SANDIA REPORT

SAND2018-12155
For Public Release
Printed

Spectral Methods in Time-dependent Data Analysis

Akshat Kumar

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

For public release

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online ordering:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2018-12155
For Public Release
Printed

Spectral Methods in Time-dependent Data Analysis

Akshat Kumar

Abstract

We aim to create a new model for time-dependent data analysis, named *dynamical learning*, that integrates data-driven manifold learning techniques with operator-theoretic methods from dynamical systems theory. This approach has the potential to deliver more efficient methods for analyzing time-dependent data, such as video streams, by naturally separating out the temporal and spatial features of the data. We aim to apply the newly developed methods to video surveillance data related to Sandia mission applications, and particularly focus on the problems of image segmentation and object tracking.

This project ended early due to the departure of the PI from Sandia about 18 months into the project. Therefore, this document reports on partial progress towards the initial goals of the project. In addition, this document reports on part of the work conducted during the project; see the Appendix for a summary of all the work conducted during the 18 months.

Acknowledgments

The work of Sections 2.1 and 3 below was performed in collaboration with Igor Mezić, Ryan Mohr, and Cory Brown from the Department of Mechanical Engineering at University of California, Santa Barbara. Additionally, some computational tasks regarding video processing were performed by Salvador Sanchez (Dept. 05346) and some numerical tasks were performed by Michael Vaiana (PhD Intern, Dept. 08759) under the mentorship of the PI.

This page intentionally left blank

1 Introduction

Modern Big Data analysis must deal with unbounded amounts of samples, each with many degrees of freedom (e.g., parameters/features). Processing these datasets is chiefly achieved through pattern recognition (PR) and prediction/inference (PI). Existing methods implement either PR or PI, but seldom unify the two; the incompatibility forces ad-hoc decisions to be made during data analysis and techniques to be patched together with no formal justification.

We aim to develop techniques that unify the PR and PI goals with focus on efficiency, automatability and robustness to incomplete data. This will be based on a combination of spectral techniques arising from geometric and dynamic data analysis: namely, we will use the method of Diffusion Maps (DM) [CLL$^+$05, CL06], which was designed for clustering and dimensionality reduction of large datasets, along with operator-theoretic methods from dynamical systems theory based on Koopman spectral techniques [Koo31, KN32]. DM has achieved tremendous success in reasoning about static (time-independent) features in several application domains [BM12, Isa09, LS08, MC13, LTZ$^+$15]. In parallel development, the rise in numerical methods for an approximate realization of Koopman spectral techniques has led to many applications in the analysis of high-dimensional dynamical systems in the model-free setting [AM17, KKS16, LDBK17]. In effect, Koopman techniques separate the spatial from temporal properties of time-dependent datasets, allow for an efficient local reconstruction and prediction, and the corresponding objects give an efficient static perspective to dynamical data. The developments in this project are based on the insight that the appropriately derived features from time-dependent datasets using Koopman methods can be combined with the geometrizing power of DM to enable efficient low-dimensional representations of large time-dependent datasets useful for attacking deeper analysis problems. From a Sandia mission-application perspective, our project is motivated by problems in video analysis, whereby these techniques will have impact on the critical needs in surveillance and intelligence, such as tracking moving objects, change detection, and the prediction of trajectories.

The following chapters mainly report on progress made during the project on applying Koopman techniques to video data for motion segmentation. Due to the early termination of the project, the development of an approach combining Koopman techniques with DM for generating low-dimensional representations and clustering was incomplete, and hence not reported on here.

2 Background

In this chapter we present some background on Koopman spectral techniques, focusing on the class of numerical algorithms known as Dynamic Mode Decomposition (DMD).

DMD is an algorithm that was first introduced to decompose numerical or experimental data coming from fluid flows. It was put on theoretically sound footing by connecting it to the Koopman operator, which is often used in the analysis of nonlinear dynamical systems. The algorithm has since been applied to problems outside of the fluids community. Given a sequence of snapshots of data (e.g. a sequence of video frames), DMD postulates the existence of a fixed linear map that generates the data. Using the data sequence an approximation of this matrix is computed and decomposed into the so-called DMD modes. Such a model represents the best linear approximation of the evolution of the dataset in a least-squares sense. These DMD modes represent (possibly nonlinear) spatial structures in the data having well-defined temporal behaviors; each DMD mode has a fixed growth rate and frequency. This nice temporal behavior lends itself well to the problem of prediction. This contrasts with other popular decompositions, such as the proper orthogonal decomposition (POD), in which the mode coefficients do not have such nice behavior, necessitating some type of statistical analysis or curve-fitting of the coefficients to time-series in order to make predictions.

In the next section, we will elaborate more on this method.

2.1 Koopman spectral methods and DMD

In the early 1930s, Bernard Koopman and John von Neumann introduced a new operator theortic view of dynamical systems [Koo31, KN32]. In the 2000s, it was discovered to be a useful framework for data-driven analysis of high-dimensional nonlinear systems [MB04, Mez05]. Since then, numerical methods used to approximate Koopman spectral objects have been used as data analysis tools in the contexts of fluid dynamics [AKM18], neurodynamics [BJOK16], energy efficiency [GM15], and molecular dynamics [WNP$^+$17]. A particular class of algorithms for approximating Koopman spectral objects, known as dynamic mode decomposition [Sch10, RMB$^+$09], has even been used in background/foreground separation in videos taken from a stationary camera [KFBE15].

In the context of deterministic dynamical systems, state-variables are the variables whose knowledge at a particular time determines their values for all of time (e.g. angle and angular velocity for the mathematical pendulum). A standard approach in dynamical systems theory is to work directly with a model governing the dynamics (*e.g.*, system of PDEs in the spatial and temporal domains, along with boundary conditions) and centered around the state-space viewpoint. When recording dynamic data from complicated systems such as video sequences, the state variables which fully describe the observed system are not known and there is no model to describe their evolution. In these circumstances, a natural view is to consider scalar-valued functions of the states (called *observables*) and their evolution by the dynamics. For the case of the pendulum, observables include kinetic energy, potential energy, or more general functions of angle and angular momentum. In this view, we can think of any measurements we take of a system of interest as a function of some unknown state. Further, measured observables can be used to generate new ones: *e.g.*, by composing on the left by scalar functions of scalar input or on the right by maps of the state space into itself. In the context of video sequences, observables then include pixel data, optical flow field components at each location, and feature point locations. In any case, it is this *model-free* perspective of dynamics with observables as the central object that is the *Koopman view*.

The simplest context in which to introduce the Koopman view is for discrete-time dynamical systems, which can be described by repeated application of a single map $T : M \to M$ on the state-space M, so that

$$x_1 = T(x) \tag{2.1}$$

is the updated state corresponding to x after a unit time-step. The more general cases of continuous-time systems and time-varying (*viz.*, non-autonomous) dynamical systems have been described in, respectively, the original introduction of the Koopman framework [Koo31] and the recent extensions [MČŽM17]; the extension to the cases of random and stochastic systems has seen recent treatment in [ČŽMM17]. Given the assumption of the dynamical system as in Eq. (2.1), the move to observables takes place through the *Koopman operator*

$$U : f \mapsto f \circ T, \tag{2.2}$$

which is a linear operator due to the bi-linearity of the composition $\circ$ and thus we can consider its eigenvalues λ and eigenfunctions ϕ following the defining equation,

$$U(\phi) = \lambda\phi.$$

Furthermore, U induces a linear dynamical system on the space of observables due to the identity $U^{k+k'} = U^k \circ U^{k'}$ and initial condition $U^0 = I$.

The cue for spectral analysis of U comes when it is *diagonalizable* and in that case, even though the *exact* evolution of observables generally entails an infinite expansion, the spectral decomposition of the projection onto an invariant subspace enables a fast evolution of observables lying there in finitely many terms[1]. Namely, let $(f_1, \dots, f_n)$ denote a finite collection of observables and assume that they lie in an invariant subspace E of U with $\dim(E) =: m < \infty$. Let $(\phi_1, \dots, \phi_m)$ be the eigenfunctions of the projection[2] of U onto E with corresponding eigenvalues $(\lambda_1, \dots, \lambda_m)$. Thus, there exist linear functionals $\psi_1, \dots, \psi_m \in E^*$ such that we have the eigendecomposition

$$U^q(f_r) = \sum_{k=1}^{m} \lambda_k^q \psi_k(f_r)\phi_k.$$

Vectorizing this equation for each observable under consideration, we arrive at a special case of

[1]Issues of convergence are explored in [KKS16, KM18b].

[2]We abuse notation a bit and use U to denote its finite-dimensional projection onto $\text{span}\{\phi_1, \dots, \phi_m\}$.

Koopman Mode Decomposition (KMD) for U:

$$\begin{bmatrix} U^q(f_1) \\ \vdots \\ U^q(f_n) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{m} \lambda_k^q \phi_k \psi_k(f_1) \\ \vdots \\ \sum_{k=1}^{m} \lambda_k^q \phi_k \psi_k(f_n) \end{bmatrix} \tag{2.3}$$

$$= \sum_{k=1}^{m} \lambda_k^q \phi_k \begin{bmatrix} \psi_k(f_1) \\ \vdots \\ \psi_k(f_n) \end{bmatrix}. \tag{2.4}$$

This component-wise eigendecomposition allows us to evolve the observables $\{f_1, \dots, f_n\}$ by any $q \in \mathbb{N}_0$ time-steps by simply multiplying each term in the sum above by the qth power of its corresponding eigenvalue λ_k. This gives us the intuitve interpretations of the magnitude and complex phase of λ_k as corresponding to rate of growth and rate of oscillation, respectivley. If we are only interested in the evolution of the observables along a single trajectory of (2.1) with initial state x, then we can absorb each $\phi_k(x)$ into the coordinate functional and write $\widetilde{\psi}_k := \phi_k(x)\psi_k$, which gives the simpler form for KMD,

$$\begin{bmatrix} [U^q(f_1)](x) \\ \vdots \\ [U^q(f_n)](x) \end{bmatrix} = \sum_{k=1}^{m} \lambda_k^q \begin{bmatrix} \widetilde{\psi}_k(f_1) \\ \vdots \\ \widetilde{\psi}_k(f_n) \end{bmatrix}. \tag{2.5}$$

The last column vector appearing in (2.4) and (2.5) is known as the *Koopman mode* corresponding to the pair (λ_k, ϕ_k) and the vector of observables $[f_1, \dots, f_n]^T$ relative to the discrete-time system described by (2.1). From (2.5) we see that the magnitude of $\psi_k(f_j)$ tells us how much the growth and oscillations rates (as a pair) given by λ_k play a role in the evolution of f_j along the single trajectory starting at x and locally in time. Similarly, the complex phase of $\psi_k(f_j)$ gives a relative phase correspsonding to the oscillations given by λ_k.

Example 2.1. Consider the special case that $M = \mathbb{C}^n$ and the dynamics on M is given by a diagonalizable linear operator $T : M \to M$. In addition, let $\{f_1, \dots, f_n\}$ be the dual basis of the standard ordered basis for $\mathbb{C}^n$. Given any basis $\{v_1, \dots, v_n\}$ of eigenvectors for T with eigenvalues $\{\lambda_1, \dots, \lambda_n\}$, we have that its dual basis $\{\phi_1, \dots, \phi_n\}$ is a set of eigenfunctions for U with eigenvalues $\{\lambda_1, \dots, \lambda_n\}$. We see that for any $x \in M = \mathbb{C}^n$, we have

$$\begin{bmatrix} [U^q(f_1)](x) \\ \vdots \\ [U^q(f_n)](x) \end{bmatrix} = \begin{bmatrix} f_1(T^q(x)) \\ \vdots \\ f_n(T^q(x)) \end{bmatrix} \tag{2.6}$$

$$= T^q(x) \tag{2.7}$$

$$= \sum_{k=1}^{m} \lambda_k^q \phi_k(x) v_k. \tag{2.8}$$

By comparing with equation (2.4) we see that $\{v_1, \dots, v_n\}$ are the Koopman modes corresponding to the pairs $\{(\lambda_1, \phi_1), \dots, (\lambda_n, \phi_n)\}$ and the vector of observables $[f_1, \dots, f_n]^T$, relative to the discrete-time system described by (2.1). With this example we see that eigendecomposition can be viewed as a special case of KMD; the extension that KMD provides is flexibilty in both $\{f_1, \dots, f_n\}$ and T, in particular, these can be nonlinear.

The DMD algorithm

Consider the case where we have a discrete-time dynamical system as in (2.1) and we have evaluated a set of observables $\{f_1, \dots, f_n\}$ along the first $m+1$ time points of a trajectory starting at x. We can put this into a matrix D such that each row corresponds to a different observable and the columns are ordered by time:

$$D = \begin{bmatrix} f_1(x) & f_1(T(x)) & \dots & f_1(T^m(x)) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(x) & f_n(T(x)) & \dots & f_n(T^m(x)) \end{bmatrix} \tag{2.9}$$

$$= \begin{bmatrix} f_1(x) & [U(f_1)](x) & \dots & [U^m(f_1)](x) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(x) & [U(f_n)](x) & \dots & [U^m(f_n)](x) \end{bmatrix}. \tag{2.10}$$

We can split D into the matrix X of the first m columns and Y of the last m columns, i.e.

$$X = \begin{bmatrix} f_1(x) & f_1(T(x)) & \dots & f_1(T^{m-1}(x)) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(x) & f_n(T(x)) & \dots & f_n(T^{m-1}(x)) \end{bmatrix}, \tag{2.11}$$

$$Y = \begin{bmatrix} f_1(T(x)) & f_1(T^2(x)) & \dots & f_1(T^m(x)) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(T(x)) & f_n(T^2(x)) & \dots & f_n(T^m(x)) \end{bmatrix}. \tag{2.12}$$

The main idea in dynamic mode decomposition (DMD) is to find a matrix A such that AX is close to Y in some sense. In this way A would be mapping the vector of measurements taken at the point $T^q(x)$ close to those taken at the point $T^{q+1}(x)$, for all $q \in \{0,\dots,m-1\}$. With this, it may seem intuitive that as our number of time points goes to infinity or as we add more observables (which could just be functions of the original ones), we should better and better approximate the Koopman operator U (and its spectral objects) by the linear operator we are representing by the matrix A. In fact, in some cases this has been proven to be the case [WKR15, KKS16, KM18b, AM17, MA]. Typically, $A = YX^{\dagger}$, which solves the optimization problem (see [BV04])

$$\|AX - Y\|_F = \inf_{B \in \mathbb{C}^{n\times n}} \|BX - Y\|_F, \tag{2.13}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Now, $A : \mathbb{C}^n \to \mathbb{C}^n$ defines a linear dynamical system, so as in Example 2.1, it has corresponding Koopman modes and spectrum. However, A is given a special structure since it is a system that, itself, forms a numerical approximation to a Koopman operator. In particular, A satisfies for $y(x) = (f_1,\dots,f_n)(x) \in \mathbb{C}^n$, $Ay(x) = y(Tx)$. Hence, we call its modes and spectrum, the *DMD modes* and *DMD spectrum*, and A itself is called a *DMD operator*.

3 Motion Segmentation in Video Data

Autonomous cars, video surveillance, multirobot collaboration, and medical imaging are all applications which involve analysis and classification of large amounts of dynamic image sequences in real time. Better approaches to these problems have largely been obtained by addressing smaller ones of computer vision, e.g. object detection, identification, tracking, motion estimation in 3D, and forecasting of future dynamics, or integrating their advancements. The idea of *motion segmentation*, the partitioning of images, or trajectories of tracked feature points, in a video into regions of similar dynamics so that each element of the partition has a corresponding element in the partition of the previous or following image, plays a key role in this area of research.

Object tracking has mostly been approached through unsupervised methods [QH07, SK04, SS16]. One might think that the problem would be greatly simplified by using object detection. However, using purely bottom up approaches in combination with detection seems to be optimal [ARS08, DB16, KMM12]. Unsupervised approaches are surely justified in contexts where tracking undetectable objects is necessary. In addition, they can be used to make object detection more robust, computationally efficient and streamline the training of object detectors by finding undetectable moving objects in large amounts of video sequences.

In the following, we approach these problems using numerical counterparts of tools from the operator-theoretic approach to dynamical systems; namely, we will rely on the variants of Dynamic Mode Decomposition (DMD) as described in [Sch10, RMB$^+$09]. The prior use of these techniques in this context has been focused mainly on background/foreground separation in videos taken from a stationary camera [KFBE15]. Herein, we consider videos taken from a moving camera and perform DMD on a different set of *observables* - measurements we take of a dynamical system (defined more precisely in the next section). Standard observables used in the context of motion segmentation include feature point coordinates [RTVM10] and optical flow [HS81] components. In combintation with DMD, Kutz et al. [KFBE15] used grayscale pixel data. In contrast to the prior DMD approaches and more close to geometric methods, we will use *pairwise distances* of feature points, tracked by an optical-flow based algorithm such as the Kanade-Lucas-Tomasi (KLT feature tracker, for our observables.

In the following two sections we demonstrate the use of Koopman spectral analysis on pairwise distance observables of feature points for several movies through the use of DMD and give some outward-looking remarks.

Acknowledgements

This section is a reproduction of original work done in collaboration with Igor Mezić, Ryan Mohr, and Cory Brown at the Department of Mechanical Engineering, University of California at Santa Barbara under contract from Sandia National Laboratories.

3.1 Koopman Mode Decomposition of Pairwise Distance Observables

In this section we apply Koopman spectral analysis to three real-life videos (with a moving camera) via the use of DMD. The first image of each sequence can be found in the first row of figure 1. From left to right the image sequences increase in their number of motions from two to four where the background is treated as a motion since the camera is moving. Note that feature points belonging to the same motion have the same color and feature point trajectories for the length of the video are plotted in the second row. The first video was taken from the data set used in [SK03, SK04]. The second video was shared by the authors of [QH07]. Finally, the last video is from the data set used in [RTVM10]. With each of these movies we have a set of feature points $\{p_1, \dots, p_r\}$, where we make the identification

$$p_i = (x_{i1}, \dots, x_{im}, y_{i1}, \dots, y_{im}) \tag{3.1}$$

with x_{ij} and y_{ij} denoting the x and y coordinates of feature point i in frame j. Thus, we are letting r and m denote the number of feature points and the number of frames, respectively. In the four motion sequence, some of the feature points are outliers - added feature points with randomized trajectories [RTVM10].

The main assumption here is that there is some underlying dynamical system which describes the real-world motions which were captured by a video camera, with a constant frame rate, and that we can represent the restriction of this dynamical system to a discrete time set (with difference between time points corresponding to the frame rate) as in (2.1). We can then use DMD by choosing a set of observables. For a single movie, we view each color or grayscale component of each pixel as providing us with an observable along a single trajectory; even though we do not know the state-space of the underlying system we are observing,i.e. the domain of T, we assume that the processes of capturing an image is a sampling of a set of observables at a single point in state space and at a single point in time. The coordinates of a feature point can then be thought of as real-valued functions composed with a vector of our pixel level observables. In this way, the coordinates of each feature point are again observables. Finally, pairwise Euclidean distances of feature points are the observables which will be subject to our Koopman spectral analysis.

We denote the value of the Euclidean distance between feature point i and j in frame k by

$$d_{ijk} = ||(x_{ik}, y_{ik}) - (x_{jk}, y_{jk})||_2. \tag{3.2}$$

This gives the matrix d_k for each frame k and by vectorizing the lower-triangular piece, we have for each k the vector

$$D_k = [d_{12k} \cdots d_{1rk} \; d_{23k} \cdots d_{2rk} \; d_{(r-1)rk}]^T \tag{3.3}$$

of length $1 + \cdots + (n-1) = \frac{n(n-1)}{2} =: M$. Since $D_k, D_{k+1} \in \mathbb{R}^M$ are related by we can use the

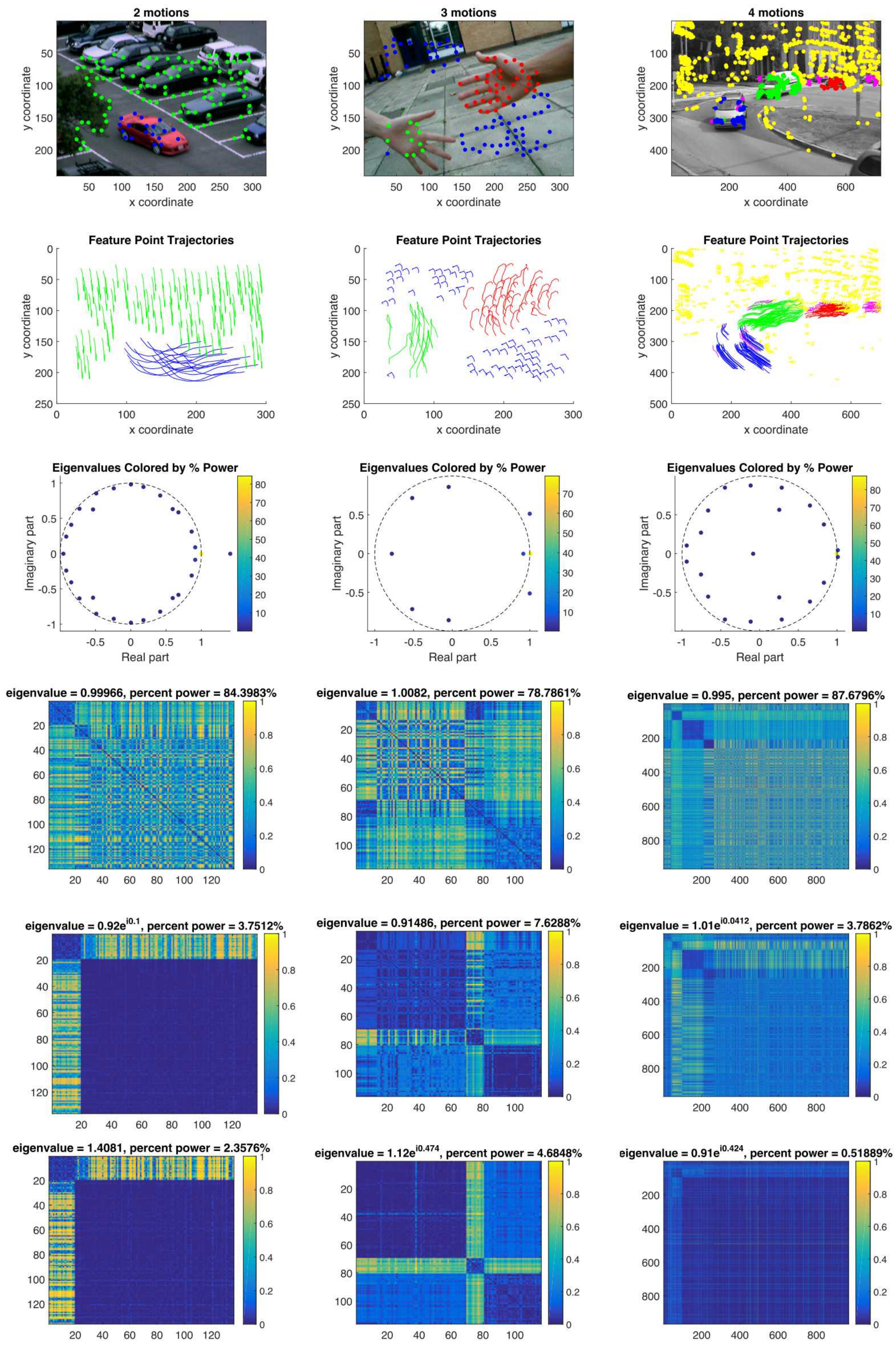

Figure 1: Caption on next page.

Figure 1: Koopman spectral quantities of the DMD matrix obtained from pairwise distance observables on three real-world video sequences. Each movie and its computed Koopman quantities correspond to a single column. The first row contains a picture from each video sequence with feature points labeled by different motions or outliers. In the second row, the paths the feature points trace are plotted, for the full video sequence, and colored to match the pictures above them. The third row contains eigenvalues of the DMD matrix colored by their percent power. The last three rows contain the three highest power eigenvectors of the DMD matrix represented as symmetric matrices; the i, j-th entry corresponds to the pairwise distance observable between the ith and jth feature points. The rows (and columns) of these matrices are ordered by motion with outliers first (if any). In terms of colors, the order of the motions is as follows: magenta, blue, green, red, and, finally yellow. Note that in the 4 motion movie, the last 15 frames were not used due to obvious mistracked feature points in those frames.

matrices

$$X = \begin{bmatrix} | & | & & | \\ D_1 & D_2 & \cdots & D_{(m-1)} \\ | & | & & | \end{bmatrix}, \text{ and} \tag{3.4}$$

$$Y = \begin{bmatrix} | & | & & | \\ D_2 & D_3 & \cdots & D_m \\ | & | & & | \end{bmatrix}. \tag{3.5}$$

in a DMD algorithm. In particular we used the DMD algorithm in [BJOK16], which includes a way to assign a ”power” that each DMD mode represents in the data matrix X. The resulting power is higher if the mode is more parallel to left singular vectors with large corresponding singular values and low otherwise. We can use this to order modes based on the level of importance they play in our data matrix as opposed to using the coefficients from the projection of a column of the data matrix onto the DMD modes. Finally, since the algorithm involves inverses of the diagonal matrix of singular values of X, we treat any singular value less than $\frac{\sigma}{10^6}$ as zero, where σ is the largest singular value.

Using the DMD algorithm mentioned, we obtain DMD eigenvalues and modes. The DMD eigenvalues are plotted in row three of figure 1 and are colored according to their associated percent total power. The DMD modes corresponding to the the eigenvalues with the top three highest powers (identifying complex conjugate pairs), are represented in the remaining three rows of figure 1. We represent the modes by symmetric matrices such that the $i, j-$th element corresponds to the pairwise distance between feature points i and j. Since these modes are in general complex, we only plot the magnitude of the components. Our feature points are already labeled according to different motions, thus we order the rows and columns of these matrices by color of the corresponding feature point: magenta (outiers), blue, green, red, and, finally, yellow. This leads to some block structure which is evident from the figure.

The first thing to notice is that the highest power DMD mode and eigenvalue pair has the eigenvalue closest to one in all three examples. We call this the stationary mode since it evolves in time by repeatedly multiplying it by one. For the two motion video (first column), this mode (fourth row) has a 19 by 19 principal submatrix in the upper left corner which is almost zero. We also see this in the lower power modes. This block appears in the stationary mode because the feature points are close together and in the remaining (dynamic) modes because these feature points do not move much relative to each other. On the other hand, the remaining motion of the two motion video is the background; the background points are fairly spreadout and so we do not see a similar block of zeros in the stationary mode as we did for the localized object (car). However, we do see such blocks in the dynamic modes. As one might expect, the dynamic modes have large components outside of the zero blocks on the diagonal; this is due to the relative motion of the car and the background. These symmetric matrix representations of the dynamic modes are static pictures which capture dynamic information.

Looking to the second column of figure 1, we notice similar structure in the DMD modes as mentioned for the two motion video. All three plotted modes have two fairly homogeneous blue blocks on the diagonal, in the bottom right corners, which correspond to the green hand (smaller block) and red hand (bottom right). A similar block for the backgound points only appears in the dynamic modes. As with the two motion case, the dynamic modes have larger components outside of these zero blocks on the diagonal. Similar structures are again seen in modes for the remaining four motion video (third column of figure 1).

3.2 Conclusion and Outlook

The main contribution of this work, distinguishing it from that in [KFBE15], is the consideration of Koopman spectral techniques on pairwise distance observables of tracked feature points as opposed to grayscale intensity observables. The coordinates of these feature points, derived from optical flow based video processing methods such as the Kanade-Lucas-Tomasi feature tracking algorithm, yield a sparser spatial dataset for processing than the full pixel-space. Moreover, using pairwise distance observables derived from these coordinates not only gives us many more meaningful observables, it also leads to the intuitive block structures seen in figure 1. In addition, the intensity-based algorithm of [KFBE15] relies on separation of slow versus fast time-scales for separating motions in different spatial regions with similar time-scales; in contrast, the pairwise distance observables separate local motions at varying speeds and directions in a single instance of the algorithm and are independent of the color/intensity space. Finally, while the results in [KFBE15] were impressive and multi-resolution DMD has many applications other than object tracking, the approach does not seem to easily extend to videos with a dynamic background. Figure 1 above shows that Koopman modes of pairwise distance observables can segment motion even with a moving background.

We have represented the absolute value of DMD modes obtained from pairwise distance observables as non-negative symmetric matrices which can be interpreted as a distance matrix for a graph whose nodes are the tracked feature points. An amazing feature of this approach is that object separation can be obtained from a single dynamic mode; this is very clearly seen, for example, in the bottom left plot of the figure. In [QH07], image segmentation is achieved by clustering data by eigenvectors of a certain distance (commute time) matrix and suggest this could also be done with feature point data from videos. Thus another possible route to automatically separating the motions based on DMD modes would be to cluster based on the eigenvectors of their non-negative symmetric matrix representations.

From the above discussion it is clear that choice of observables is key to the success of applying Koopman methods to data. We merely note that observables can be choosen using a neural network approach [YKH17, LDBK17].

In applications involving non-rigid motions, such as tracking people, there may be significant relative motion of feature points on a single object to be tracked. In the case of tracking a few people walking, this may result in several dynamic modes, represented as in figure 1, with a single nonzero block on the diagonal and zeros elsewhere; we would expect such modes to have corresponding eigenvalues with oscillation rates given by the different cadences of the walkers.

In applications such as autonomous cars, estimating and forecasting motions of tracked objects is also relevant. We note that Koopman techniques can be used for prediction of future dynamics [Gia17, KM18a] and coordinates of estimated motions of tracked objects would be the natural observables to use.

4 Appendix: Summary of Project

4.1 Project activities and highlights

FY17 activities

1. Developed general methods for identifying distinct motions in time-dependent datasets, motivated by motion segmentation problem in video analysis.
2. Compared various forms of DMD and DM on feature point tracks in video data.
3. Identified applications in quantum dynamics based on prior model reduction work and new ideas in thermalization.
4. Mentored PhD intern: develop fast-paced course to teach fundamentals of ergodic theory and functional analysis.
5. Summer intern project: used ergodic theory and ergodic quotients [BM12] work to develop DM kernels for directly embedding time-dependent datasets into geometrically meaningful low-dimensional representation, spatially organized by trajectories and attractors.
6. Developed motion segmentation technique and algorithm (PWD-DMD): constructs objects that enable clustering of dataset by distinct motions (DMD modes) and capable of classifying datasets by the types of motions present (DMD spectra).

FY18 activities

1. Applied motion segmentation algorithm to new datasets with increasing number of distinct motions and identify signatures in spectra for classifying datasets by motion properties.
2. This leads to the idea of "spectral characterization": namely, that multiple time-dependent datasets can be organized by their dynamical properties through manifold learning applied over Koopman spectra. An outline of the approach is:
   - Run DMD with fixed good observables on datasets within the same class (e.g., partition large dataset into smaller *local* pieces) and use spectra as feature vectors for DM.
   - Use diffusion distance in embedding to identify datasets with similar dynamical properties: each DMD mode captures spatial aspects of dynamics linking the full class of datasets and can be used for classification, prediction, and data reduction purposes.
   - In the context of a single large dataset, this corresponds to extracting dynamical features locally and stitching them together via manifold learning to organize the dataset based on spatial and dynamical features.
   - Much of the ongoing theory work is concerned with the analysis of these ideas in the dynamical systems framework.

3. Studied challenges and state-of-the-art in analysis of Wide Area Motion Imagery and specifically, the WPAFB 2009 dataset.
    - Optical flow is not directly available and feature trackers fail quickly on this dataset, so previous algorithm is not applicable as-is.
    - Direct application of DM or DMD is computationally intractable due to the various features of this dataset.
    - This leads to studying *local* methods in manifold learning: e.g., patch-based DM and local diffusion folders.
    - Another approach: partition the view space into patches, treated as separate time-dependent datasets and apply spectral characterization. Using just intensity initially, this reveals segmentation of the view space by similar dynamics and sparsifies the dataset to identify large-scale regions of motion, such as highways (compare to [Lea13]).
4. Connected with Dept. 6300 to identify prior projects attempting motion segmentation in datasets related to the WPAFB 2009.
    - Closest match is Kyle Merry's project on SDMS Minor Area Motion Imagery (MAMI) dataset that could achieve optical flow and feature tracking, but not motion segmentation.
    - Acquire simulations with feature tracks from this project and determine path to doing motion segmentation on their results: FY17 techniques can be directly applied with some coding modifications (restrict tracked features to those that remain on screen through all frames).

4.2 Programmatic summary

What worked well?

- Motion segmentation in the usual setting of large-scale structures with multiple types of motions with a priori feature-tracking.
- Generally, the methods are useful in applications with explicit time-dependence in data (meaning, trajectories can be directly sampled) where clustering based on distinct motions is necessary

What posed challenges?

- Applications with implicit time-dependence in data: the accessible observables have no information about trajectories, but rather a passive view of pieces of all trajectories that flow through a fixed spatial region; *e.g.,* operating on intensity data without pre-processing to produce feature tracks/optical flow.

- Massive scale data with large number of relatively small scale of distinct motions/targets; spatial under-sampling becomes challenging.
- "Background" contains multiple types of distinct motions, coupled with low/sparse temporal resolution, very small target size and mix of sparse and densely packed objects; *e.g.,* in video data: parallax with large displacement over small number of frames and wide-area motion imagery.
- These are all features of the WPAFB 2009 dataset that, in accumulation, make it especially challenging.

How have you deviated from your original plan?

- Original planning was to spend FY17 – 18 doing more theory work, specifically developing more fundamental ideas at the intersection of manifold learning and dynamical systems theory, both in the numerical/approximative and asymptotic/theoretical settings.
- Upon seeing the positive FY17 results in motion segmentation, committee requested further focus on video applications to mission-specific datasets such as WPAFB 2009 *etc.* and to identify "limitations" of DMD/DM based approaches.
- This resulted in less focus on theory work and more focus on understanding the challenging WPAFB 2009 dataset and developing methods to affect it using new techniques.
- In this direction, the first half of FY18 was focused on ideas around spectral characterization: local-to-global methods for motion classification in time-dependent datasets.

Summary of next steps

- Develop algorithm or heuristics for DMD mode selection for motion segmentation from feature tracks algorithm;
  1. How many DMD modes are necessary to identify all distinct motions?
  2. Identify "good" DMD modes for motion segmentation.
  3. Construct block-diagonalization of good DMD modes
- Spectral characterization of time-dependent datasets
  1. Automate search criteria for identifying relevant piece of spectrum (domain-specific problem).
  2. Explore coupling with Multi-Resolution DMD (mrDMD) [KFBE15].
- Tackle WPAFB 2009 dataset (assuming no a priori optical flow/feature track information)
  1. Pre-process by solving scene alignment problem (this enables dynamical methods to be localized and state-space to be sparsified).

- Most techniques in literature on WAMI datasets rely on training from ground-truths or pulling external data like mapping road networks onto the scene.
- For an approach agnostic to context-specific external information and ground-truth training, consider warping with respect to homography computed from a RANSAC-based algorithm through matching SIFT-like features across two frames. This dataset poses challenges in addition to strong parallax: the temporal resolution is very low and there is a large variation in geometry and intensity profiles [LZS17]. An approach might be to use this in conjunction with exact DMD on 'fat' data matrices over large number of observables and few time-steps, with frames chosen for best scene alignment results.

2. Spectral characterization approach with stronger observables: e.g., local binary patterns, or derivatives of intensity profile (*i.e.,* $(I,\, p_x(I),\, p_y(I))$, with $p_\cdot(I) = \partial_t I / \partial_\cdot(I)$ or just $p_\cdot(I) = \partial_\cdot(I)$ as suggested by [WSI07, GK13].
3. Patch-based DM amd DM for optical flow [WB10]: localized DM over patches, correlating diffusion distance across frames to give global optical flow. Image completion results as in [GK13] show DM eigenfunctions from texture descriptors can capture non-smooth structures/geometry profiles.
4. Possible work-flow: subsample frames, optimizing for scene alignment; do spectral characterization with iterative refinement such as mrDMD to identify regions of similar activity (highways vs parking lots vs background etc.); build optical flow over regions of interest and acquire feature tracks; apply PWD-DMD for object-based motion segmentation.

- Motion segmentation on Dept. 6300 (Kyle Merry) results on SDMS MAMI dataset.

  1. This dataset along with results of prior LDRD's efforts has several features making the PWD-DMD algorithm a good match. Background motion is rotational; most objects stay within view space for long time-spans; high temporal resolution; strong separation of dynamical characteristics between background and foreground. Sparsified optical flow and feature tracks exist; prior efforts have resolved some scene alignment issues.
  2. Apply PWD-DMD on temporal windows where feature points are contiguously present in the view space. DMD modes may be linked across temporal windows using their structure and observable properties. Along with modes, the spectra can also be tracked,

Application partners

1. Dept. 6300 (Kyle Merry)
   - `kmerry@sandia.gov`
   - Use case: Motion segmentation based on feature tracks acquired from results in previous project to push what is possible for dataset impacting surveillance mission area needs.

- Provided initial exemplary dataset produced through simulations for motion segmentation.
- Techniques and algorithms from project can be applied with some programming modifications (not needing any theory advances).

2. Timothy Draelos / Sapan Agarwal / Deep Neural Net team

   - tjdrael@sandia.gov, sagarwa@sandia.gov
   - Use case: Need to do analysis on time-varying datasets.
   - Deep Learning has many issues when applied directly to changing datasets, but can provide good observables from training that can be used with the techniques in this project to enable time-dependent analysis (*c.f.* [YKH17]).
   - Based on the setup of techniques in this project, specifically the separation of application-specific observables and the processing step done on them, there is a clear point of interaction with DNN-based methods that can enable a wider mission application space.

3. Dept. 5346 (SAR Group)

   - Salvador Sanchez worked on the project to produce video data and gained some understanding of the types of capabilities we can produce.
   - SAR team has analysis and data reduction needs on satellite imagery: close match for applications of this project.
   - Salvador can help to identify more specific applications in the SAR domain.

4. Dept. 1463 (Kiran Lakkaraju)

   - klakkar@sandia.gov
   - Brief discussions with Kiran indicated strong interest in doing manifold learning for clustering and data reduction on data produced from Cognitive Systems Models.

Team/Collaborators/Consultant

1. Mohan Sarovar

   - mnsarov@sandia.gov
   - Role and future collaboration: physics applications, especially intersection with quantum information.

2. Salvador Sanchez

   - sgsanch@sandia.gov
   - Role: video processing, acquiring test data and preprocessing target data.
   - Future collaboration: SAR applications, coding for application to MAMI dataset.

3. Tian Ma
    - `tma@sandia.gov`
    - Role: consultant on tracking applications; helped to identify certain directions for future applications and connections to acquire potentially relevant video data.
    - Several discussions on the state of video analysis and tracking capabilities at Sandia.
    - Discussions indicated the cutting-edge nature of this LDRD, the lack of similar capabilities, further confirmed the difficulty of the WPAFB 2009 dataset, and identified several issues with the current/existing approaches outside of this project that the techniques in this LDRD can overcome (especially regarding ineffectiveness of deep learning methods, PCA and other SVD-based techniques).
4. Fredrick Rothganger
    - `frothga@sandia.gov`
    - Role: consultant/advice on tracking and motion segmentation methods in video.
5. Richard Lehoucq
    - `rblehou@sandia.gov`
    - Role: consultant/advice on optical flow methods.
6. Dept. Mechanical Eng., UCSB (Igor Mezic)
    - `mezic@ucsb.edu`
    - Role: collaborator on DMD applications.
    - Future collaboration: video analysis and DMD applications

References

[AKM18] Hassan Arbabi, Milan Korda, and Igor Mezic. A data-driven koopman model predictive control framework for nonlinear flows. *arXiv preprint arXiv:1804.05291*, 2018.

[AM17] Hassan Arbabi and Igor Mezic. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, 2017.

[ARS08] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. People-tracking-by-detection and people-detection-by-tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[BJOK16] Bingni W Brunton, Lise A Johnson, Jeffrey G Ojemann, and J Nathan Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of neuroscience methods*, 258:1–15, 2016.

[BM12] Marko Budišić and Igor Mezić. Geometry of the ergodic quotient reveals coherent structures in flows. *Physica D: Nonlinear Phenomena*, 241(15):1255–1269, 2012.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CL06] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[CLL+05] Ronald R Coifman, Stephane Lafon, Ann B Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7426–7431, 2005.

[ČŽMM17] Nelida Črnjarić-Žic, Senka Maćešić, and Igor Mezić. Koopman operator spectrum for random dynamical system. *arXiv preprint arXiv:1711.03146*, 2017.

[DB16] Benjamin Drayer and Thomas Brox. Object detection, tracking, and motion segmentation for object-level video segmentation. *arXiv preprint arXiv:1608.03066*, 2016.

[Gia17] Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 2017.

[GK13] Sergei Gepshtein and Yosi Keller. Image completion by diffusion maps and spectral relaxation. *Image Processing, IEEE Transactions on*, 22(8):2983–2994, 2013.

[GM15] Michael Georgescu and Igor Mezić. Building energy modeling: A systematic approach to zoning and model reduction using koopman mode analysis. *Energy and buildings*, 86:794–802, 2015.

[HS81] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.

[Isa09] Jason C Isaacs. *Diffusion map kernel analysis for target classification*. IEEE, 2009.

[KFBE15] J Nathan Kutz, Xing Fu, Steve L Brunton, and N Benjamin Erichson. Multi-resolution dynamic mode decomposition for foreground/background separation and object tracking. In *Computer Vision Workshop (ICCVW), 2015 IEEE International Conference on*, pages 921–929. IEEE, 2015.

[KKS16] Stefan Klus, Péter Koltai, and Christof Schütte. On the numerical approximation of the perron-frobenius and koopman operator. *Journal of Computational Dynamics*, 3(1):51–79, 2016.

[KM18a] Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.

[KM18b] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, 2018.

[KMM12] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE transactions on pattern analysis and machine intelligence*, 34(7):1409–1422, 2012.

[KN32] BO Koopman and J v Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, 18(3):255–263, 1932.

[Koo31] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[LDBK17] Qianxiao Li, Felix Dietrich, Erik M Bollt, and Ioannis G Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.

[Lea13] Pengpeng Liang and et al. Vehicle detection in wide area aerial surveillance using temporal context. In *16th International Conference on Information Fusion (FUSION)*. IEEE, 2013.

[LS08] Linh Lieu and Naoki Saito. High dimensional pattern recognition using diffusion maps and earth movers distance. *submitted to Signal Processing, preprint available at http://www. math. ucdavis. edu/llieu/pub. html*, 2008.

[LTZ+15] Wenzhao Lian, Ronen Talmon, Hitten Zaveri, Lawrence Carin, and Ronald Coifman. Multivariate time-series analysis and diffusion maps. *Signal Processing*, 116:13–28, 2015.

[LZS17] Rodney LaLonde, Dong Zhang, and Mubarak Shah. Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information. *arXiv preprint arXiv:1704.02694*, 2017.

[MA] Igor Mezic and Hassan Arbabi. On the computation of isostables, isochrons and other spectral objects of the koopman operator using the dynamic mode decomposition.

[MB04] Igor Mezić and Andrzej Banaszuk. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1-2):101–133, 2004.

[MC13] Gal Mishne and Israel Cohen. Multiscale anomaly detection using diffusion maps. *Selected Topics in Signal Processing, IEEE Journal of*, 7(1):111–123, 2013.

[MČŽM17] Senka Maćešić, Nelida Črnjarić-Žic, and Igor Mezić. Koopman operator family spectrum for nonautonomous systems-part 1. *arXiv preprint arXiv:1703.07324*, 2017.

[Mez05] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[QH07] Huaijun Qiu and Edwin R Hancock. Clustering and embedding using commute times. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11), 2007.

[RMB+09] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127, 2009.

[RTVM10] Shankar Rao, Roberto Tron, Rene Vidal, and Yi Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1832–1845, 2010.

[Sch10] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[SK03] Yasuyuki Sugaya and Kenichi Kanatani. Outlier removal for motion tracking by subspace separation. *IEICE TRANSACTIONS on Information and Systems*, 86(6):1095–1102, 2003.

[SK04] Yasuyuki Sugaya and Kenichi Kanatani. Multi-stage unsupervised learning for multi-body motion segmentation. *IEICE Transactions on Information and Systems*, 87(7):1935–1942, 2004.

[SS16] Reza Sabzevari and Davide Scaramuzza. Multi-body motion estimation from monocular vehicle-mounted cameras. *IEEE Transactions on Robotics*, 32(3):638–651, 2016.

[WB10] Szymon Wartak and Adrian G. Bors. Optical flow estimation using diffusion distances. In *20th International Conference on Pattern Recognition (ICPR)*. IEEE, 2010.

[WKR15] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[WNP+17] Hao Wu, Feliks Nüske, Fabian Paul, Stefan Klus, Péter Koltai, and Frank Noé. Variational koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations. *The Journal of Chemical Physics*, 146(15):154104, 2017.

[WSI07] Yonatan Wexler, Eli Shechtman, and Michal Irani. Space-time completion of video. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 3:463 – 476, 2007.

[YKH17] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for koopman operators of nonlinear dynamical systems. *arXiv preprint arXiv:1708.06850*, 2017.

DISTRIBUTION:

2	MS 9018	Central Technical Files, 8944
1	MS 0899	Technical Library, 9536

Sandia National Laboratories