Modeling and Estimating Persistent Motion with Geometric Flows

Dahua Lin
CSAIL, MIT
dhlin@mit.edu

Eric Grimson
CSAIL, MIT
welg@csail.mit.edu

John Fisher
CSAIL, MIT
fisher@csail.mit.edu

Abstract

We propose a principled framework to model persistent motion in dynamic scenes. In contrast to previous efforts on object tracking and optical flow estimation that focus on local motion, we primarily aim at inferring a global model of persistent and collective dynamics. With this in mind, we first introduce the concept of geometric flow that describes motion simultaneously over space and time, and derive a vector space representation based on Lie algebra. We then extend it to model complex motion by combining multiple flows in a geometrically consistent manner. Taking advantage of the linear nature of this representation, we formulate a stochastic flow model, and incorporate a Gaussian process to capture the spatial coherence more effectively. This model leads to an efficient and robust algorithm that can integrate both point pairs and frame differences in motion estimation. We conducted experiments on different types of videos. The results clearly demonstrate that the proposed approach is effective in modeling persistent motion.

1. Introduction

Modeling and analysis of motion patterns in video is an important topic in computer vision. While extensive efforts have been devoted the problem of local motion estimation, such as tracking individual objects or estimating the optical flow between two consecutive frames, research on modeling persistent motion patterns has received less attention. Persistent motions are ubiquitous. In many applications, such as scene understanding and crowd surveillance, one is primarily interested in collective and persistent motion patterns rather than the motions associated with individual entities. Figure 1 depicts the frames in three different video sequences. In such scenes, characterizations such as *the vehicles are moving towards bottom right corner with a slight rotation* and *the athletes are running along a circular path* are more pertinent than the velocities of individual objects.

In modeling persistent motion, our primary goal is to *infer a global model of collective behavior over space and time*. Whereas we can extract persistent motions base on

Figure 1. This figure shows the frames respectively captured in three different dynamic scenes that exhibit obvious persistent motion patterns: the flow of water in a spring, cars running on a high way, and athletes running along a circular path.

trajectories or optical flow fields using existing local estimation techniques, this approach is challenged by various practical difficulties as we discuss in section 2. In addition, current methods do not provide a natural mechanism for aggregating potentially sparse observations over space and time into a collective model. This arises from their narrow focus on temporally and spatially local motion (e.g. the velocity of a particular object or at a particular time). While enforcing smoothness may be useful in reducing the estimation error of local motion, it doesn't help to truly capture the global motion patterns. In this paper, we address this problem by exploring a new methodology under the notion of *geometric flows*. It allows the incorporation of observed motions that may be separated spatially and temporally into a globally consistent model. Moreover, this model can be efficiently and robustly estimated without explicit reliance on continuous tracking or optical flow estimation.

For introducing a new representation, we first review two conventional ways to describe motion. The first is to represent the motion of an object by its trajectory, that is, over a narrow path through space and time. Alternately, one might use a geometric transform to describe how a region deforms at a particular instance of time. While this captures the common behavior of an entire region, it only does so over a small temporal window. This motivates the need to establish a temporally and spatially global representation that unifies trajectories and geometric transforms. Conse-

quently, we introduce the notion of *geometric flow*, which characterizes a motion pattern as both a collection of trajectories driven by common rules and a continuous geometric transform process.

The geometric flow originates from the concept of *flow* in differential geometry. It describes motion as a continuous geometric transform constrained by global consistency. Each family of geometric flows is associated with a Lie algebra, i.e. a vector space comprised of the infinitesimal generators of the flows, and each geometric flow can be parameterized by a vector in this space. A Lie algebraic representation allows decomposing a flow into a linear combination of base flows, thus greatly simplifying statistical estimation. Furthermore, we extend this representation to describe complex flows using triangle mesh and derive a *consistent subspace* of the joint Lie algebra to preserve the consistency between the flows associated with adjacent triangle cells.

With the Lie algebraic representation, learning persistent motion patterns reduces to linear coefficient estimation. In particular, we estimate the coefficients by integrating two types of observations: *point pairs* and *frame sequences*. Here, point pairs are extracted by associating SIFT feature points between two consecutive frames. Additionally, we directly utilize frame-to-frame differences by taking advantage of the linear isomorphism between the flow space and the space of frame differences, without the need to explicitly solve for optical flows. Taking observation errors into account, we formulate a generative stochastic model of flows, and incorporate a Gaussian process as a prior on the flow parameters so as to capture the global coherence more effectively. The stochastic formulation is then generalized to admit multiple concurrent flows through an MRF.

The main contributions are summarized as follows: (1) we introduce geometric flow to model persistent motions that unifies trajectories and geometric transforms through their intrinsic connections, (2) we derive a Lie algebraic representation that simplifies the modeling of flows, and (3) we formulate a stochastic model that integrates different types of observations for robust motion estimation.

2. Related Work

Existing work on motion analysis can be roughly categorized into two groups. The first group [4, 7, 19] relies on tracking of individual objects. The set of trajectories derived from the tracking algorithm are then used to construct a statistical model enabling further motion analysis. Despite extensive efforts [13, 20] such analysis under difficult conditions (e.g. crowded scenes, low-resolution) remains challenging for lack of reliable and persistent multi-object tracking algorithms. Errors in track association are common and can bias results. The second group are based on analysis of dense maps of optical flow [1, 16, 17, 18]. These methods suffer due to the sensitivity of optical flow estima-

tion to occlusions, noise, and varying illumination. Again, for complex scenes, state-of-the-art methods [6, 12, 14] are challenged to attain reliable estimates in practice.

The work presented in this paper is distinguished from both tracking and optical flow estimation in two aspects: (1) Our work targets a different goal, that of deriving generalizable motion patterns that capture the underlying regularities over space and time, while previous work mainly focus on accurate estimation of local velocities (e.g. the velocity of individual objects or points at a particular time). (2) Our framework has a global parametric model in its heart that allows spatially and temporally separated observations be incorporated into the estimation of motion models, while in most previous work, only local information is utilized in estimating each velocity.

Recently, there have been some efforts devoted to exploiting global motion patterns. Ali and Shah [2] proposed a force field model for tracking individuals in crowded scenes. Jacobs et al [8] proposed a method to learn location-specific prior to assist object tracking. While global motion patterns were explored in these papers, they used these patterns to help tracking individuals rather than taking them as the main subject of modeling. Moreover, the models of global patterns were tailored to specific applications, and it is unclear whether they can be generalized to other cases. In contrast, our approach is a generic methodology, which can be readily applied to a broad range of scenarios.

The most relevant work to this paper is our earlier work in [10], which proposed to use Lie algebra for motion modeling. However, the framework has been substantially extended in this paper. The differences lie in two aspects: (1) The representation of geometric flow in this paper is based on the decomposition of the infinitesimal generators, which is very general and can be applied to modeling virtually any continuous flows. In comparison, [10] relies on a particular form of matrix parameterization, restricting its application to affine fields. (2) We explore the intrinsic relation between flows and image sequences that it generate. Using this relation, we derive a method that can directly estimate flows by decomposing the image differences without the need of making point correspondence. Instead, [10] employs a tracker to acquire short-time tracks as input.

3. Theory of Geometric Flow

In this section, we first discuss the concept of geometric flow, including its connections with trajectories, geometric transforms and velocity fields. We then construct a Lie algebraic representation based on the infinitesimal generators of a flow and further extend it to multiple flows combined in a geometrically consistent way.

There are two primary representations used for motion description. **Trajectory-based descriptions**, often used in person or vehicle tracking systems, collect the kinematic
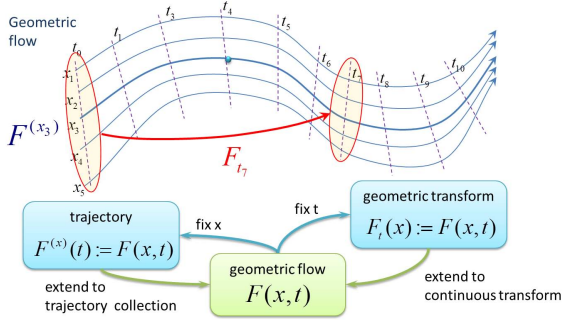
Figure 2. Conceptually, a flow can be obtained in either of the following two ways: (1) By inspecting the full motion of a collection of points whose initial locations differ, we get a set of trajectories, or (2) By integrating the geometric transforms terminating at different times t, we get a continuous transform process, which describes how the region changes over time. In this sense, geometric flows unify trajectory sets and continuous geometric transforms. Conversely, from a flow one can derive the trajectory starting at x, defined by $F^{(x)}(t) := F(x, t)$ or a geometric transform terminated at time t, defined by $F_t(x) := F(x, t)$.

state of an individual object over time, typically independent of other objects in the scene. **Geometric transforms**, often used in object alignment and image registration applications, describe the transformation of points over an entire region. Whereas trajectory representations describe the motion of a single point over a long time duration, geometric transforms describe the motion of all points over a spatial region, but only over a short time window. While useful for many applications, by themselves they are lacking when used for modeling persistent flows as neither simultaneously describes motion over *both space and time*.

To develop a methodology for inferring global motion patterns over space and time, we unify the descriptions above in a function F that governs the motion of *each* point in a region over time. Given the initial position x and time duration t, F yields the location of the point at time t. Mathematically, a geometric flow must satisfy two identities:

$$F(x, 0) = x. \tag{1}$$

$$F(F(x, t_1), t_2) = F(x, t_1 + t_2). \tag{2}$$

Eq.(1) simply states that at time $t = 0$ the point is at its initial position while Eq.(2) states that the geometric flow is *associative*, i.e. that a point moving along the flow for time t_1 and then for time t_2 is equivalent to moving for time $t_1 + t_2$. Note that t can be negative, allowing "backward tracing". Figure 2 illustrates a geometric flow and its connections with trajectories and geometric transforms.

3.1. Flow and Velocity Field

Consider a point driven by a geometric flow F that starts at y and suppose it passes x at time t, i.e. $F^{(y)}(t) = x$.

The velocity of the point at t can be obtained by taking the derivative of $F^{(y)}$. A geometric flow has an important property with regards to velocity: Given any x in the flow domain, any point driven by the flow passes through x with the same velocity independent of its initial location. The property implies that *each geometric flow F induces a time-invariant velocity field, denoted by V_F, which can be expressed by*

$$\frac{\partial F(x, t)}{\partial t} = V_F(F(x, t)). \tag{3}$$

Alternately, given a velocity field V_F, one can reconstruct the flow F by solving the differential equation in Eq.(3). This is equivalent to the process of generating the trajectories with the velocities specified by V_F. The *Fundamental Theorem of Flows* [9] states that under mild conditions, each velocity field induces a unique geometric flow.

3.2. Lie Algebraic Representation

The notion of geometric flow leads naturally to a Lie algebraic representation. Consider a transform $F_{\Delta t}$ derived from a flow F. As it induces motion in each point along the velocity given by $V_F(x)$, we have $F_{\Delta t}(x) \simeq T_{V, \Delta t} := x + V_F(x)\Delta t$, when the time interval Δt is sufficiently small. Due to the associativity in Eq.(2), we can express each derived transform F_t as a composition of many short time transforms as $F_t = F_{\Delta t} \circ \cdots \circ F_{\Delta t}$. Taking the limit as $\Delta t \to 0$ results in the following equation:

$$F_t = \lim_{N \to \infty} (T_{V_F, \frac{t}{N}})^N. \tag{4}$$

This central result in Lie algebra theory connects geometric transforms to the driving velocity field. Intuitively, it reflects the observation that *a geometric transform is formed by accumulating the small changes due to the underlying velocity field*. Hence, the velocity field V_F is often called the *infinitesimal generator*.

Let G be a transformation group, the set of all infinitesimal generators of the transforms in G constitutes a vector space, called the *Lie algebra* associated with G, denoted by $\mathrm{Lie}(G)$. As each infinitesimal generator induces a geometric flow, $\mathrm{Lie}(G)$ gives rise to a family of geometric flows, of which the induced transforms all lie in G.

Suppose $\mathrm{Lie}(G)$ is an L-dimensional vector space. Given a basis $(E_1, \ldots, E_L)$, the infinitesimal generator of each flow in the corresponding family can be decomposed into a linear combination of the basis and uniquely characterized by a coefficient vector $\boldsymbol{\alpha} = [\alpha^1, \ldots, \alpha^L]^T$, as

$$V_F = \sum_{l=1}^{L} \boldsymbol{\alpha}^l E_l. \tag{5}$$

The vector $\boldsymbol{\alpha}$ is called the *Lie algebraic representation* of the flow F with respect to the given basis and describes the F as a combination of basic motion patterns.

A Lie algebraic representation has two advantages: (1) The functional form F of a geometric flow is in general nonlinear. As many statistical models presume an underlying vector space, this complicates a statistical model of flows. Exploiting the linear nature of the infinitesimal generator, the Lie algebraic representation largely overcomes such difficulties. (2) Geometric constraints of a flow which typically restrict the induced transforms to a particular subgroup are often nonlinear in functional form. Such constraints become linear with the Lie algebraic representation as each subgroup of transforms is described by a linear subspace of the Lie algebra.

3.3. Flow Stitching and Consistent Subspace

We extend the Lie algebraic representation via flow stitching in order to model the complex motion patterns one observes in natural scenes. Specifically, we partition the scene using a triangle mesh with m cells and n vertices, and attach each cell with a flow from a basic flow family. Collectively, the local flows defined within each mesh element form a global flow. It can be shown that, in the limit of a large number of mesh elements, such a representation can describe all diffeomorphic flows.

Suppose the basic flow family is associated with a K-dimensional Lie algebra, with a basis $(B_1, \ldots, B_K)$. Then the local flow of the i-th cell can be represented by a K-dimensional Lie algebraic representation $\boldsymbol{\beta}_i = [\beta_i^1, \ldots, \beta_i^K]^T$. Thus, the velocity field induced by this local flow is given by $V_{F_i}(x) = \sum_{k=1}^{K} \beta_i^k B_k(x)$.

Local flows may generate different velocities at shared vertices, however, consistency across mesh elements can be imposed by an additional subspace constraint as we describe here. Consider a vertex x shared by the i-th and j-th cells. In general, $V_{F_i}(x)$ may not equal $V_{F_j}(x)$, leading to discontinuities at a cell boundary. To avoid such inconsistencies, we require that the local flows yield the same velocities at shared vertices, i.e. $V_{F_i}(x) = V_{F_j}(x)$, resulting in the *consistency constraints* of the local coefficients as

$$\sum_{k=1}^{K} (\beta_i^k - \beta_j^k) B_k(x) = 0. \tag{6}$$

As the consistency constraints are linear, they give rise to a subspace of the joint Lie algebra, called the *consistent subspace*. In general, its dimension d_C in depends on the choice of both the basic flow family and the mesh topology.

4. Stochastic Model Estimation

With the Lie algebraic representation, the estimation of a geometric flow F naturally reduces to the estimation of the coefficients of its infinitesimal generator V_F with respect to a given set of basis $(E_1, \ldots, E_L)$. This also simplifies

incorporating a motion prior. In the following, we present a robust and efficient algorithm to estimate the flows from noisy observations. We first discuss the generative model of observations and the prior model of geometric flows, respectively. Subsequently, we combine these two parts to derive a joint model for flow estimation.

4.1. Generative Model of Observations

Consider a geometric flow F, whose infinitesimal generator is $V_F = \sum_{l=1}^{L} \alpha^l E_l$. We model the position of a point at time t by a random variable X_t and add a Brownian motion B_t to model noisy deviations arising in real scenes. This leads to the *stochastic flow* as

$$F_\mathbf{G} : \quad dX_t = V_F(X_t) dt + \mathbf{G} dB_t. \tag{7}$$

Here, $\mathbf{G}$ is a coefficient matrix. Let $\mathbf{x}_t$ be a sample trajectory of the stochastic process $F_\mathbf{G}$, whose position at time t is known. Then, if $|\Delta t|$ is sufficiently small, we have

$$p(\mathbf{x}_{t+\Delta t} | \mathbf{x}_t) \sim \mathcal{N}(\mathbf{x}_t + V_F(\mathbf{x}_t) \Delta t, \boldsymbol{\Sigma}_G |\Delta t|), \tag{8}$$

where $\mathcal{N}$ denotes the Gaussian distribution, and $\boldsymbol{\Sigma}_G = \mathbf{G}\mathbf{G}^T$. Note that while the geometric flow model is global, it can be efficiently estimated based on a collection of local observations extracted over the entire scene. In this paper, we utilize two types of observations, as described below.

1. Point pairs. We first extract SIFT feature points[11] in each frame, and then establish the correspondence between the feature points in consecutive frames by comparing their descriptors. Each pair of the positions of matched points is called a *point pair*. Following Eq.(8), we derive the generative likelihood of a point pair $(\mathbf{x}, \mathbf{x}')$ based on $F_\mathbf{G}$ as

$$p((\mathbf{x}, \mathbf{x}') | F_\mathbf{G}) = \mathcal{N}\left(\mathbf{x}' - \mathbf{x} \left| \sum_{l=1}^{L} \alpha^l E_l(\mathbf{x}), \boldsymbol{\Sigma}_G \Delta t + \boldsymbol{\Sigma}_s \right.\right). \tag{9}$$

where Δt is the time interval between consecutive frames, and $\boldsymbol{\Sigma}_s$ is the covariance matrix of measurement noise.

2. Image sequences: Each image sequence is denoted by $\mathcal{I}_S = (I_{t_0}, \ldots, I_{t_J})$, where I_{t_i} is the frame captured at time t_i. As the stochastic flow is a Markov process, we have

$$p(\mathcal{I}_S | F_\mathbf{G}) = \prod_{j=1}^{J} p(I_{t_j} | I_{t_{j-1}}; F_\mathbf{G}). \tag{10}$$

Assuming that observed pixels are independent conditioned on the previous frame, we get

$$p(I_{t_j} | I_{t_{j-1}}; F_\mathbf{G}) = \prod_{\mathbf{x} \in \mathcal{D}_I} p(I_{t_j}(\mathbf{x}) | I_{t_{j-1}}; F_\mathbf{G}). \tag{11}$$

Here $I_{t_j}(\mathbf{x})$ is the pixel value of I_{t_j} at location $\mathbf{x}$, and $\mathcal{D}_I$ is the set of all observed pixel locations in the flow
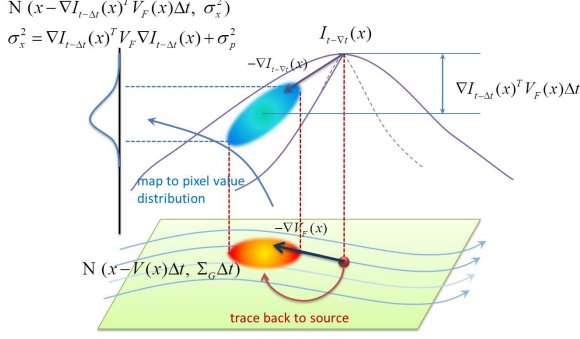
Figure 3. Each pixel in current frame is modeled as generated by moving a source pixel along the flow to current position. To get its distribution, we first trace the pixel backward along the flow to obtain the distribution of source point location, and then map it to the distribution of pixel values through the image. The additional term σ_p^2 is to capture the measurement noise of pixel values.

domain. Through back tracing (see Figure 3), we obtain $p(I_{t_j}(\mathbf{x})|I_{t_{j-1}}; F_{\mathbf{G}}) = \mathcal{N}(I_{t_{j-1}}(\mathbf{x}) - \mu_{\mathbf{x}}, \sigma_{\mathbf{x}}^2)$, with

$$\mu_{\mathbf{x}} = \nabla I_{t_{j-1}}(\mathbf{x})^T V_F(\mathbf{x})\Delta t, \tag{12}$$

$$\sigma_{\mathbf{x}}^2 = \nabla I_{t_{j-1}}(\mathbf{x})^T \Sigma_G \nabla I_{t_{j-1}}(\mathbf{x})\Delta t + \sigma_p^2. \tag{13}$$

Note here that Eq.(13) suppresses the influence of the pixels with high contrast neighborhood. Using this result together with Lie algebraic representation, we can further expand each factor in Eq.(11) as

$$p(I_{t_j}(\mathbf{x})|I_{t_{j-1}}; F_{\mathbf{G}}) =$$
$$\mathcal{N}\left(I_{t_j}(\mathbf{x}) - I_{t_{j-1}}(\mathbf{x}) \middle| -\sum_{l=1}^{L} \alpha^l \mu_{\mathbf{x}}^{(l)}, \sigma_{\mathbf{x}}^2 \right). \tag{14}$$

Here, $\mu_{\mathbf{x}}^{(l)} = \nabla I_{t_{j-1}}(\mathbf{x})^T E_l(\mathbf{x})\Delta t$. Point pairs and image sequences are complementary in flow estimation. While the changes in smoothly varying region can be reliably captured by pixel differences, the motion in high contrast area can be more effectively captured by the point pairs that are extracted using SIFT detectors.

4.2. Gaussian Process Prior

Persistent flows in natural scenes exhibit complex yet spatially coherent variations. Effective modeling of such flows requires a fine-grained mesh that may compromise the spatial coherence. Consequently, we incorporate a Gaussian process (GP) [15] as a prior on flows to enforce long-range spatial coherence while retaining modeling flexibility.

Let $\mathbf{c}_i$ be the circumcenter of the i-th cell, and $\beta_i = [\beta_i^1, \ldots, \beta_i^K]^T$ be the associated local Lie algebraic representation. The covariance function is defined over the local representations, such that

$$\text{cov}(\beta_i^k, \beta_j^k) = \sigma_\beta^2 \exp\left(-\frac{1}{2}\frac{||\mathbf{c}_i - \mathbf{c}_j||^2}{\sigma_{gp}^2} \right). \tag{15}$$
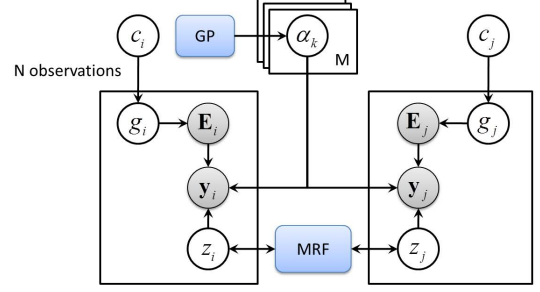
Figure 4. This graphical model incorporates the generative likelihood of each observation (discussed in section 4.1) and the GP-prior of the flows (discussed in section 4.2). In addition, each observation entry is associated with a label variable z_i and an inlier indicator g_i. The label variables are connected to each other through an MRF, while the distribution of g_i is specified by a confidence value c_i, i.e. the prior probability of $g_i = 1$.

This leads to a Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{G}_\beta)$ of the concatenated local representation, in which $\mathbf{G}_\beta$ is an $mK \times mK$ matrix. Recall that under consistency constraints, we can write $\boldsymbol{\beta} = \mathbf{U}\boldsymbol{\alpha}$ with a d_C-dimensional coefficient vector $\boldsymbol{\alpha}$. Hence, the GP-prior of a consistently stitched flow can be further derived as $p(\boldsymbol{\alpha}) = \mathcal{N}\left(\mathbf{0}, (\mathbf{U}^T\mathbf{G}_\beta^{-1}\mathbf{U})^{-1} \right)$.

4.3. Robust Estimation of Concurrent Flows

From Eq.(9) and (14), we see that the likelihood of both point pairs and image sequences has a similar form, in spite of their different generation processes. We can write them uniformly as products of $\mathcal{N}(\mathbf{y}_i|\mathbf{E}_i\boldsymbol{\alpha}, \Sigma_i)$. The triple $(\mathbf{y}_i, \mathbf{E}_i, \Sigma_i)$ is called an *observation entry*.

Multiple coexisting flows are common in natural scenes. Consequently, we need not only estimate each individual flow, but also determine its spatial domains. In doing so, we model the scene with M flow models, including a "background flow" with its Lie algebraic representation fixed to zero. In addition, we associate each observation entry with a hidden variable z_i for assigning that observation to a particular flow. An MRF among z_i is incorporated to encourage assigning the same label to observations that are spatially closed to each other. Moreover, the errors due to mismatch of point pairs or the gradient computed across sharp boundary may severely bias estimation results. Therefore, we incorporate a binary variable g_i for each observation entry, which indicates whether the observation is an inlier.

The graphical model of the joint probabilistic framework is shown in figure 4. The estimation is done via variational EM based on a mean-field approximation of the posteriori. The algorithm iteratively re-estimates the flow coefficients

using the relabeled observation entries, re-assigns each observation to the updated models, and updates the inlier probabilities. Graphcut[5] is used for re-labeling in the variational E-steps.

5. Experiments

In contrast to much of the prior work on motion analysis where accurate estimation of local velocities is the major concern, this paper aims at discovering the persistent patterns. Hence, the capability of being generalized to model unseen frames is an important aspect to examine. To test this ability, we conducted experiments on videos from Dyn-Tex database [3] and UCF Crowd Analysis database [1]. These videos represent a wide spectrum of dynamic scenes.

5.1. Experiment Settings

We use consistently stitched flows to model the motion patterns in each video. Each flow is established on a triangle mesh that covers the entire scene and parameterized by the Lie algebraic representation. To generate the triangle mesh, we first make a rectangle mesh with 5 rows and 6 columns, and divide each rectangle cell into two triangles. Affine flow family is chosen as the basic flow family for describing the motion within each cell. This family is associated with a 6-dimensional Lie algebra. While we found that this setting suffices to obtain good results on the testing videos, generally there is no restriction to the choice of basic flow family and mesh topology. One can also use other flow families with our framework to model more complex motion. The representation dimension (the dimension of the consistent subspace) is $L = 84$, which is much smaller than that of the dense velocity map. In addition, we use GP-Prior to enforce long-range spatial coherence, where σ_{gp} is set to 100.

We use multiple concurrent flows to model a dynamic scene, including one that corresponds to the static background. The number of flows in each scene is given. To initialize the algorithm, we manually specify a "seed area" for each flow, and the observations contained in these areas are used to estimate the initial flow models. Besides, we set the prior confidence of inlier to 0.9 for each observation entry. Flow segmentation, inlier probabilities, and flow models are then iteratively updated until convergence.

To compare our approach with traditional local motion estimation methods, we implement an optical flow estimation algorithm with multi-scale search and local smoothness. Moreover, we adapt the algorithm to incorporate multiple frames in estimation for fair comparison. This is accomplished by assuming a time-invariant velocity v at each location, and integrating the term $||\frac{\partial I}{\partial t} + v^T \nabla I||^2$ for every pair of consecutive frames into the objective function. The design parameters of the optical flow, including the local search range and the coefficients in the smoothness terms, are optimized by cross validation.

5.2. Collective Behavior of Moving Objects

The algorithms are first evaluated on the scenes that comprise groups of moving objects, such as people and vehicles. To test the generalization performance, for each video, we first manually trace 20 trajectories, whose nominal duration is 100 frames, as ground truth, and then use the first 20 frames to estimate the motion models. These models are then used to simulate the trajectories starting from the initial positions as those in the manually traced ones. The performance is measured by the deviation of the simulated trajectories from the ground truth.

Figure 5 shows the results. The results shown in the first row are obtained on a scene with cars moving along a high way. We see that the optical flow is over-fitted to the short-time behavior of individual cars: (1) it only extracts motion in the places where cars are passing by during first 20 frames; (2) the velocity map lacks spatial coherence. For the same example, geometric flow accurately capture the collective behavior of the cars, while preserving spatial coherence. Note that the subtle variation of the moving direction of the cars is precisely captured in the flow model.

We also evaluate the trajectory prediction performance, observing that the predicted trajectories simulated on the optical flow field quickly deviate from the ground truth; while the ones yielded by geometric flow are much more accurate. The plotted error curves indicate that the average deviation due to the optical flow is more than 8 times larger than that due to the geometric flow.

The second row shows the scene with a crowded group of athletes running along a circular path. Similar observations are obtained in this example. Again, due to its local focus, the motion field produced by optical flow lacks spatial coherence and doesn't generalize well, while geometric flow yields much better generalization performance.

5.3. Continuous Motion Patterns

The tests are also done on modeling continuous motion patterns, such as flowing water and deforming objects. In Figure 6, the first column shows a mountain spring comprised of several sections with different motion patterns. To model this spring, we use four concurrent flows. The second column shows a disc rotating in a very high speed, whose appearance is severely blurred. The water transparency and the blurred texture on the disc lead to great challenges for motion estimation. In the face of such difficulties, optical flow performs poorly, resulting in meaningless motion patterns. Nonetheless, the geometric flow still works well. The subtle variation of the water flowing direction is precisely modeled while the spatial coherence is well preserved. The rotation of the disc in the right column is also successfully captured by the geometric flow.
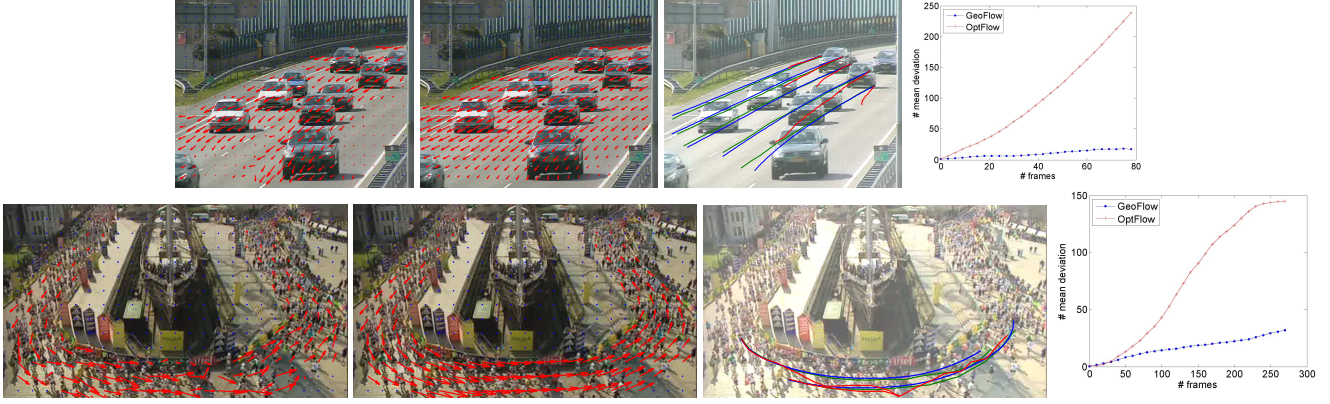
Figure 5. The first row shows the scene with cars moving on a high way; the second row shows a group of athletes running along a circular path. In each row, from left to right, the first two pictures show the results obtained using optical flow and geometric flow respectively. The estimated motion patterns are visualized as velocity fields. The third picture shows a subset of the trajectories predicted by the models. The blue curves are those yielded by geometric flow, while the red ones are yielded by optical flow. The ground-truth trajectories are also shown, which are in green color. The fourth picture compares the trajectory-prediction errors quantitatively.

	Fitting Error	Generalized Error
OptFlow	0.0213	0.0785
GeoFlow	0.0228	0.0242

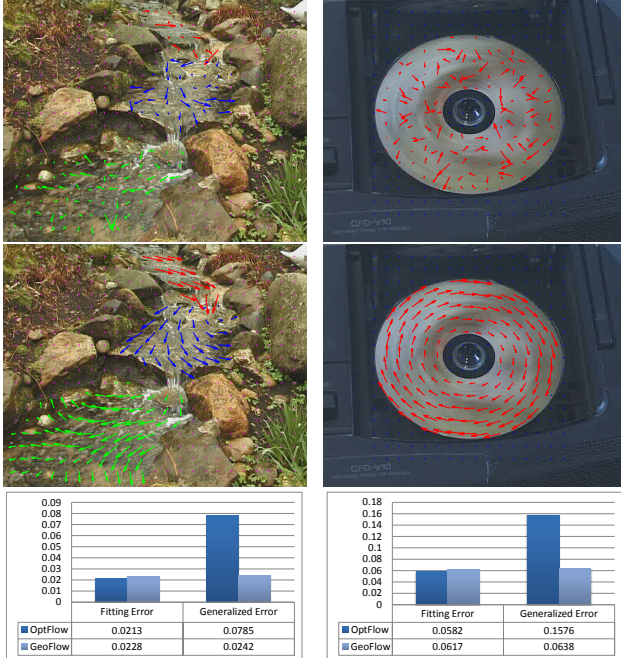|	Fitting Error	Generalized Error
OptFlow	0.0582	0.1576
GeoFlow	0.0617	0.0638

Figure 6. In this figure, the first column shows the results obtained on modeling the flowing water in a mountain spring. The second column shows that on a rotating disc. The bottom row are two charts, giving the average fitting errors and generalization errors obtained from the corresponding example.

As there are no discrete targets that can be tracked in these scenes, we use frame prediction to measure the performance. Concretely, we generate the frames from their preceding frames based on the predicted motion, which are then compared with the actual frames, in terms of average pixel-wise frame prediction error. The performance is mea-
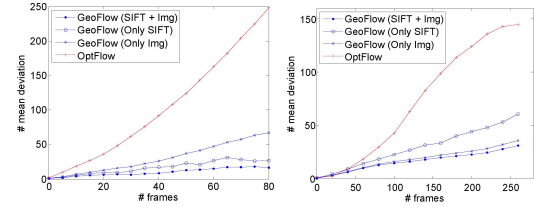

Figure 7. The trajectory prediction errors with different types of observations. The left and right charts are respectively obtained from the scene with moving cars and that with running athletes.

sured respectively for training frames and testing frames, which respectively reflect the fitting accuracy and generalization performance. We see that while the geometric flows fit slightly less accurately to the training frames, their generalize remarkably better than the optical flow.

5.4. The Roles of Different Factors

Observations: To study how the use of different observations affects the performance, we test three settings: (1) only using image frames, (2) only using SIFT pairs, and (3) using both. Figure 7 compares the trajectory prediction errors under these settings, as well as that due to optical flow estimation. The results show that these two types of observations have relative strength in different scenes. Point pairs perform better in the scenes with structured appearance where they can be accurately located; while pixel differences are more reliable for the scenes with smooth textures. However, in either case, the combination of both leads to further improvement, which reflects their complementary nature.

GP-Priors: We test the framework with different values of σ_{gp} to study how GP-prior influences the estimation. In Fig-
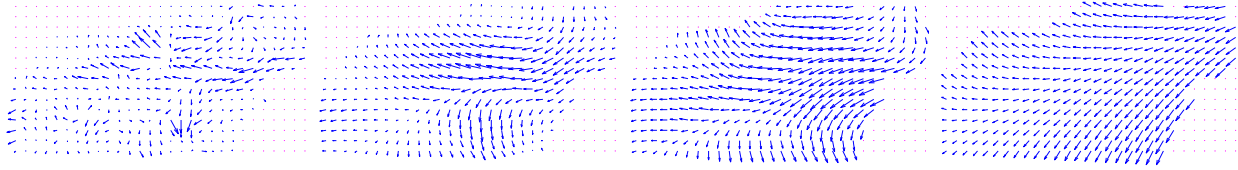
Figure 8. The figure shows the motion patterns of the bottom-right part of the mountain spring estimated under different settings. From left to right, the results are obtained by optical flow, geometric flows with σ_{gp} set to 0, 100, and 10000 respectively.

ure 8, we see that the motion pattern becomes more coherent as σ_{gp} increases. When σ_{gp} approaches infinity, local flow of every cell is restricted to be the same, resulting in a uniformly affine field. When σ_{gp} approaches zero, long-range consistency is no longer enforced. While the result obtained under this setting is less coherent than that with GP utilized, it still preserves the coherence within each cell and the consistency between neighboring cells, and thus is better than that of the optical flow.

Training Frames: In general, for both geometric flows and optical flows, the generalization performance increases as one uses more frames to estimate the model. However, owning to its intrinsic geometric structure, geometric flows can reliably capture the underlying motion patterns with much less training frames, while for optical flow, it requires one third of the entire sequence as training frames to achieve comparable generalization performance.

6. Conclusion

We developed a new model of persistent motion based on geometric flow that establishes a global perspective of the dynamics, and introduced the Lie algebraic representation that simplifies the model estimation. The comparative experiments conducted on a variety of dynamic scenes show that the presented approach achieve remarkably better generalization performance than optical flow estimation.

Acknowledgment

References

[1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proc. of CVPR'07*, 2007.

[2] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *ECCV'08*, 2008.

[3] T. Amiaz, S. Fazekas, D. Chetverikov, and N. Kiryati. Detecting regions of dynamic textures. In *Conf. on Scale Space and Variational Methods in Computer Vision*, 2007.

[4] Bastian, K. Schindler, and L. V. Gool. Coupled detection and trajectory estimation for multi-object tracking. In *Proc. of ICCV'07*, 2007.

[5] Y. Boycov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*, 23(11):1222–1239, 2001.

[6] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proc. of ECCV'04*, 2004.

[7] A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Trans. on PAMI*, 30(5):909–926, May 2008.

[8] N. Jacobs, M. Dixon, and R. Pless. Location-specific transition distributions for tracking. In *WMVC'08*, 2008.

[9] J. M. Lee. *Introduction to Smooth Manifolds*. Springer, 2002.

[10] D. Lin, E. Grimson, and J. Fisher. Learning visual flows: A lie algebraic approach. In *CVPR'09*, 2009.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int'l J. of Computer Vision*, 60(2), 2004.

[12] B. D. Lucas and T. Kanade. An iterative image registration technique with an application in stereo vision. In *IJCAI*, 1981.

[13] H. T. Nguyen, Q. Ji, and A. W. Smeulders. Spatio-temporal context for robust multitarget tracking. *IEEE Trans. on PAMI*, 29(1):52–64, Jan. 2007.

[14] N. Papadakis, T. Corpetti, and E. Memin. Dynamically consistent optical flow estimation. In *Proc. of ICCV'07*, 2007.

[15] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[16] X. Ren. Local grouping for optical flow. In *Proc. of CVPR'08*, 2008.

[17] S. Roth and M. J. Black. On the spatial statistics of optical flow. In *Proc. of ICCV'05*, 2005.

[18] T. Schoenemann and D. Cremers. High resolution motion layer decomposition using dual-space graph cut. In *Proc. of CVPR'08*, 2008.

[19] X. Wang, K. T. Ma, G.-W. Ng, and E. Grimson. Trajectory analysis and semantic region modeling using a nonparametric bayesian model. In *Proc. of CVPR'08*, 2008.

[20] T. Zhao, R. Nevatia, and B. Wu. Segmentation and tracking of multiple humans in crowded environments. *IEEE Trans. on PAMI*, 30(7):1198–1211, July 2008.