Dynamics-based Extraction of Information Sparsely Encoded In High Dimensional Data Streams

Mario Sznaier   Octavia Camps

Electrical and Comp. Engineering Department,
Northeastern University,
Boston, MA 02115.

Abstract— **A major roadblock in taking full advantage of the recent exponential growth in data collection and actuation capabilities stems from the *curse of dimensionality*. Simply put, existing techniques are ill-equipped to deal with the resulting volume of data. The goal of this paper is to show how the use of simple dynamical systems concepts can lead to tractable, computationally efficient algorithms for extracting information sparsely encoded in extremely large data sets. In addition, as shown here, this approach leads to non–entropic information measures, better suited than the classical, entropy–based information theoretic measure, to problems where the information is by nature dynamic.**
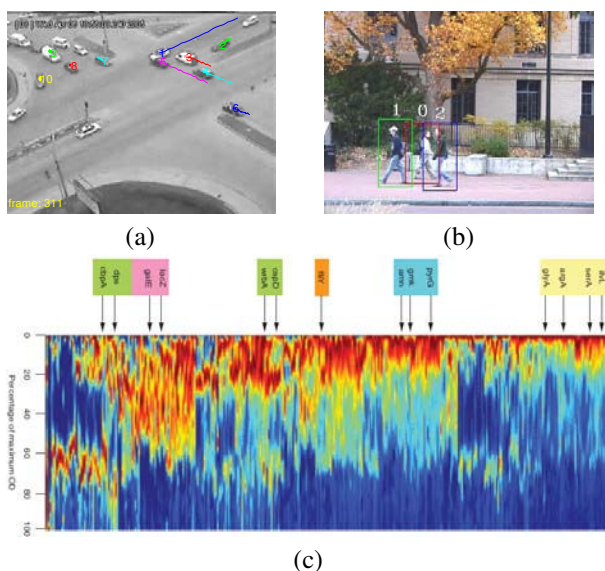
I. INTRODUCTION

(a)   (b)

(c)

Fig. 1. Examples of sparsely encoded information. (a) Detecting a traffic accident. (b) Tracking a person in a crowd.(c) *gfp*–visualized promoter activation during a diauxic shift experiment in E. coli [49]. In all cases fewer than $\mathcal{O}(10^{-5})$ of the data is relevant.

The recent exponential growth in data collection and actuation capabilities has the potential to profoundly impact society, with benefits ranging from safer, self aware environments, to enhanced image-based therapies. A major road-block to realizing this vision stems from the curse of dimensionality. Simply put, existing techniques are ill-equipped to deal with the resulting volume of data.

This paper discusses the key role that dynamics can play in timely extracting and exploiting actionable information that is very sparsely encoded in high dimensional data streams. Its central theme is the use of dynamical models as information encoding paradigms. Our basic premise is that spatio-temporal dynamic information can be compactly encapsulated in dynamic models, whose rank, a measure of the dimension of useful information, is often far lower than the raw data dimension. Embedding problems in the conceptual world of dynamical systems makes available a rich, extremely powerful resource base, leading to robust solutions, or, in cases where the underlying problem is intrinsically hard, to computationally tractable approximations with sub-optimality certificates. For instance, in this context, changes in the underlying process can be detected by simply computing the rank of a Hankel matrix constructed from the data and missing information can be recovered by solving a rank minimization problem that can be relaxed to a tractable semi-definite program. Finally, the use of dynamic models leads naturally to non-entropic information measures, better suited for problems where the information is by nature dynamic and changes as it propagates through a network where the nodes themselves are dynamical systems. These ideas are illustrated with several examples from different applications, including change detection in video sequences, motion segmentation, and uncovering co-promoted genes.

II. KEY SUBPROBLEMS ARISING IN THE CONTEXT OF DYNAMIC INFORMATION EXTRACTION

The challenges entailed in exploiting dynamic information sparsely encoded in very large data sets are illustrated in Figure 1: In all cases, decisions must be taken based on events discernible only in a small fraction of a very large data record: a short video sequence adds up to megabytes, yet the useful information (a change of behavior of a single target), may be encoded in just a portion of a few frames, e.g., less than 10^{-6} of the total data. Similarly, the data from the diauxic shift experiment shown in Figure 1(c) consists of 342×10^3 data points from the time traces of 1,920 promoters, (e.g., a total of 19 Mb of data), yet only a few critical time instants and promoter correlations are of interest. Additional challenges arise from the quality of the data, often fragmented and corrupted by noise. Addressing these challenges requires solving the following subproblems:

A: Nonlinear embedding of dynamic data. Finding low dimensional manifold structures in data, a hallmark of machine learning, is a key precursor to both dimensionality reduction and robust information extraction. Existing static techniques ([22] and references therein) provide low dimensional embeddings, but fail to exploit the large gap between data dimension and dynamic rank. As we will show in this paper, this can be accomplished by employing low rank dynamic models to capture time/parameter dependence on low dimensional manifolds that maximally absorb stationary high dimensions and nonlinearities.

B: Uncovering structures embedded in data. A key step in information extraction is the ability to find structures embedded in the data. For example, when analyzing data generated by an unknown number N_o of sources, it is of interest to identify the number of sources, associated substreams, and the individual dynamics. This is commonly accomplished by searching for statistical correlations or exploiting *a priori* known structural constraints. For example, independently moving objects in a video clip are efficiently detected by factorizing a matrix of feature trajectories [8], [44], [50]. However, methods based on correlations and (application dependent) *a priori* information alone are fragile to missing/corrupted data and have trouble disambiguating structures with overlapping *kinematic* or *statistical properties*. As shown here, all these difficulties can be avoided by seeking *dynamically coherent* substreams, e.g., subsets that can be jointly explained by low rank models. Further, this task can be efficiently carried out *without* explicitly finding these models, by estimating ranks of Hankel matrices constructed from time traces. A second example of the use of dynamics to uncover structure is given by the problem of recovering the 3D Euclidean geometry of a rigid object from a set of 2D image coordinates of its points in a sequence of images. Traditional approaches to this problem that rely only on geometrical constraints can recover 3D geometry only up to a projectivity transformation. On the other hand, exploiting the underlying dynamics leads to a convex algorithm that recovers the (provably correct) 3D Euclidian geometry.

C: Dynamic data segmentation. The goal here is to partition the data record into maximal, disjoint sets within which the data satisfies a given predicate. Examples include segmenting a video sequence of a person into its constituent activities, or identifying time periods where a given group of gene promoters is active. While this problem has been the object of considerable research in the past decade, it remains very challenging in cases involving noisy data, where most existing methods lead to computationally demanding problems [26], [2], with poor scaling properties. As we will show in the sequel, the use of dynamics provides a unified, efficient approach to robust segmentation. In its simplest form, the idea is to group data according to the complexity of the model that explains it. Intuitively, models associated with homogeneous data, e.g., a single activity or metabolic stage, have far lower complexity than those jointly explaining multiple datasets. Boundaries are thus characterized by a step increase in model complexity. In turn, these jumps in

model complexity can be efficiently detected by examining the singular values of a matrix directly constructed from the data.

D: Dynamic interpolation. Data streams are often fragmented: clinical trial patients may miss appointments, targets may be momentarily occluded. The challenges here are to (i) identify fragments belonging to the same data sets (for instance "tracklets" corresponding to a track of a single target, fragmented due to occlusion), and (ii) interpolate the missing data while preserving relevant dynamical invariants embedded in it. The latter is particularly important in cases where a transition is mediated by the missing data. An example is detecting an activity change from video data, when the transition point is occluded. Formulating the problem as a minimum order dynamical interpolation one leads to computationally attractive solutions, whereby values for missing data are selected as those that do not increase the complexity - or rank - of the model underlying the data record.

E: Hypothesis testing and distributed information sharing. Examples include determining whether (possibly nonoverlapping) data streams correspond to the same process or assessing whether a data set is a realization of a given process. In turn, this entails computing worst-case distances between data and model predictions, a task that can be efficiently accomplished by combining concepts from dynamical systems and information based complexity. Situations involving multiple information sources and users require the ability to (i) maintain consistent data labeling across sources, and (ii) mitigate the communications and computational burdens entailed in sharing very large datasets. Both issues can be efficiently addressed by exploiting the dynamical models underlying the data. Briefly, the idea is to identify a dynamic operator mapping the dynamic evolution of data projections over individual manifolds, amounting to a *dynamical registration* between sources. Sharing/comparing data streams then entails transmitting only the (low order) projections of dynamic variables and running these projections through the interconnecting operator.

In the remainder of this paper we show how the use of systems theory concepts provides a unified set of tools leading to computationally efficient solutions to problems A –E above. In all cases, these solutions are illustrated with practical examples.

III. Nonlinear embedding of dynamic data.

In the past few years, considerable research has been devoted to the problem of non-linear dimensionality reduction via manifold embedding. Briefly, the idea is to obtain lower complexity data representations by embedding it into low dimensional non-linear manifolds while preserving spatial neighborhoods. Commonly used methods include locally linear embeddings (LLE) [35], Isomap [40], Laplacian Eigenmaps [1], Hessian LLE [13], and Semidefinite Embedding [46], [45]. These methods successfully exploit spatial correlations to achieve (often substantial) dimensionality reduction. However, they fail to take advantage of the

temporal correlations that are characteristic of dynamic data. As we show next, constraining target manifolds to those spanned by feasible *dynamical* trajectories enables additional (substantial) dimensionality reduction and provides robustness against missing data and outliers.
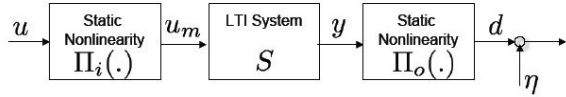

Fig. 2. Hammerstein/Wiener System Structure. Wiener system.

The starting point is the realization that since projections to/from manifolds can be modeled as memoryless nonlinearities, the problem of jointly identifying the embedding manifold, the dynamics characterizing the evolution of the data on this manifold, and the projection operators can be recast into the Hammerstein/Wiener system identification problem illustrated in Figure 2. Here $\Pi_i(.)$ and $\Pi_o(.)$ are memoryless nonlinearities, S is a linear time invariant (LTI) system that describes the temporal evolution of the data on the manifold, and $\mathbf{u} \in R^{n_u}$, $\mathbf{d} \in R^{n_d}$, $\mathbf{u}_m \in R^{n_{u_m}}$ and $\mathbf{y} \in R^{n_y}$, with $n_d \gg n_y$, $n_u \gg n_{u_m}$ represent the respective input (for instance a vector composed of past values of the output and a stochastic driving signal), the raw data, and their projections on the low dimensional manifold. A potential difficulty here stems from the fact that, as recently shown in [31], robust identification of Hammerstein/Wiener systems is generically NP–hard. However, efficient, computationally tractable relaxations that scale polynomially with problem size (both manifold dimension and number of temporal data points) can be obtained by pursuing a *risk-adjusted* approach. The main idea is to identify first the (piecewise) linear dynamics by sampling the set of signals $(\mathbf{u}_m, \mathbf{y})$, and attempting to find, for each sample (typically a subset of a ball in ℓ^2) an LTI operator S (the dynamics on the manifold) compatible with existing *a priori* information and such that $\mathbf{y} = S\mathbf{u}_m$. As shown in [38], [37] both steps reduce to a convex optimization problem via the use of Parrot's Theorem on norm-preserving matrix expansions and standard results on interpolation. The effectiveness of this approach is illustrated in Figure 3, where the application of these ideas enabled sustained tracking of multiple subjects in a cluttered outdoor scene [23]. Here, recasting the problem into a nonlinear identification form allowed for reducing the problem to an identification/prediction one in a 3-dimensional manifold.

It is worth emphasizing that this approach has the ability to exploit the synergy between the data embedding and dynamic modeling problems to improve robustness and computational properties. Robustness is improved by *automatically* discarding manifolds incompatible with *a priori* existing information on the dynamics, while computationally attractive models result from maximally absorbing nonlinearities in the manifold structure. Further, the *consistency set* [34] associated with the identification problem provides the means to (in)validate assumptions about the geometry
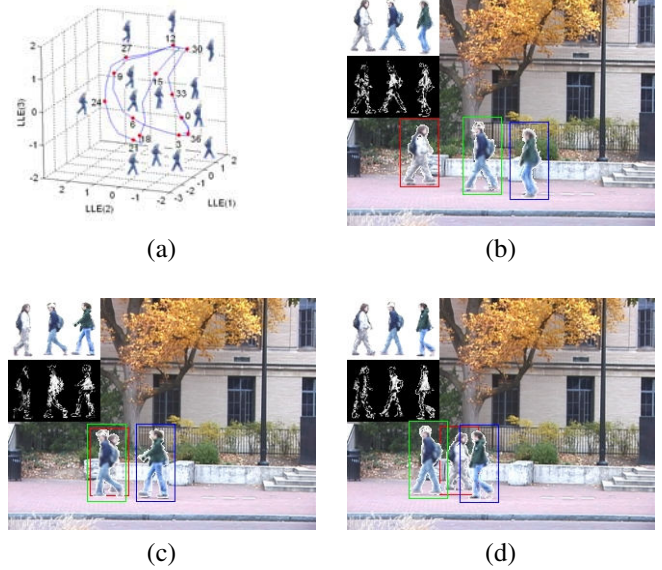

Fig. 3. (a): Sample 3 dimensional manifold extracted from a walking sequence. (b)-(d): use of dynamics on this manifold to predict target position and appearance.

of the manifolds and to quantify the approximation error. Thus, viewing data as a manifestation of hidden dynamics allows a synergy between machine learning (the manifold structure), identification theory (theoretical underpinnings, computational framework) and information based complexity (worst case prediction–error bounds).

IV. STRUCTURE EXTRACTION FROM HIGH DIMENSIONAL DATA STREAMS.

Structure extraction methods based on correlations and (application dependent) *a priori* information alone are often fragile to missing/corrupted data and have trouble disambiguating structures with overlapping *kinematic* or *statistical properties*. As an illustrative example, consider time traces $\mathbf{p}_{ti} = (u_{ti}, v_{ti})^T, t = 1 \ldots n$ of n_p features P_i, $i = 1, \ldots, n_p$, from a single rigid object. Kinematic constraints imply that the rank of the "measurements" matrix $\mathcal{W}_{1:F} \doteq [\mathbf{p}_{ti}] \in R^{2n \times n_p}$ is at most 4 [41]. The number N_o of *independent* rigid bodies can thus be estimated by factorizing that matrix into rank 4 submatrices. Yet this approach fails to disambiguate objects with partially shared motion, as illustrated in Figure 4: Here rank$(\mathcal{W}) = 7$ due to shared propeller rotations; hence any segmentation based solely on factorizing $\mathcal{W}$ will fail to distinguish this case from the case of just two independently moving propellers. The root-cause is that properties that are invariant under row permutations in $\mathcal{W}$ are limited to revealing geometric dependencies but *ignore dynamic constraints*[1]. As shown next, these ambiguities can be solved through the use of dynamical models that exploit both sets of constraints.

The starting point is the realization that for two points $\mathbf{p}_r, \mathbf{p}_s$ belonging to the same source, the time evolution of

---

[1]Any permutation of the rows of $\mathcal{W}$ satisfies the same geometric constraints, but corresponds to different time trajectories.

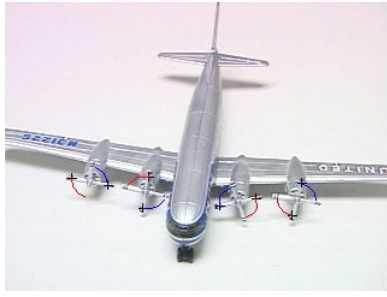Fig. 4. Objects with shared motion modes. Right and left wing propellers move in opposite directions at the same speed.

(a)   (b)
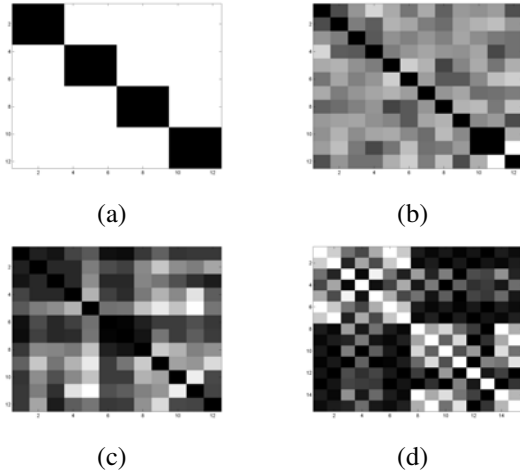
(c)   (d)

Fig. 5. Segmentation results for the airplane example. (a) Dynamics based segmentation. (b) Costeira-Kanade segmentation. (c) Zelnik-Manor-Irani segmentation. (d) GPCA segmentation.

$y_{r,s}(k) \doteq \mathbf{p}_r(k) - \mathbf{p}_s(k)$ does not carry information about the overall *group* motion of the source. Equivalently, states associated with group motion are *unobservable* from $y_{r,s}$ if $\mathbf{p}_r$ and $\mathbf{p}_s$ belong to the same dynamic cluster. Hence the associated Hankel matrix is rank deficient [34] vis-a-vis the case of points from different sources. This leads to the following simple dynamic clustering algorithm [25]:

(i) For each pair (r,s), form the Hankel matrix $\mathsf{H}_{y_{r,s}}$ of pairwise differences $y_{r,s}(k) = \mathbf{p}_r(k) - \mathbf{p}_s(k)$:

$$\mathbf{H}_y = \begin{bmatrix} \mathbf{y}(1) & \mathbf{y}(2) & \cdots & \mathbf{y}(\frac{n}{2}) \\ \mathbf{y}(2) & \mathbf{y}(3) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}(\frac{n}{2}) & \cdots & \cdots & \mathbf{y}(n) \end{bmatrix} \quad (1)$$

(ii) Group points according to the minimum value of rank[$\mathsf{H}_{y_{r,s}}$].

In this context, robust handling of noisy measurements $\hat{y}(k) = y(k) + \eta(k)$, is accomplished by simply replacing "rank" by the number of singular values above the covariance of the measurement noise[2], leading to an algorithm computationally no more expensive than a sequence

---

[2]In this case $\mathsf{H}_{\hat{y}} = \mathsf{H}_y + \mathsf{H}_\eta$, and, under ergodicity assumptions, $\mathsf{H}_\eta^T \mathsf{H}_\eta$ is an estimate of the covariance matrix of the noise.

---

of SVDs. The effectiveness of this approach is illustrated in Figure 5 where darker matrix elements indicate higher correlations: As shown there, the dynamics based approach achieves perfect segmentation, while methods relying solely on factorizations of $\mathcal{W}$ [50], [47], [43] fail.
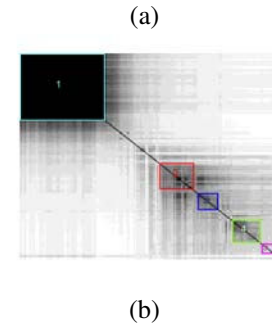

(a)

(b)

Fig. 6. (a) Sample frame. (b) Structures found using dynamic rank (darker color indicates higher dynamic correlation). The hierarchy in the lower right corner corresponds to different portions of the body.
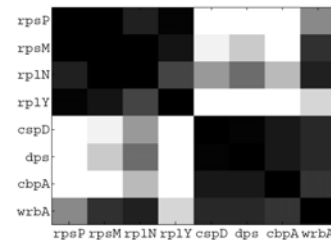

Fig. 7. Dynamic correlation between genes in the diauxic shift experiment of Figure 1(c). The two identified groups correspond to growth related (top left) and stationary (bottom right) genes. The fainter correlation between wrbA and (rpsM,rplN) was unexpected.

An interesting property of the dynamics based approach to segmentation, illustrated in Figures 6 and 7 is the ability to provide a *hierarchical* segmentation according to the complexity of the joint dynamics. This is key to model the behavior of a target composed by several components acting in a dynamically correlated fashion, e.g., the limbs of a walking person or co-regulated genes. The aggregate behaves as a non-rigid object, whose components share motion modes.

A. Euclidean Geometry from Dynamics

Another example illustrating the power of dynamics to uncover structure embedded in data is given by the problem

Fig. 8. Sample data for a dynamic segmentation example: Frames 311 and 341 from a car crash sequence.
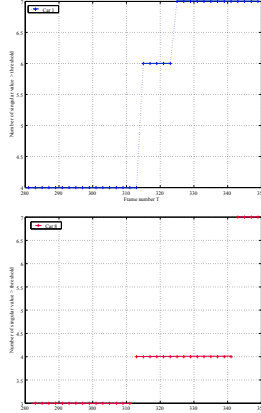

Fig. 9. Hankel rank time traces for the crash detection example. Top: Car 1. Bottom: Car 8.

of recovering 3D shape of an object from a video sequence of perspective images. Traditional approaches to 3D structure from perspective images rely only on geometrical constraints (reflecting the rigid nature of the object) but can only recover shape up to a projectivity transformation [20], [15]. In contrast, using the *dynamical* information encapsulated in the *temporal ordering* of the frames [36], leads to a provably correct algorithm that finds the Euclidean structure (up to a single scaling factor) by minimizing the rank of the Hankel matrix associated with its trajectory, subject to one linear and two rank constraints. From its definition, it is clear that the rank of the Hankel matrix encapsulates temporal correlations, since it is not invariant under a permutation of the ordering of the frames. The surprising result is that this rank also encapsulates rigidity, since it can be proved that the correct
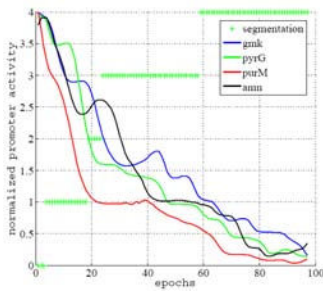

Fig. 10. Detecting transitions in an E. coli culture via Hankel rank. Jumps at 20 and 57 correspond to shifts from metabolizing glucose to lactose, to stationary phase, respectively.

3D rigid geometry, up to an overall constant scaling factor, is precisely the one that minimizes it, subject to the additional constraints.

More formally, consider the image trajectories $\mathbf{p}_{ki} = (u_{ki}, v_{ki})^T$, $i = 1, 2, 3$, $k = 1, \ldots, F$ of the perspective projections of three points $\mathbf{P}_{ki}$, $i = 1, 2, 3$, belonging to a rigid moving under some point-wise rigid motion operator $\mathcal{L}$. Then, it can be shown [36] that the 3D camera Cartesian coordinates of $\mathbf{P}_{ki}$ $i = 1, 2, 3$, $k = 1, \ldots, F$ are given by:

$$\mathsf{P}_{ki} = \begin{bmatrix} X_{ki} \\ Y_{ki} \\ Z_{ki} \end{bmatrix} = \frac{1}{\lambda_o \rho^k} Z_{ki}^* \begin{bmatrix} \frac{1}{f} u_{ki} \\ \frac{1}{\alpha f} v_{ki} \\ 1 \end{bmatrix} \quad (2)$$

where λ_o and $\rho > 0$ are constant factors (point and frame independent), and where $\{Z_{k1}^*, Z_{k2}^*, Z_{k3}^*\}_{k=1,\ldots,F}$ solve the following rank minimization problem

$$\min_{\{Z_{k1}^*, Z_{k2}^*, Z_{k3}^*\}_{k=1,\ldots,F}} \mathrm{rank}\left(\begin{bmatrix} \mathsf{H}_{\mathbf{y}^{13}} & \mathsf{H}_{\mathbf{y}^{23}} \end{bmatrix}\right) \quad (3)$$
$$\text{subject to:} \ Z_{ki} \geq 1$$

where

$$\mathbf{y}_k^{ij} = \begin{bmatrix} \frac{1}{f}(Z_{ki}^* u_{ki} - Z_{kj}^* u_{kj}) \\ \frac{1}{\alpha f}(Z_{ki}^* v_{ki} - Z_{kj}^* v_{kj}) \\ Z_{ki}^* - Z_{kj}^* \end{bmatrix}$$

and where $\mathsf{H}_{\mathbf{y}}$ denotes the Hankel matrix of the sequence $\{\mathbf{y_k}\}_{k=1}^F$.

Thus, the correct *relative* 3D structure can be recovered by solving a rank–minimization problem, since $Z_{ki}^* = \lambda_o \rho^k Z_{ki}$, then $\frac{Z_{ki}^*}{Z_{ki}} = \frac{Z_{kj}^*}{Z_{kj}}$ for all (i, j), where Z and Z^* denote the actual and recovered depths, respectively. While in many situations this may suffice, in others it is of interest to recover the geometry up to an overall, frame-independent scaling. This can be accomplished by the algorithm below which incorporates one linear and two rank constraints imposing rigidity between the first and last frame [36]:

Algorithm 1: RANK MINIMIZATION
BASED 3D-DEPTH RECOVERY

Data: Camera Intrinsic Parameters.
Input: (u_{ki}, v_{ki}), the temporally ordered 2-D coordinates of N points in F frames.
Output: 3D depths Z_{ki} up to an overall scaling constant.
  1. Form the *difference* vectors $\mathbf{y}_k^{iN} \doteq \mathbf{P}_{ki}^* - \mathbf{P}_{kN}^*$,
$i = 1, \ldots, N-1$ where $\mathsf{P}_{ki}^* \doteq Z_{ki}^* \begin{bmatrix} \frac{u_{ki}}{f} & \frac{v_{ki}}{\alpha f} & 1 \end{bmatrix}^T$,
and the corresponding Hankel matrices $\mathsf{H}_{\mathbf{y}^{iN}}$
  2. Solve: $\min_{Z_{ki}^* \geq 1} \mathrm{rank}\begin{bmatrix} \mathsf{H}_{\mathbf{y}^{1N}} \ldots \mathsf{H}_{\mathbf{y}^{N-1 N}} \end{bmatrix}$ subject to (4) and (6)

$$m_1^{20}(\frac{u_{11}^2}{f^2} + \frac{v_{11}^2}{f^2\alpha^2} + 1) + m_1^{02}(\frac{u_{12}^2}{f^2} + \frac{v_{12}^2}{f^2\alpha^2} + 1) -$$
$$2 * m_1^{11}(\frac{u_{11}u_{12}}{f^2} + \frac{v_{11}v_{12}}{f^2\alpha^2} + 1) -$$
$$m_F^{20}(\frac{u_{F1}^2}{f^2} + \frac{v_{F1}^2}{f^2\alpha^2} + 1) - m_F^{02}(\frac{u_{F2}^2}{f^2} + \frac{v_{F2}^2}{f^2\alpha^2} + 1) +$$
$$2 * m_F^{11}(\frac{u_{F1}u_{F2}}{f^2} + \frac{v_{F1}v_{F2}}{f^2\alpha^2} + 1) = 0 \tag{4}$$

where

$$m_t^{20} \doteq Z_{t1}^2, \ m_t^{11} \doteq Z_{t1}Z_{t2}, \ m_t^{02} \doteq Z_{t2}^2 \tag{5}$$

and

$$\text{rank}\left\{\begin{bmatrix} m_t^{20} & m_t^{11} \\ m_t^{11} & m_t^{02} \end{bmatrix}\right\} = 1, \ t = \{1, F\} \tag{6}$$

Figures 11 to 13 show that the dynamics based 3D recovery algorithm (HankelSFM) does not introduce perspective distortion and that it has better accuracy than using purely geometrical algorithms (MHSFM [27] and HTSFM [21]).

(a)   (b)
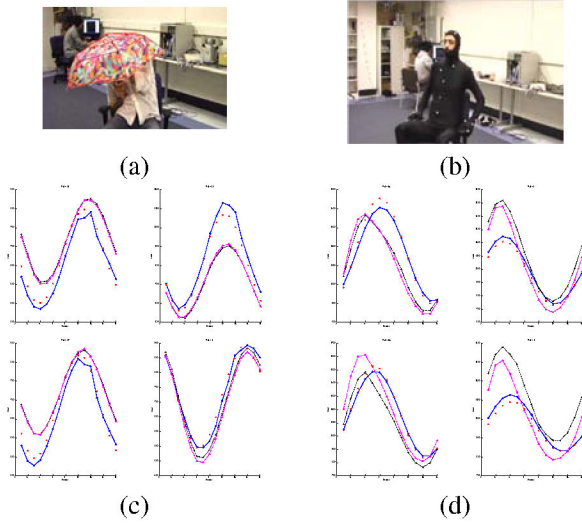
(c)   (d)

Fig. 11. Sample frames of two sequences: Umbrella (a) and Human on a chair (b). Estimated depth trajectories for four basis points. Red stars: ground truth data; solid blue line: HankelSFM; dotted black line: MHSFM; and dashed magenta line: HTSFM. (c) Umbrella experiment. (d) Human on a chair experiment.

V. ROBUST DYNAMIC DATA SEGMENTATION.

In principle, changes in the processes underlying a given data record can be detected by a two tiered approach: identification of an underlying set of models (the consistency set) followed by a model (in)validation step to detect points at which new data are inconsistent with all the models in the set. However, the entailed computational complexity is high, roughly n^5 for n data points. A fast, computationally efficient alternative can be obtained by searching for points where the complexity of the underlying model changes [10]. The main idea behind this approach is the fact that models associated with homogeneous data have far lower complexity than those jointly explaining multiple datasets. Further, the complexity of the (unknown) model can be estimated from the experimental data by computing the number $N_{sv,\sigma}(H_y)$
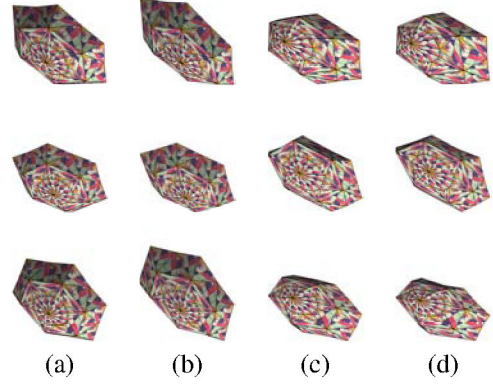

(a)   (b)   (c)   (d)

Fig. 12. 1, 6 and 12 of the umbrella sequence. (a) Ground truth data, and 3D structure recovered using (b) HankelSFM, (c) MHSFM and (d) HTSFM

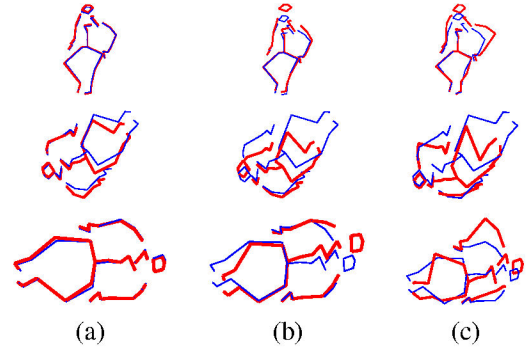(a)   (b)   (c)

Fig. 13. Frames 1, 7, 14 frames of the human on a chair sequence with ground truth data (blue) superimposed with 3D structure (red) recovered using (a) HankelSFM, (b) MHSFM and (c) HTSFM.

of (significant) singular values of a Hankel matrix similar to H_y in (1). Hence the data record can be segmented according to discontinuities in $N_{sv,\sigma}(H_y)$. Figures 8 and 9 illustrate the effectiveness of this approach in detecting contextually abnormal behavior - an accident - evidenced by a jump in the Hankel rank. An application of this technique to detecting changes in promoter activity in E. coli is shown in Fig. 10.

The approach outlined above works well for cases where the noise is moderate and adequately characterized as an ℓ^2 bounded signal. Cases where these conditions do not hold (for instance ℓ^∞ noise) can be handled by a modification of this idea (detecting mode changes in piecewise affine models) as follows [29], [31]. The starting point is the assumption that the data record has been generated by a piecewise affine model of the form:

$$\mathcal{H} : f\left(\mathbf{p}_{\sigma(t)}, \{\mathbf{x}(k)\}_{k=t-i}^{t+j}\right) = \eta_f \tag{7}$$

where f is an affine function[3] of the parameter vector $\mathbf{p}_{\sigma(t)}$ which takes values from a finite unknown set according to a piecewise constant function $\sigma(t)$, and η_f denotes an unknown noise signal. Here i and j are positive integers

[3]That is: $f\left(\mathbf{p}_{\sigma(t)}, \{\mathbf{x}(k)\}_{k=t-i}^{t+j}\right) = A(\mathbf{x})\mathbf{p}_{\sigma(t)} + \mathbf{b}(\mathbf{x})$

that account for the memory of the model (e.g., $j = 0$ corresponds to a causal model, or $i = j = 0$ corresponds to a memoryless model). Next, consider the sequence of *first order differences* of the parameters $\mathbf{p}_{\sigma(t)}$, given by

$$\mathbf{g}(t) = \mathbf{p}_{\sigma(t)} - \mathbf{p}_{\sigma(t+1)} \qquad (8)$$

Clearly, a non-zero element of this sequence corresponds to a *change* in the underlying model. Hence, partitioning the data record into maximal homogeneous sequences is equivalent to finding a hybrid model of the form (7), consistent with the *a priori* information (e.g., a bound on $\|\eta\|_{\ell^\infty}$) and experimental data, such that the number of non-zero elements of the vector $\mathbf{g}(.)$ is minimized. Formally, defining $\delta(t) = \|\mathbf{g}(t)\|_\infty$, the objective is to minimize $\|\delta\|_{\ell^o}$, the number of non-zero elements of δ, subject to (7). Using the fact that the *convex envelope* of $\|\cdot\|_{\ell^0}$ in R^N is the ℓ^1-norm [16], this non-convex problem can be relaxed to:

$$
\begin{aligned}
\text{minimize}_{\mathbf{p}(t),\eta(t)} \quad & \|\{\mathbf{g}\}\|_{\ell_1} \\
\text{subject to} \quad & f\left(\mathbf{p}(t), \{\mathbf{x}(k)\}_{k=t-i}^{t+j}\right) = \eta(t) \quad \forall t \\
& \|\{\eta\}\|_* \leq \epsilon
\end{aligned}
$$
$$(9)$$

Since f is an affine function of $\mathbf{p}(t)$, (9) has a convex feasibility set $\mathcal{F}$. Thus, using the ℓ_1 norm leads to a convex, computationally tractable relaxation. The resulting solution can be further improved using the iterative procedure proposed in [17], based on solving, at each iteration, the following weighted ℓ_1-norm minimization over the convex feasible set $\mathcal{F}$:

$$
\begin{aligned}
\text{minimize}_{z,g,p,\eta} \quad & \sum_{t=1}^{T-1} w_t^{(k)} z_t \\
\text{subject to} \quad & \|\mathbf{g}(t)\|_\infty \leq z_t \qquad\qquad\quad \forall t \\
& f\left(\mathbf{p}(t), \{\mathbf{x}(k)\}_{k=t-i}^{t+j}\right) = \eta(t) \quad \forall t \\
& \|\{\eta\}\|_* \leq \epsilon
\end{aligned}
$$
$$(10)$$

where $w_i^{(k)} = (z_i^{(k)} + \delta)^{-1}$ are weights with $z_i^{(k)}$ being the arguments of the optimal solution at the kth iteration and $z^{(0)} = [1, 1, .., 1]^T$; and where δ is a (small) regularization constant that determines what should be considered zero.

The choice of $*$, the norm characterizing the noise, is application dependent. For instance the ℓ^∞-norm performs well in finding anomalies, since in this case the change detection algorithm looks for *local* errors, highlighting outliers. On the other hand, when a bound on the ℓ^1 or ℓ^2-norm of the noise is used, the change detection algorithm is more robust to outliers and it favors the continuity of the segments (i.e., longer subsequences). In addition, when using these norms, the optimization problem automatically adjusts the noise distribution among the segments, better handling the case where the noise level is different in different segments.

A. Example 1: Video Segmentation.

Segmenting and indexing video sequences have drawn significant attention due to the increasing amounts of data in digital video databases. Systems that are capable of
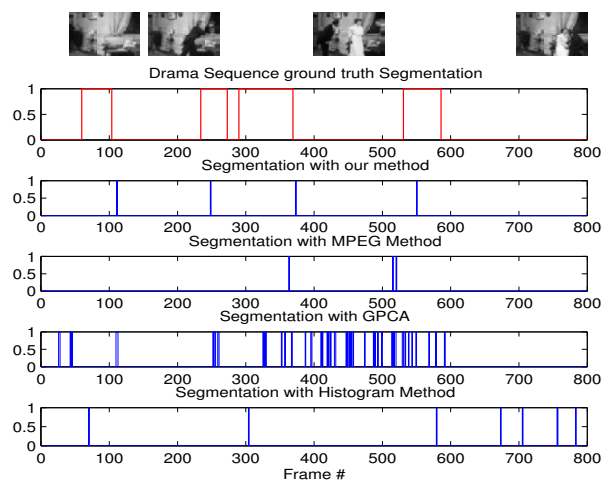

Fig. 14. Video segmentation as a hybrid system identification

segmenting video and extracting key frames that summarize the video content can substantially simplify browsing these databases over a network and retrieving important content. An analysis of the performances of early shot change detection algorithms is given in [18]. The methods analyzed in [18] can be categorized into two major groups: i) methods based on histogram distances, and ii) methods based on variations of MPEG coefficients. A comprehensive study is given in [48] where a formal framework for evaluation is also developed. Other methods include those where scene segmentation is based on image mosaicing [30], [32] or frames are segmented according to underlying subspace structure [24].

Given a video sequence of frames $\left\{\mathcal{I}(t) \in \mathbb{R}^D\right\}_{t=1}^T$, the video segmentation problem can be solved by first projecting the data into a lower dimensional space, using for instance Principal Component Analysis (PCA), and then applying the sparsification algorithm described above to the projected data (to exploit the fact that the number of pixels D is usually much larger than the dimension of the subspace where the frames are embedded):

$$\mathcal{I}(t) \longmapsto \mathbf{x}(t) \in \mathbb{R}^d.$$

Assuming that each $\mathbf{x}(t)$ within the same segment lies on the same hyperplane not passing through the origin[4] leads to the following hybrid model:

$$\mathcal{H}_1 : f\left(\mathbf{p}_{\sigma(t)}, \mathbf{x}(t)\right) = \mathbf{p}_{\sigma(t)}^T \mathbf{x}(t) - 1 = 0 \qquad (11)$$

Thus, in this context algorithm (10) can be directly used to robustly segment the video sequence. It is also worth stressing that as a by-product this method also performs *key frame extraction* by selecting $\mathcal{I}(t)$ corresponding to the minimum $\|\eta(t)\|$ value in a segment (e.g., the frame with the

---

[4]Note that this always can be assumed without loss of generality due to the presence of noise in the data.

	Roadtrip	Mountain	Drama	Family
Sparsification	0.9373	0.9629	0.9802	0.9638
MPEG	1	0.9816	0.9133	0.9480
GPCA	0.6965	0.9263	0.7968	0.8220
Histogram	0.9615	0.5690	0.8809	0.9078

TABLE I

RAND INDICES

Fig. 15. Comparison of segmentation results for the Smoke sequence concatenated with its transposed dynamics.

smallest fitting error) as a good representative of the entire segment.

The content of a video sequence usually changes in a variety ways: For instance: the camera can switch between different scenes (e.g., shots); the activity within the scene can change over time; objects or people can enter or exit the scene, etc. There is a hierarchy in the level of segmentation one would require. The noise level ϵ can be used as a tuning knob in this sense.

Figure 14 shows the results of applying this approach to a video sequence, `drama.avi`, available from `http://www.open-video.org`. The original mpeg files were decompressed, converted to grayscale and title frames were removed. Each sequence shows a different characteristic on the transition from one shot to the other. The camera is mostly non-stationary, either shaking or moving. For comparison, results using GPCA, a histogram based method and an MPEG method for segmenting the sequences with optimal parameters (found by trial and error) are also shown. Table I shows the Rand indices [33] corresponding to the clustering results obtained for this sequence and three others from the same database (`roadtrip.avi`, `mountain.avi`, and `family.avi`) using the different methods, providing a quantitative criteria for comparison. Since the Rand index does not handle dual memberships, the frames corresponding to transitions were neglected while calculating the indices. These results show that indeed the sparcity method does well, with the worst relative performance being against MPEG and B2B in the sequence Roadtrip. This is mostly due to the fact that the parameters in both of these methods were adjusted by a lengthy trial and error process to yield optimal performance in this sequence. Indeed, in the case of MPEG based segmentation, the two parameters governing cut detection were adjusted to give optimal performance in the Roadtrip sequence, while the five gradual transition parameters were optimized for the Mountain sequence.

B. Example 2: Segmentation of Dynamic Textures.

Modeling, recognition, synthesis and segmentation of dynamic textures have drawn a significant attention in recent years [14], [5], [7], [19]). In the case of segmentation tasks, the most commonly used models are mixture models, which are consistent with the hybrid model framework.

In the sparsification framework described earlier in this section, the problem of temporal segmentation of dynamic textures reduces to the same mathematical problem as the
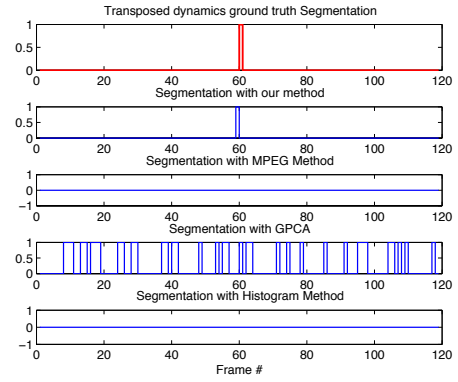
video segmentation problem, with the difference that now the underlying hybrid model should take the dynamics into account. First, dimensionality reduction is performed via PCA ($\mathcal{I}(t) \longmapsto \mathbf{y}(t) \in \mathbb{R}^d$) and then the reduced-order data is assumed to satisfy a simple causal autoregressive model similar to the one in [7]. Specifically, in this case the hybrid model is given by:

$$\mathcal{H}_2 : f\left(\mathbf{p}_{\sigma(t)}, \{\mathbf{y}(k)\}_{k=t-n}^t\right) = \mathbf{p}_{\sigma(t)}^T \begin{bmatrix} \mathbf{y}(t-n) \\ \vdots \\ \mathbf{y}(t) \end{bmatrix} - 1 = 0 \tag{12}$$

where n is the regressor order. This model, which can be considered as a step driven autoregressive model, was found to be effective experimentally[5]. The power of this approach is illustrated in Figures 15 and 16 where two very challenging sequences were segmented. The first sequence consists of a patch of dynamic texture (smoke) appended in time to another patch from the same texture but transposed. Thus, the two subsequences have the same photometric properties but differ in the main motion direction. The second sequence was generated using another dynamic texture (river) by sliding a window both in space and time (by going forward in time in the first half and by going backward in the second), thus reversing the dynamics due to the river flow.

VI. CONSTRAINED INTERPOLATION OF HIGH DIMENSIONAL SIGNALS:

Consider first the simpler case of interpolating noiseless data, generated by a single LTI system, with McMillan degree bounded by some known n_o. Formally, given a partial sequence $\mathbf{d}_g = \{d_1, \cdots, d_q, d_{q+r}, \cdots, d_n\}$, the goal is to estimate the missing elements $\mathbf{d}_x = \{d_{q+1}, \ldots, d_{q+r-1}\}$ that optimally fit the given data. Intuitively, the best fitting missing elements are those that require adding the least

---

[5]The independent term 1 here accounts for an exogenous driving signal. Normalizing the value of this signal to 1, essentially amounts to absorbing its dynamics into the coefficients $\mathbf{p}$ of the model. This allows for detecting both changes in the coefficients of the model and in the statistics of the driving signal.
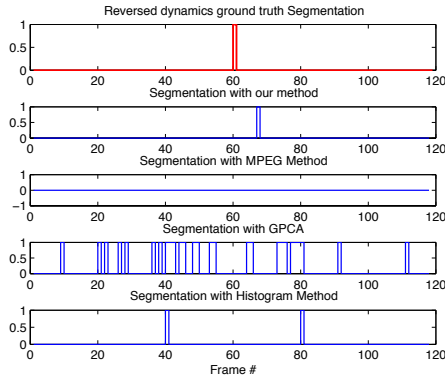
Fig. 16. Comparison of segmentation results for the River sequence concatenated with its reversed dynamics.

number of modes to the existing model in order to explain the new data. Using the fact that the order of the underlying model is given by the rank of the corresponding Hankel matrix (under the assumption that $n \gg n_o$), this problem can be recast into the following rank minimization form: $\mathbf{d}_{xo} = argmin_{\mathbf{d}_x} \text{rank}(\mathsf{H})$ where H is the Hankel matrix associated with the completed sequence $\mathbf{d} = \{d_i\}$. In this context, noise can be readily handled by simply adding a new variable v such that the measured data $y = d + v$, and a suitable noise description of the form $v \in \mathcal{N}$, a convex, compact set. Finally, since rank minimization is NP–hard [42], using the convex relaxation proposed in [17] leads to the following algorithm [9]:

Algorithm 2: HANKEL RANK MINIMIZATION BASED INTERPOLATION/PREDICTION

Input at time k: N_h: Horizon length;
$\mathcal{I}_a \subseteq [k - N_h, k]$: set of indices of available measurements (with $\text{card}(\mathcal{I}_a) \geq n$);
$\mathcal{I}_e \subseteq [k - N_h, k + 1]$: set of indices of data to be estimated; with $\mathcal{I}_a \cup \mathcal{I}_e = \mathcal{I}$; available data y_ℓ, $\ell \in \mathcal{I}_a$;
set membership description
of the measurement noise $v \in \mathcal{N}$.
Output: Estimates $\hat{\zeta}_\ell$ of ζ_ℓ, $\forall \ell \in \mathcal{I}_e \cup \mathcal{I}_a$

1. Let ζ^* denote the following sequence:
$$\zeta_i^* = \begin{cases} y_i - v_i \text{ if } i \in \mathcal{I}_a \\ x_i \text{ if } i \in \mathcal{I}_e \end{cases} \text{ where } v, x \text{ are free}$$
variables, and form the matrix
$$\mathsf{H}_\zeta \doteq \begin{bmatrix} \zeta_{i_1}^* & \zeta_{i_2}^* & \cdots & \zeta_{i_{n+n_u+1}}^* \\ \zeta_{i_2}^* & \zeta_{i_3}^* & \cdots & \zeta_{i_{n+n_u+2}}^* \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_{i_{n+1}}^* & \zeta_{i_{n+2}}^* & \cdots & \zeta_{i_{2n+n_u+1}}^* \end{bmatrix}$$

2. (Approximately) minimize $\text{rank}[\mathsf{H}(x, v)]$ by solving the following convex problem in x, v, R, S:
minimize $\quad Tr(\mathsf{R}) + Tr(\mathsf{S})$
subject to $\quad \begin{bmatrix} \mathsf{R} & \mathsf{H}(x) \\ \mathsf{H}(x)^T & \mathsf{S} \end{bmatrix} \geq 0, \ \{v_\ell\} \in \mathcal{N}.$

3. Estimate/predict the output ζ_ℓ from the noisy measurements y_ℓ by:
$$\hat{\zeta}_i = \begin{cases} y_i - v_i \text{ if } i \in \mathcal{I}_a \text{ (estimation)} \\ x_i \text{ if } i \in \mathcal{I}_e \text{ (interpolation/prediction)} \end{cases}$$
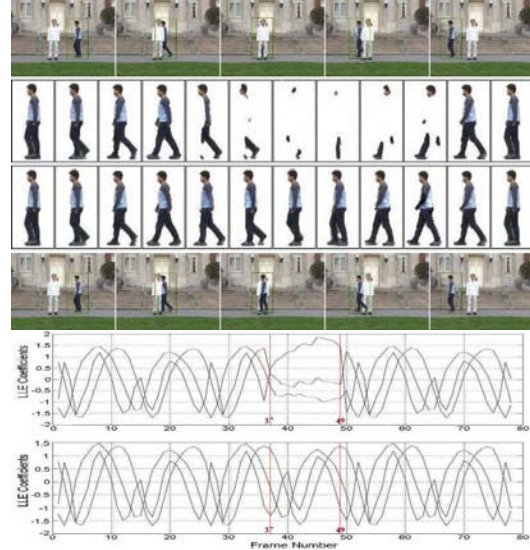

Fig. 17. Missing data (second and fifth rows) interpolated by rank minimization on 3D manifolds (third and sixth rows).

Examples of application of this idea are shown in Figures 17 and 18. In Figure 17 nonlinear embeddings were used first to map the data to a low order manifold where the rank-minimization based interpolation was performed, followed by a remapping of the data to pixel space. Fig. 18 shows how a combination of dynamic interpolation and Hankel–rank based segmentation is able to detect occluded events.
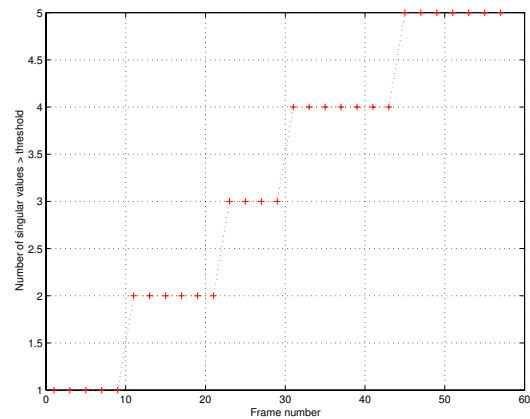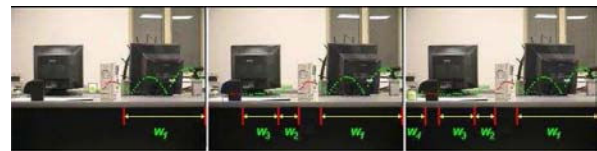

Fig. 18. Occluded event detection. Top: Dynamic data interpolation. Bottom: Hankel rank plot showing events.

It is worth mentioning that the ideas discussed in this section are directly applicable to hybrid models of the form (7). In this case, minimizing the rank is roughly equivalent to interpolating the data so that the resulting underlying model exhibits the minimum number of jumps.

VII. Hypothesis testing and data sharing.

A salient feature of the dynamics–based information extraction framework is its ability to furnish application–relevant *worst–case bounds* on distances between data and model predictions and, significantly, between non-overlapping data streams, in terms of their respective models. These bounds lead to computationally efficient hypothesis testing techniques. Consider a data stream $\{y_k\}_{k=0}^{N-1}$ generated by an underlying model of the form:

$$y_{k+1} = \mathcal{F}[\mathbf{y}_k, \mathbf{e}_k], \ \mathbf{y}_k \doteq [y_k, \ldots, y_{k-n}], \mathbf{e}_k \doteq [e_k, \ldots, e_{k-n}] \tag{13}$$

where e is a stochastic input, and $\mathcal{F}$ is the (unknown) evolution operator. Collecting all available *a priori* information about $\mathcal{F}$ and e (e.g., known dynamic modes and noise statistics) into sets $\mathcal{S}$ and $\mathcal{N}$ reduces the problem of finding $\mathcal{F}$ to a finite dimensional optimization via an extended Caratheodory-Fejer interpolation framework [39]. This method is *interpolatory*, hence it generates a model

$$\mathcal{F}_{id} \in \mathcal{T}(\mathbf{y}) \doteq \{\mathcal{F} \in \mathcal{S} : y_{k+1} = \mathcal{F}[\mathbf{y}, \mathbf{e}], \ \mathbf{e} \in \mathcal{N}\}$$

the set of all models consistent with both the *a priori* information and the experimental data. Since the actual (unknown) model that generated the data must also belong to $\mathcal{T}(y)$, a bound on the worst case prediction error of the identified model $\mathcal{F}_{id}$ is given by:

$$\|\hat{\mathbf{y}} - \mathbf{y}\|_* \leq \sup_{\mathcal{F}_1, \mathcal{F}_2 \in \mathcal{T}(\mathbf{y})} \|\mathcal{F}_1[\mathbf{y}, \mathbf{e}] - \mathcal{F}_2[\mathbf{y}, \mathbf{e}]\|_* = \mathcal{D}[\mathcal{T}(\mathbf{y})] \tag{14}$$

where $\|.\|_*$ is a suitable norm and $\mathcal{D}(.)$ denotes the diameter of the set $\mathcal{T}(\mathbf{y})$. When the sets $\mathcal{S}$ and $\mathcal{N}$ are convex, computing this bound reduces to a convex optimization problem [34, Lemma 10.3]. Note that these bounds are computed only once and remain valid as long as the underlying dynamics do not change.
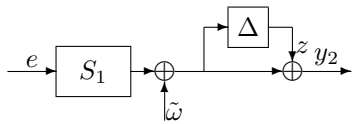
Fig. 19. Distance between data streams as a model (in)validation problem.

The worst case bounds provided by $\mathcal{D}[\mathcal{T}(\mathbf{y})]$ can be used to robustly assess the distance between non-overlapping data streams [28]. The idea is to measure this quantity in terms of the distance between the corresponding (model) consistency sets. Intuitively, two partial data streams are considered to be manifestations of the same underlying process if they can be generated by the same dynamic model. The introduction of the consistency set in this context allows for taking into consideration data–quality issues (relatively few observations,
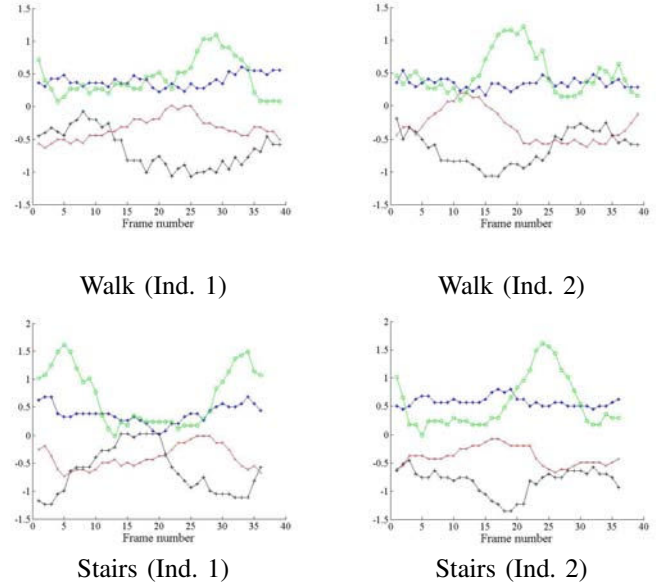
Fig. 20. Sample joint traces (*, shoulder; +, elbow; x, hip; o, knee) for different activities

Person	Walking	Running	Staircase
A	1, 2	16 to 18	25 to 27
B	3 to 8	11 to 15	21 to 24
C	9, 10	none	28 to 30
D	none	19	none
E	none	20	none

TABLE II
Experimental Data

corrupted by noise) and *a priori* information. Computing the exact distance between consistency sets is costly, but it can be relaxed to the model (in)validation form shown in Figure 19. Given data streams y_1, y_2, the idea is to identify a nominal model S_1 associated with y_1 and a deformation operator Δ so that the pair (S_1, Δ) generates y_2. As shown in [6], computing the minimum norm $\gamma_{min} \doteq \min \|\Delta\|_\infty$ over the set of all operators with this property reduces to a convex Linear Matrix Inequality optimization problem. Thus, the value γ_{min} provides a computationally tractable upper bound on the distance between consistency sets.

This idea is illustrated next, using as an example the problem of gait classification. The experimental data listed in Table II and plotted in Figure 20, consists of 30 vector sequences, taken from 5 different persons, named A, B, C, D and E. Each sequence contains measurements of the angles of the shoulder, elbow, hip and knee joints of a person walking, running or walking up a staircase. For illustrative sake, these sequences are numbered from 1 to 30 so that the first 10 correspond to walking, the second set of 10 to running and the third set of 10 to walking up a staircase.

Table III shows the distance from each data set to the dynamic model representing each activity. For each sequence, these nominal models were obtained by first finding a model associated with each of the remaining sequences and then

Sequence	S_{walk}	S_{run}	S_{stair}
y_1	$0.1743^\dagger$	0.6758	0.5973
y_2	$0.2333^\dagger$	0.5818	0.2427
y_3	$0.0305^\dagger$	0.6843	0.6866
y_4	$0.0410^\dagger$	0.6217	0.5305
y_5	$0.0819^\dagger$	0.6915	0.6069
y_6	$0.0001^\dagger$	0.6879	0.7688
y_7	$0.0900^\dagger$	0.6892	0.9188
y_8	$0.2068^\dagger$	0.6037	0.7883
y_9	$0.0001^\dagger$	0.6926	0.6028
y_{11}	0.9265	$0.3415^\dagger$	1.0000
y_{12}	0.9676	$0.2452^\dagger$	0.9325
y_{13}	1	$0.0002^\dagger$	0.9323
y_{14}	1	$0.0002^\dagger$	0.9903
y_{15}	1	$0.0002^\dagger$	0.8999
y_{16}	1	$0.0005^\dagger$	0.5707
y_{17}	0.9220	$0.0532^\dagger$	0.5437
y_{18}	1	$0.0004^\dagger$	0.6961
y_{19}	1	$0.3545^\dagger$	0.8374
y_{21}	0.9631	0.5002	$0.3174^\dagger$
y_{22}	0.7952	0.4122	$0.0577^\dagger$
y_{23}	0.7215	0.4089	$0.0936^\dagger$
y_{24}	0.8499	0.4456	$0.0805^\dagger$
y_{25}	0.7252	0.5928	$0.3962^\dagger$
y_{26}	$0.6828^\dagger$	0.7127	0.8827
y_{27}	0.5553	0.5818	$0.4682^\dagger$
y_{28}	0.2650	0.6801	$0.1699^\dagger$
y_{29}	$0.0391^\dagger$	0.6102	0.1470

TABLE III

DISTANCE BETWEEN DATA AND MODELS (IN THE $\mathcal{H}_\infty$ NORM). IN EACH ROW $\dagger$ DENOTES THE MODEL WHOSE OUTPUT BEST MATCHES THE GIVEN SEQUENCE

Fig. 21. Mapping manifolds between 2 sensors used to recreate an occluded person.

VIII. CONCLUSIONS

Arguably, one of the hardest challenges entailed in exploiting actionable information sparsely encoded in high volume data streams is the development of scalable, tractable methods capable of dealing with the extremely large volume of data [11]. Recent work (manifold embedding [22], compressive sensing, [3], [12], [4]) have led to substantial progress in addressing this issue. However, these methods stop short of fully exploiting the gap between data dimensionality and the rank of the *dynamical* system underlying the data record.

As shown in this paper, the use of dynamic models as an information encoding paradigm, can lead to both, substantial dimensionality reduction and computationally attractive algorithms for data extraction/interpretation. Dynamic structures can be tractably discovered from the data in a way which leverages their inherently lower dimensionality. One key feature is the ability of dynamic representations to produce quantifiable measures of uncertainty as *provable error bounds* on the validity of the data interpretation suggested by the model. Another is their relative computational simplicity: in many cases postulating the existence of such a model and associated invariants (e.g., model order) is enough to develop computationally attractive, robust solutions to problems such as segmentation, interpolation and event detection. We believe that these techniques hold the key to render practical several applications, ranging from self-aware environments to automatic discovery of co-regulated genes, that are currently at the proof–of–concept stage, and where the major roadblock is precisely the lack of techniques to robustly handle the extremely high volume of (often relatively low quality) data.

IX. ACKNOWLEDGMENTS

selecting as representative of each class the model closest to its center (e.g., the one solving $\min_i \max_j \|S_i - S_j\|_\infty$). Note that nearest neighbor classification using this metric can successfully recognize 25 out of the 27 sequences under consideration; it only confuses 2 sequences, (y_{26} and y_{29}, belonging to persons A and C walking up a staircase) as walking sequences. The failure is due to the fact that in these instances the experimental data record is too short to disambiguate between activities.

In the case of distributed data sources, the high costs (both in bandwidth and computational cost) entailed in sharing information can be avoided by (i) associating to each source a set of intrinsic coordinates on a low dimensional manifold, and (ii) using robust identification techniques [6] to identify dynamic models for the mappings between the projections of the different local data sources (e.g., sensors) onto the respective manifolds (see Figure 21). Then, only these low dimensional projections need to be exchanged between nodes, and each node can reconstruct the data observed by other nodes, simply by applying the interconnecting models. Figure 21 shows an application of these ideas to the problem of tracking and disambiguating two virtually identical targets.
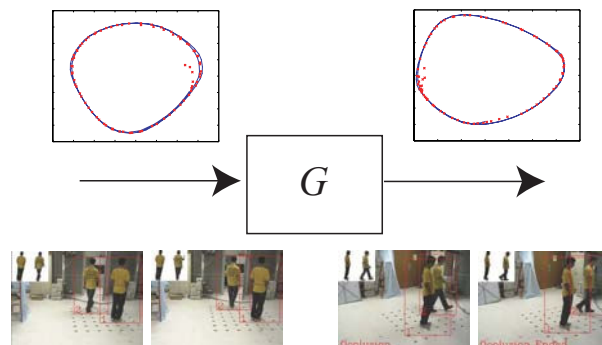

REFERENCES

[1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[2] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Trans. on Automatic Control*, 50(10):1567–1580, Oct. 2005.

[3] E. Candes, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, Aug 2006.

[4] L. Carin, D. Liu, and B. Guo. In situ compressive sensing. *Inverse Problems*, 24(1):015023–1–015023–23, Feb 2008.

[5] A. B. Chan and N. Vasconcelos. Mixtures of dynamic textures. In *ICCV05*, volume 1, pages 641–647, 2005.

[6] J. Chen and G. Gu. *Control Oriented System Identification, An $\mathcal{H}_\infty$ Approach*. John Wiley, New York, 2000.

[7] L. Cooper, J. Liu, and K. Huang. Spatial segmentation of temporal texture using mixture linear models. In *WDV05*, pages 142–150, 2005.

[8] J. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *IJCV*, 29(3):159–179, 1998.

[9] T. Ding, M. Sznaier, and O. Camps. Rank minimization approach to video inpainting. In *Proc. 2007 ICCV*, Oct. 2007.

[10] T. Ding, M. Sznaier, and O. Camps. Fast track matching and event detection. In *IEEE CVPR*, June 2008.

[11] D. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. In *AMS Conf. on Math Challenges of the 21st Century*, 2000. Plenary presentation.

[12] D. Donoho. Compressed sensing. *IEEE Trans Info Theory*, 52(4):1289–13006, April 2006.

[13] D. L. Donoho and C. E. Grimes. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. In *Proc. Nat. Academy of Arts and Sciences*, vol 100, pp. 5591–5596, 2003.

[14] G. Doretto, A. Chiuso, Y.N. Wu, and S. Soatto. Dynamic textures. *IJCV*, 51(2):91–109, February 2003.

[15] O. D. Faugeras, Q. T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, 2001.

[16] M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proc. 2001 ACC*, vol. 6, pp. 4734–4739.

[17] M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proc. 2003 ACC*, vol. 3, pp. 2156–2162.

[18] U. Gargi, R. Kasturi, and S. H. Strayer. Performance characterization of video-shot-change detection methods. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(1):1–13, 2000.

[19] A. Ghoreyshi and R. Vidal. Segmenting dynamic textures with ising descriptors, arx models and level sets. In *WDV06*, pp. 127–141, 2006.

[20] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[21] Y.S. Hung and W.K. Tang. Projective reconstruction from multiple views with minimization of 2d reprojection error. *International Journal of Computer Vision*, 66(3):305–317, 2006.

[22] J. A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer, 2007.

[23] H. Lim, O. I. Camps, M. Sznaier, and V. Morariu. Dynamic appearance modelling for human tracking. In *IEEE CVPR*, pp. 751–757, 2006.

[24] L. Lu and R. Vidal. Combined central and subspace clustering for computer vision applications. In *ICML*, pp. 593–600, 2006.

[25] R. Lublinerman, M. Sznaier, and O. I. Camps. Dynamics based robust motion segmentation. In *IEEE CVPR*, pp. 1176–1184, 2006.

[26] Y. Ma and R. Vidal. A closed form solution to the identification of hybrid arx models via the identification of algebraic varieties. In *Hybrid Systems Computation and Control*, pp. 449–465, 2005.

[27] S. Mahamud and M. Hebert. Iterative projective reconstruction from multiple views. In *IEEE CVPR*, vol. 2, pp. 430–437, 2000.

[28] M. C. Mazzaro, M. Sznaier, and O. Camps. A model (in)validation approach to gait classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(11):1820–1825, 2005.

[29] Ozay N, M. Sznaier, and O. Camps. Sequential sparsification for change detection. In *IEEE CVPR*, June 2008.

[30] M. Osian and L.J. Van Gool. Video shot characterization. *Machine Vision and Applications*, 15(3):172–177, July 2004.

[31] N. Ozay, M. Sznaier, C. Lagoa, and O Camps. A sparsification approach to set membership identification of a class of affine hybrid systems. In *IEEE CDC*, pp. 123–130, Dec. 2008.

[32] N. Petrovic, A. Ivanovic, and N. Jojic. Recursive estimation of generative models of video. In *IEEE CVPR*, pages I: 79–86, 2006.

[33] W.M. Rand. Objective criteria for the evaluation of clustering methods. *J. American Statistical Association*, 66:846–850, 1971.

[34] R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.

[35] Lawrence K. Saul and Sam T. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J. on Machine Learning Research*, 4:119–155, 2003.

[36] M. Sznaier, M. Ayazoglu, and O. Camps. sing dynamics to recover euclidian 3–dimensional structure from 2–dimensional perspective projections. In *Proc. 48th IEEE CDC*, pp. 2414–2419, Dec. 2009.

[37] M. Sznaier, C. Lagoa, and W. Ma. Risk adjusted identification of a class of nonlinear systems. In *Proc. 46 IEEE Conf. Dec. Control*, pp. 5117–5122, Dec. 2007.

[38] M. Sznaier, C. Lagoa, and M. C. Mazzaro. An algorithm for sampling balls in $\mathcal{H}_\infty$ with applications to risk–adjusted performance analysis and model (in)validation. *IEEE Trans. Autom. Control*, 50(3):410–416, 2005.

[39] M. Sznaier, C. Mazzaro, and O. I. Camps. Open-loop worst-case identification of nonschur plants. *Automatica*, 39(6):1019–1025, 2003.

[40] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[41] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. of Computer Vision*, 9(2):137–154, 1992.

[42] L. Vandenberghe and S. P. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[43] R. Vidal and R. Hartley. Motion segmentation with missing data using powerfactorization and gpca. In *IEEE CVPR*, vol. 2, pp. 310–316, 2004.

[44] R. Vidal, Y. Ma, S. Soatto, and S. Sastry. Two–view multibody structure from motion. *Int. J. of Computer Vision*, 68(1):7–25, 2006.

[45] Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised learning of image manifolds by semidefinite programming. In *IEEE CVPR*, pp. 988–995, 2004.

[46] Kilian Q. Weinberger, Fei Sha, and Lawrence K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proc. 2004 ICML*.

[47] Jing Xiao, Jinxiang Chai, and Takeo Kanade. A closed-form solution to non-rigid shape and motion recovery. In *Proc. 2004 ECCV*.

[48] J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin, and B. Zhang. A formal study of shot boundary detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(2):168–186, February 2007.

[49] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette, and U. Alon. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods*, 3(623 – 628), 2006.

[50] L. Zelnik-Manor and M. Irani. Degeneracies, dependencies and their implications in multi-body and multi-sequence factorization. In *IEEE CVPR*, pp. 287–293, 2003.