Go with the Flow: Optical Flow-based Transport Operators for Image Manifolds

Aswin C. Sankaranarayanan, Chinmay Hegde, Sriram Nagaraj and Richard G. Baraniuk
ECE Department, Rice University

Abstract—Image articulation manifolds (IAMs) play a central conceptual role in a host of computer vision and image understanding problems. The core premise is that we can view a collection of images, each of which is indexed by a small number of degrees of freedom (3D camera pose, motion/deformation, etc.), as a low-dimensional nonlinear manifold. In order to perform parameter estimation and navigation on an IAM, we require a *transport operator* that traverses the manifold from image to image. The two current approaches to manifold transport suffer from major shortcomings that have limited the practical impact of manifold methods. First, algebraic methods require that the IAM possess an unrealistic algebraic structure. Second, locally linear methods based on a tangent plane approximation cannot cope with the non-differentiability of IAMs containing images with sharp edges. In this paper, we demonstrate that the *optical flow* between pairs of images on an IAM is a valid transport operator with a number of attractive properties. In particular, we establish that the optical flow forms a low-dimensional smooth manifold. Several experiments involving novel-view synthesis, geometric clustering, and manifold charting validate that the optical flow manifold approach both offers performance significantly superior to current approaches and is practical for real-world applications.

Fig. 1. Image articulation manifolds (IAMs) are non-differentiable; therefore, locally linear models such as tangent spaces provide an inaccurate approximation to the manifold geometry. In contrast, the optical flow manifold (OFM) associated with a point on the IAM accurately captures the intrinsic curved geometric structure of the IAM.

I. INTRODUCTION

A host of problems in vision, machine learning, and pattern recognition involve the efficient analysis, modeling, and processing of image ensembles. In this paper, we consider image ensembles generated by varying a small number of parameters, such as three-dimensional (3D) camera pose, or motion/deformation of an object in a scene. The geometric notion of an *image articulation manifold* (IAM) [1] provides a powerful model for such ensembles: a collection of N-pixel images, with each image indexed by K degrees of freedom, can be modeled as a K-dimensional nonlinear manifold embedded in $\mathbb{R}^N$.

Manifold models provide a powerful framework for analyzing and processing parameterized images that naturally accounts for the nonlinearity of image ensembles. For instance, problems such as parameter estimation, supervised and semi-supervised classification, and novel view synthesis can be cast as navigating to appropriate points/regions on an IAM. Navigation can often be achieved by first learning a "scaffold" for the IAM of interest (e.g., from a sparse sampling) and then

constructing *transport operators* that traverse the totality of the IAM.

Two families of methods for IAM transport have been developed. First, *algebraic* methods for IAM transport exploit the precise geometric relationships between the images comprising an IAM. Using analytic tools from differential geometry, these methods support tasks such as classification, recognition, and clustering [2]. However, these approaches are limited to a small class of IAMs with a well-defined algebraic structure (such as a Lie group structure); such structure occurs only in some very special cases, such as affine articulations [3], [4] and diffeomorphisms [5], [6]. For most practical situations, including manifolds that characterize pose changes, algebraic transport operators do not apply.

Second, *locally linear* methods for IAM transport rely on the geometric notion of a tangent space to the manifold at a given point (see Figure 1). Local linearity implicitly assumes that over small neighborhoods the manifold can be modeled as approximately planar. Unfortunately, this assumption is seldom of much practical use for IAMs. The rub is that IAMs containing images containing sharp edges and textures are *nowhere differentiable* under the standard L^2 metric [1], [7]. This makes locally linear models for IAMs both inaccurate and fragile.

In this paper, we propose a new framework for modeling and

Fig. 2. A comparison of locally linear transport (LLT) and optical flow-based transport for synthesizing new images on an IAM. We aim to synthesize images on the IAM that lie between two given input images. The non-differentiability of IAMs cannot be accurately captured by locally linear tangent spaces and transport; hence the corresponding synthesized images exhibit severe blurring and cross-fading artifacts. In contrast, optical flow-based transport results in sharp, realistic images.

navigating image manifolds. Our key observation is that the *optical flow* [8] between a pair of images is a natural instance of transport on an IAM. We argue that, for a large class of IAMs, the set of optical flows with respect to a base (reference) image forms a low-dimensional smooth manifold. Optical flow transport inspires a new approach to high-dimensional computational tools such as isometric manifold embedding, Karcher mean estimation, and local charts that have immediate applicability for image processing and vision tasks such as pose-invariant object recognition, geometric clustering, and novel-view synthesis. A cartoon illustration of our approach is depicted in Figure 1.

Our specific contributions are as follows. We first establish that the space of optical flows $\mathcal{F}$ between images on a K-dimensional IAM is itself a K-dimensional manifold. We prove that the space $\mathcal{F}$ (which we call the *optical flow manifold* (OFM)) is a *differentiable* manifold in the case of affine and pose transformations; we also show empirical evidence of OFM smoothness for a nonrigid deformations. We study the geometrical structure of the OFMs of the translation and pose IAMs in detail. By interpreting the OFM as a collection of nonlinear "tangent spaces" that accurately capture the intrinsic curved structure of the IAM, we develop analogies to the concepts of the *Log* and *Exp* maps from differential geometry. The maps enable new high-performance algorithms for manifold learning, image synthesis, parameter estimation, and charting. See Figure 2 for a image synthesis example.

II. BACKGROUND

We will view an N-pixel image I both as a vector in $\mathbb{R}^N$ and as a function defined over a rectangular lattice on $\mathbb{R}^2$. In the latter case, $I(\mathbf{x})$ denotes the pixel intensity at the 2D spatial location $\mathbf{x} = (x, y)$.

A. Image articulation manifolds

In this paper, we are interested in image ensembles that are generated by varying an articulation parameter $\theta \in \Theta$. If Θ is a space of dimension K, then the ensemble of images forms a

K-dimensional *image articulation manifold* (IAM) $\mathcal{I} \subset \mathbb{R}^N$:

$$\mathcal{I} = \{I_\theta : \theta \in \Theta\}. \tag{1}$$

For certain IAMs, the mapping $I : \theta \mapsto I_\theta$ is *locally isometric*, i.e., local distances along the manifold are equal to the corresponding distances in the articulation space [1]:

$$d_{\mathcal{I}}(I_{\theta_1}, I_{\theta_0}) := \|I_{\theta_1} - I_{\theta_0}\|_2 = C\|\theta_1 - \theta_0\|_2,$$

where C is a normalization constant. Many problems in computer vision (including object recognition, image registration, and pose estimation) can be cast as inference problems over $\mathcal{I}$.

A key focus of this paper is the study of *transport operators* $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$ that enable us to move on an IAM from one image I_0 to another I_1 as follows:

$$I_1(\mathbf{x}) = I_0(\mathbf{x} + f(\mathbf{x})). \tag{2}$$

Note carefully from (2) that, given I_0, the transport operator f serves as a *generative model* for images on the IAM. Existing approaches to IAM-based transport can be broadly classified into algebraic and locally linear methods.

B. Algebraic transport

For certain classes of articulations, it is possible to analytically compute transport operators that capture the curved geometrical structure of an IAM. As an example, consider the space of *affine* image articulations, where (2) takes the form

$$I(\mathbf{x}) = I_0(A\mathbf{x} + \mathbf{t}). \tag{3}$$

In this case, the transport operator f is of the form $f(\mathbf{x}) = (A - \mathbb{I})\mathbf{x} + \mathbf{t}$; this can be modeled as the group of 2D affine transformations.

A rich body of work on algebraic transport operators exists in the literature. Miao and Rao [9] have developed a learning framework for affine algebraic transport operators using a matrix exponential-based generative model and have demonstrated improved performance over locally linear approximations. Culpepper and Olshausen [4] have extended this framework using a more complex model on the transport operator in order to model paths on arbitrary image manifolds. In addition to the affine group, other common examples of algebraic transport operators are the projective group (used to model homographies and projective transformations), and the diffeomorphism group (used to model 1D warping functions and density functions) [2], [10], [11]. However, while algebraic transport methods are mathematically elegant, they are applicable only to a very restrictive class of IAMs. Many IAMs of practical interest, including the manifolds corresponding to 3D pose articulations and non-rigid deformations, possess no analytical algebraic structure.

C. Locally linear transport

For most classes of articulations, an algebraic description of transport on the manifold $\mathcal{I}$ is unavailable, and we only have access to a discrete sampling of images from the manifold. In

such cases, there are few principled approaches for constructing transport operators. One common heuristic is to use locally linear models to approximate and traverse the manifold. A locally linear transport (LLT) operator uses an approximation of (2) to express an image I_0 as a linear combination of its k-nearest neighbors. LLT is geometrically equivalent to traversing the *linear tangent space* at I_0. Although they have not been traditionally introduced this way, a variety of manifold-based learning and processing techniques fall into this category. One well-known technique is local linear embedding (LLE) for manifold learning [12].

An critical requirement for successful application of LLT is that the IAM be smooth. However, it has been shown that IAMs containing images with textures and sharp edges are *nowhere differentiable* [1], [7]. Approximating the non-differentiable structure of an IAM by local tangent planes often leads to undesirable results (see Fig. 3). For instance, given a training set of sample images from an IAM, LLE-based synthesis of a novel image from its nearest neighbors in the training set tends to yield highly blurred edges and strong cross-fading artifacts (see Fig. 2).

D. Optical flow

Our key insight in this paper is that the space of *optical flows* forms a natural set of transport operators for IAMs. Given two images I_0 and I_1, the optical flow from I_0 to I_1 (if it exists) is defined as the tuple $(\mathbf{v}_x, \mathbf{v}_y)$, where $\mathbf{v}_x \in \mathbb{R}^N$ and $\mathbf{v}_y \in \mathbb{R}^N$, such that

$$I(x,y) = I_0(x + \mathbf{v}_x(x,y), y + \mathbf{v}_y(x,y)). \quad (4)$$

For N-pixel images, the *optical flow field* $\mathbf{v}$ defined over the domain of the images consists of N ordered pairs; thus, we may regard $\mathbf{v}$ as a point in $\mathbb{R}^{2 \times N}$. Since the pioneering work of Horn and Schunck [8], significant progress has been made on reliable optical flow estimation given a pair of images [13]–[16].

A few remarks about the theory and practice of optical flow are in order. The optical flow, as defined in (4), is not unique. Optical flow estimation suffers for inherent ambiguities, such as the aperture problem [8], as well as degeneracies due to occlusions and boundary effects. In a theoretical setup, we can avoid these degenerate cases by assuming that the optical flow is the projection of a 3D motion flow, so that its existence and uniqueness is axiomatically guaranteed. In practice, we implicitly assume that (i) the images possess rich texture and (ii) satisfy the brightness constancy assumption [8]. For the rest of the paper, we will assume that we can robustly compute the optical flow field given a pair of images.

III. OPTICAL FLOW MANIFOLDS: THEORY

The key goal of this paper is to construct new nonlinear transport operators for IAMs using the optical flow machinery. Towards this end, we begin by defining the space of optical flows between image pairs on an IAM and then characterize the geometric properties of this space.
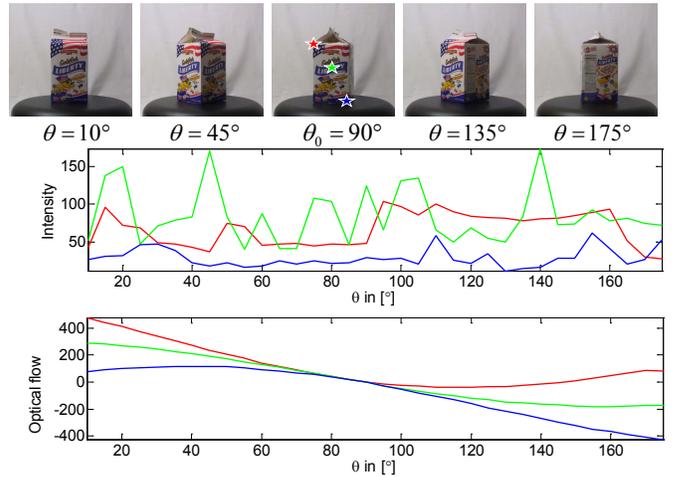
Fig. 3. The IAM and OFM associated with a textured object rotating about a fixed axis. (Top) Images from the $1D$ rotation IAM. (Middle) Image intensity plots for three different pixels as a function of the rotation parameter θ. The colors of the plots correspond to the pixels marked in middle image above. The intensity profiles are highly irregular, which is evidence of the non-differentiability of the IAM. Hence, local linear approximations are inaccurate even for very small intervals. (Bottom) In contrast, the optical flow field at the three pixels varies smoothly, and the OFM (in this case referenced to $\theta_0 = 90$) can be well-modeled as locally linear. Images are from [17].

A. Definition

Consider an IAM $\mathcal{I}$ defined according to (1). Given a reference image I_{θ_0} corresponding to an articulation parameter $\theta_0 \in \Theta$, define $\mathcal{F}_{\theta_0}$ as the set of optical flows between the images in a neighborhood of I_{θ_0}:

$$\mathcal{F}_{\theta_0} = \{f_{\theta_0}(\theta) \in \mathbb{R}^{2N}, \theta \in B(\theta_0)\}, \quad (5)$$

where $f_{\theta_0}(\theta) = [\mathbf{v}_x \ \mathbf{v}_y]$ is the optical flow as defined in (4). $B(\theta_0) \subset \Theta$ denotes a neighborhood of θ_0 where the optical flow can be computed reliably. Note that each $f \in \mathcal{F}_{\theta_0}$ defines an operator on I_{θ_0} such that

$$I(x) = I_{\theta_0}(\mathbf{x} + f(\mathbf{x})).$$

We prove below that in many practical scenarios the set of optical flows $\mathcal{F}_{\theta_0}$ itself forms a *low-dimensional manifold*, which we dub the optical flow manifold (OFM) at I_{θ_0}. We will drop the subscript θ_0 when the context is clear.

Thus, an OFM is the space of optical flow-based transport operators associated with one point (image) on the IAM (see Fig. 1). (Note that the OFM $\mathcal{F}_{\theta_0}$ changes for each reference image I_{θ_0}.) In many ways, the OFM is a nonlinear analogue to the notion of the tangent space at a point on a manifold. However, unlike linear tangent space transport, which is valid only for infinitesimal transformations, nonlinear OFM transport is valid for much larger neighborhoods, as discussed in more detail below (see Fig. 3 for an example).

B. Example: Translation manifold

Consider the IAM $\mathcal{I}$ of images of an object undergoing arbitrary 2D translations. Assume that the domain of each

image is infinite (hence, there are no image boundary artifacts). Then, the space of articulations $\Theta = \mathbb{R}^2$ consists of all possible 2D translations. Any pair of images $I_{\theta_1}, I_{\theta_0}$ corresponding to the articulation parameters $\theta_0, \theta_1 \in \Theta$ can be related by the expression:

$$I_{\theta_1}(\mathbf{x}) = I_{\theta_0}(\mathbf{x} + \theta_1 - \theta_0).$$

Hence, the optical flow between the two images is given in closed form by $[\mathbf{v}_x \ \mathbf{v}_y] = [(\theta_1 - \theta_0)_x \mathbf{1}_N, (\theta_1 - \theta_0)_y \mathbf{1}_N]$, where $(\theta_1 - \theta_0)_x$ represents displacement along the x-axis, $(\theta_1 - \theta_0)_y$ represents displacement along the y-axis, and $\mathbf{1}_N \in \mathbb{R}^N$ denotes the all-ones vector. Thus, the OFM at I_{θ_0} (5) can be described by the closed form expression

$$\mathcal{F}_{\theta_0} = \{(c_x \mathbf{1}_N, c_y \mathbf{1}_N), c_x, c_y \in \mathbb{R}\} \equiv \mathbb{R}^2.$$

It easily follows that the OFM corresponding to every point I_{θ_0} is equivalent to 2D Euclidean space. Consequently, the OFM $\mathcal{F}_{\theta_0}$ is both infinitely smooth as well as isometric to the parameter space $\mathbb{R}^2$. An identical argument extends this result to the more general class of *affine transformations*, where pairs of images are linked by a relation of the form (3) introduced in Section II.

C. Example: Pose manifold

The pose manifold is the IAM corresponding to the rigid body motion of a camera observing a static scene. It is well-known that the articulation space is 6D, with 3 degrees of camera rotation and 3 degrees of camera translation, i.e., $\Theta = SO(3) \times \mathbb{R}^3$. Without loss of generality, assume that the optical flow is the (unique) 2D projection of the motion flow of the scene induced due to the camera motion. Additionally, assume the reference articulation as $\theta_0 = (R_0, \mathbf{t}_0) = (\mathbb{I}, \mathbf{0})$ and that the camera's internal calibration is known and accounted for. At the reference articulation θ_0, denote the depth map of the scene at pixel $\mathbf{x}$ by the function $\lambda_0(\mathbf{x})$. Under a different articulation $\theta = (R_\theta, \mathbf{t}_\theta)$, the optical flow observed at the pixel $\mathbf{x}$ is given by

$$f(\theta)(\mathbf{x}) = P\left(\lambda_0(\mathbf{x})R_\theta \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} + \mathbf{t}\right) - P\left(\lambda_0(\mathbf{x}) \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}\right), \tag{6}$$

where P is a projection operator such that $P((x, y, z)) = (x/z, y/z)$.

We note that in (6), the optical flow $\{f(\theta)(\mathbf{x}), \ \theta \in \Theta\}$ depends on the depth value $\lambda_0(\mathbf{x})$. When $\lambda_0(\mathbf{x}) \neq 0$, there exists a neighborhood of I_{θ_0} on the IAM such that the flow $f(\theta)(\mathbf{x})$ is a *smooth* function of θ. This follows from the fact that the projection operator P is a well-defined smooth function provided $z \neq 0$. Denote this neighborhood by $B(\mathbf{x}; \theta_0)$. Then, the size of $B(\mathbf{x}; \theta_0)$ is determined by the maximum articulation required to move the scene point such that it lies on the plane $z = 0$. Consequently, if all pixels have depth values bounded away from zero, then it follows that there exists a neighborhood $B(\theta_0) = \bigcap_{\mathbf{x}} B(\mathbf{x}; \theta_0)$ where the variations in optical flow for all pixels in the image are *uniformly smooth*. Converting this pixel-domain smoothness

into a high-dimensional smoothness condition, we arrive at the following key result.

Proposition: Consider the pose manifold corresponding to a scene such that at reference articulation θ_0, $\lambda_0(\mathbf{x}_0) > \lambda$. In the absence of occlusion, there exists a neighborhood on the IAM such that the OFM at this reference point is a 6D *infinitely smooth* submanifold of $\mathbb{R}^{2N}$.

In addition to smoothness, the OFM associated with the pose manifold can also be shown to be *locally isometric* to Euclidean space. This has important practical implications, especially for the problem of *manifold learning*, which seeks to develop a low-dimensional Euclidean embedding of point samples from a manifold lying in a high-dimensional ambient space. Typical manifold learning algorithms (e.g., ISOMAP [18]) implicitly rely on the assumption that the underlying manifold exhibits local isometry, which is not true for most IAMs. Learning the manifold structure not from the point samples on the IAM but rather from the OFM computed from the point samples can lead to a significant performance boost. We reserve a detailed study of the local isometry of the OFM for the extended version of this paper [19].

Surprisingly, the Proposition guarantees the smoothness of the OFM without any assumption on the smoothness (or even continuity) of the depth map $\lambda_0(\mathbf{x})$. Our only assumption is that no occlusions occur, which is violated in many practical scenarios of interest. While explicitly modeling occlusions within the theoretical framework can be tricky, in practice we can detect pixels that exhibit occlusions and apply appropriate heuristics to handle such situations (see Section IV).

The Proposition analytically demonstrates the global differentiability/smoothness of the OFM for the pose (and hence affine transformation) IAMs. Section V below empirically demonstrates that the OFM is smooth for an even larger class of IAMs, including images of an object undergoing non-rigid deformations (see Fig. 6). We defer a full analytical treatment of OFM smoothness for more general classes of IAMs for future research; in the rest of the paper, we will restrict our attention to IAMs with smooth OFMs.

IV. OPTICAL FLOW MANIFOLDS: PRACTICE

In the previous section, we showed that, at least for the pose and affine transformation IAMs, the corresponding OFMs are smooth. In this section, we leverage this smoothness to construct new nonlinear transport operators for IAMs that first transport linearly on the OFM and then map that transport to the IAM via (4). In this section, develop a set of computational tools and tricks that are useful for many practical applications.

A. IAM transport via the OFM

Consider an IAM $\mathcal{I} \subset \mathbb{R}^N$ and a point $I_{\theta_0} \in \mathcal{I}$. Our goal is to develop a transport operator that can accurately navigate from any point $I_\theta \in \mathcal{I}$ to the points its neighborhood. As discussed in Section I, the classical choice of transport uses vectors from the IAM tangent space at I_{θ_0}; however, this often leads to undesirable results due to the non-differentiability of

(a) Images

(b) Ground Truth   **(c) IAM**   **(d) OFM**
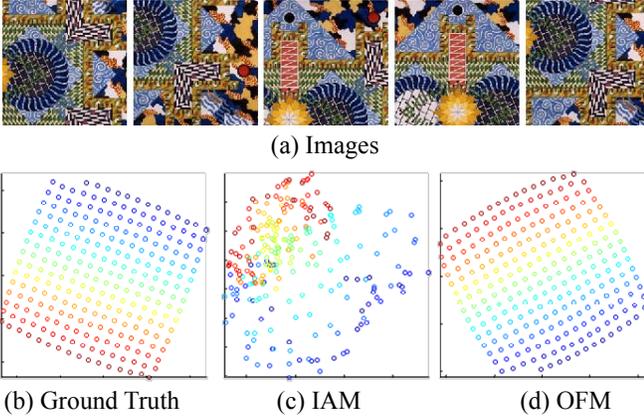
Fig. 4. Dimensionality reduction of an IAM vs. its OFM. The IAM is generated by cropping 200×200 pixel patches at random from a larger image, thereby generating a 2D translation manifold. (a) Sample images from the IAM showing several images at various translations. (b) Sampling of the parameter space. 2D embeddings obtained on (c) the IAM vs. (d) the OFM. Note the near perfect isometry to the parameter space in (d).

$\mathcal{I}$. We develop our new OFM-based transport operator in two steps. Assume that the OFM $\mathcal{F}$ at I_{θ_0} is both smooth and isometric.

Step 1: Given a set of images $\{I_{\theta_i}, i = 1, \ldots, M\}$ in a neighborhood of the reference image I_{θ_0}, estimate the optical flows $f_i = (\mathbf{v}_x, \mathbf{v}_y)_i$ from I_{θ_0} to all of the images I_{θ_i}. Given the set of optical flows $\{f_1, \ldots, f_M\} \in \mathcal{F}$, construct its low-dimensional Euclidean embedding $\{\mathbf{e}_1, \ldots, \mathbf{e}_M\} \in \mathcal{E} \equiv \mathbb{R}^K$ using a dimensionality reduction tool such as (LLE [12], ISOMAP [18], etc. If the OFM is smooth and locally isometric, then the embedding of the optical flows will be isometric to the IAM's parameter space. Figure 4 demonstrates that the OFM is much more amenable to accurate dimensionality reduction than its corresponding IAM.

Step 2: Using the low-dimensional embedding, we generate the IAM transport operator from θ_0 to a new point $\mathbf{e}' \in \mathcal{E}$ in the parameter space as follows. Represent $\mathbf{e}'$, using a convex combination of its k-nearest neighbors:

$$\mathbf{e}' = \sum_{j \in \mathcal{N}(k, \mathbf{e}')} w_j \mathbf{e}_j, \quad \sum_j w_j = 1,$$

where $\mathcal{N}(k, \mathbf{e}')$ is the index set of the k-nearest neighbors of $\mathbf{e}'$. Then, the neighborhood on the OFM associated with $\mathbf{e}'$ can be constructed by using the same coefficients to generate the corresponding optical flow vectors in $\mathcal{F}$:

$$f' = \sum_{j \in \mathcal{N}(k, \mathbf{e}')} w_j f_j,$$

where both the neighborhood $\mathcal{N}$ and the weights $\{w_j\}$ are inferred from $\mathbf{e}'$ and its neighbors. Using the new optical flow f', we can use (2) to navigate on the IAM to the correct image.

In summary, we transport nonlinearly on the IAM by transporting linearly on the OFM and then applying (2). The success of this approach depends critically on the twin

assumptions of local linearity and local isometry of the OFM which, as we have shown, are valid for several classes of interesting manifolds.

B. Connections to the exponential map

In the differential geometric treatment of analytical manifolds, it is often useful to define a one-to-one mapping from the set of all vectors belonging to the tangent space at $I_{\theta_0} \in \mathcal{I}$ to the set of all points belonging to a neighborhood of I_{θ_0}. This functional relation is dubbed the exponential (*Exp*) map; the inverse function is called the logarithmic (*Log*) map [20]. Unfortunately, the non-differentiable structure of IAMs has blocked the formal development and application of these tools in practical applications.

We observed in Section III that the OFM at a point on an IAM can be interpreted as the nonlinear analogue of the tangent space to the IAM at that point. Therefore, using the OFM-based navigation framework from above, we can construct two new operators that are analogous to the exponential and logarithmic maps. We dub them the *OFExp* and the *OFLog* operators. As discussed above in Section IV-A, given a reference image I_{θ_0} and a sampling of images $\{I_{\theta_i}, i = 1, \ldots, I_M\} \subset \mathcal{I}$, we can obtain the optical flows $\{f_1, \ldots, f_M\} \subset \mathcal{F}_0$ and compute their low-dimensional Euclidean embeddings $\{\mathbf{e}_1, \ldots, \mathbf{e}_M\} \subset \mathcal{E} \equiv \mathbb{R}^K$. The space $\mathcal{E}$ is analogous to the tangent space at the reference image I_{θ_0} and can be equipped with a suitable orthonormal basis B.

Formally, the *OFExp* and *OFLog* operators are defined as follows. Given a query image I_θ, the *OFLog* operator applied to I_θ is defined as the vector in $\mathcal{E}$ obtained by computing the optical flow from I_{θ_0} to I_θ and then projecting it onto the orthornormal basis B. Thus, the *OFLog* operator is a mapping from the IAM to the low-dimensional embedding $\mathcal{E}$ of the smooth OFM $\mathcal{F}$.

Given a query optical flow vector $\mathbf{e}' \in \mathcal{E}$, the *OFExp* operator applied to ($\mathbf{e}'$) is obtained by expressing $\mathbf{e}'$ as a linear combination of the basis vectors B to obtain the optical flow vector f'. Once f' is obtained, we apply (4) to the reference image I_{θ_0} to obtain a image $I' = f'(I_0)$. Thus, the *OFExp* operator is a mapping from the embedded OFM representation $\mathcal{E}$ to the IAM $\mathcal{I}$.

For analytic manifolds, the Exp and Log maps enable extremely useful computational tools for geodesic generation, Karcher means, and other quantities. In Section V, we demonstrate that our *OFExp* and *OFLog* enable analogous tools for non-differentiable IAMs.

C. Occlusions and boundary effects

Occlusions and boundary effects are invariably encountered in practice imagery. They can lead to undefined and/or inconsistent optical flow estimates at various pixels in the image. In practice, a small amount of occlusion between the reference and query image can be tolerated by not using the part of the image that becomes occluded. A formal theoretical treatment of occlusion phenomena in the optical flow framework is beyond the scope of the paper; rather, we will mitigate these

undesirable effects using a simple heuristic to detect and discard bad optical flow estimates.

Suppose we are given a pair of images I_1 and I_2. We use (4) to compute the flow f from I_1 to I_2, and the flow f' from I_1 to I_2. Then, we add these flow vectors to obtain the composite flow f'', and apply this flow to I_1 using (4) to obtain a new image I_1'. Ideally, $I_1 = I_1'$; however, the images might be mis-aligned. We identify pixels where the drift is more than 1 pixel. Such pixels are discarded and the flow values at the pixel are not used in subsequent computations.

Finally, it should be intuitively clear that a clever choice of reference image I_{θ_0} can alleviate occlusion problems. Given an unorganized database of images sampled from an IAM, picking the "best" reference image is a nontrivial problem. However, in Section V, we describe an algorithm for Karcher mean estimation based on the *OFExp* operator that often yields a good choice of reference.

V. APPLICATIONS AND EXPERIMENTS

We now present the results of a number of numerical experiments that illustrate the utility and applicability of our proposed OFM-based manifold modeling framework. We used the Brox and Malik toolbox [15] for estimating the optical flows between pairs of images with the default parameter settings. The results do not change significantly when we use other optical flow estimation algorithms, such as [14], [16].

A. Generating geodesics

Geodesics (informally, shortest manifold paths) are useful for defining meaningful relations between images on an IAM. However, Donoho and Grimes [1] have argued that the non-differentiable nature of most IAMs renders moot the study of their geodesics; indeed, for IAMs containing images with sharp edges, the length of any geodesic is shown to be infinite.

However, shortest paths in the *parameter* space Θ can be well-defined, particularly when Θ can be identified with Euclidean space. Since in many cases the OFM is isometric to the parameter space, we can *induce* a geodesic on the IAM by first generating the geodesic in the space of optical flows and then applying (2) to obtain a path on the IAM.

Formally, given a reference image I_{θ_0}, the associated OFM $\mathcal{F}$ and its low dimensional embedding $\mathcal{E}$, the geodesic path between two images I_1 and I_2 within a neighborhood I_{θ_0} can be analytically generated as follows. First, compute $\mathbf{e}_1 = OFLog(I_1)$ and $\mathbf{e}_2 = OFLog(I_2)$ as defined in Section IV. Then, the geodesic on the IAM between I_1 and I_2 is given by $\Gamma = \{\gamma(t) \mid 0 \leq t \leq 1\}$, where $\gamma(t) = OFExp((1-t)\mathbf{e}_1 + t\mathbf{e}_2)$. Figure 2 shows experimental results for geodesic generation on a pose IAM of a scene containing some household objects. It is visually evident that our proposed OFM-based synthesis approach outperforms LLT-based methods, yielding crisp images with sharp edges.

B. Karcher mean estimation

The *Karcher mean* of a set of points on a manifold is defined as the point $I_{\mathrm{KM}} \in \mathcal{I}$ such that the sum of geodesic

Sampling of 10 images from an IAM

OFM-based Karcher Mean estimation

Initialization   Intermediate points   Converged

Gr. Truth   OFM   LLT on IAM
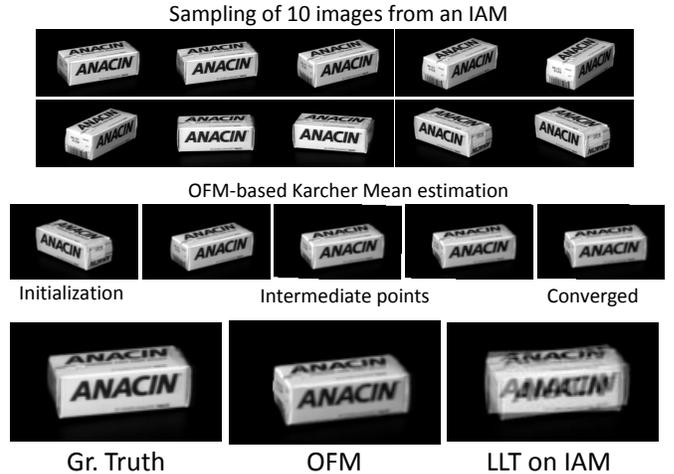
Fig. 5. OFM-based Karcher mean estimation for 10 images from the COIL dataset [21]. The top and bottom rows correspond to two trials with widely different initialization points chosen from the image samples. In both cases, the converged Karcher mean estimates are identical. The Karcher mean estimates are also vastly superior to those obtained using LLT on the IAM.

distances from I_{KM} to each point in the set is minimized [22]. In principle, the Karcher mean corresponds to a point on the manifold that is most representative of a given point set. This can be invaluable in applications ranging from robust statistical inference and model-building to data visualization. To date, since geodesics for IAMs are hard to estimate, the study of Karcher means for image data has been limited.

We can modify a simple Karcher mean estimation algorithm [22] to use the *OFExp* operator defined in Section IV. Given a finite set of images $\{I_i, i = 1, \ldots, M\}$, we initialize by randomly choosing a base (reference) image I_θ from the set and then computing the set of optical flows $\{f_i, i = 1, \ldots, M\}$ and their low-dimensional Euclidean embedding $\{\mathbf{e}_i, 1, \ldots, M\}$. The *OFExp* map applied to the sample mean $\mathbf{e}' = (1/M) \sum_i \mathbf{e}_i$ gives a new reference image $I_{\theta'}$. Note that this new, intermediate reference image need not coincide with any element in the given set of images. Using this new reference image, we again perform the computation of optical flows, their Euclidean embedding, the mean estimation, and the exponential map. This procedure is repeated until convergence. While convergence to the true Karcher mean is not guaranteed, we have empirically observed in our experiments that the algorithm always converged to a solution that can be interpreted as a "mean" image.

Figures 5 and 6 display the results of our new OFM-based Karcher mean estimation algorithm on different standardized datasets [21], [23]. Figure 6, in particular, demonstrates the stability of the method with respect to the choice of the initial reference image. The images in Fig. 6 come from a highly nonlinear IAM generated by extracting image frames from a video sequence of an object undergoing a non-rigid deformation. Despite initializing the estimation algorithm with very different starting reference images, the output Karcher mean estimates are identical; consequently, our algorithm can
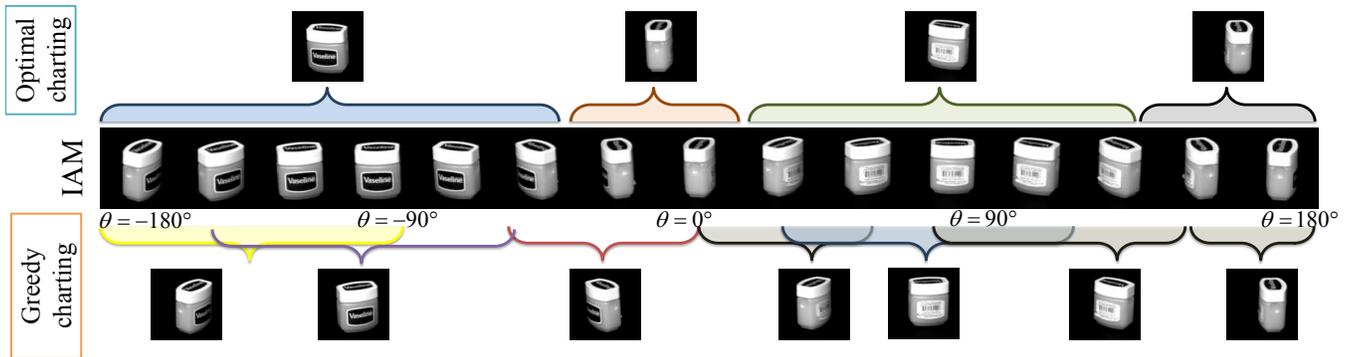
Fig. 7. OFM-based charting of a 1D IAM of a cuboid rotating about an axis using (top) optimal and (bottom) greedy methods. Each reference image charts a neighborhood (marked with braces) of the IAM. While the optimal chart provides maximum parsimony in representation (in this case, 4 base images), the computation time of the algorithm is prohibitive. In contrast, the greedy algorithm returns more reference points (in this case, 7 base images) but in a computationally efficient manner. Images are from the COIL dataset [21].
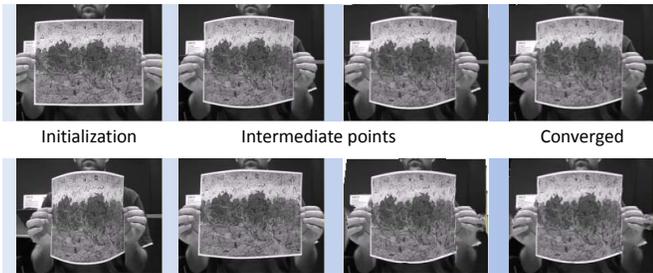
Fig. 6. OFM-based Karcher mean estimation for images from video of a non-rigidly deforming object [23]. The top and bottom rows correspond to two trials with widely different initialization points (the first and last frames of the video) chosen from the image samples. In both cases, the converged Karcher mean estimates are identical.

be interpreted as being robust with respect to the starting reference point. Finally, to pick up the thread begun in Section III, Fig. 6 demonstrates empirically that the OFM concept and toolset is applicable beyond the class of manifolds (pose, affine transformation) that we can currently prove to have smooth OFMs.

C. Charting IAMs

OFMs enable us to build a concise analytic description for a given IAM using the concept of a transport operator from a reference image. In practice a single reference point I_{θ_0}, together with the associated OFM $\mathcal{F}_{\theta_0}$ at that point, rarely suffices to specify an entire IAM. This is due to phenomena such as occlusions (see Section IV-C) as well as the appearance of novel, unseen image features as we traverse from one neighborhood to another neighborhood on the IAM. We can overcome this issue using the technique of *charting* an IAM by inferring a "good" set of reference points and then constructing the OFMs at these reference points [24]. Then, the complete IAM can be concisely described using this collection of reference points (or landmarks) and the OFM (or map) associated with each of the landmarks.

It is intuitively clear that there is no one unique way to build a chart for an IAM. Indeed, a fundamental tradeoff exists between the number of reference points in the chart and the sizes of the maps in the chart. This tradeoff results in differences in computational considerations. If we desire maximum parsimony (i.e., the minimum number of reference points), then the resulting algorithm to find the optimal chart is computationally demanding, since it requires optical flow estimation between all pairs of images from the dataset. For M images on the IAM, this algorithm has computational complexity $O(M^2)$. Alternatively, we can greedily build chart-based representations in $O(M)$ time using a simple modification of the algorithm described in Section V-B to encompass the well-known concept of k-means. Figure 7 demonstrates an optimal (parsimonious) and a greedily-constructed chart over an 1D pose IAM of a rotating household object.

VI. DISCUSSION

In this paper, we have advocated for and developed a new transport operator-based viewpoint for the analysis, modeling, and processing of IAMs. Our key result is that the space of optical flow-based transport operators forms a smooth manifold for a large class of interesting IAMs. Working on the space of optical flows provides a much better match to the curved geometry of an IAM than traditional methods based on LLT. We have proved that our optical flow-based approach is exact for the IAMs generated by affine articulations and camera pose changes and a good approximation for several other IAMs of interest. It is our hope that the OFM concept and its associated toolset can deliver state-of-the-art performance in image processing and vision problems where IAMs have come up short.

Related work: The approach described in this paper follows a general principle in image understanding that proceeds by modeling image deformations, as opposed to modeling image intensities. [2], [25]–[27]. Indeed, the notion of transport operator is the same as a deformation (or a diffeomorphism, when it is invertible). However, our ideas should be interpreted as a specialization to the case of manifold-valued data. A key distinction is our characterization of the space of transport

operators as a K-dimensional manifold, whereas diffeomorphisms as a group [2] are an infinite-dimensional construct. Here, the reduced dimensionality is important, as it leads to a concise model underlying the image ensemble.

Morphlets [27] are also closely related to the ideas espoused in this paper. Morphlets are motivated by the idea that ensembles of image deformations are smoother and easier to model and analyze than ensembles of raw images. The ideas in this paper rely on this same general premise. However, while morphlets provide tools to represent image deformations and to interpolate across image deformations, the treatment of [27] is limited to image pairs. In contrast, the ideas in this paper apply to image ensembles consisting of a potentially large number of images.

In the context of image ensembles, Beymer and Poggio [25] argue for the use of motion-based representations for learning problems. However, their goal is image synthesis, and their method offers no insights into the geometric nature of manifold-valued data. In contrast, our approach naturally leads to tools such as *OFExp* and *OFLog*; to the best of our knowledge these are completely novel.

Jojic et al. [26] use a layered representation to represent videos by separating the appearance of moving objects from their motion and then representing each using subspace and manifold models. This enables a simple, yet powerful, representation that is capable of modeling and synthesizing complex scenes using simpler primitives such as subspaces and translation manifolds. Our goal is to study the properties of transport operators/deformations associated with image ensembles; in this regard, we go beyond this simple manifolds induced by translations and affine transformations.

Beyond optical flow: The use of optical flow as a transport operator is meaningful only when the pixel brightness is constant across the images in the ensemble and when the articulation manifests itself as a projection of the motion field. When either condition is violated, our approach needs to be appropriately modified. An extreme example is the case of IAMs formed by varying illumination; we defer a discussion of the necessary modifications to future research.

Depending on the application at hand, it is possible to construct transport operators other than optical flow. For example, we can build a similar theory for sparse correspondences as opposed to dense optical flow for applications such as detection and recognition rather than synthesis. Sparse correspondences can be estimated more efficiently and reliably than dense flow. The success of the SIFT [28], for example, cannot be overstated. Given that sparse point correspondences form the basis of the current state-of-the-art algorithms for object detection and recognition, the allure of constructing transport operators from such correspondences is considerable.

Finally, much of the work on unsupervised manifold learning is insensitive to the modality of the data. While our development of transport operators is strongly tied to the fact that we are analyzing image ensembles generated by imaging objects, the ideas in this paper can potentially be extended to non-image modalities given an appropriate definition of transport operator.

References

[1] C. Grimes and D. Donoho, "Image manifolds which are isometric to euclidean space," *J. Math. Imag. and Vision*, vol. 23, no. 1, pp. 5–24, 2005.

[2] M. Miller and L. Younes, "Group actions, homeomorphisms, and matching: A general framework," *Intl. J. Comp. Vision*, vol. 41, no. 1, pp. 61–84, 2001.

[3] D. Lin, E. Grimson, and J. Fisher, "Learning visual flows: A Lie algebraic approach," in *Proc. Comp. Vis. Patt. Recog.*, 2009.

[4] B. Culpepper and B. Olshausen, "Learning transport operators for image manifolds," in *Proc. Neural. Info. Proc. Sys.*, 2009.

[5] E. Klassen, A. Srivastava, W. Mio, and S. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Trans. Patt. Anal. Mach. Intelligence*, pp. 372–383, 2004.

[6] A. Trouvé and L. Younes, "Local geometry of deformable templates," *SIAM J. Math. Anal.*, vol. 37, no. 1, pp. 17–59, 2006.

[7] M. Wakin, D. Donoho, H. Choi, and R. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. SPIE Wavelets XI*, 2005.

[8] B. Horn and B. Schunck, "Determining optical flow," *Artif. Intel.*, vol. 17, no. 1-3, pp. 185–203, 1981.

[9] X. Miao and R. Rao, "Learning the lie groups of visual invariance," *Neural Comp.*, 2007.

[10] A. Srivastava, M. Miller, and U. Grenander, "Ergodic algorithms on special Euclidean groups for ATR," *Systems And Control In The Twenty-first Century*, 1997.

[11] O. Tuzel, F. Porikli, and P. Meer, "Learning on lie groups for invariant detection and tracking," in *Proc. Comp. Vis. Patt. Recog.*, 2008.

[12] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[13] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," in *Proc. Intl. Conf. Comp. Vision*, 2007.

[14] C. Liu, "Beyond pixels: exploring new representations and applications for motion analysis," Ph.D. dissertation, MIT, 2009.

[15] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE Trans. Patt. Anal. Mach. Learning*, vol. 33, no. 3, pp. 500–513, 2011.

[16] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *Proc. Comp. Vis. Patt. Recognition*, June 2010.

[17] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *Intl. J. Computer Vision*, vol. 73, no. 3, pp. 263–284, 2007.

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

[19] A. C. Sankaranarayanan, C. Hegde, S. Nagaraj, and R. G. Baraniuk, "Go with the flow: Optical flow-based transport operators for image manifolds," *(In preparation)*, 2011.

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

[21] S. Nene, S. Nayar, and H. Murase, "Columbia object image library (COIL-100)," *Tech. Rep. CUCS-006-96, Columbia University*, 1996.

[22] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Comm. Pure and Appl. Math.*, vol. 30, no. 5, pp. 509–541, 1977.

[23] M. Salzmann, R. Hartley, and P. Fua, "Convex optimization for deformable surface 3-d tracking," in *Proc. Intl. Conf. Comp. Vision*, 2007.

[24] M. Brand, "Charting a manifold," *Advances in Neural Info. Proc. Sys.*, pp. 985–992, 2003.

[25] D. Beymer and T. Poggio, "Image representations for visual learning," *Science*, vol. 272, no. 5270, pp. 1905–1909, 1996.

[26] N. Jojic and B. Frey, "Learning flexible sprites in video layers," in *Proc. Comp. Vis. Patt. Recog.*, 2001.

[27] J. Kaplan and D. Donoho, "The morphlet transform: A multiscale representation for diffeomorphisms," in *Workshop on Image Reg. Deformable Envir.*, 2006.

[28] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.