligned}$$

Thus we are guaranteed that $d(y,\mathcal{N}) > \delta(\mathcal{M},\mathcal{N})/2$. Therefore, $d(y,\mathcal{M}) < d(y,\mathcal{N})$ and the classifier defined by (9) satisfies $P_{\mathcal{MN}} = 0$. We can refine this result in two possible ways. A first possible refinement is to note that the amount of noise that we can tolerate without making an error depends on x. Specifically, for a given $x \in \mathcal{M}$, provided that $\|n\| \leq d(x,\mathcal{N})/2$ we still have that $P_{\mathcal{MN}} = 0$. Thus, for a given $x \in \mathcal{M}$ we can tolerate noise bounded by $d(x,\mathcal{N})/2 \in [\delta(\mathcal{M},\mathcal{N})/2, D(\mathcal{M},\mathcal{N})/2]$.

A second possible refinement is to ignore this dependence on x while extending our noise model to the case where $\|n\| > \delta(\mathcal{M},\mathcal{N})/2$ with non-zero probability. We can still bound $P_{\mathcal{MN}}$, since

$$P_{\mathcal{MN}} \leq P(\|n\| > \delta(\mathcal{M},\mathcal{N})/2). \tag{10}$$

Thus, we now provide lower bounds on these separation distances. The corresponding upper bounds are available in [27]. In the interest of space, the proof of this and subsequent theorems are omitted and can be found in [27].

Theorem 4. *Consider the joint manifolds $\mathcal{M}^* \subset \boldsymbol{\mathcal{M}}$ and $\mathcal{N}^* \subset \boldsymbol{\mathcal{N}}$. Then, the following bounds hold:*

$$\delta^2(\mathcal{M}^*,\mathcal{N}^*) \geq \sum_{j=1}^{J} \delta^2(\mathcal{M}_j,\mathcal{N}_j), \tag{11}$$

$$\Delta^2(\mathcal{M}^*,\mathcal{N}^*) \geq \max_{1\leq k\leq J} \left(\Delta^2(\mathcal{M}_k,\mathcal{N}_k) + \sum_{j\neq k} \delta^2(\mathcal{M}_j,\mathcal{N}_j) \right), \tag{12}$$

$$D^2(\mathcal{M}^*,\mathcal{N}^*) \geq \max_{1\leq k\leq J} \left(D^2(\mathcal{M}_k,\mathcal{N}_k) + \sum_{j\neq k} \delta^2(\mathcal{M}_j,\mathcal{N}_j) \right). \tag{13}$$

As an example, if we consider the case where the separation distances are constant for all j, then the joint minimum separation distance satisfies $\sqrt{J}\delta(\mathcal{M}_1,\mathcal{N}_1) \leq \delta(\mathcal{M}^*,\mathcal{N}^*)$, and using the upper bound for $\delta(\mathcal{M}^*,\mathcal{N}^*)$ from [27], we obtain

$$\delta(\mathcal{M}^*,\mathcal{N}^*) \leq \sqrt{\delta^2(\mathcal{M}_1,\mathcal{N}_1) + (J-1)\Delta^2(\mathcal{M}_1,\mathcal{N}_1)} \leq \delta(\mathcal{M}_1,\mathcal{N}_1) + \sqrt{J-1}\Delta(\mathcal{M}_1,\mathcal{N}_1).$$

In the case where $\delta(\mathcal{M}_1,\mathcal{N}_1) \ll \Delta(\mathcal{M}_1,\mathcal{N}_1)$, we observe that $\delta(\mathcal{M}^*,\mathcal{N}^*)$ can be considerably larger than $\sqrt{J}\delta(\mathcal{M}_1,\mathcal{N}_1)$. This means that we can potentially tolerate much more noise while ensuring $P_{\mathcal{M}^*\mathcal{N}^*} = 0$. To see this, let $\boldsymbol{n}$ denote a noise vector and recall that we require $\|n_j\| < \epsilon = \delta(\mathcal{M}_j,\mathcal{N}_j)/2$ to ensure that $P_{\mathcal{M}_j\mathcal{N}_j} = 0$. Thus, if we require that $P_{\mathcal{M}_j\mathcal{N}_j} = 0$ for all j, then we have that

$$\|\boldsymbol{n}\| = \sqrt{\sum_{j=1}^{J} \|n_j\|^2} < \sqrt{J}\epsilon = \sqrt{J}\delta(\mathcal{M}_1,\mathcal{N}_1)/2.$$

However, if we instead only require that $P_{\mathcal{M}^*\mathcal{N}^*} = 0$, then we only need $\|\boldsymbol{n}\| < \delta(\mathcal{M}^*,\mathcal{N}^*)/2$, which can be a significantly less stringent requirement.

The benefit of classification using the joint manifold is more apparent when we extend our noise model to the case where we allow $\|n_j\| > \delta(\mathcal{M}_j,\mathcal{N}_j)/2$ with non-zero probability and apply (10). To bound the probability in (10), we will make use of the following adaptation of Hoeffding's inequality [28].

Lemma 4. *Suppose that $n_j \in \mathbb{R}^N$ is a random vector that satisfies $\|n_j\|^2 \leq \epsilon$, for $j = 1, 2, \ldots, J$. Suppose also that the n_j are independent and identically distributed (i.i.d.) with $E[\|n_j\|^2] = \sigma^2$. Then if $\boldsymbol{n} = (n_1, n_2, \ldots, n_J) \in \mathbb{R}^{JN}$, we have that for any $\lambda > 0$,*

$$P\left(\|\boldsymbol{n}\|^2 > J(\sigma^2 + \lambda)\right) \leq \exp\left(-\frac{2J\lambda^2}{\epsilon^2}\right).$$

Note that this lemma holds for any distribution on the noise that is bounded. For Gaussian noise it is possible to establish a very similar result (differing only in the constant inside the exponent) by using

standard tail bounds. Using this lemma we can relax the assumption on ϵ so that we only require that it is finite, and instead make the weaker assumption that $E[\|\boldsymbol{n}\|^2] = J\sigma^2 \leq \delta^2(\mathcal{M},\mathcal{N})/4$ for a particular pair of manifolds $\mathcal{M}$, $\mathcal{N}$. This assumption ensures that $\lambda = \delta^2(\mathcal{M},\mathcal{N})/4 - \sigma^2 > 0$, so that we can combine Lemma 4 with (10) to obtain a bound on $P_{\mathcal{MN}}$. Note that if this condition does not hold, then this is a very difficult classification problem since the *expected* norm of the noise is large enough to push us closer to the other manifold, in which case the simple classifier given by (9) makes little sense.

We now illustrate how Lemma 4 can be be used to compare error bounds between classification using a joint manifold versus using a pair of component manifolds. The proof can be found in [27].

Theorem 5. *Suppose that we observe $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ where $\boldsymbol{x} \in \mathcal{M}^*$ and $\boldsymbol{n}$ is a random vector such that $\|n_j\|^2 \leq \epsilon$, for $j = 1, 2, \ldots, J$, and that the n_j are i.i.d. with $E[\|n_j\|^2] = \sigma^2 \leq \delta^2(\mathcal{M}_k,\mathcal{N}_k)/4$. Define*

$$c_k = \delta^2(\mathcal{M}_k,\mathcal{N}_k)/4 - \sigma^2 \qquad \text{and} \qquad c^* = \delta^2(\mathcal{M}^*,\mathcal{N}^*)/4 - \sigma^2.$$

If

$$\delta(\mathcal{M}_k,\mathcal{N}_k) \leq \frac{\delta(\mathcal{M}^*,\mathcal{N}^*)}{\sqrt{J}}, \tag{14}$$

and we classify the observation y according to (9) then $c^* > c_k$ and

$$P_{\mathcal{M}^*\mathcal{N}^*} \leq \exp\left(-2\left(\frac{c^*}{\epsilon}\right)^2\right) \qquad \text{and} \qquad P_{\mathcal{M}_k\mathcal{N}_k} \leq \exp\left(-2\left(\frac{c_k}{\epsilon}\right)^2\right). \tag{15}$$

This result can be weakened slightly to obtain the following corollary [27].

Corollary 1. *Suppose that we observe $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ where $\boldsymbol{x} \in \mathcal{M}^*$ and $\boldsymbol{n}$ is a random vector such that $\|n_j\|^2 \leq \epsilon$, for $j = 1, 2, \ldots, J$ and that the n_j are i.i.d. with $E[\|n_j\|^2] = \sigma^2 \leq \delta^2(\mathcal{M}_k,\mathcal{N}_k)/4$. If*

$$\delta^2(\mathcal{M}_k,\mathcal{N}_k) \leq \frac{\sum_{j\neq k}\delta^2(\mathcal{M}_j,\mathcal{N}_j)}{J-1}, \tag{16}$$

and we classify according to (9), then (15) holds with the same constants as in Theorem 5.

Corollary 1 shows that we can expect a classifier based on the joint manifold to outperform a classifier based the k-th component manifold whenever the squared separation distance for the k-th component manifolds is comparable to the average squared separation distance among the remaining component manifolds. Thus, we can expect the joint classifier to outperform most of the individual classifiers, but it is still possible that some individual classifiers will do better. Of course, if one knew in advance which classifiers were best, then one would only use data from the best classifiers. We expect that more typical

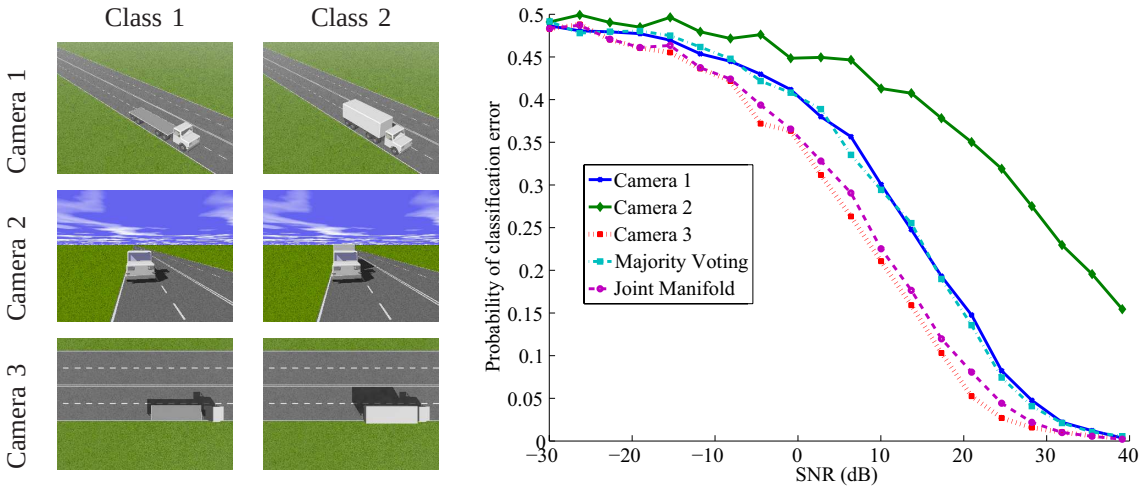


Fig. 4. *Sample images of 2 different trucks from multiple camera views and SNR vs. probability of error for individual cameras, the joint manifold, and majority voting. The number of pixels in each camera image* $N = 240 \times 360 = 76800$. *Joint manifold-based classification outperforms majority voting and performs nearly as well as the best camera.*

situations include the case where the best classifier changes over time or where the separation distances are nearly equal for all component manifolds, in which case the condition in (16) is true for all k.

B. Experiments

In this section, we apply the random projection-based fusion algorithm to perform binary classification. Suppose a number of synthetic cameras, each with resolution N, observe the motion of a truck along a straight road.[4] This forms a 1-D manifold in the image space $\mathbb{R}^N$ pertaining to each camera; the joint manifold is also a 1-D manifold in $\mathbb{R}^{JN}$. Suppose now that we wish to classify between two types of trucks. Example images from three camera views for the two classes are shown in Fig. 4. The resolution of each image is $N = 240 \times 320 = 76800$ pixels. In our experiment, we convert the images to grayscale and sum $M = 200$ random projections for the three camera views. The sample camera views suggest that some views make it easier to distinguish between the classes than others. For instance, the head-on view of the two trucks is very similar for most shift parameters, while the side view is more appropriate for discerning between the two classes of trucks.

The probability of error, which in this case is given by $\frac{1}{2}P_{\mathcal{MN}} + \frac{1}{2}P_{\mathcal{NM}}$, for different manifold-based

[4]Our synthetic images were generated using POVRAY (*http://www.povray.org*), an open-source ray tracing software package.

classification approaches as a function of the signal-to-noise ratio (SNR) is shown in Fig. 4. It is clear that the joint manifold approach performs better than majority voting and is comparable in performance to the best camera. While one might hope to be able to do even better than the best camera, Theorem 5 suggests that in general this is only possible when no camera is significantly better than the average camera. Moreover, in the absence of prior information regarding how well each camera truly performs, the best strategy for the central processor would be to fuse the data from *all* cameras. Thus, joint manifold fusion proves to be more effective than high-level fusion algorithms like majority voting.

This example highlights two scenarios when our proposed approach should prove useful. First, our method acts as a simple scheme for data fusion in the case when most cameras do not yield particularly reliable data (and thus decision fusion algorithms like majority voting are ineffective.) Second, due to the high dimensionality of the data, transmitting the images could be expensive in terms of communication bandwidth. Our method ensures that the communication cost is reduced to be proportional only to the number of degrees of freedom of the signal.

V. Joint Manifold Learning

In contrast to the classification scenario described above, in which we knew the manifold structure *a priori*, we now consider *manifold learning* algorithms that attempt to learn the manifold structure from a set of samples of a manifold. This is accomplished by constructing a (possibly nonlinear) embedding of the data into a subset of $\mathbb{R}^L$, where $L < N$. If the dimension K of the manifold is known, then L is typically set to K. Such algorithms provide a powerful framework for navigation, visualization and interpolation of high-dimensional data. For instance, manifold learning can be employed in the inference of articulation parameters (e.g., 3-D pose) from a set of images of a moving object. In this section we demonstrate that in a variety of settings, the joint manifold is significantly easier to learn than the individual component manifolds. This improvement is due to both the kind of increased robustness to noise noted in Section IV and to the fact that, as was shown in Theorem 2, the joint manifold can be significantly better-conditioned than the component manifolds, meaning that it is easier to learn the structure of the joint manifold from a finite sampling of points.

A. Theory

Several algorithms for manifold learning have been proposed, each giving rise to a nonlinear map with its own special properties and advantages (e.g., Isomap [29], Locally Linear Embedding (LLE) [30], Hessian Eigenmaps [31], etc.) Of these approaches, we devote special attention here to the Isomap

algorithm, which assumes that the point cloud consists of samples from a data manifold that is (at least approximately) isometric to a convex subset of Euclidean space. In this case, there exists an isometric mapping f from a parameter space $\Theta \subseteq \mathbb{R}^K$ to the manifold $\mathcal{M}$ such that $d_{\mathcal{M}}(f(\theta_1), f(\theta_2)) = \|\theta_1 - \theta_2\|_2$ for all $\theta_1, \theta_2 \in \Theta$. In essence, Isomap attempts to discover the inverse mapping $f^{-1} : \mathcal{M} \to \Theta$.

Isomap works in three stages:

1) Construct a graph G that contains a vertex for each data point; an edge connects two vertices if the Euclidean distance between the corresponding data points is below a specified threshold.
2) Weight each edge in the graph G by computing the Euclidean distance between the corresponding data points. We then estimate the geodesic distance between each pair of vertices as the length of the shortest path between the corresponding vertices in the graph G.
3) Embed the points in $\mathbb{R}^K$ using multidimensional scaling (MDS), which attempts to embed the points so that their Euclidean distance approximates the estimated geodesic distances.

A crucial component of the MDS algorithm is a suitable linear transformation of the matrix of squared geodesic distances; the rank-K approximation of this new matrix yields the best possible K-dimensional coordinate structure of the input sample points in a mean-squared sense. Further results on the performance of Isomap in terms of geometric properties of the underlying manifold can be found in [32].

We examine the performance of Isomap for learning the joint manifold as compared to learning the J isometric component manifolds separately. We assume that we have noiseless samples from $\{\mathcal{M}_j\}_{j=1}^J$. In order to judge the quality of the embedding learned by Isomap, we will observe that for any pair of samples p, q from a manifold $\mathcal{M}$ whose vertices are linked within the graph G, we have that

$$\rho \leq \frac{\|p - q\|}{d_{\mathcal{M}}(p, q)} \leq 1 \tag{17}$$

for some $\rho \in [0, 1]$ that will depend on the samples of $\mathcal{M}$ and the graph G. Isomap will perform well if the largest value of ρ that satisfies (17) for any pair of samples that are connected by an edge in the graph G is close to 1. Using this fact, we can compare the performance of manifold learning using Isomap on samples from the joint manifold $\mathcal{M}^*$ to using Isomap on samples from a particular component manifold $\mathcal{M}_k$. The proof of this theorem can again be found in [27].

Theorem 6. *Let $\mathcal{M}^*$ be a joint manifold from J isometric component manifolds. Let $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}^*$ and suppose that we are given a graph G that contains one vertex for each sample obtained from $\mathcal{M}^*$. For*

each $k = 1, \ldots, J$, define ρ_j as the largest value such that

$$\rho_j \leq \frac{\|p_j - q_j\|}{d_{\mathcal{M}_j}(p_j, q_j)} \leq 1 \tag{18}$$

for all pairs of points connected by an edge in G. Then we have that

$$\sqrt{\frac{\sum_{j=1}^{J} \rho_j^2}{J}} \leq \frac{\|\boldsymbol{p} - \boldsymbol{q}\|}{d_{\mathcal{M}^*}(\boldsymbol{p}, \boldsymbol{q})} \leq 1. \tag{19}$$

From Theorem 6 we see that, in many cases, the joint manifold estimates of the geodesic distances will be more accurate than the estimates obtained using one of the component manifolds. If for a particular component manifold $\mathcal{M}_k$ we observe that $\rho_k \leq \sqrt{\frac{\sum_{j=1}^{J} \rho_j^2}{J}}$, then we know that the joint manifold leads to better estimates. Essentially, we may expect that the joint manifold will lead to estimates that are better than the average case across the component manifolds.

We now consider the case where we have a dense sampling of the manifolds so that the $\rho_j \approx 1$, and examine the case where we obtain noisy samples. We will assume that the noise is i.i.d. and demonstrate that any distance calculation performed on $\mathcal{M}^*$ serves as a better estimator of the pairwise (and consequently, geodesic) distances between any two points $\boldsymbol{p}$ and $\boldsymbol{q}$ than that performed on any component manifold using the points p_j and q_j. Again, the proof of this theorem can be found in [27].

Theorem 7. *Let $\mathcal{M}^*$ be a joint manifold from J isometric component manifolds. Let $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}^*$ and assume that $\|p_j - q_j\| = d$ for all j. Assume that we acquire noisy observations $\boldsymbol{s} = \boldsymbol{p} + \boldsymbol{n}$ and $\boldsymbol{r} = \boldsymbol{q} + \boldsymbol{n}'$, where $\boldsymbol{n}$ and $\boldsymbol{n}'$ are independent noise vectors with $\mathbb{E}[\|n_j\|^2] = \mathbb{E}[\|n_j'\|^2] = \sigma^2$, $\|n_j\|^2 \leq \epsilon$, and $\|n_j'\|^2 \leq \epsilon$ for $j = 1, \ldots, J$. Then,*

$$P\left(1 - \delta \leq \frac{\|\boldsymbol{s} - \boldsymbol{r}\|^2}{\|\boldsymbol{p} - \boldsymbol{q}\|^2 + 2J\sigma^2} \leq 1 + \delta\right) \geq 1 - 2c^{-J^2},$$

where $c = \exp\left(2\delta^2 \left(\frac{d^2 + 2\sigma^2}{d\sqrt{\epsilon} + \epsilon}\right)^2\right)$.

We observe that the estimate of the true distance suffers from a constant small bias; this can be handled using a simple debiasing step.[5] Theorem 7 indicates that the probability of large deviations in the estimated distance decreases *exponentially* in the number of component manifolds J; thus we should observe significant "denoising" even in the case where J is relatively small.

[5]Manifold learning algorithms such as Isomap deal with biased estimates of distances by "centering" the matrix of squared distances, i.e., removing the mean of each row/column from every element.

B. Practice

The theoretical results derived above assume that the acquired data arises from J isometric component manifolds. As noted in Section III-A, barring controlled or synthetic scenarios, this is very rarely the case. In practice, the isometric assumption breaks down due to two reasons: *(i)* the cameras may be at different distances from the scene, nonidentical cameras may possess different dynamic ranges, or the cameras may be of different modalities (such as visual versus infrared cameras or even visual plus audio data), and thus the component manifolds may be scaled differently; *(ii)* due to occlusions or partially-overlapping fields of view, certain regions of some component manifolds may be ill-conditioned.

In order to handle such non-idealities, we make two modifications to the Isomap algorithm while performing joint manifold learning. Recall that that in order to find the nearest-neighbor graph G, Isomap must first calculate the matrix of squared pairwise Euclidean distances. We denote this matrix $\mathbf{D}$ for the joint manifold $\mathcal{M}^*$ and D_j for the component manifold $\mathcal{M}_j$. Note that $\mathbf{D} = \sum_{j=1}^{J} D_j$. Thus, if a particular component manifold is scaled differently than the others, by which we mean that $d_{\mathcal{M}_j}(f_j(\theta_1), f_j(\theta_2)) = C_j\|\theta_1 - \theta_2\|_2$ with $C_j \neq 1$, then all the entries of the corresponding D_j will be reweighted by C_j^2, so that D_j will have a disproportionate impact on $\mathbf{D}$. This corresponds to the first non-ideality described above, and can be alleviated by *normalizing* each D_j by its Frobenius norm, which can be interpreted as scaling each manifold so that an Eulerian path through the complete graph has unit length.

The second non-ideality can be partially addressed by attempting to adaptively detect and correct for occlusion events. Consider the case of large-scale occlusions, in which we make the simplifying assumption that for each camera the object of interest is either entirely within the camera's view or entirely occluded. In this case, the non-occluded component manifolds are still locally isometric to each other, i.e., there exists a neighborhood U such that $d_{\mathcal{M}_j}(f_j(\theta_1), f_j(\theta_2)) = \|\theta_1 - \theta_2\|_2$ for all $\theta_1, \theta_2 \in U$ and for all j corresponding to the non-occluded component manifolds. Thus, if we knew which cameras were occluded for a pair of points, say $\boldsymbol{x}^m$ and $\boldsymbol{x}^n$, then we could simply ignore those cameras in computing $\mathbf{D}_{m,n}$ and rescale $\mathbf{D}_{m,n}$ so that it is comparable with the case when no cameras exhibit occlusions. More specifically, we let $\widetilde{J}$ denote the index set for non-occluded component manifolds and set $\mathbf{D}_{m,n} = (|J|/|\widetilde{J}|)\sum_{j\in\widetilde{J}} \|x_j^m - x_j^n\|_2^2$. To do this automatically, we *threshold* $\|x_j^m - x_j^n\|_2^2$ to zero when it is below a certain value, i.e., we set $\widetilde{J} = \{j : \|x_j^m - x_j^n\|_2^2 \geq \epsilon\}$ for some parameter ϵ, since this for the component manifolds in which the object of interest is occluded this distance will be relatively small. The parameter ϵ can be reasonably inferred from the data. $\mathbf{D}$ is used by subsequent steps in Isomap to learn an improved low-dimensional embedding of the high-dimensional acquired data. Note that while

this approach does not rigorously handle boundary cases where objects are only partially occluded, our experimental results below indicate that the algorithms are robust to such cases.

C. Experiments

We provide a variety of results using both simulated and real data that demonstrate the significant gains obtained by using the joint manifold model, both with and without the use of random projections. The manifold learning results have been generated using Isomap [29]. For ease of presentation, all of our experiments are performed on 2-D image manifolds isomorphic to a closed rectangular subset of $\mathbb{R}^2$. Thus, ideally the 2-D embedding of the data should resemble a rectangular grid of points that correspond to the samples of the joint manifold in high dimensional space.

1) Manifolds isometric to Euclidean space: As a first example, we consider three different manifolds formed by $N = 64 \times 64 =$ 4096 pixel images of an ellipse with major axis a and minor axis b translating in a 2-D plane, for $(a, b) = (7, 7),\ (7, 6)$ and $(7, 5)$; an example point is shown in Fig. 5. The eccentricity of the ellipse directly affects the condition number $1/\tau$ of the image articulation manifold; in fact, it can be shown that manifolds associated with more eccentric ellipses exhibit higher values for the condition number. Consequently, we expect that it is "harder" to learn such manifolds. Figure 5 shows that this is indeed the case. We add a small amount of i.i.d. Gaussian noise to each image and apply Isomap to both the individual datasets as well as the concatenated dataset. We observe that the 2-D rectangular embedding is poorly learnt for each of the component manifolds but improves visibly for the joint manifold.

2) Gaussian noise in realistic images: We demonstrate how using joint manifolds can help ameliorate imaging artifacts such as Gaussian noise in a more realistic setting. We test our proposed joint manifold learning approach on a set of synthetic truck images. The data comprises a set of 540 views of a truck on a highway from 3 vantage points. Each image is of size $N = 90 \times 120 =$ 10800. The images are parametrized by the 2-D location of the truck on the road; thus, each of the image data sets can be modeled by a 2-D manifold. Sample views are shown in Fig. 4; for this experiment, we only use images from Class 2. We convert the images to grayscale, so that the ambient dimension of the data from each camera lies in $\mathbb{R}^{10800}$. Next, we add i.i.d. Gaussian noise to each image and attempt to learn the 2-D manifold. The noise level is quite high (PSNR = 3.5dB), as evidenced by the sample images in Fig. 6. It is visually clear from the 2-D embedding results that the learning performance improves markedly when the data is modeled using a joint manifold, thus providing numerical evidence for Theorem 7.

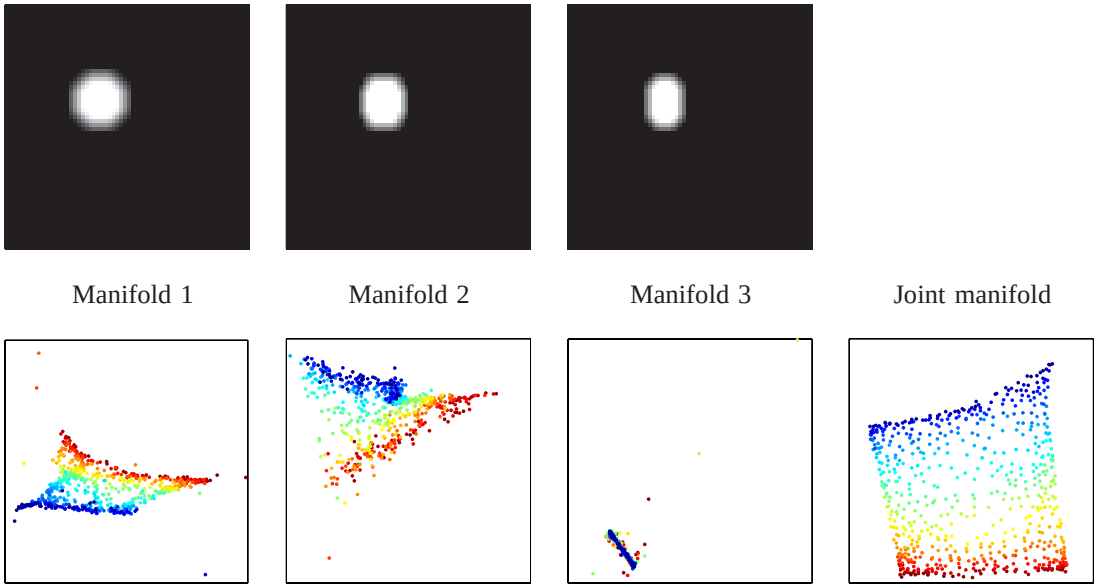


Fig. 5. *(top) Articulation manifolds sharing a common 2-D parameter space Θ. Images simulate viewing a translating disc from $J = 3$ viewing angles. (bottom) 2-D embedding of individual and joint manifolds learned via Isomap.*

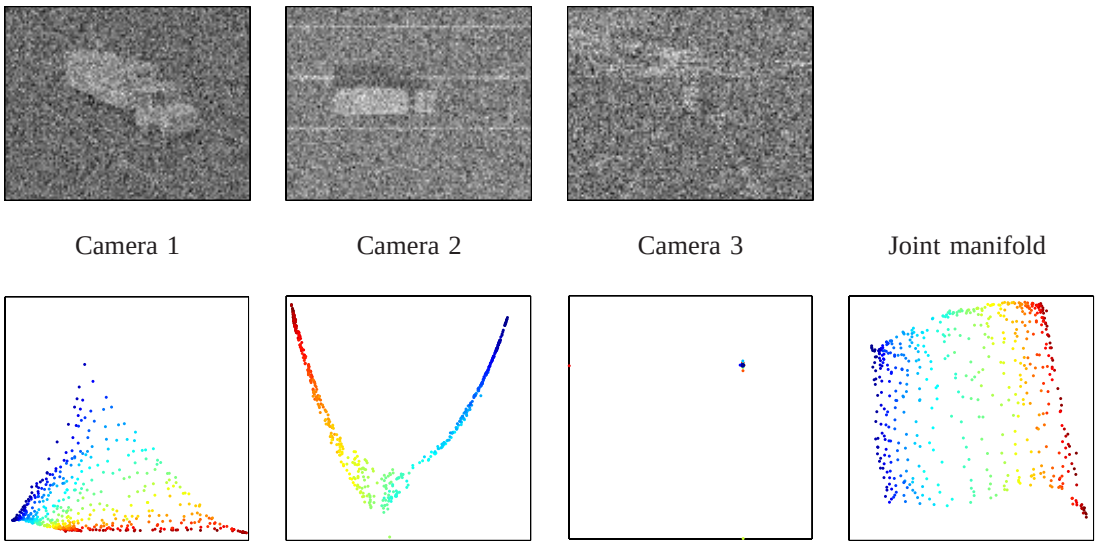


Fig. 6. *(top) Noisy captured images. SNR $\approx$ 3.5 dB ; (bottom) 2-D embeddings learned via Isomap from noisy images. The joint manifold model helps ameliorate the effects of noise.*

3) Real data experiment—learning with occlusions: We now test our methods on data from a camera network; the images are obtained from a network of four Unibrain Fire-i$^{\text{TM}}$ OEM Firewire board cameras. Each camera has resolution $N = 320 \times 240 = 76800$ pixels. The data comprises $J = 4$ different views of

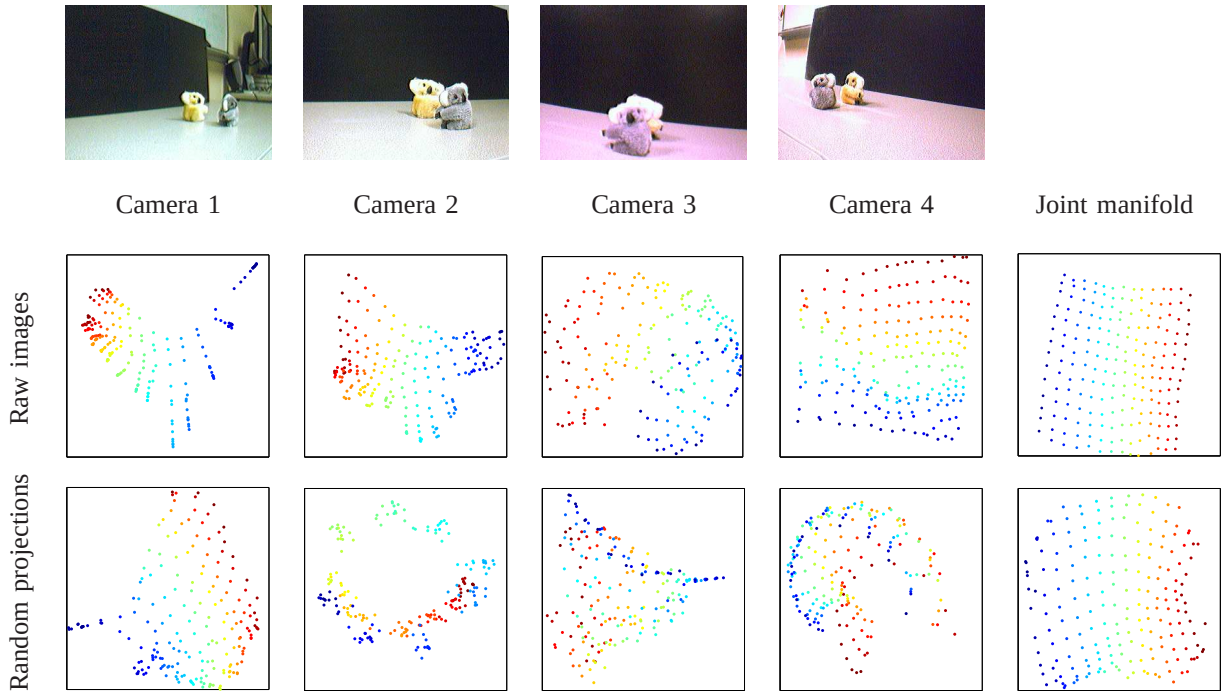


Fig. 7. *(top) Sample images of 2 koalas moving along individual 1-D paths, yielding a 2-D manifold; (middle) 2-D embeddings of the dataset learned via Isomap from $N = 76800$ pixel images; (bottom) 2-D embeddings of the dataset learned from $M = 2400$ random projections. Learning the joint manifold yields a much improved 2-D embedding.*

the independent motions of 2 toy koalas along individual 1-D paths, yielding a 2-D combined parameter space. This data suffers from real-world artifacts such as fluctuations in illumination conditions and variations in the pose of the koalas; further, the koalas occlude one another in certain views or are absent from certain views depending on the particular vantage point. Sample images and 2-D embedding results are displayed in Fig. 7. We observe that the best embedding is obtained by using the modified version of Isomap for learning the joint manifold. To test the effectiveness of the data fusion method described in Section III-B, we compute $M = 2400$ random projections of each image and sum them to obtain a randomly projected version of the joint data and repeat the above experiment. The dimensionality of the projected data is 3.125% of the original data; yet, we see very little degradation in performance, thus displaying the effectiveness of random projection-based fusion.

4) Real data experiment—unsupervised target tracking: As a practical application of manifold learning, we consider a situation where we are given a set of training data consisting of images of a target moving through a region along with a set of test images of the target moving along a particular trajectory. We do not explicitly incorporate any known information regarding the locations of the cameras or the parameter

space describing the target's motion. The training images comprise $J = 4$ views of a coffee mug placed at different positions on an irregular rectangular grid. Example images from each camera are shown in Fig. 8. For the test data, we translate the coffee mug so that its 2-D path traces out the shape of the letter "R". We aim to recover this shape using both the test and training data. To solve this problem, we attempt to learn a 2-D embedding of the joint manifold using the modified version of Isomap detailed in Section V-B. The learned embedding for each camera is shown in Fig. 8. As is visually evident, learning the data using any one camera yields very poor results; however learning the joint manifold helps discern the 2-D structure to a much better degree. In particular, the "R" trajectory in the test data is correctly recovered only by learning the joint manifold. Finally, we repeat the above procedure using $M = 4800$ random projections of each image, and fuse the data by summing the measurement vectors. While the recovered trajectory of the anomalous (test) data suffers some degradation in visual quality, we observe comparable 2-D embedding results for the individual and joint manifolds as with the original data set. Since the dimensionality of the projected data is merely 6.25% that of the original data set, this would translate to significant savings in communication costs in a real-world camera network.

VI. Discussion

Joint manifolds naturally capture the structure present in the data produced by camera networks. We have studied topological and geometric properties of joint manifolds, and have provided some basic examples that illustrate how they can improve the performance of common signal processing algorithms. We have also introduced a simple framework for data fusion for camera networks that employs independent random projections of each image, which are then accumulated to obtain an accurate low-dimensional representation of the joint manifold. Our fusion scheme can be directly applied to the data acquired by such devices. Furthermore, while we have focused primarily on camera networks in this paper, our framework can be used for the fusion of signals acquired by many generic sensor networks, as well as multimodal and joint audio/video data.

References

[1] H. Lu, *Geometric Theory of Images*, Ph.D. thesis, University of California, San Diego, 1998.

[2] D. L. Donoho and C. Grimes, "Image manifolds which are isometric to Euclidean space," *J. Math. Imaging and Computer Vision*, vol. 23, no. 1, July 2005.

[3] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, August 2005.

[4] M. Belkin and P. Niyogi, "Using manifold structure for partially labelled classification," in *Advances in NIPS*, vol. 15. MIT Press, 2003.

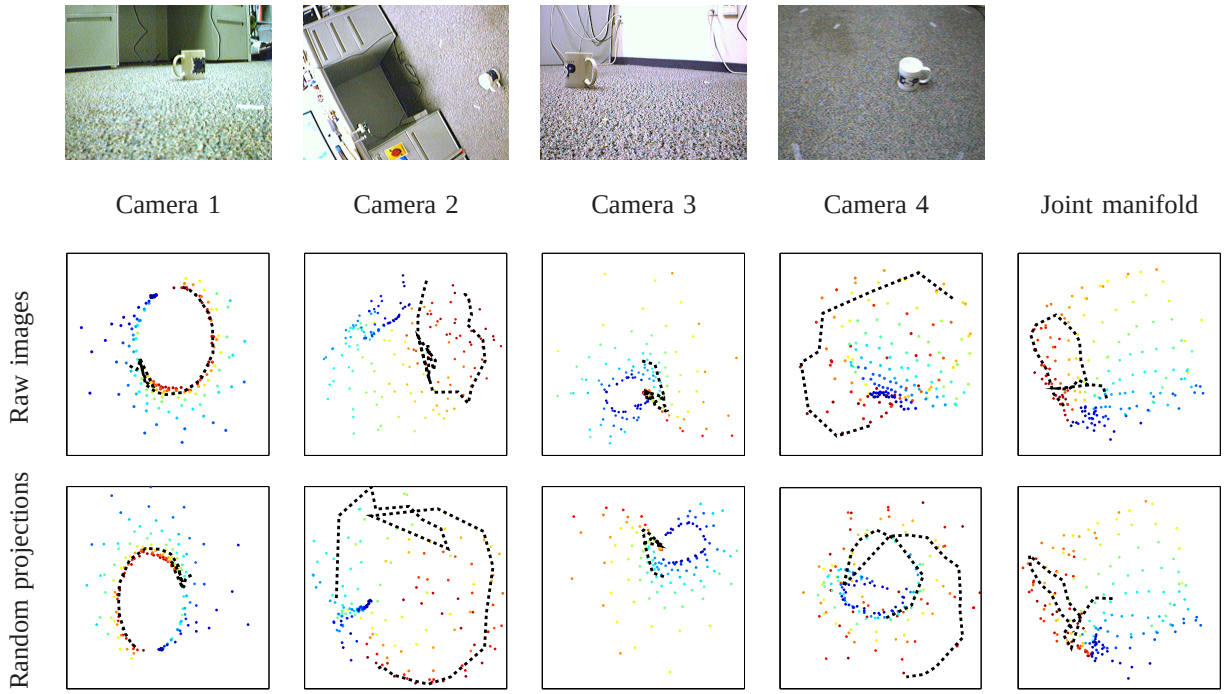


Fig. 8. *(top) Sample images of the 2-D movement of a coffee mug; (middle) 2-D embeddings of the dataset learned via Isomap from $N = 76800$ pixel images; (bottom) 2-D embeddings of the dataset learned via Isomap from $M = 4800$ random projections. The black dotted line corresponds to an "R"-shaped trajectory in physical space. Learning the joint manifold yields a much improved 2-D embedding of the training points, as well as the "R"-shaped trajectory.*

[5] J. A. Costa and A. O. Hero., "Geodesic entropic graphs for dimension and entropy estimation in manifold learning," *IEEE Trans. Signal Proc.*, vol. 52, no. 8, pp. 2210–2221, August 2004.

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[7] G. E. Hinton, P. Dayan, and M. Revow, "Modelling the manifolds of images of handwritten digits," *IEEE Trans. Neural Networks*, vol. 8, no. 1, 1997.

[8] D. S. Broomhead and M. J. Kirby, "The Whitney Reduction Network: A method for computing autoassociative graphs," *Neural Computation*, vol. 13, pp. 2595–2616, 2001.

[9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, September 2006.

[10] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[11] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 83–91, March 2008.

[12] S. Lafon, Y. Keller, and R. R. Coifman, "Data fusion and multicue data matching by diffusion maps," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1784–1797, Nov. 2006.

[13] C. Wang and S. Mahadevan, "Manifold alignment using Procrustes analysis," in *Proc. Int. Conf. on Machine Learning (ICML)*, Helsinki, Finland, July 2008, pp. 1120–1127.

[14] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, London, England, 2003.

[15] P. Niyogi, S. Smale, and S. Weinberger, "Finding the homology of submanifolds with confidence from random samples," Tech. Rep. TR-2004-08, University of Chicago, 2004.

[16] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, no. 1, pp. 51–77, Feb. 2009.

[17] M. F. Duarte, M. A. Davenport, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Multiscale random projections for compressive classification," in *IEEE Int. Conf. on Image Processing (ICIP)*, San Antonio, TX, Sept. 2007.

[18] C. Hegde, M. B. Wakin, and R. G. Baraniuk, "Random projections for manifold learning," in *Advances in Neural Information Processing Systems*, Vancouver, BC, Dec. 2007, vol. 20, pp. 641–648.

[19] W. B Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. in Modern Analysis and Probability*, 1984, pp. 189–206.

[20] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Proc. Int. Work. on Info. Proc. in Sensor Networks (IPSN)*, Apr. 2006, pp. 51–59.

[21] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. Int. Work. on Info. Proc. in Sensor Networks (IPSN)*, Apr. 2006, pp. 134–142.

[22] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *European Conf. Comp. Vision (ECCV)*, Marseille, France, Oct. 2008, pp. 155–168.

[23] S.K. Nayar, V. Branzoi, and T. Boult, "Programmable imaging using a digital micromirror array," in *European Conf. Comp. Vision (ECCV)*, Washington, DC, Jun. 2004, pp. 436–443.

[24] U. Grenander, A. Srivastava, and M.I. Miller, "Asymptotic performance analysis of Bayesian object recognition," *IEEE Trans. Info. Theory*, vol. 46, no. 4, pp. 1658–1666, July 2000.

[25] U. Grenander, M.I. Miller, and A. Srivastava, "Hilbert–Schmidt lower bounds for estimators on matrix Lie groups," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 20, no. 8, pp. 790–802, August 1998.

[26] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *SPIE Symp. on Elec. Imaging: Comp. Imaging*, 2007.

[27] M. A. Davenport, C. Hegde, M. F. Duarte, and R. G. Baraniuk, "A theoretical analysis of joint manifolds," Tech. Rep. TREE0901, Rice University ECE Department, Jan. 2009.

[28] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. of the American Statistical Association*, vol. 58, no. 301, March 1963.

[29] J. B. Tenenbaum, V.de Silva, and J. C. Landford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[30] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[31] D. Donoho and C. Grimes, "Hessian eigenmaps: locally linear embedding techniques for high dimensional data," *Proc. National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.

[32] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," Tech. Rep., Stanford University, Stanford, CA, Dec. 2000.