[DESCRIPTION]

[Title of Invention]

METHOD AND SYSTEM FOR DETERMINING PEDESTRIAN FLOWS

[Technical Field]

[0001]

This invention relates generally to computer vision, an more particularly segmenting pedestrians in flows as depicted in videos.

[Background Art]

[0002]

The segmentation of motion flows within dense crowds of pedestrians in videos is an essential tool in crowd safety and crowd control applications. Videos of crowded scenes can exhibit complex crowd behaviors even under normal situations. For example, crowd flows in large congested areas, such as train stations, can initially appear chaotic. However, it is often the case that low dimensional dynamical structures exist in the flow that is desirable to identify and segment from unstructured flows. Moreover, the automatic segmentation of independent crowd flows aids in the monitoring and prediction of hazardous situations in crowded environments.

[0003]

Of particular interest is the detection and estimation of crowd flows using motion information extracted from the videos. Motion vectors can be determined using optical flow estimation applied to texture, that is pixel intensities, in videos, or the motion vectors can be extracted directly from a bitstream. The bitstream can be encoded using any of the well known coding standards, e.g., MPEG, H.264, HEVC, etc.

[0004]

Considering pedestrians in a crowded scene as particles in a flow, the motion vectors of a video frame correspond to observations of the velocities of particles at a time instance in the flow. Processing motion vectors, instead of the video texture, protects the privacy of individuals observed in surveillance videos.

where $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ are system parameters determining maximal velocities in the horizontal and vertical directions, and $\bar{\rho}$ is a maximal density in the scene.
[0009]

The solution to (1) results in a crowd density map ρ, and the velocity fields $(\mathbf{u}, \mathbf{v})$ for all (x, y, t) that satisfy given initial and boundary conditions. Although the dimensionality of the differential equations governing the evolution of ρ and $(\mathbf{u}, \mathbf{v})$ can be infinite, it is often the case that the flows exhibit low dimensional behavior.
[0010]

A low dimensional state variable at time t is $\mathbf{x}(t)$, for which an observable vector $\mathbf{y}(t) = G(\mathbf{x}(t))$ corresponds to stacking of the density and velocity fields for all positions x and y at time t. The function G is a mapping from the low dimensional manifold on which $\mathbf{x}$ evolves to a space of observables. Then, the solution to (1) determines the transient response and stability of the corresponding dynamical system generally characterized by

$$\dot{\mathbf{x}}(t) = F(\mathbf{x}(t)), \quad (3)$$

where $F(\cdot)$ is some mapping in the low dimensional manifold on which the dynamical system evolves. For discrete time systems, the dynamical system evolution is characterized by

$$\mathbf{x}_{k+1} = F(\mathbf{x}_k), \quad (4)$$

where k is a time index.
[0011]

The Koopman Operator and Dynamic Mode Decomposition

The Koopman operator is a linear operator K that satisfies

$$\begin{aligned} & G(F(\mathbf{x}_k)) = KG(\mathbf{x}_k) \\ \Leftrightarrow\ & \mathbf{y}_{k+1} = K\mathbf{y}_k. \end{aligned} \quad (5)$$

[0012]

Although this dynamical system is nonlinear and evolves a finite dimensional manifold, the Koopman operator is linear and infinitely dimensional. Spectral analysis of the Koopman operator can be used to decompose the flow

in terms of Koopman modes and associated Koopman eigenvalues that determine the temporal behavior of the corresponding Koopman mode.

[0013]

A dynamic mode decomposition (DMD) can be used to estimate the Koopman modes. The DMD has been used in fluid dynamics as a data-driven and equation-free method for identifying system dynamics. Consider data matrices

$$\mathbf{Y}_1 = \begin{bmatrix} | & | & & | \\ \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{m-1} \\ | & | & & | \end{bmatrix}; \ \mathbf{Y}_2 = \begin{bmatrix} | & | & & | \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_m \\ | & | & & | \end{bmatrix}. \quad (6)$$

[0014]

The DMD determine the best fit matrix $\mathbf{K}$ that satisfies the relation

$$\mathbf{Y}_2 \approx \mathbf{K}\mathbf{Y}_1. \quad (7)$$

[0015]

The eigenvectors and eigenvalues of $\mathbf{K}$ approximate the Koopman modes and Koopman eigenvalues. Herein, the terms Koopman modes and DMD modes are used interchangeably.

[Summary of Invention]

[0016]

The embodiments of the invention provide a method and system for segmenting pedestrian flows in a scene based on motion vectors in an video acquired of a scene by a video camera. The motion vectors are treated as observables of an underlying dynamical system that defines the flows in the scene as organized in an input data matrix. The method determines a low rank Koopman operator that characterizes the dynamical system using the motion vectors. Then, the spectrum of the Koopman operator is analyzed to segment independent flows in the scene.

[0017]

In one embodiment of the invention, a low rank Koopman operator is determined by solving a nuclear norm minimization problem constrained by a one step prediction equation of a dynamic mode decomposition (DMD). The

effect of noise can be minimized by selecting an appropriate regularizer for the distribution of the noise. For example, a Frobenius norm or Euclidean norm regularizer can be used for Gaussian distributed noise. A sum of absolute values or one-norm regularizer can be used for sparse impulsive noise. In the case both types of noise exist, two regularizers can be used during the optimization.

[0018]

One embodiment of the invention performs a singular value decomposition (SVD) of the Koopman operator. Row space (left) singular vectors corresponding to the nonzero singular values are then used to cluster blocks in the video as segments corresponding to independent flows. In another embodiment, the eigenvectors of the Koopman operator are determined, and the input data matrix is projected onto the eigenvectors. The eigenvectors that have magnitude projection coefficients greater than a threshold are then used to cluster blocks and segment the independent flows.

[0019]

Another embodiment of the invention characterizes the segmented pedestrian flows by identifying a dynamical regime according to eigenvalues of the eigenvectors that describe the dynamics for each flow. The segmented flow data are projected onto the eigenvectors of the Koopman operator and the eigenvalues that are associated with the eigenvectors that have magnitude projection coefficients greater than a threshold are used to define the dynamical regime. The dynamical regime characterizes the type of the flow as, for example, stable or unstable.

[Brief Description of Drawings]

[0020]

[Fig. 1]

Fig. 1 is a schematic of a system and method for segmenting pedestrians flows in a video acquired of a scene according to the embodiments of the invention;

[Fig. 2]

Fig. 2 is a flow diagram of a method for determining Koopman modes according to embodiments of the invention;

[Fig. 3]

Fig. 3 is a flow diagram of a method for segmenting independent flows using the Koopman modes according to embodiments of the invention; and

[Fig. 4]

Fig. 4 is a flow diagram of a method for characterizing independent segmented pedestrian flows.

[Description of Embodiments]

[0021]

As shown in Fig. 1, the embodiments of our invention provide a system and method for segmenting pedestrians flows in a video acquired of a scene. The embodiments can use motion vectors extracted from an encoded bitstream, although the motion vectors can be obtained by other means. The motion vectors are treated as observables of an underlying dynamical system composed of several independent dynamics that characterize the flows in the scene. The method determines a low rank Koopman operator that characterizes the dynamical system using the motion vectors. The spectrum of the Koopman operator is then analyzed using singular value decomposition (SVD) or eigenvalue decomposition. The independent crowd flows are determined by clustering the dominant singular vectors or the dominant eigenvectors, depending on the decomposition that is used.

[0022]

System and Method Overview

As shown in Fig. 1, a video 101 is acquired of pedestrians in a scene 102 with a video camera 103. The video includes a set of frames. Each frame includes a set of blocks that typically encode motion vectors. Each frame k 104 is processed by the method operating in a processor 100 connected to memory and the camera by input/output buses as known in the art. The video is in a form of an encoded bitstream.

[0023]

As an advantage, an encoded bitstream is usually compressed by at least two orders of magnitude or more. Hence, the bitstream takes much less storage and time to process. In addition, the method can operate at the block level so that the bitstream does not need to be decoded to obtain pixels. However, the method can operate in a camera and obtain the motion vectors by other means.

[0024]

The method extracts 110 motion vectors 105 from the video, e.g., blocks in the bitstream, and constructs a data matrix 111 by stacking the motion vectors as described below in greater detail.

[0025]

A low-rank Koopman operator 121 is determined 120 from the data matrix 111. The spectrum of a low-rank Koopman operator is analyzed to determine dominant Koopman modes 131. The Koopman modes can then be used to segment 300 blocks in frames of the video to obtain independent flows 109.

[0026]

Fig. 2 shows the steps 200 used to analyze the spectrum and determine the Koopman modes 131. The steps include the SVD 210 to obtain singular vectors 211 or an eigenvalue decomposition to determine eigenvectors, or both. Step 240 determines projection coefficients, which are then used to select 250 eigenvectors. The singular vectors or selected eigenvectors, or both, can then be used to determine the Koopman modes 131.

[0027]

Fig. 3 shows the steps 200 of segmenting the pixels, which take as input the Koopman modes 131. Motion vectors are clustered using k-means clustering 310 or graph spectral clustering 320 to obtain the independent flows 109.

[0028]

Fig. 4 shows the steps 400 of characterizing each segmented flow by determining a dynamical regime of the flow. The dynamical regimes essentially characterizes the flows. The steps take as input the Koopman eigenvectors 221, the data matrix 111 and the segmented flows 109. A subset of rows 401 of the

data matrix 111 are extracted using the segmented flows 109. The subset of rows 401 are projected 410 onto the Koopman eigenvectors 221 to determine projection coefficients 411. An eigenvalue selection step 420 chooses the eigenvalues that correspond to the eigenvectors that have projection coefficients with magnitudes larger than a threshold. The dynamical regime of the flow or flow characteristics 421 is identified by the selected eigenvalues.

[0029]

We now describe the details of the above steps.

[0030]

Motion Vectors and Observable data matrix

The horizontal and vertical motion vectors of all spatial blocks in a video frame k 104 are $\mathbf{u}_k$ and $\mathbf{v}_k$, respectively. We construct the data matrix $\mathbf{z}_k = [\mathbf{u}_k, \mathbf{v}_k]$ 111 by stacking the horizontal and vertical motion vectors in the data matrix 111. Then, the temporal evolution of the motion vectors can be modeled, using the Koopman operator $\mathbf{K}$, by fitting the data to a one step prediction model

$$\mathbf{z}_k = \mathbf{K}\mathbf{z}_{k-1}. \qquad (8)$$

[0031]

Notice that (8) ensures that the same operator $\mathbf{K}$ models the temporal evolution of the motions vectors $\mathbf{u}_k$ and $\mathbf{v}_k$.

[0032]

If we have $m+1$ observations of motion vectors from $m+1$ video frames, then we can determine the operator $\mathbf{K}$ using a dynamic mode decomposition (DMD) by forming the data matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ similar to (6) and determining the operator $\mathbf{K}_{\text{ls}}$ that achieves the least squares fit, i.e.,

$$\begin{aligned} \mathbf{K}_{\text{ls}} &= \operatorname{argmin}_K \tfrac{1}{2} \| \mathbf{Z}_2 - K\mathbf{Z}_1 \|_F^2 \\ &= \mathbf{Z}_2\mathbf{Z}_1^{\dagger}, \end{aligned} \qquad (9)$$

where the superscript † indicates the Moore-Penrose pseudo-inverse of a matrix.

[0033]

If the motion vectors correspond to the true motion of pedestrians in the video, then the operator $\mathbf{K}_{ls}$ characterizes the full dynamics of the flow in the system. However, compressed domain motion vectors are generally very noisy and often contain motion vectors that do not correspond to the true motion. The noise depends on the specific rate-distortion optimization by the video encoder.

[0034]

Low-rank DMD

A noise-free velocity observations $\mathbf{y}_k$ measure true motion of pedestrians in the scene. and the motion vectors $\mathbf{z}_k$ correspond to noisy observations, such that

$$\mathbf{z}_k + \mathbf{e}_k = \mathbf{y}_k, \qquad (10)$$

where $\mathbf{e}_k$ is additive noise. Let $\mathbf{E}_1, \mathbf{E}_2$ and $\mathbf{Y}_1, \mathbf{Y}_2$ be as in (6) for $\mathbf{e}_k$ and $\mathbf{y}_k$, respectively. Then, the Koopman operator corresponding to the noise-free dynamics is given by a total least squares constraint

$$\begin{aligned} & \mathbf{Y}_2 && = && \mathbf{K}\mathbf{Y}_1 \\ \Leftrightarrow \quad & \mathbf{Z}_2 + \mathbf{E}2 && = && \mathbf{K}(\mathbf{Z}_1 + \mathbf{E}_1). \end{aligned} \qquad (11)$$

[0035]

However, the problem (11) is non-convex and ill posed because only $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are observed, and $\mathbf{K}$, $\mathbf{E}_1$ and $\mathbf{E}_2$ are unknowns. To remedy this situation, we invoke the following prior conditions in our problem. Highly complex systems can be accurately modeled by low-order dynamics. This translates to the prior condition that the operator $\mathbf{K}$ modeling the evolution of a noise-free system has a low rank.

[0036]

The second prior condition derives from the definition of the matrices $\mathbf{E}_1$ and $\mathbf{E}_2$. If we denote by $\mathbf{I}_{m-1}$ the identity matrix of size $(m-1) \times (m-1)$, and let $\mathbb{I}^0 = [\mathbf{0}|\mathbf{I}_{m-1}]^T$ and $\mathbb{I}_0 = [\mathbf{I}_{m-1}|\mathbf{0}]^T$ be selection operators that respectively remove the first column and the last column of a matrix with m columns, then $\mathbf{E}_1$ and $\mathbf{E}_2$ satisfy a relation

$$\mathbf{E}_1\mathbb{I}^0 = \mathbf{E}_2\mathbb{I}_0. \qquad (12)$$

[0037]

$$\begin{aligned} \min_{\mathbf{L},\mathbf{R},\mathbf{E}_1,\mathbf{E}_2,\mathbf{S}_1,\mathbf{S}_2} \quad & \frac{1}{2}(\| \mathbf{L} \|_F^2 + \| \mathbf{R} \|_F^2) + \frac{\gamma}{2}(\| \mathbf{E}_2 \|_F^2 + \| \mathbf{e}_0 \|_2^2) + \lambda \| \mathbf{S}_2 \|_1, \\ \textit{subject to} \quad & \mathbf{Z}_2 + \mathbf{E}_2 + \mathbf{S}_2 = \mathbf{L}\mathbf{R}^T(\mathbf{Z}_1 + \mathbf{E}_1 + \mathbf{S}_1), \\ & \mathbf{E}_1 \mathbb{I}^0 = \mathbf{E}_2 \mathbb{I}_0, \\ & \mathbf{S}_1 \mathbb{I}^0 = \mathbf{S}_2 \mathbb{I}_0, \end{aligned} \tag{16}$$

where the variables $\mathbf{S}_1$ and $\mathbf{S}_2$ represent the outlier noise in the data, and γ and λ are regularization parameters. The solution to (16) can also be obtained using the ADMM procedure.

[0041]

Flow Segmentation

In order to segment 300 different flows having independent dynamics, we evaluate the spectrum of the estimated low-rank Koopman operator $\mathbf{K} = \mathbf{L}\mathbf{R}^T$. Here, we consider different scenarios for performing the segmentation.

[0042]

Segmetation by Singular Value Decomposition

In a first approach, we can perform the SVD 210 of the Koopman opertor $\mathbf{K}$ to extract row space singular vectors $\mathbf{U}_r$ 211, such that $\mathbf{K} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^T$. The matrix $\mathbf{U}_r$ is an $N \times r$ matrix, where r is the rank of the operator $\mathbf{K}$. We treat the rows of $\mathbf{U}_r$ as feature vectors for every pixel in the video frame to which we apply *k*-means clustering 310, as shown in Fig. 3, to segment the pixels into a predetermined number of clusters n. Then, the clustered pixels can be used to determine the spatial segmentation of independent flows 109 in the scene 102 as observed in the video 101.

[0043]

Alternatively, we can apply spectral clustering to the rows of the matrix $\mathbf{U}_r$. Here, we normalize the rows of $\mathbf{U}_r$ to produce a matrix $\bar{\mathbf{U}}_r$ and construct an affinity matrix $\mathbf{A} = |\bar{\mathbf{U}}_r \bar{\mathbf{U}}_r^T|^2$, where the exponent is applied element-wise in this case. We then apply the following steps to determine the principal eigenvectors of the corresponding graph Laplacian.

[CLAIMS]

[Claim 1]

A method for determining flows, comprising steps:

acquiring a video of the flows with a camera, wherein the flows are pedestrians in a scene, wherein the video includes a set of frames;

extracting motion vectors from each frame in the set;

constructing a data matrix from the motion vectors in the set of frames;

determining a low rank Koopman operator from the data matrix;

analyzing a spectrum of the low rank Koopman operator to determine a set of Koopman modes; and

segmenting the frames into the flows according to a clustering of the Koopman modes, wherein the flows are independent pedestrian flows, and wherein the steps are performed in a processor.

[Claim 2]

The method of claim 1, wherein the motion vectors include Gaussian distributed noise, and further comprising:

determining the low rank Koopman operator by solving a factorized nuclear norm minimization problem with least squares constraints or total least squares constraints; and

minimizing the Gaussian distributed noise by a Frobenius norm regulizer.

[Claim 3]

The method of claim 1, wherein the motion vectors contain sparse outlier noise, and further comprising:

minimizing the sparse outlier noise by a one-norm regulizer.

[Claim 4]

The method of claim 1, further comprising:

performing a singular value decomposition (SVD) of the low rank Koopman operator to produce singular vectors; and

applying a k-means clustering or a graph spectral clustering to row space singular vectors of the low rank Koopman operator.

[Claim 5]