MIT Libraries | DSpace@MIT

MIT Open Access Articles

Learning Visual Flows: A Lie Algebraic Approach

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Dahua Lin, E. Grimson, and J. Fisher. "Learning visual flows: A Lie algebraic approach." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. 2009. 747-754. © 2009 IEEE

As Published: http://dx.doi.org/10.1109/CVPRW.2009.5206660

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: http://hdl.handle.net/1721.1/59336

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

Learning Visual Flows: A Lie Algebraic Approach

Dahua Lin
CSAIL, MIT
dhlin@mit.edu

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

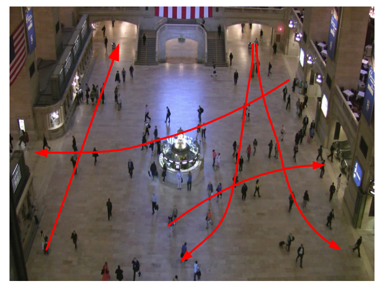
John Fisher
CSAIL, MIT
fisher@csail.mit.edu

Abstract

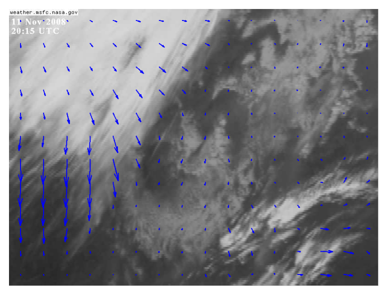
We present a novel method for modeling dynamic visual phenomena, which consists of two key aspects. First, the integral motion of constituent elements in a dynamic scene is captured by a common underlying geometric transform process. Second, a Lie algebraic representation of the transform process is introduced, which maps the transformation group to a vector space, and thus overcomes the difficulties due to the group structure. Consequently, the statistical learning techniques based on vector spaces can be readily applied. Moreover, we discuss the intrinsic connections between the Lie algebra and the Linear dynamical processes, showing that our model induces spatially varying fields that can be estimated from local motions without continuous tracking. Following this, we further develop a statistical framework to robustly learn the flow models from noisy and partially corrupted observations. The proposed methodology is demonstrated on real world phenomenon, inferring common motion patterns from surveillance videos of crowded scenes and satellite data of weather evolution.

1. Introduction

Dynamical analysis plays a crucial role in the understanding and interpretation of visual phenomena. Of particular interest are representations that capture underlying dynamics in a parsimonious manner while enabling tractable analysis. Figure 1 illustrates two applications where modeling of visual dynamics has important utility. While seemingly different, they share a common embedded problem at their core: inferring a model of dynamics that govern the scene from observations. In the rail station scene, *common* motion patterns of individual persons are of primary interest, while in the weather data we are interested in dynamic changes in intensity patterns. With these aims in mind, we develop a method for probabilistic modeling of motions in which we adopt the formalism of Lie groups and Lie algebras. As a result, the use of well known probabilistic models is made simpler. While the existing literature devoted to the study of Lie Groups [2] in disciplines like control and system identification is extensive, we focus on statistical modeling of collective motions for describing image sequences.



(a) people motion pattern analysis



(b) Weather evolution modeling

Figure 1. **(a)** is a video frame captured in a rail station. From the video, we can find out that there are only a small number of motion patterns that most people follow, which may contain useful information for surveillance, anomaly detection, and station management. **(b)** illustrates an infrared satellite picture of atmosphere, and the corresponding optical flow. The dynamic information of the flow would be useful in weather modeling and prediction.

There has been much work in dynamic modeling in computer vision in the past decade. Most approaches fall roughly into one of two categories. The first category makes use of graphical models and their associated Markov structure to describe temporal dependency across frames for continuous tracking of individual objects [1, 3, 6, 14, 17, 19, 22]. Such methods suffer in multi-object tracking scenarios such as in Figure 1(a). The second category explicitly incorporates multiple objects into the tracking framework [7, 12, 13, 23]. These methods typically rely on analysis of full trajectories rather than local temporal motions and as such are susceptible to incorrect measurement-to-track association as well as combinatorial complexity.

Recent methods exploit local motions in which temporally local changes are extracted as features and then grouped by proximity-based clustering algorithms [15] or generative models [17, 20]. These methods have the advantage of circumventing difficulties of full trajectory analysis by avoiding the measurement-to-track association problem, but their proximity-based strategy tends to divide the space into local clusters while missing global scene regularities. Additionally, they assume a single prototypical motion for each group precluding spatial variation in group behavior. As we shall show in subsequent analysis, spatial variation in common group behavior is not unusual.

Motivated by the analysis above, we propose a new approach to dynamic modeling. Rather than grouping objects by proximity, we instead group them by common *flow fields* parameterized by geometric transformations. The objects subjected to the same flow field may have (and typically do have) substantially different locations and velocities, but are consistent in a geometric sense.

One difficulty in developing statistical models of parameterized geometric transformations is that their inherent group structure, which complicates the application of statistical methods that rely on underlying vector spaces. While one may ignore the multiplicative nature of the underlying group structure, treating transform matrices as if they lie in a linear space, this leads to undesired effects and complexities when incorporating geometric constraints. Lie algebra theory mitigates many of these issues. Its main utility is to construct *equivalent* representation in a vector space while maintaining connections to the geometric transform group. Thereby, we acquire a vector representation of each transform to which many widely used statistical models can readily be applied.

Furthermore, we consider the theoretical relations between Lie algebraic representations and linear dynamical processes, resulting in a velocity field perspective *constrained by the group structure*. Thus, we develop a statistical approach for flow-field estimation utilizing observations of local motions that avoids the complications associated with trajectory maintenance. As we will show, robustness as well as outlier motion identification is also attained as a consequence. We test the methodology in two candidate applications. The first involves analysis of motion patterns of people in Grand Central Station from overhead video while the second infers motions within satellite weather image sequences.

The use of geometric transforms is not a new development in computer vision. However, most prior work [4, 11, 18] utilizes geometric transformations and associated Lie group analysis in rigid body tracking or alignment. Our use of geometric transformations is novel in a variety of ways. Transformations are used to describe the collective motion of separate objects, many of whom, though bound by a common motion descriptor, are moving independently of each other. Consequently, (possibly overlapping) flows are associated with regions rather than objects. Additionally, flow fields may persist in a region despite objects entering and leaving. Finally, the geometric characterization of the Lie algebraic representation and its relation to a linear dynamical process, which plays an essential role in our approach, has not been explored.

2. Lie Algebraic Representation

We provide a brief discussion of the groups structure of affine transformations. While the discussion is restricted to two-dimensional affine transformations, we note that the methodology is straightforwardly extended to higher dimensions as well as to other families of transforms. Two-dimensional affine deformations, however, represent a sufficiently rich class for purposes of discussion and to demonstrate the methodology in two candidate applications.

2.1. Lie Algebra of Affine Transforms

Affine transforms, parameterized by $\mathbf{A}$ and $\mathbf{b}$, have the following form

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}. \tag{1}$$

and can be expressed in homogeneous coordinates as

$$\bar{\mathbf{x}}' = \begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \mathbf{T}\bar{\mathbf{x}}. \tag{2}$$

All matrices in this form comprise the elements of an algebraic structure known as a Lie group. While widely used in many vision applications, its use in statistical methods presents some difficulties. It is not a vector space, and thus is not closed under vector addition nor scalar multiplication, complicating the use of statistical learning methods with implicit vector space assumptions. Moreover, it is often the case that one would like to impose geometric constraints upon the transformation. For example, restriction to volume-preserving deformations corresponds to a determinant constraint, i.e. $\det(\mathbf{T}) = 1$. This and a variety of geometric constraints are nonlinear and can be difficult to incorporate into statistical models. The difficulty essentially arises from the fact that the affine group has a multiplicative rather than additive structure. It is desirable to establish a mapping from the multiplicative structure to an *equivalent* vector space representation. This is precisely what the Lie algebra accomplishes in a local sense. There is rich theory of Lie groups and Lie algebras. Interested readers are referred to [8, 10] for an elaborated introduction.

The Lie algebraic representation of a 2D affine transform is a 3×3 matrix with all zeroes on the bottom row. It is related to the homogeneous matrix representation through matrix exponentiation and the matrix logarithm. If $\mathbf{X}$ denotes the Lie algebra representation of $\mathbf{T}$ then

$$\mathbf{T} = \exp(\mathbf{X}) = \mathbf{I} + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{X}^k, \tag{3}$$

$$\mathbf{X} = \log(\mathbf{T}) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (\mathbf{T} - \mathbf{I})^k. \tag{4}$$

As we shall see, by working with this representation, the algebraic structure becomes additive and as a consequence many statistical learning methods may be readily applied.

2.2. Geometric Characterization and Constraints

One advantage of the Lie algebraic representation is that transformation *subgroups* are mapped to linear *subspaces*. Within the 2D affine group, there are many subgroups that correspond to particular families of transforms. This gives rise to a linear parameterization of them. Consider rotations by an angle θ of which the transform matrix $\mathbf{T}_{R(\theta)}$ and the corresponding Lie algebraic representation $\mathbf{X}_{R(\theta)}$ are

$$\mathbf{T}_{R(\theta)} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathbf{X}_{R(\theta)} = \begin{bmatrix} 0 & -\theta & 0 \\ \theta & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{5}$$

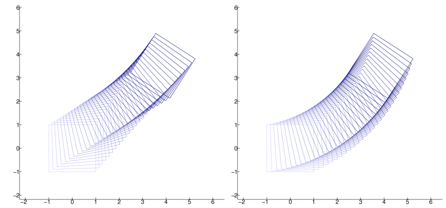
It can be easily seen that the Lie algebraic representation of all rotations lies in a one dimensional subspace. Similarly, the Lie algebraic representations of many other important transforms such as scaling, shearing, and translation, correspond to subspaces of the Lie algebra, as well. This property in turn allows for linear characterization of a variety of geometric constraints. Consider the volume-preserving constraint discussed above. Since the composition of two volume-preserving transforms is also volume-preserving. All volume-preserving transforms constitute a subgroup of the affine group. Consequently, their Lie algebraic representations form a subspace. The associated constraint is captured by the simple expression

$$\mathrm{tr}(\mathbf{X}) = 0 \quad \Leftrightarrow \quad X_{11} + X_{22} = 0. \tag{6}$$

2.3. The Perspective of Processes and Field

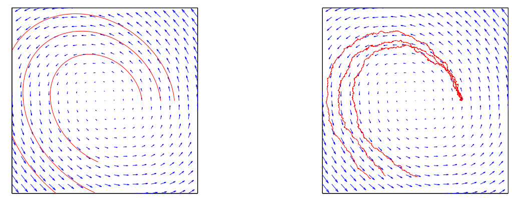
We are interested in describing transformations as a continuous-time process. Assuming that the transformation occurs over a unit of time, the motion of each point can then be expressed as a function of time $\bar{\mathbf{x}}(t)$ such that $\bar{\mathbf{x}}(0) = \bar{\mathbf{x}}_0$ and $\bar{\mathbf{x}}(1) = \mathbf{T}\bar{\mathbf{x}}_0 = \exp(\mathbf{X})\bar{\mathbf{x}}_0$.

Two ways to parameterize the process lead to very different transformation sequences. A common approach, using the homogeneous representation, is to define $\bar{\mathbf{x}}(t) = \bar{\mathbf{x}}_0 + t(\mathbf{T} - \mathbf{I})\bar{\mathbf{x}}_0$. Alternatively, utilizing the Lie algebraic representation one can define $\bar{\mathbf{x}}(t) = \exp(t\mathbf{X})\bar{\mathbf{x}}_0$. Figure 2 compares the effect of these two approaches, in which the terminal transformation is a translation plus rotation, i.e. a rigid transformation. In contrast to the conventional approach, when using the Lie algebra representation, each intermediate transform is also a rigid transformation, i.e. the elements of the sequence remain within the subgroup. Generally, when the process is parameterized using the Lie algebra representation, if the initial and final transformations are within a subgroup, then intermediate transformations along the path will also lie in that subgroup. Moreover, the process defined using the Lie algebra is optimal in the sense that it corresponds to the shortest geodesics connecting the two transforms on the affine manifold.



(a) Traditional way (b) Lie algebraic way

Figure 2. This illustrates the sequence of intermediate transforms in the continuous time process from initial one to terminal one.



(a) with noiseless processes (b) with noisy processes

Figure 3. Illustration of velocity field of an affine transform.

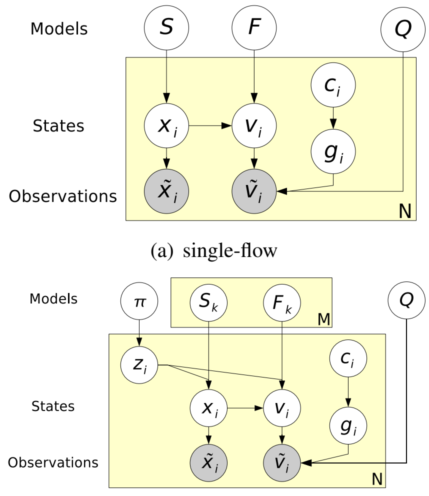
The process $\bar{\mathbf{x}}(t) = e^{t\mathbf{X}}\bar{\mathbf{x}}_0$ has intrinsic relations with the Linear dynamical systems in control theory. It is the solution to the following system of differential equations:

$$\frac{d\bar{\mathbf{x}}(t)}{dt} = \mathbf{X}\bar{\mathbf{x}}(t), \quad \bar{\mathbf{x}}(0) = \bar{\mathbf{x}}_0. \tag{7}$$

Therefore, a geometric transform can be equivalently represented as a dynamic process governed by the above differential equation. The Lie algebraic representation is in form of $\mathbf{X} = \begin{bmatrix} \mathbf{Y} & \mathbf{u} \\ \mathbf{0} & 0 \end{bmatrix}$, hence equation (7) can be rewritten as

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{Y}\mathbf{x} + \mathbf{u}, \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{8}$$

Since $\frac{d\mathbf{x}(t)}{dt}$ represents the velocity, this equation expresses the velocity as a function of location, or in other words, establishes the time-invariant velocity field, where there is a unique velocity at each location. Figure 3(a) illustrates a velocity field with several processes driven by it. In this system, the entire trajectory of a point is uniquely determined



(a) single-flow

(b) multi-flow

Figure 4. The graphical models of the statistical formulation.

by its initial position. This is too restrictive for our purposes in which we assume that objects in the same flow may follow different paths even when starting from the same location. This issue can be addressed by adding a noise term to the state equation as

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{Y}\mathbf{x} + \mathbf{u} + \varepsilon_t, \tag{9}$$

where ε is a white noise process. The introduction of the noise term converts the ordinary differential equation to a stochastic differential equation, which has remarkably different statistical characteristics. The stochastic nature of the extended system offers substantially greater flexibility. Figure 3(b) shows several processes starting with the same initial position that end up with distinct trajectories.

3. Statistical Modeling and Inference

As a consequence of adopting the Lie algebra representation, a statistical method for inference of affine flow fields can be developed. Importantly, we rely solely on local velocity observations, avoiding the difficulties of trajectory maintenance.

3.1. Statistical Formulation

In (9), the velocity of each point depends solely on its location. While the equation can be used to synthesize full trajectories, when the locations are observed the estimation procedure only requires local velocity observations eliminating the need for continuous tracking.

Observations are location-velocity pairs $\{\mathbf{x}_i, \mathbf{v}_i\}_{i=1}^N$ that can be acquired from an image sequence with a variety of algorithms (e.g. image alignment, feature matching, and optical flow estimation).

Figure 4(a) shows the graphical model of the above formulation. The generative process is described as follows:

1) The i-th location is drawn from a Gaussian distribution $p(\mathbf{x}_i|S) = \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_S, \boldsymbol{\Sigma}_S)$ where $S = (\boldsymbol{\mu}_S, \boldsymbol{\Sigma}_S)$ is the mean and covariance (sufficiently large to cover a region). From this we obtain a noisy location $\tilde{\mathbf{x}}_i$ drawn from $p(\tilde{\mathbf{x}}_i|\mathbf{x}_i) = \mathcal{N}(\tilde{\mathbf{x}}_i|\mathbf{x}_i; \boldsymbol{\Sigma}_X)$.

2) The i-th velocity $\mathbf{v}_i$ is drawn from the flow model (the core of our framework) utilizing the stochastic field equation (9). Assuming Gaussian driving noise yields $p(\mathbf{v}_i|\mathbf{x}_i; F) = \mathcal{N}(\mathbf{v}_i|\mathbf{Y}_F\mathbf{x}_i + \mathbf{u}_F, \boldsymbol{\Sigma}_F)$ where $F = (\mathbf{Y}_F, \mathbf{u}_F, \boldsymbol{\Sigma}_F)$ is comprised of a mean transformation and the covariance of the driving noise.

3) The measurement model of velocity accounts for two types of errors, additive noise and incorrect matching (e.g. from frame-to-frame). We introduce a hidden variable g_i to account for the latter ($g_i = 1$ for correct matching and $g_i = 0$ for incorrect matching). The prior probability on g_i is denoted c_i and depends on the data acquisition procedure. When $g_i = 1$, $\tilde{\mathbf{v}}_i$ is a noisy observation of $\mathbf{v}_i$; $p(\tilde{\mathbf{v}}_i|\mathbf{v}_i, g_i = 1) = \mathcal{N}(\tilde{\mathbf{v}}_i|\mathbf{v}_i; \boldsymbol{\Sigma}_V)$. When $g_i = 0$, $\tilde{\mathbf{v}}_i$ is drawn from an uninformative uniform distribution Q accounting for incorrect matches as $p(\tilde{\mathbf{v}}_i|\mathbf{v}_i, g_i = 0; Q) = Q(\tilde{\mathbf{v}}_i)$.

The formulation involves several covariance matrices: $\boldsymbol{\Sigma}_S, \boldsymbol{\Sigma}_F, \boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_V$, which are set in the design phase, rather than estimated from the data. Assuming that all observations are independently sampled from the above model, the joint probability is written as

$$\prod_{i=1}^N p(\mathbf{x}_i|S)p(g_i|c_i)p(\mathbf{v}_i|\mathbf{x}_i; F)p(\tilde{\mathbf{x}}_i|\mathbf{x}_i)p(\tilde{\mathbf{v}}_i|\mathbf{v}_i, g_i; Q). \tag{10}$$

To account for the scenes governed by multiple flows, we extend the statistical formulation to a mixture model as shown in figure 4(b). The multi-flow formulation comprises M submodels. Each submodel contains a spatial model S_k and a flow model F_k such that each flow covers a particular region in the scene and where regions may overlap, resulting in a region-dependent mixture model. Accordingly, we introduce an additional hidden variable z_i with multinomial distribution π to indicate which sub-model the observation is generated from. The joint probability of the multi-flow model is thus

$$\prod_{i=1}^N p(z_i|\pi)p(\mathbf{x}_i|z_i; S)p(\mathbf{v}_i|\mathbf{x}_i, z_i; F)p(\tilde{\mathbf{x}}_i|\mathbf{x}_i)p(\tilde{\mathbf{v}}_i|\mathbf{v}_i, g_i; Q), \tag{11}$$

where $S = (S_1, \ldots, S_M)$ and $F = (F_1, \ldots, F_M)$ refer to a collection of models.

3.2. The Prior of Flow Model

As mentioned above, the Lie algebraic representation lies in a vector space and geometric constraints can be expressed as systems of linear equations, simplifying the incorporation of domain-specific knowledge. If the transform is in a particular Lie subgroup (e.g. due to some geometric constraints), then the corresponding Lie algebra representation lies in a linear subspace of the full Lie algebra. We denote the dimension of the representation space by L and its basis by$\{(\mathbf{Y}^l_B, \mathbf{u}^l_B)\}_{l=1}^L$. Any element in this space can be expressed as a linear combination of its basis elements

$$(\mathbf{Y}, \mathbf{u})_{\boldsymbol{\alpha}} = \sum_{l=1}^{L} \alpha_l (\mathbf{Y}^{(l)}_B, \mathbf{u}^{(l)}_B), \tag{12}$$

where the coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)^T$ parameterizes the transform. The estimation of the flows is thus transformed to the estimation of the vector $\boldsymbol{\alpha}$.

As the parameters also lie in a vector space, we can choose from a large group of well-studied models whose underlying measure is Euclidean in nature for their prior distribution. Here, we simply use a Gaussian distribution as their prior: $p(\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\alpha}|\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_\alpha)$. We emphasize that the incorporation of prior probabilistic models and constraints is a result of adopting the Lie algebraic representation.

3.3. Learning by Variational Inference

We would like to infer the flow models by maximizing the following objective

$$J_{map}(S, F, \pi) = \sum_{i=1}^{N} \log p(\tilde{\mathbf{x}}_i, \tilde{\mathbf{v}}_i | S, F, Q, c) + \log p(\boldsymbol{\alpha}). \tag{13}$$

that is, via MAP estimation. Here, the corrupted measurement model Q, prior probability on correct matches c are assumed to be given. Exact evaluation of (13) is generally intractable as it involves marginalization of $z_i, \mathbf{x}_i, \mathbf{v}_i$ and g_i. This motivates approximate inference techniques where there are two primary strategies MCMC sampling and variational inference. As our formulation conforms to the conjugate-exponential family, we use a variational inference approach to maximizing the following lower bound of Equation (13):

$$\begin{aligned} J_{var}(S, F, \pi) = &\sum_{i=1}^{N} \mathbb{E}_{q_i} \log p(\tilde{\mathbf{x}}_i, \tilde{\mathbf{v}}_i, \mathbf{x}_i, \mathbf{v}_i, z_i, g_i | S, F, Q, c) \\ &- \sum_{i=1}^{N} \mathbb{E}_{q_i} \log q_i(\mathbf{x}_i, \mathbf{v}_i, z_i, q_i) + \log p(\boldsymbol{\alpha}). \end{aligned} \tag{14}$$

Here, q_i is referred to as the variational distribution and is chosen so as to incorporate hidden variables in a tractable way. As is commonly done, we choose q_i as a product distribution

$$\begin{aligned} q_i(\mathbf{x}_i, \mathbf{v}_i, z_i, g_i) = &\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_{x_i}, \mathbf{R}_{x_i}) \mathcal{N}(\mathbf{v}_i|\boldsymbol{\mu}_{v_i}, \mathbf{R}_{v_i}) \\ &\text{Mult}(z_i|\tau_i) \text{Bernoulli}(g_i|\rho_i). \end{aligned} \tag{15}$$

The procedure iterates between variational E-steps and M-steps until convergence. In E-steps, we update the parameters of the variational distributions as

$$\hat{\mathbf{R}}^{-1}_{x_i} = \sum_{k=1}^{M} \tau_{ik} \left(\boldsymbol{\Sigma}^{-1}_{S_k} + \mathbf{Y}_k^T \boldsymbol{\Sigma}^{-1}_{F_k} \mathbf{Y}_k \right) + \boldsymbol{\Sigma}^{-1}_X;$$

$$\hat{\mathbf{R}}^{-1}_{x_i} \hat{\boldsymbol{\mu}}_{x_i} = \sum_{k=1}^{M} \tau_{ik} \left(\boldsymbol{\Sigma}^{-1}_{S_k} \boldsymbol{\mu}_{S_k} + \mathbf{Y}_k^T \boldsymbol{\Sigma}^{-1}_{F_k} (\boldsymbol{\mu}_{v_i} - \mathbf{u}_k) \right) + \boldsymbol{\Sigma}^{-1}_X \tilde{\mathbf{x}}_i;$$

$$\hat{\mathbf{R}}^{-1}_{v_i} = \sum_{k=1}^{M} \tau_{ik} \boldsymbol{\Sigma}^{-1}_{F_k} + \rho_i \boldsymbol{\Sigma}^{-1}_V;$$

$$\hat{\mathbf{R}}^{-1}_{v_i} \hat{\boldsymbol{\mu}}_{v_i} = \sum_{k=1}^{M} \tau_{ik} \boldsymbol{\Sigma}^{-1}_{F_k} (\mathbf{A}_k \boldsymbol{\mu}_{x_i} + \mathbf{b}_k) + \rho_i \boldsymbol{\Sigma}^{-1}_V \tilde{\mathbf{v}}_i;$$

$$\begin{aligned} \log \hat{\tau}_{ik} = &\log \pi_k + \log \mathcal{N}(\boldsymbol{\mu}_{x_i} | \boldsymbol{\mu}_{S_k}, \boldsymbol{\Sigma}_{S_k}) + \log \mathcal{N}(\boldsymbol{\mu}_{v_i} | \mathbf{f}_k(\boldsymbol{\mu}_{x_i}), \boldsymbol{\Sigma}_{F_k}) \\ &- \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}_{S_k} \mathbf{R}_{x_i}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}_{F_k} (\mathbf{R}_{x_i} + \mathbf{R}_{v_i})) + \text{const}; \end{aligned}$$

with $\sum_{i=1}^{k} \hat{\tau}_{ik} = 1$ and $\mathbf{f}_k(\boldsymbol{\mu}_{x_i}) = \mathbf{Y}_k \mathbf{x}_i + \mathbf{u}_k$.

$$\eta(\hat{\rho}_i) = \eta(c_i) + \log \mathcal{N}(\tilde{\mathbf{v}}_i | \boldsymbol{\mu}_{v_i}, \boldsymbol{\Sigma}_V) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}_V \mathbf{R}_{v_i}) - \log Q(\tilde{\mathbf{v}}_i);$$

$$\hat{\rho}_i = \frac{e^{\eta(\hat{\rho}_i)}}{1 + e^{\eta(\hat{\rho}_i)}}.$$

While the M-steps update the spatial models as

$$\hat{\boldsymbol{\mu}}_{S_k} = \frac{1}{w_k} \sum_{i=1}^{n} \tau_{ik} \boldsymbol{\mu}_{x_i};$$

$$\hat{\boldsymbol{\Sigma}}_{S_k} = \frac{1}{w_k} \sum_{i=1}^{n} \tau_{ik} \left(\mathbf{R}_{x_i} + (\boldsymbol{\mu}_{x_i} - \hat{\boldsymbol{\mu}}_{S_k})(\boldsymbol{\mu}_{x_i} - \hat{\boldsymbol{\mu}}_{S_k})^T \right);$$

where $w_k = \sum_{i=1}^{n} \tau_{ik}$ and update the flow models as

$$\begin{aligned} \hat{\boldsymbol{\alpha}} = &\left(\boldsymbol{\Sigma}_\alpha + \sum_{i=1}^{N} \tau_{ik} (\mathbf{G}_i^T \boldsymbol{\Sigma}^{-1}_F \mathbf{G}_i + \mathbf{H}_i) + \right)^{-1} \\ &\left(\boldsymbol{\Sigma}^{-1}_\alpha \boldsymbol{\alpha}_0 + \sum_{i=1}^{N} \tau_{ik} (\mathbf{G}_i^T \boldsymbol{\Sigma}^{-1}_F \boldsymbol{\mu}_{v_i}) \right), \end{aligned}$$

where $\mathbf{G}_i = [\mathbf{g}^{(1)}_i, \ldots, \mathbf{g}^{(L)}_i]$ with $\mathbf{g}^{(l)}_i = \mathbf{Y}^{(l)} \mathbf{x}_i + \mathbf{u}^{(l)}$, $\mathbf{H}_i$ is a $L \times L$ matrix with $\mathbf{H}_i(l, k) = \text{tr}(\boldsymbol{\Sigma}^{-1}_F \mathbf{Y}^{(l)} \mathbf{R}_{x_i} \mathbf{Y}^{(l)})$.

4. Experiments

We present two applications of our method for statistical modeling of transformations. The first analyzes the aggregate motion fields of people moving in a busy train station, while the second analyzes dense flow from weather data.

4.1. Analyzing People's Motion Patterns

Figure 1(a) shows a single frame from an video sequence captured in New York's Grand Central station. The video sequence is 15 minutes duration captured at 24 fps at an image resolution of 1440×1080. The first 1000 frames are used to initialize the model. The ensuing 18000 frames are processed using the tracking algorithm of [16]. In such scenes, one expects a degree of regularity of motion due to a variety of factors including the movement of large crowds of people negotiated in a confined space or common entrances and exits. Our aim is to capture the aggregate motion patterns solely from local motion observations.

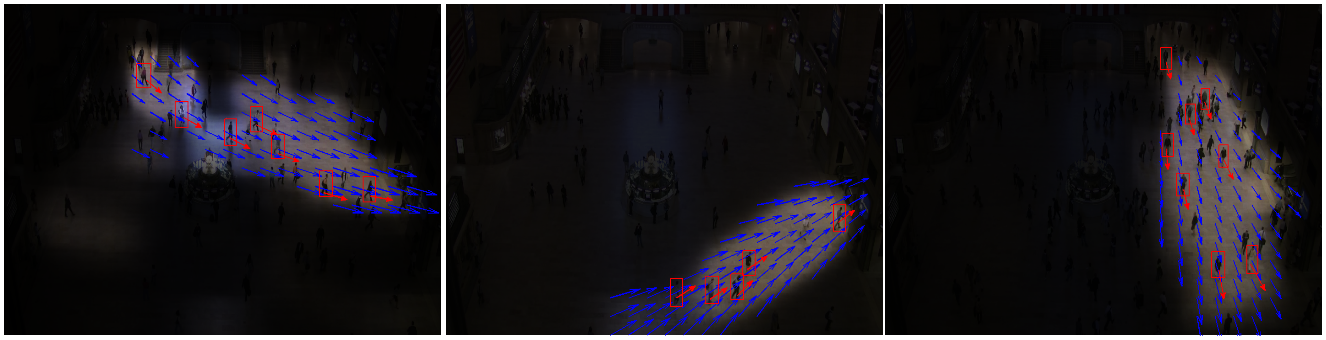


Figure 6. Three are three representative flows discovered by the Lie algebra based flow model. The region that is not covered by the flow is masked. The blue arrows indicate the flow field, and a subset of persons governed by the flow is highlighted with red boxes.

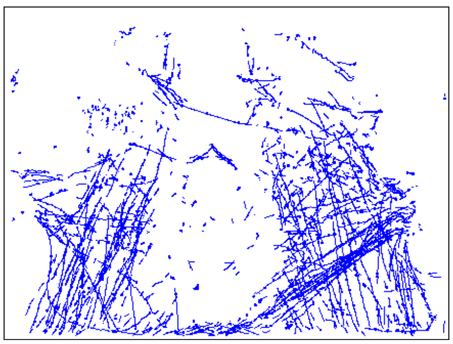


Figure 5. The plot of all extracted local motions.

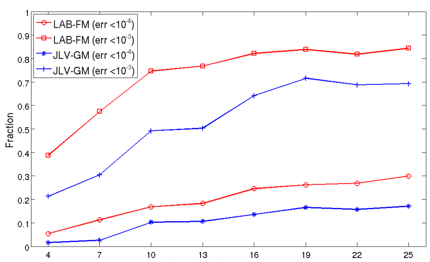


Figure 7. The performance comparison on people motion analysis.

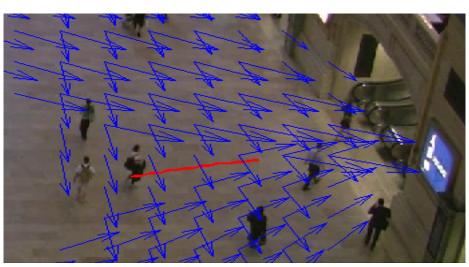


Figure 8. An example of outlier detected by the flow model.

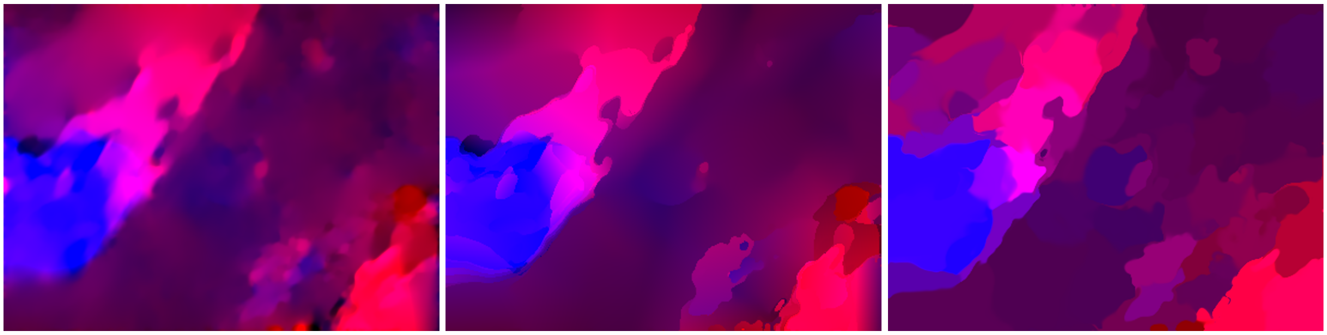
In such scenes, tracking errors commonly occur due to a variety of issues such as occlusions and individuals crossing paths. In Figure 5 we show roughly 10% of the extracted local motions. It can be observed that, even in the presence of errors, there is observable structure in the motions. Additionally, one can observe that structured motion fields overlap with each other.

We apply our flow model in order to recover these flows, setting the number of flows to $M = 16$. The noise covariance matrices are all set to $3^2\mathbf{I}$ (a rough estimation, final performance is relatively insensitive to these settings). Figure 6 illustrates several of the learned flows, where affine deformations are represented as flow fields (blue). Individual motions associated with this flow (red) demonstrate that the affine model is able to capture aggregate motion over a large region, despite the fact that individuals following these patterns appear distinct locations and times and walk along different paths to different locations.

We compare our results to the modeling strategy in other work [9, 20, 21], which groups the locations and local motions based on their proximity and models each groups with a prototypical motion. For conciseness, we refer to our method as "LAB-FM" (Lie Algebra Based Flow Model) and the comparison model as "JLV-GM"(Joint Location-Velocity Group Model). In order to have a fair comparison, JLV-GM is formulated similarly, so as to cope with noise and outliers. The consequence being that the essential difference arises from exploiting the group structure in the Lie Algebra space.

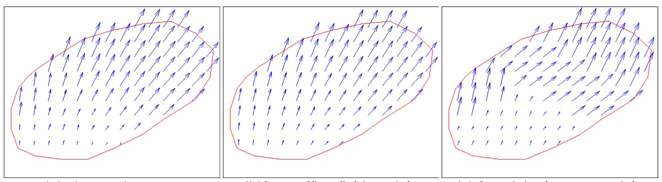
In order to mitigate the influence of outliers, we compare their performance in terms of the fraction of samples whose squared errors are below some thresholds. This measures the ability of the model in describing the observed motion in the scene. Setting the thresholds to 10^{-4} and 10^{-3}, and the number of mixing models M to different values, we obtain the performance curve shown in figure 7. The results clearly show that the performance increases as we add more components, and our LAB-FM consistently outperforms JLV-GM, that is at a given threshold, the fraction of motions which is below this error threshold is higher for the LAB-FM than for JLV-GM.

While outliers may be indicative of many things, they primarily correspond to motions which differ from the typical behavior of individuals in the scene. Figure 8 shows one



(a) The observed optical flow (b) The field yielded by LAB-FM (c) The field yielded by JLV-GM

Figure 9. The colorize visualization of velocity fields of the weather image. It uses red to to represent horizontal velocity and blue to represent vertical velocity. The color of an arbitrary velocity is derived by combining red and blue with the their projection on x and y axes.



(a) observations (b)from affine field model (c) from joint loc-vec model

Figure 10. Comparison of the velocity fields at a local region.

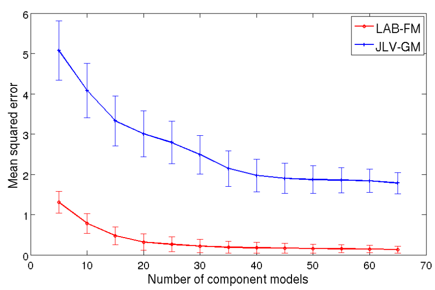


Figure 11. The performance comparison on weather data.

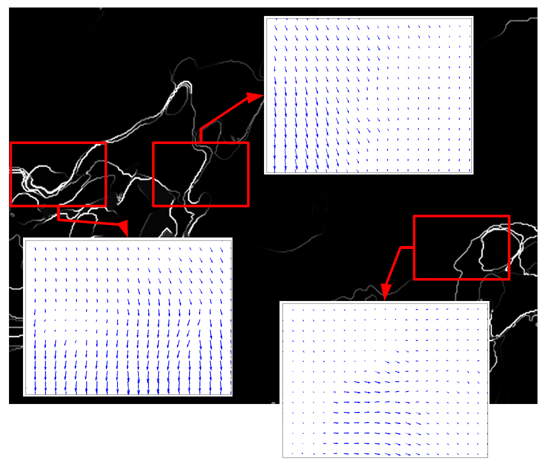


Figure 12. The background image is the map of gradients of the flow coefficients, upon which there are three graphs showing the the optical flows at the high-contrast local regions. There may be interesting weather phenomena in these regions.

of the outliers. There are three dominant flow fields in the scene. The "outlier", however, walks towards the escalator (a converging destination for two of the flows from either top or bottom of the scene) from a horizontal direction. The implication is that during the observation period, the majority of individuals in this region either enter the escalator from one of two directions or pass it by.

4.2. Modeling the Dynamics of Weather Data

Our second application considers analysis of dense optical flow derived from a sequence of satellite images available at NASA's GOES website. The acquisition times are separated by 30 minutes. We apply the optical flow algorithms of [5] to the images which is shown in figure 1.

Here we have set the number of mixture models to be $M = 40$ for both LAB-FM and JLV-GM, and all noise variances to 0.1. Whereas the dense flow algorithm computes arbitrary flow fields, LAB-FM approximates the dense flow as a collection of spatially varying fields. JLV-GM, on other hand, groups flows using a homogeneous representation. In both cases, the dense flow is described in a more parsimonious manner (model order 40). For comparison, the results are visualized in figure 9 using colorization. We can see that for a large portion of the observed flow, the "color" varies smoothly. LAB-FM captures smoothly varying fields while JLV-GM approximates the flow with constant-value blocks leading to noticeable artifacts. This can be seen more clearly in figure 10, which compares the reconstruction results of a local region in detail.

As the dense optical flow are assumed to be free of outliers, we can quantitatively compare the methods using mean-square error of the approximation. This is shown in figure 11 for 50 Monte-Carlo runs using a range of model orders. LAB-FM clearly outperforms JLV-GM by a significant factor (about 10).

Furthermore, the geometric information captured by the flow models provides insight at a high level. Here we are interested in using the flow analysis to detect regions in which the flow is changing dramatically. This analysis is performed simply by examining the gradient of the α-vector parameterizing the affine flow in the Lie Algebra space. Large gradients indicate that the flow is changing abruptly at a location. Figure 12 shows a map of the gradients where high contrast regions are indicative of interfaces between dynamic weather phenomenon.

5. Conclusion

We presented a new approach to modeling dynamic visual scenes comprised of two key aspects: the concept of flow to unify the modeling of moving elements at different locations with geometric consistency, and a Lie algebraic representation that enables application of statistical inference techniques to geometric transforms residing in a group. Furthermore, we developed a variational inference method exploiting the Lie algebraic representation.

These models are evaluated on two applications and compared with the joint location-velocity model widely used in motion analysis. In both applications, the proposed method exhibits superior performance. The experimental results demonstrate that our approach is suitable for modeling real phenomena typically involving spatially varying motion fields, and that it is capable of discovering underlying regularities governing the dynamics.

It should be emphasized that the notion of flow and the Lie algebraic representation together constitute a general methodology that can be incorporated into any algebraic and statistical framework for dynamic modeling. In addition to affine transforms, it can also be extended to work with other transforms including nonlinear families. We believe that with proper generalization, a broad spectrum of applications can benefit from it.

Acknowledgement

This research was partially supported by HSN (Heterogeneous Sensor Networks), which receives support from Army Research Office (ARO) Multidisciplinary Research Initiative (MURI) program (Award number W911NF-06-1-0076).

References

[1] V. Ablavsky, A. Thangali, and S. Sclaroff. Layered graphical models for tracking partially-occluded objects. In *Proc. of CVPR'08*, 2008.

[2] A. A. Agrachev and Y. L. Sachkov. *Control Theory from the Geometric Viewpoint*. Springer, 2004.

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

[4] E. Bayro-Corrochano and J. Ortegon-Aguilar. Lie algebra approach for tracking and 3d motion estimation. *Image and Vision Computing*, 25:907–921, 2007.

[5] 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.

[6] 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.

[7] Y. Guo, S. Hsu, H. S. Sawhney, R. Kumar, and Y. Shan. Robust object matching for persistent tracking with heterogeneous features. *IEEE Trans. on PAMI*, 29(5):824–839, May 2007.

[8] B. C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer, 2003.

[9] I. N. Junejo and H. Foroosh. Trajectory rectification and path modeling for video surveillance. In *Proc. of ICCV'07*, 2007.

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

[11] X. Miao and R. P. N. Rao. Learning the lie groups of visual invariance. *Neural Computation*, 19:2665–2693, 2007.

[12] 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.

[13] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. Segmenting, modeling, and matching video clips containing multiple moving objects. *IEEE Trans. on PAMI*, 29(3):477–491, Mar. 2007.

[14] V. Sharma and J. W. Davis. Integrating appearance and motion cues for simultaneous detection and segmentation of pedestrians. In *Proc. of ICCV'07*, 2007.

[15] E. Shechtman and M. Irani. Space-time behavior-based correlation or how to tell if two underlying motion fields are similar without computing them? *IEEE Trans. on PAMI*, 29(11):2045–2056, Nov. 2007.

[16] C. Stauffer. Adaptive background mixture models for real-time tracking. In *Proc. of CVPR'99*, 1999.

[17] A. Thayananthan, M. Iwasaki, and R. Cipolla. Principled fusion of high-level model and low-level cues for motion segmentation. In *Proc. of CVPR'08*, 2008.

[18] O. Tuzel, F. Porikli, and P. Meer. Learning on lie groups for invariant detection and tracking. In *Proc. of CVPR'08*, 2008.

[19] J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models for human motion. *IEEE Trans. on PAMI*, 30(2):283–298, Feb. 2008.

[20] 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.

[21] X. Wang, X. Ma, and E. Grimson. Unsupervised activity perception by hierarchical bayesian models. In *Proc. of CVPR'07*, 2007.

[22] T. Xiang and S. Gong. Beyond tracking: Modeling activity and understanding behavior. *Int'l. J. Comp. Vision*, 67(1):21–51, 2006.

[23] 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.