A Tube-and-Droplet-Based Approach for Representing and Analyzing Motion Trajectories

Weiyao Lin, *Senior Member, IEEE*, Yang Zhou, Hongteng Xu, Junchi Yan, Mingliang Xu, Jianxin Wu, and Zicheng Liu, *Fellow, IEEE*

Abstract—Trajectory analysis is essential in many applications. In this paper, we address the problem of representing motion trajectories in a highly informative way, and consequently utilize it for analyzing trajectories. Our approach first leverages the complete information from given trajectories to construct a thermal transfer field which provides a context-rich way to describe the global motion pattern in a scene. Then, a 3D tube is derived which depicts an input trajectory by integrating its surrounding motion patterns contained in the thermal transfer field. The 3D tube effectively: 1) maintains the movement information of a trajectory, 2) embeds the complete contextual motion pattern around a trajectory, 3) visualizes information about a trajectory in a clear and unified way. We further introduce a droplet-based process. It derives a droplet vector from a 3D tube, so as to characterize the high-dimensional 3D tube information in a simple but effective way. Finally, we apply our tube-and-droplet representation to trajectory analysis applications including trajectory clustering, trajectory classification & abnormality detection, and 3D action recognition. Experimental comparisons with state-of-the-art algorithms demonstrate the effectiveness of our approach.

Index Terms—Trajectory representation, trajectory analysis, 3D tube, abnormality detection, 3D action recognition

✦

1 INTRODUCTION

MOTION information, which reflects the temporal variation of visual contents, is essential in depicting the semantic contents in videos. As the motion information of many semantic contents is described by motion trajectories, trajectory analysis is of considerable importance to many applications including video surveillance, object behavior analysis, and video retrieval [1], [2], [3], [4], [5]. Formally, trajectory analysis can be defined as the problem of deciding the class of one or more input trajectories according to their shapes and motion routes [4], [6], [7], [8].

A motion trajectory is in general obtained by tracking an object over frames and linking object positions into a position sequence [1], [2], [9]. Although trajectories contain detailed information of object movements, reliable trajectory analysis remains challenging due to the uncertain nature of object motion and the ambiguity from similar motion patterns.

One major challenge for trajectory analysis is to differentiate trajectory classes with only subtle differences. For example, Fig. 1a shows three trajectory classes CP_1, CP_2, and CP_3, where CP_1 and CP_2 include vehicle trajectories following two adjacent leftward street lanes and CP_3 includes vehicle trajectories following a left-turn street lane. Since trajectories in CP_1 and CP_2 are similar in both motion direction and location, the original position sequence representation is insufficient to differentiate them. This necessitates the development of more informative motion trajectory representation. However, most existing trajectory representation methods [3], [4], [10], [11] focus on performing transformation or parameterization on the original position sequence, while the problem of more informative representation is not well addressed. Although some trajectory-modeling or local-modeling methods [4], [6], [7], [12], [13] increase the informativeness of trajectories by including the contextual information among multiple trajectories, they only model partial contextual information from trajectories with similar patterns or trajectories in the same class. Thus, they still have limitations when differentiating ambiguous trajectories, such as trajectories near the boundary of similar trajectory classes (e.g., trajectories A and B in Fig. 1a).

We argue that due to the stable constraint from a scene, the *complete* contextual motion pattern around a trajectory provides an important cue for trajectory depiction. By *complete*, we refer to the contextual motion information from *all* given trajectories in the neighborhood of an input trajectory. For example, Fig. 1b shows the complete contextual motion patterns of two easily-confused trajectories A and B which are extracted from two similar trajectory classes CP_1 and CP_2. If we look at the contextual motion information in the neighborhood of trajectory A, there is a strong left and

----

- W. Lin and Y. Zhou are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.
  E-mail: {wylin, zhz128ly}@sjtu.edu.cn.
- H. Xu is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332.
  E-mail: hxu42@gatech.edu.
- J. Yan is with IBM Research, Shanghai 200240, China.
  E-mail: yanjc@cn.ibm.com.
- M. Xu is with the Center for Interdisciplinary Information Science Research, Zhengzhou University, Zhengzhou 100044, China.
  E-mail: iexumingliang@zzu.edu.cn.
- J. Wu is with National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China.
  E-mail: wujx2001@nju.edu.cn.
- Z. Liu is with Microsoft Research, Redmond, WA 98052.
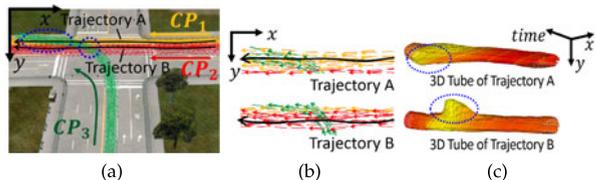 E-mail: zliu@microsoft.com.

Fig. 1. (a) An example of ambiguous trajectories: The orange, red, and green curves labeled by CP_1-CP_3 indicate three trajectory classes. The black curves labeled by *Trajectory A* and *B* are two trajectories from CP_1 and CP_2, respectively. (b) Complete contextual motion patterns of trajectories A and B in (a): Contextual motion patterns are described by the motion flows of all trajectory points in the neighborhood of A and B (Note: We use color to differentiate motion flows from different trajectory classes only for a clearer illustration. In our approach, we do not differentiate motion flows' classes and directly use *all* motion flows in the neighborhood of an input trajectory to model its contextual motion pattern). (c) Use our 3D tube to model A and B's complete contextual motion patterns. (Best viewed in color.)

slightly upward motion pattern (in green color) towards the end of A. In contrast, if we look at the contextual motion information in the neighborhood of trajectory B, there is an obvious upward motion pattern in the middle of B. Thus, the ambiguity between trajectories is expected to be reduced if the above-mentioned contextual motion information is properly modeled (cf. Fig. 1c).

1.1 Our Work

Based on this intuition, we develop a novel framework which utilizes the global motion trends in a scene to describe a trajectory. Specifically, for each point in a trajectory, we derive the dominant scene motion pattern around the point and utilize it to depict the point's complete contextual motion pattern. By integrating the dominant scene motion patterns for all points in a trajectory, we are able to describe a trajectory in a highly informative way and make fine distinctions among similar trajectories. The framework of our approach is shown in Fig. 2.

Given a set of trajectories, thermal transfer fields are first constructed to describe the global motion pattern in the scene (cf. module 'constructing thermal transfer fields' in Fig. 2). Then, for an input trajectory, we expand it over the time domain to construct a 3D spatio-temporal curve, and derive an equipotential line for each point in this curve (cf. module 'deriving equipotential lines' in Fig. 2). The equipotential lines are decided by the locations of spatio-temporal curve points and their surrounding dominant scene motion patterns defined in thermal transfer fields. In this way, the complete contextual motion pattern can be captured.

After obtaining equipotential lines, a 3D tube is constructed which concatenates equipotential lines according to the temporal order of spatio-temporal curve points (cf. module 'constructing 3D tubes' in Fig. 2). This 3D tube is able to depict an input trajectory in a highly informative way, where the movement of a trajectory and the contextual motion pattern around a trajectory are effectively captured by the route and shape of the 3D tube.

Finally, a droplet-based process is applied which 'injects' water in one end of a 3D tube and achieves a water droplet flowed out from the other end. This water droplet is further sampled to achieve a low-dimensional droplet vector to characterize the 3D tube shape (cf. module 'droplet-based process for feature extraction' in Fig. 2). Since different trajectories are depicted by 3D tubes with different shapes, by suitably modeling the water flow process, the droplet vector can precisely catch the unique characteristics of a 3D tube. The droplet vector will serve as the final trajectory representation format and is applied in trajectory analysis.

In summary, our contributions are three folds.

- We construct a thermal transfer field to describe the global motion pattern in a scene, derive equipotential lines to capture the contextual motion information of trajectory points, and introduce a 3D tube by concatenating equipotential lines to represent a motion trajectory. These components establish a novel framework for addressing the informative trajectory representation problem.
- Under this framework, we develop a droplet-based process which derives a simple but effective low-dimensional droplet vector to characterize the high-dimensional information in a 3D tube. The derived droplet vector can not only capture the characteristics of a 3D tube, but also suitably reduce the disturbance from trajectory noises.
- We investigate our tube-and-droplet representation to various trajectory analysis applications including trajectory clustering, trajectory classification & abnormality detection, and 3D action recognition, and achieve the state-of-the-art performance.

2 RELATED WORKS

2.1 Trajectory Representation and Modeling

Properly representing motion trajectories is crucial to trajectory analysis. Many algorithms [1], [5], [6], [9], [14] have been proposed for trajectory representation. Most of them
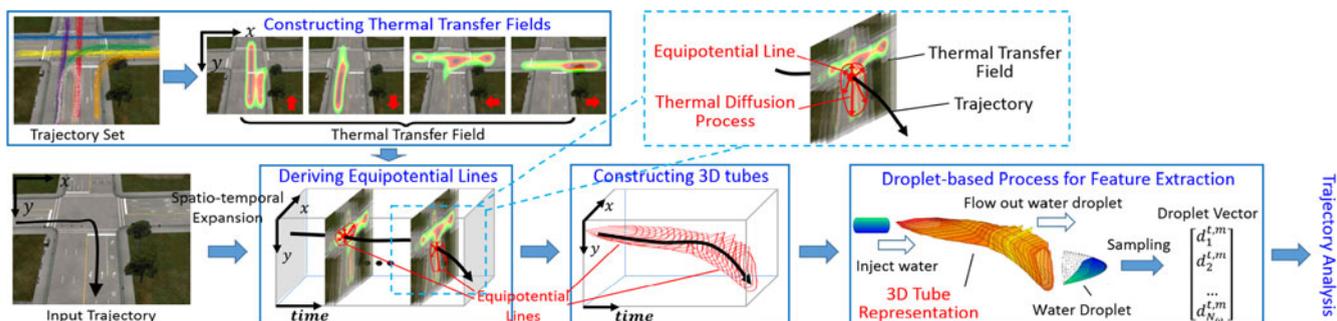

Fig. 2. The framework of the proposed approach. It constructs a scene-specific thermal transfer field via trajectory training data; for a test trajectory sample, it builds a 3D tube based on the constructed thermal transfer field, and generates a feature vector via a droplet process. The obtained feature vector is applicable to various trajectory analysis applications. (Best viewed in color).

[3], [10], [11], [14] aim to find suitable parameter sets to describe trajectories. Discrete Fourier transform (DFT) coefficients [10] and polynomial curve fitting [3] are two examples. Rao et al. [14] extracted dynamic instants and used them as the key points to represent the spatio-temporal curvature of trajectories. These methods focus more on the effective representation of a trajectory's position sequence, while the issue of more *informative* representation is not addressed. Although some works [1], [2], [3], [12], [14] increase the informativeness of trajectory representation by introducing additional information such as motion velocity, temporal order, or object size, the added information is still restricted within a single trajectory and cannot be used to discriminate trajectories with similar patterns.

Trajectory-modeling methods that include the contextual information of multiple trajectories are also proposed [4], [6], [7]. These methods construct probability models for each trajectory class and utilize them to guide the trajectory analysis process. Kim et al. [6] introduced Gaussian process regression flows to model the location and velocity probability for each trajectory class. Morris and Trivedi [7] clustered trajectories into spatial routes and encoded the spatio-temporal dynamics of each cluster by hidden Markov models (HMM). Hu et al. [4] built a time-sensitive Dirichlet process mixture model (tDPMM) to represent the spatio-temporal characteristics of each trajectory class. These methods only focus on modeling the contextual information inside each individual class, while ignoring information from external trajectory classes. Therefore, they still have limitations when differentiating ambiguous trajectories, such as trajectories near the boundary of similar trajectory classes.

Another line to integrate contextual information is the local-modeling methods [13], [15]. These methods aim to utilize dynamic models or topic models to describe trajectories' local motion patterns. Nascimento et al. [15] introduced low-level dynamic models to decompose trajectories into basic displacement patterns, and recognized trajectories based on the switch among these low-level models. Wang et al. [13] proposed a non-parametric topic model called Dual Hierarchical Dirichlet Processes (Dual-HDP), which constructs local semantic regions including similar trajectory segments and performs trajectory analysis accordingly. Although these models share the contextual information from multiple trajectory classes, only the contextual information from similar trajectory segments is considered. Therefore, they are less effective when trajectories have large variations or reliable local models cannot be constructed due to insufficient similar trajectory segments.

The proposed 3D tube representation differs with existing approaches in the following aspects:

- We model the *complete* contextual information around a trajectory, not only the contextual information from partial trajectories. This enables us to precisely catch the subtle changes among different trajectory classes.
- Most existing works handle trajectory analysis with complex probability models, which require sufficient trajectory data to construct reliable models. We establish a novel framework to model trajectories in a simple but effective way, which can work robustly under relatively small data size. Experimental evaluation

shows we can achieve state-of-the-art performance on benchmark datasets with this simple procedure.
- Existing methods focus on the abstract modeling of trajectory information where the modeled trajectory information cannot be easily visualized. Our 3D tube representation is able to visualize a variety of trajectory information, including spatial-temporal movements, contextual motion patterns, and possible motion directions, in a clear and unified way.

2.2 Handling High-Dimensional Representations

Since highly informative trajectory representation often leads to a complex and high-dimensional representation format, it is non-trivial to find suitable ways to handle this high-dimensional representation for trajectory analysis.

In [16], Euclidean distance and dynamic time wrapping (DTW) distance were utilized to measure the distance of two time-series trajectory sequences. Vlachos et al. [17] further introduced the longest common subsequence (LCSS) distance. While these methods can handle the original position sequence format of a trajectory, they are not suitable to process higher dimensional representations.

In order to handle higher-dimensional data representations, some dimension reduction approaches were developed [18], [19]. However, due to the large variation and ambiguity among trajectories, trajectory representations often have complex distributions. Thus, simply applying dimension reduction cannot achieve satisfactory results. Moreover, some advanced manifold approaches were also introduced [20], [21], [22]. In [20], Lui modeled high-dimensional inputs as high-order tensors. The similarity between inputs were measured by the intrinsic distance between tensors, which is estimated through manifold charting. Although these methods can achieve better results when processing high-dimensional trajectory representations, they have considerably high computation complexity. Thus they are difficult to be applied on large scale trajectory analysis.

Besides dimension reduction, other works [12], [23], [24] aim to develop proper distance metrics to measure the similarity between high-dimensional inputs. Lin et al. [12] introduced a surface matching distance to measure the similarity between high-dimensional surface shapes. Sangineto [23] and Gao et al. [24] developed advanced Hausdorff distance which treats each high-dimensional input as a set of points and estimates the similarity between inputs from the distance between point sets. These methods are dependent on the quality of high-dimensional inputs, and noisy inputs will adversely affect the performance of these methods.

Different from the previous methods, we develop a novel droplet-based process which simulates the physical *water flow* process and derives a low-dimensional droplet vector to characterize the high-dimensional 3D tube shape. This process has low complexity and can suitably reduce the disturbance from trajectory noises.

3 3D TUBE REPRESENTATION

In order to include the complete contextual motion information in trajectory representation, it is important to properly model and embed the global motion pattern information in a scene. More specifically, assuming that there are M

trajectories available from a scene, denoted as $\{\Gamma^m\}_{m=1}^M$. Each trajectory is represented as $\Gamma^m = \{\mathbf{p}_n^m\}_{n=1}^{\mathcal{L}^m}$, where $\mathbf{p}_n^m \in \mathbb{R}^2$ is the *position* of the nth point of the trajectory and $\mathcal{L}^m$ is the length of trajectory Γ^m. Accordingly, the *speed* of the point in the position is calculated as $\mathbf{u}_n^m = \frac{\mathbf{p}_{n+1}^m - \mathbf{p}_n^m}{\Delta t} \in \mathbb{R}^2$, where Δt is the sampling interval between adjacent points. We aim to find the scene's global motion pattern which best describes the motion trends provided by M given trajectories[1], and use this global motion pattern to derive the complete contextual motion patterns for positions along the route of an input trajectory. In this way, we are able to construct a 3D tube from these contextual motion patterns and obtain an informative representation for the input trajectory.

In this paper, we borrow the idea from thermal propagation [25], [26], [27] and introduce a trajectory model, which finds global motion pattern, derives contextual motion patterns, and constructs 3D tube representations with thermal propagation processes.

3.1 Constructing Thermal Transfer Fields

First, we model the global motion pattern in a scene based on the works in fluid thermal propagation [25], [26], [27]. Specifically, the aggregation of M given trajectories is modeled as a '*fluid*' in the scene, where each $\mathbf{p}_n^m$ is a sample of the fluid and the corresponding $\mathbf{u}_n^m$ refers to the movement of the fluid which results in the transfer of thermal energies in position $\mathbf{p}_n^m$.

According to thermal transmission theories [25], [26], the thermal diffusion result of the entire fluid is affected by a scene-related *thermal transfer field* which decides the thermal propagation strengths at different positions and in different directions. Therefore, by constructing an optimal thermal transfer field that best suits the thermal dynamics of the fluid defined by given trajectories, the fluid's thermal dynamics, which characterize the motion pattern from M given trajectories, can be effectively embedded in the thermal transfer field.

In this work, we construct a scene's thermal transfer field based on the strategy of finite-element analysis [27]. Formally, we first segment the scene into grids of positions $\mathcal{G} = [1, \ldots, W] \times [1, \ldots, H]$, where W and H are the width and height of the scene. Then, the thermal transfer field of the scene can be represented as

$$\mathbf{K} = [k(\mathbf{p}, \mathbf{a})]_{W \times H \times |\mathcal{A}|}, \tag{1}$$

where $k(\mathbf{p}, \mathbf{a})$ ($\mathbf{p} \in \mathcal{G}$, $\mathbf{a} \in \mathcal{A}$) is the thermal transfer coefficient indicating the thermal propagation strength along direction $\mathbf{a}$ at position $\mathbf{p}$. Here $\mathbf{a}$ is a normalized vector ($\|\mathbf{a}\|_2 = 1$) indicating thermal propagation directions, which is selected from a pre-defined direction set $\mathcal{A}$. $|\mathcal{A}|$ counts the number of directions in the set. In this paper, $\mathcal{A}$ contains four directions, which depicts a scene's global motion pattern in upward (y^-), downward (y^+), leftward (x^-), and rightward (x^+) directions, respectively (cf. module 'constructing thermal transfer fields' in Fig. 2).

Assuming that the given trajectories correspond to a stable fluid, we can construct an optimal thermal transfer field

$\mathbf{K}$ by minimizing the total amount of thermal energies being transferred during the fluid flow process, as

$$\min_{\mathbf{K}} \sum_{\mathbf{a} \in \mathcal{A}} \sum_{\mathbf{p} \in \mathcal{G}} \Delta E(k(\mathbf{p}, \mathbf{a}))$$
$$s.t. \quad k(\mathbf{p}, \mathbf{a}) \geq 0, \sum_{\mathbf{p} \in \mathcal{G}} k(\mathbf{p}, \mathbf{a}) = \kappa, \text{ for } \mathbf{a} \in \mathcal{A}. \tag{2}$$

where $\Delta E(k(\mathbf{p}, \mathbf{a}))$ is the amount of thermal energy transported by the fluid from position $\mathbf{p}$ along direction $\mathbf{a}$ within a unit time interval. κ ($\kappa > 0$) is a constant. The first constraint in (2) guarantees to achieve physically-meaningful thermal transfer fields (i.e., avoid negative transfer fields), and the second constraint guarantees to obtain proper distributions of $k(\mathbf{p}, \mathbf{a})$ (i.e., avoid transfer fields becoming all infinite values). $\Delta E(k(\mathbf{p}, \mathbf{a}))$ can be calculated from thermal transmission theories [25], [26]:

$$\Delta E(k(\mathbf{p}, \mathbf{a})) = \eta \frac{\rho(\mathbf{p}) u(\mathbf{p}, \mathbf{a})}{k(\mathbf{p}, \mathbf{a})}. \tag{3}$$

Here η is a parameter related to the temperature difference between a position $\mathbf{p}$ and its neighbors [25], [26]. In our paper, since we want to focus on the relationship between ΔE and $k(\mathbf{p}, \mathbf{a})$, we simply assume the temperature difference condition to be the same when calculating ΔE at different positions, and set η as a constant. $\rho(\mathbf{p})$ is the *density* of fluid at position $\mathbf{p}$, $u(\mathbf{p}, \mathbf{a})$ is the moving *velocity* along direction $\mathbf{a}$ at position $\mathbf{p}$. Physically, $\rho(\mathbf{p}) u(\mathbf{p}, \mathbf{a})$ measures the number of fluid particles passing through position $\mathbf{p}$ along direction $\mathbf{a}$ within a unit time interval. $k(\mathbf{p}, \mathbf{a})$ indicates the efficiency of thermal energy transfer along direction $\mathbf{a}$ at $\mathbf{p}$. As a result, $k(\mathbf{p}, \mathbf{a}) \Delta E$ refers to the amount of thermal energies actually received by $\mathbf{p}$'s neighboring position when an amount of energy ΔE is transferred out from $\mathbf{p}$ [25], [26].

According to (2) and (3), we let a fluid flow along the aggregated routes of M given trajectories, and measure the total amount of thermal energy transfers over all positions. The thermal transfer field that leads to the smallest total transferred energy will be the optimal field that best suits the scene.

More specifically, from (3), the amount of energy transfer ΔE is jointly decided by $k(\mathbf{p}, \mathbf{a})$ (thermal transfer coefficient), $\rho(\mathbf{p})$ (fluid's density), and $u(\mathbf{p}, \mathbf{a})$ (fluid's velocity). When $\rho(\mathbf{p}) u(\mathbf{p}, \mathbf{a})$ increases, the amount of fluid flowing along direction $\mathbf{a}$ at $\mathbf{p}$ becomes stronger, which leads to larger chances of energy transfer. Thus, by minimizing $\sum_{\mathbf{a}, \mathbf{p}} \Delta E$ in (2), $k(\mathbf{p}, \mathbf{a})$ is proportionally adjusted with $\rho(\mathbf{p}) u(\mathbf{p}, \mathbf{a})$ such that $k(\mathbf{p}, \mathbf{a})$ with higher energy transfer efficiencies are assigned to positions/directions with stronger fluid flows. In this way, the resulting thermal transfer field can properly suit the thermal dynamics of the fluid.

$\rho(\mathbf{p})$ in (3) is calculated based on the nonparametric estimation of sample points $\mathbf{p}_n^m$ in given trajectories:

$$\rho(\mathbf{p}) = \sum_{m=1}^M \sum_{n=1}^{\mathcal{L}^m} \exp\left(-\frac{\|\mathbf{p} - \mathbf{p}_n^m\|}{2\sigma^2}\right). \tag{4}$$

---

1. Note that we do not differentiate given trajectories' class labels, and directly use *all* given trajectories to find global motion patterns.
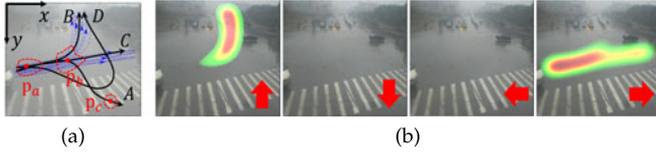
Fig. 3. (a) Input trajectories (black solid curves) and given trajectories (blue dashed curves). (b) Thermal transfer fields constructed from the blue dashed curves in (a). Figures from left to right correspond to thermal transfer fields in directions upward (y^-), downward (y^+), leftward (x^-), rightward (x^+), respectively. (Best viewed in color.)

Similarly, $u(\mathbf{p}, \mathbf{a})$ in (3) is the nonparametric estimation of sample points and their corresponding projected speeds:

$$u(\mathbf{p}, \mathbf{a}) = \sum_{m=1}^{M} \sum_{n=1}^{\mathcal{L}^m} \max(\mathbf{a}^{\top} \mathbf{u}_n^m, 0) \exp\left(-\frac{\|\mathbf{p} - \mathbf{p}_n^m\|}{2\sigma^2}\right), \quad (5)$$

where $\mathbf{a}^{\top} \mathbf{u}_n^m$ is the projection of speed $\mathbf{u}_n^m$ in $\mathbf{a}$th direction. $\max(\mathbf{a}^{\top} \mathbf{u}_n^m, 0)$ ensures that only the directions with positive velocity are considered.

Eqs. (4) and (5) ensure that $\rho(\mathbf{p})u(\mathbf{p}, \mathbf{a})$ can correctly reflect the global motion pattern contained in given trajectories. For example, when a large number of trajectories pass through a position $\mathbf{p}$ with relatively high speed, $\rho(\mathbf{p})u(\mathbf{p}, \mathbf{a})$ will become a large number. On the contrary, for positions where trajectories rarely pass, $\rho(\mathbf{p})u(\mathbf{p}, \mathbf{a})$ will be small. Moreover, note that $\rho(\mathbf{p})u(\mathbf{p}, \mathbf{a})$ in (4) and (5) is estimated by integrating the information of *all* trajectories without differentiating their classes. Therefore, the resulting thermal transfer fields in fact include the *complete* contextual motion information among given trajectories. This enables the embedding of the complete contextual motion pattern in later 3D tube construction and trajectory representation steps.

Based on (3), (4), and (5), (2) can be solved by the *Cauchy–Schwarz inequality* and the optimal solutions turns out to be:

$$k(\mathbf{p}, \mathbf{a}) \propto \sqrt{\rho(\mathbf{p})u(\mathbf{p}, \mathbf{a})}. \quad (6)$$

Figs 2 and 3b show some examples of the constructed thermal transfer fields from given trajectories. We can see that the thermal transfer fields constructed by our approach can effectively capture the four-direction global motion patterns contained in given trajectories. For example, in Fig. 3, both the upward and rightward motion patterns of the blue trajectories in Fig. 3a are effectively embedded in the *upward* and *rightward* thermal transfer fields in Fig. 3b.

3.2 Deriving Equipotential Lines

Based on the constructed thermal transfer field $\mathbf{K}$, we can perform thermal propagation for each point in an input trajectory, so as to embed the input trajectory's contextual motion information. More specifically, we first obtain a *thermal diffusion map* for an input trajectory point $\mathbf{p}_n^m$, which depicts the dominant scene motion pattern around $\mathbf{p}_n^m$ by propagating thermal energies from $\mathbf{p}_n^m$ to the entire scene through $\mathbf{K}$. Then we derive a constant-energy line called *equipotential line* from this map, and use it to capture the contextual motion information for the trajectory point.

Formally, for each trajectory point $\mathbf{p}_n^m$, we denote its corresponding thermal diffusion map as $\mathbf{E}(\mathbf{p}_n^m) =$
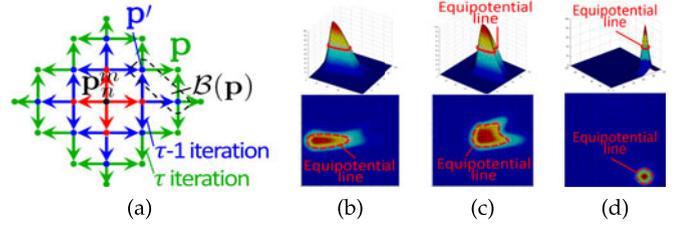

Fig. 4. Examples of equipotential lines. (a) The process of iterative thermal diffusion. (b)-(d) Lower figures: The thermal diffusion maps and the equipotential lines for trajectory points $\mathbf{p}_a$, $\mathbf{p}_b$, and $\mathbf{p}_c$ in Fig. 3a; upper figures: The thermal diffusion maps and the equipotential lines displayed in 3D. (Best viewed in color.)

$[E(\mathbf{p}_n^m, \mathbf{p})]_{W \times H}$, $\mathbf{p} \in \mathcal{G}$, where $E(\mathbf{p}_n^m, \mathbf{p})$ is the thermal energy at position $\mathbf{p}$ in $\mathbf{p}_n^m$'s map. Taking $\mathbf{E}(\mathbf{p}_n^m)$ as a function of time t and direction $\mathbf{a} \in \mathcal{A}$, we calculate $\mathbf{E}(\mathbf{p}_n^m)$ via an iterative method. Specifically, the initial map $\mathbf{E}_0(\mathbf{p}_n^m)$ is defined as follows:

$$E_0(\mathbf{p}_n^m, \mathbf{p}) = \begin{cases} E_\epsilon, & \text{if } \mathbf{p} = \mathbf{p}_n^m, \\ 0, & \text{otherwise} \end{cases}, \quad (7)$$

where $E_0(\mathbf{p}_n^m, \mathbf{p})$ is the initialized thermal energy at position $\mathbf{p}$. $E_\epsilon = 100$ is a constant. According to (7), the thermal diffusion map of trajectory point $\mathbf{p}_n^m$ is initialized by assigning a large thermal energy to $\mathbf{p}_n^m$'s position and assigning zero energies to other positions. Then, a thermal diffusion process is applied which transfers thermal energies from $\mathbf{p}_n^m$ to other positions through the constructed thermal transfer field, and creates an energy-propagated thermal diffusion map. The thermal diffusion process is defined by [25]:

$$\frac{\partial E(\mathbf{p}_n^m, \mathbf{p})}{\partial t} = \sum_{\mathbf{a} \in \mathcal{A}} k(\mathbf{p}, \mathbf{a}) \frac{\partial^2 E(\mathbf{p}_n^m, \mathbf{p})}{\partial \mathbf{a}^2}. \quad (8)$$

Here $\frac{\partial^2 E(\mathbf{p}_n^m, \mathbf{p})}{\partial \mathbf{a}^2}$ means the 2nd order partial derivative of $E(\mathbf{p}_n^m, \mathbf{p})$ along direction $\mathbf{a}$. $k(\mathbf{p}, \mathbf{a})$ is a thermal transfer coefficient in the constructed thermal transfer field $\mathbf{K}$ (cf. (1), (2)). From (8), the thermal diffusion process is mainly controlled by the thermal energy difference among neighboring positions and the thermal transfer coefficients from thermal transfer field. Thus, by properly setting the initial energy difference (cf. (7)), the resulting thermal diffusion map can effectively capture the desired contextual motion pattern contained in the thermal transfer field.

Eq. (8) is difficult to solve because: 1) The thermal diffusion process is performed by integrating the information from multiple directions $\mathbf{a}$; 2) The thermal transfer fields are non-homogeneous whose thermal transfer coefficients $k(\mathbf{p}, \mathbf{a})$ vary over different locations. We propose an approximation solution which obtains thermal diffusion map in an iterative way, as illustrated in Fig. 4a.

From Fig. 4a, after initializing a thermal diffusion map by (7), the thermal energy at $\mathbf{p}_n^m$ is diffused iteratively outward to other positions. During each iteration τ, positions which were diffused in the previous iteration $\tau - 1$ will diffuse energies to their outside neighboring positions (as indicated by arrows in Fig. 4a). According to the thermal diffusion process in (8), thermal diffusion
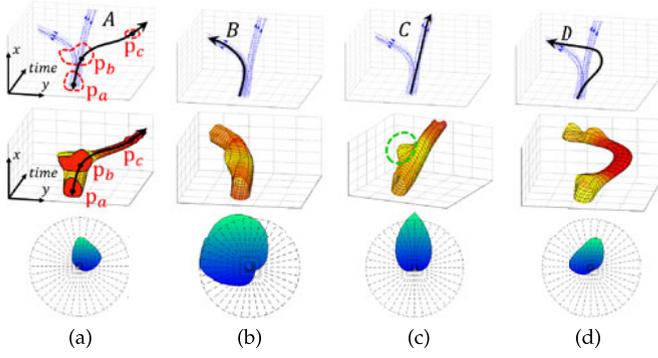
Fig. 5. Examples of 3D tubes and water droplet results. First row: results by expanding trajectories in Fig. 3a into 3D spatio-temporal curves; second row: 3D tube representations for the black input trajectories A-D in Fig. 3a; third row: water droplet results derived from 3D tubes. (Note: In the middle row of (a)-(d), the thickness of a tube is represented by different colors where yellow indicates thick and red indicates narrow. Best viewed in color.)

between neighboring positions in one iteration can be approximated by:

$$E_\tau(\mathbf{p}_n^m, \mathbf{p}) \approx \frac{\sum_{\mathbf{p}' \in \mathcal{B}(\mathbf{p})} E_{\tau-1}(\mathbf{p}_n^m, \mathbf{p}') \exp\left(-\frac{\|\mathbf{p}-\mathbf{p}'\|}{k(\mathbf{p}, \frac{\mathbf{p}-\mathbf{p}'}{\|\mathbf{p}-\mathbf{p}'\|})}\right)}{|\mathcal{B}(\mathbf{p})|}, \quad (9)$$

where $\mathcal{B}(\mathbf{p})$ is the neighborhood of $\mathbf{p}$ and $|\mathcal{B}(\mathbf{p})|$ is the size of the neighborhood. For $\mathbf{p}$ and its neighbors $\mathbf{p}' \in \mathcal{B}(\mathbf{p})$, $\|\mathbf{p} - \mathbf{p}'\|$ is their distance and $k(\mathbf{p}, \frac{\mathbf{p}-\mathbf{p}'}{\|\mathbf{p}-\mathbf{p}'\|})$ is the thermal transfer coefficient along direction $\frac{\mathbf{p}-\mathbf{p}'}{\|\mathbf{p}-\mathbf{p}'\|}$ at position $\mathbf{p}$.

Eq. (9) reveals that: 1) The diffused thermal energy in a position $\mathbf{p}'$ is summed over all energies propagated from its neighboring positions in both vertical and horizontal directions. In this way, the complete contextual information can be included. 2) The thermal transfer coefficients $k(\mathbf{p}, \mathbf{a})$ control the thermal diffusion results. In this way, the motion pattern information from the thermal transfer field can be properly reflected in the resulting thermal diffusion map. 3) The thermal diffusion map $\mathbf{E}(\mathbf{p}_n^m)$ for position $\mathbf{p}_n^m$ is fully decided by the thermal transfer field $\mathbf{K}$ (cf. (7) and (9)). This implies that the contextual motion information $\mathbf{E}(\mathbf{p}_n^m)$ for each position in a scene is fixed after a scene's global motion pattern $\mathbf{K}$ is determined. Therefore, if two input trajectories pass through position $\mathbf{p}_n^m$ with different routes, they will have the same thermal diffusion map at $\mathbf{p}_n^m$.

Fig. 4 shows some examples of thermal diffusion map results derived from the thermal transfer fields in Fig. 3b. In Figs. 4, 4b, 4c, and 4d show the thermal diffusion maps of three points $\mathbf{p}_a$, $\mathbf{p}_b$, $\mathbf{p}_c$ on the black trajectory A in Fig. 3a.

According to Fig. 3a, since moving rightward from $\mathbf{p}_a$ appears frequently in the scene (as there are lots of dashed blue trajectories moving rightward around $\mathbf{p}_a$), large *rightward*-direction thermal transfer coefficients are obtained around $\mathbf{p}_a$. This allows more thermal energies being propagated to $\mathbf{p}_a$'s right side, thus leading to a long rightward tail in the thermal diffusion map of $\mathbf{p}_a$ (cf. Fig. 4b). Similarly, since $\mathbf{p}_b$ is located in a region including frequent movements in both rightward and upright-ward directions, $\mathbf{p}_b$'s thermal diffusion map includes big tails in both directions

and displays a V-like shape (cf. the lower figure in Fig. 4c). Comparatively, since $\mathbf{p}_c$ is located in a region where seldom trajectories pass, the constructed thermal transfer coefficients are small in all directions around $\mathbf{p}_c$. This makes $\mathbf{p}_c$'s thermal diffusion map decay quickly around $\mathbf{p}_c$, as in Fig. 4d. From the example of Figs. 4b, 4c, and 4d, our thermal diffusion map indeed provides a reliable way to capture the complete and unique contextual motion patterns for input trajectory points.

In order to represent thermal diffusion maps in a more effective way, we further derive equipotential lines to capture the fundamental information of thermal diffusion maps. An equipotential line can be easily achieved by finding a constant-energy line on a thermal diffusion map. In this paper, we acquire constant-energy line whose energy decreases to half of the initial energy E_ϵ, as indicated by the red circles in Figs. 4b, 4c, and 4d.

3.3 Constructing 3D Tubes

After deriving equipotential lines for all points in a trajectory, a 3D tube can be constructed to represent this trajectory by concatenating these equipotential lines according to their temporal order in the trajectory.

Fig. 5 shows some 3D tube examples for the black input trajectories A-D in Fig. 3. The first row in Fig. 5 shows the results by expanding trajectories in Fig. 3a into 3D spatio-temporal curves. The second row of Fig. 5 illustrates the 3D tube representations for trajectories A-D in Fig. 3a. Furthermore, the equipotential lines for three points on one input trajectory (i.e., trajectory A in Fig. 3a) are also displayed by red slices in Fig. 5a. These red slices clearly show that a 3D tube is constructed by sequentially concatenating a trajectory's equipotential lines in a 3D spatio-temporal space.

From Fig. 5, we can observe that:

- The constructed 3D tube contains rich information about a trajectory, where both the movement and the contextual motion pattern are effectively embedded. For example, the route of a 3D tube represents the movement. The thickness variation of a 3D tube indicates whether there are frequent motion patterns in the context around a trajectory (e.g., a 3D tube will become narrow if a trajectory goes through a region where trajectories rarely pass, such as trajectory A in Fig. 5a). Moreover, the shapes of equipotential lines in a 3D tube also indicate possible motion trends provided by the contextual motion patterns. For example, the convex part circled by the green dashed line in the second row of Fig. 5c indicates that moving upleft-ward around $\mathbf{p}_b$ (i.e., turn left in the original 2D scene) is another possible motion trend which also appears frequently in the scene. More discussions about the informativeness of 3D tube representation will be provided in the experimental results.

- Different from the previous trajectory modeling methods [4], [7] whose modeled trajectory information cannot be easily visualized, our 3D tube representation is able to visualize information of a trajectory in a clear and unified way. For example, one can easily observe a trajectory's movement and contextual motion pattern from the route and shape
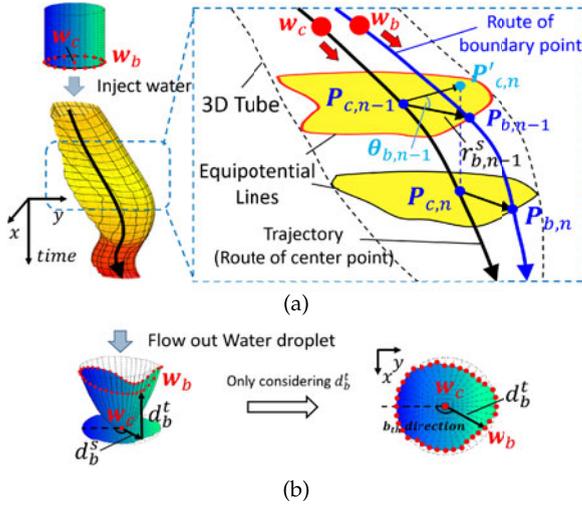
(a)

(b)

Fig. 6. The water droplet process. (a) Illustrations of the input water droplet, $\mathbf{P}_{b,n}$, $\mathbf{P}_{c,n}$, $\theta_{b,n}$, and routes of water droplet points $\mathbf{w}_c$, $\mathbf{w}_b$ when passing through a 3D tube. (b) The resulting 3D water droplet (left) and the simplified 2D droplet by only considering d_b^t (right).

variation of its 3D tube representation. This in fact provides a useful tool for people to intuitively observe and analyze trajectory information.

4 THE WATER-DROPLET PROCESS

After constructing 3D tubes for input trajectories, we need to find suitable ways to effectively handle the high-dimensional information included in 3D tubes. In this paper, we introduce a droplet-based process which simulates the physical water flow process [26] and derives a low-dimensional droplet vector to characterize a high-dimensional 3D tube shape.

The process of the droplet-based process is displayed in Fig. 6. We inject a drop of water with fixed shape in one end of a 3D tube (cf. Fig. 6a) and achieve a water droplet flowed out from the other end (cf. Fig. 6b, note that the water is flowed along the time-axis in a 3D spatial-temporal space). According to fluid mechanic theories [26], when a water droplet passes through a 3D tube, its shape will be affected by the friction and extrusion forces from the boundary of the tube. 3D tubes with different shapes will provide different impacts to water passing through them, and create different droplet outputs. Therefore, by properly designing the water flow process, the derived droplet can effectively capture the characteristics of a high-dimensional 3D tube shape.

A water droplet is described by a center point $\mathbf{w}_c$ and a set of boundary points $\{\mathbf{w}_b\}_{b=1,\dots,N_w}$ where $N_w = 36$ is the total number of boundary points being considered. Moreover, two distances are defined to represent the relative location between a boundary point $\mathbf{w}_b$ and $\mathbf{w}_c$: 1) d_b^t is the distance between $\mathbf{w}_b$ and $\mathbf{w}_c$ in the time axis; 2) d_b^s is the distance between $\mathbf{w}_b$ and $\mathbf{w}_c$ in the spatial plane, as in Fig. 6b. In order to simplify calculation, we only calculate d_b^t at the output of a 3D tube, such that a 3D droplet can be simply described by d_b^t in a 2D plane, as in Fig. 6b. Therefore, in the following, we focus on discussing d_b^t.

Before water starts to flow in a 3D tube, d_b^t are initialized as 0 to make all boundary points located on the same spatial plane as the center point $\mathbf{w}_c$, as in the upper figure in

Fig. 6a. Then, during the water flow process, we let the center point $\mathbf{w}_c$ follow the route of a 3D tube's input trajectory (i.e., the input trajectory that creates the 3D tube), as indicated by $\{\mathbf{P}_{c,n}\}_{n=1}^{\mathcal{L}}$, where $\mathcal{L}$ is the length of the input trajectory and $\mathbf{P}_{c,n}$ is the location of the nth input-trajectory point in the spatial-temporal space (cf. the black line in Fig. 6a). Similarly, a boundary point $\mathbf{w}_b$ ($b \in \{1,\dots,N_w\}$) will follow the route constructed by the bth boundary points on a 3D tube: $\{\mathbf{P}_{b,n}\}_{n=1}^{\mathcal{L}}$, where $\mathbf{P}_{b,n}$ is the position located on bth direction of $\mathbf{P}_{c,n}$ on $\mathbf{P}_{c,n}$'s equipotential line (cf. the blue line in Fig. 6a).

When the center point $\mathbf{w}_c$ of a water droplet passes through a 3D tube with a constant velocity in the time-axis, the velocity of the droplet's boundary points $\mathbf{w}_b$ ($b = 1,\dots,N_w$) will be jointly affected by two forces: the viscosity force from $\mathbf{w}_c$ which pulls $\mathbf{w}_b$ close to $\mathbf{w}_c$, and the friction force from a 3D tube's boundary which resists $\mathbf{w}_b$ from approaching $\mathbf{w}_c$ [26]. Since both forces are controlled by the shape of a 3D tube, by calculating the relative distances d_b^t between $\mathbf{w}_c$ and $\mathbf{w}_b$ at a tube's output, the characteristics of a 3D tube can be properly captured.

The time-axis distance d_b^t between points $\mathbf{w}_c$ and $\mathbf{w}_b$ at the output of a tube (cf. Fig. 6b) can be approximated by:

$$d_b^t \propto \frac{1}{\mathcal{L}} \sum_{n=1}^{\mathcal{L}} (v_{c,n} - v_{b,n}), \qquad (10)$$

where $\mathcal{L}$ is the length of a 3D tube. $v_{c,n}$ and $v_{b,n}$ are the time-axis velocities when points $\mathbf{w}_c$ and $\mathbf{w}_b$ pass through positions $\mathbf{P}_{c,n}$ and $\mathbf{P}_{b,n}$, respectively. $v_{c,n} = v_{\mathcal{C}}$ meaning that $\mathbf{w}_c$ passes through different $\mathbf{P}_{c,n}$ with a constant time-axis velocity $v_{\mathcal{C}}$. $\frac{1}{\mathcal{L}}$ is included to normalize d_b^t.

From (10), the time-axis distance d_b^t is measured by accumulating the velocity difference $v_{c,n} - v_{b,n}$ between points $\mathbf{w}_c$ and $\mathbf{w}_b$ when passing through a 3D tube. By performing velocity accumulation, we are able to characterize high-dimensional tube information in a relative simple way while reducing the disturbance from trajectory noises.

According to fluid viscosity theories [28], $v_{c,n} - v_{b,n}$ in (10) can be approximated by (see Appendix in the supplementary material for details, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2016.2608884):

$$v_{c,n} - v_{b,n} \approx \left(1 - \frac{\lambda_1}{r_{b,n-1}^s}\right)(v_{\mathcal{C}} - v_{b,n-1}) + \lambda_2 \cos\theta_{b,n-1}, \quad (11)$$

where λ_1 and λ_2 are constant coefficients. $v_{b,n}$ is the time-axis velocity of $\mathbf{w}_b$ when passing through position $\mathbf{P}_{b,n}$. $\mathbf{w}_b$'s velocity is initialized as 0 before the water flow process (i.e., $v_{b,0} = 0$). $r_{b,n-1}^s$ is the distance from $(n-1)$th trajectory position $\mathbf{P}_{c,n-1}$ to its bth direction boundary position $\mathbf{P}_{b,n-1}$ on the equipotential line (cf. Fig. 6a). $\theta_{b,n-1}$ is the angle between lines $\overrightarrow{\mathbf{P}_{c,n-1}\mathbf{P}_{b,n-1}}$ and $\overrightarrow{\mathbf{P}_{c,n-1}\mathbf{P}_{c,n}'}$, where $\mathbf{P}_{c,n}'$ is the projected position of $\mathbf{P}_{c,n}$ on $\mathbf{P}_{c,n-1}$'s spatial plane, as in Fig. 6a.

Basically, $r_{b,n-1}^s$ in (11) reflects the size variation of a water droplet, while $\theta_{b,n-1}$ evaluates the relative location of a boundary position $\mathbf{P}_{b,n-1}$ with respect to the motion direction $\overrightarrow{\mathbf{P}_{c,n-1}\mathbf{P}_{c,n}'}$ of a water droplet. Since the size and motion

direction of a water droplet is controlled by the shape and route of the 3D tube it passes through, by introducing $r_{b,n-1}^{s}$ and $\theta_{b,n-1}$, a 3D tube's shape and route information can be effectively embedded.

Moreover, the terms $\frac{\lambda_1}{r_{b,n-1}^{s}}$ and $\cos\theta_{b,n-1}$ in (11) reflect the impact of a 3D tube's shape to the viscosity and friction forces applied on a water droplet point $\mathbf{w}_b$. When $\mathbf{w}_b$ passes through a boundary position $\mathbf{P}_{b,n-1}$ located in front of the motion route of $\mathbf{w}_b$'s water droplet (i.e., small $\theta_{b,n-1}$), the friction force on $\mathbf{w}_b$ becomes large due to the increased normal force that $\mathbf{w}_b$'s water droplet applies on tube boundary $\mathbf{P}_{b,n-1}$. This will reduce the velocity of $\mathbf{w}_b$ and lead to larger d_b^{t}. Similarly, when a 3D tube becomes slim on bth direction (i.e., small $r_{b,n-1}^{s}$), the viscosity force on $\mathbf{w}_b$ will be enlarged which pulls $\mathbf{w}_b$ closer to $\mathbf{w}_c$ and creates smaller d_b^{t}. This point will be further discussed in the experimental results.

The third row of Fig. 5 shows the droplet results derived from the 3D tubes in the second row of Fig. 5 by using (10) and (11). From Fig. 5, we can observe the effectiveness of our water-droplet process:

- The major motion directions of 3D tubes are properly captured by the large sectors in droplet results. For example, the water droplet of trajectory C has a large sector in the upward direction since trajectory C moves forward only. Comparatively, the water droplet of trajectory B has large sectors in both top and left directions. This indicates the 'forward+left turn' movement of B.

- Droplets derived from thick 3D tubes have larger sizes than those derived from slim tubes. For example, the droplet for trajectory C has large size since C follows a frequent motion pattern in the scene and has a thick 3D tube. Comparatively, since trajectory A turns to an irregular region in the middle, its 3D tube becomes narrow in the later part. This leads to a small size in its corresponding droplet. Fig. 5 implies that the size of a droplet can effectively differentiate regular and irregular motion patterns. Therefore, in this paper, droplet size is utilized as a major feature to detect abnormalities in trajectory analysis (cf. (13)).

Finally, the obtained water droplet is sampled to achieve a low-dimensional droplet vector. In this paper, we simply concatenate time-axis distances d_b^{t} in a water droplet as the low-dimensional droplet vector:

$$\mathbf{V}^m = [d_1^{t,m}, d_2^{t,m}, \ldots, d_{N_w}^{t,m}], \qquad (12)$$

where $\mathbf{V}^m$ is the droplet vector for trajectory Γ^m. $d_b^{t,m}$ is the time-axis distance in bth direction of Γ^m's water droplet. N_w is the length of the vector and is set as 36 in our experiments.

5   IMPLEMENTATION IN TRAJECTORY ANALYSIS

With the tube-and-droplet representation, trajectories can be depicted by droplet vectors and analyzed accordingly. In this section, we discuss the implementations of our tube-and-droplet representation in three trajectory analysis applications: trajectory clustering, trajectory classification & abnormality detection, and 3D action recognition.

5.1   Trajectory Clustering

When performing trajectory clustering, we first utilize all trajectories being clustered as the given trajectories to construct thermal transfer fields (cf. Section 3.1). Then, a 3D tube and a droplet vector are derived based on the constructed thermal transfer fields to represent each trajectory (cf. Sections 3.3 and 4). Finally, we measure the distance between trajectories by calculating the distance between their corresponding droplet vectors, and perform trajectory clustering according to these droplet-vector distances. In this paper, we utilize euclidean distance to measure the distance between droplet vectors, and utilize spectral clustering [29] to cluster trajectories.

5.2   Trajectory Classification and Abnormality Detection

In trajectory classification and abnormality detection, a set of normal training trajectories are provided whose class labels are given. We aim to recognize the classes of input test trajectories with the guidance of training trajectories, and identify abnormal test trajectories which do not follow the regular motion patterns provided by training trajectories.

Similar to trajectory clustering, we utilize all normal trajectories in the training set to construct thermal transfer fields, and derive a droplet vector for each individual training trajectory. Note that since our approach utilizes the *complete* contextual information, we do not differentiate trajectory classes, that is, normal training trajectories from different trajectory classes are utilized indiscriminatively when constructing thermal transfer fields.

During testing, we first obtain a droplet vector for an input test trajectory. Then, the abnormality of a test trajectory is evaluated by its corresponding droplet vector. Since the size of a droplet can effectively differentiate regular and irregular motion patterns, we detect a test trajectory Γ^m to be abnormal if:

$$\max_{b}\{d_b^{t,m}\} + \frac{1}{N_w}\sum_b d_b^{t,m} < TH, \qquad (13)$$

where $d_b^{t,m}$ is the bth element in Γ^m's droplet vector. N_w is the length of the droplet vector (cf. (12)). TH is a threshold decided by specific scenario. In the experiments of this paper, we simply calculate $\max_b\{d_b^{t,m}\} + \frac{1}{N_w}\sum_b d_b^{t,m}$ value for all normal trajectories in the training set, select the smallest value T from them, and set $0.9T$ as the threshold. In (13), the term $\frac{1}{N_w}\sum_b d_b^{t,m}$ is calculated to measure the size of a droplet while the term $\max_b\{d_b^{t,m}\}$ is used to evaluate the normality in a trajectory's major motion direction.

Finally, if a test trajectory Γ^m is evaluated as normal by (13), a one-against-all linear SVM classifier [30] trained directly from the droplet vectors of training trajectories is applied to classify Γ^m into one of the trajectory classes.

5.3   3D Action Recognition

We also extend our tube-and-droplet approach into the application of 3D action recognition. In 3D action recognition, 3D skeleton sequences are provided which depict human actions in a 3D x-y-$depth$ space [31], [32], [33], [34], [35]. An example skeleton sequence is shown in Fig. 7.
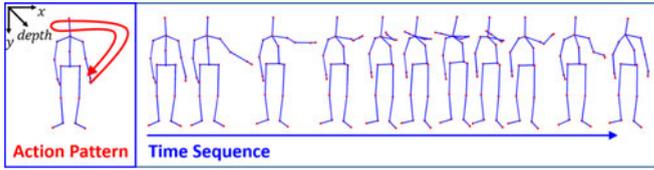
Fig. 7. An example of 3D skeleton sequence from MSR-Action3D dataset. Left: The 3D trajectory of 'horizontal wave' action of a 'left hand' body point. Right: The skeleton sequence for 'horizontal wave' action.

Fig. 9. An example of equipotential surfaces (left) and the resulting water droplet sphere (right) for 3D trajectories.

Since skeleton sequences are described by the trajectories of multiple body points (e.g., the red curve in Fig. 7), they are able to be represented and analyzed by the proposed tube-and-droplet approach. However, a 3D skeleton sequence differs from a regular motion trajectory in: 1) A 3D skeleton sequence includes multiple trajectories for different body points of a human; 2) The trajectory of a body point is located in 3D space (x-y-$depth$) instead of a 2D plane. Therefore, we extend our tube-and-droplet representation into higher dimension to depict 3D trajectories. Besides, in order to handle multiple trajectories in a 3D skeleton sequence, we represent each trajectory independently and then combine them together to achieve a unified depiction of the entire skeleton sequence.

The detailed implementation of our approach on 3D action recognition is described in the following.

First, thermal transfer fields are constructed for each body point based on the trajectories from training skeleton sequences. Since trajectories are located in 3D space, thermal transfer fields need to be modeled by 3D volumes. Therefore, we introduce two additional directions (i.e., z^- and z^+ representing forward and backward directions in depth axis) and extend the process of constructing thermal transfer fields (cf. (1)) into 3D space. Furthermore, in order to make thermal transfer fields more powerful for different 3D actions, we construct a set of 3D thermal transfer fields for each individual action class under each body point. Therefore, if there are N_A action classes being recognized and N_B body points in a skeleton, a total of $N_A \times N_B$ sets of thermal transfer fields are constructed.

Fig. 8 shows an example set of thermal transfer fields which represents the motion pattern of 'horizontal wave' action for a 'left hand' body point (cf. Fig. 7). According to Fig. 8, six thermal transfer fields are constructed to describe global motion patterns in six directions inside a 3D space (cf. Figs. 8c, 8d, 8e, 8f, 8g, and 8h). Besides, each thermal

transfer field is modeled as a 3D volume indicating the thermal propagation strength at different positions.

After thermal transfer fields are constructed, a tube and a droplet vector is constructed to depict the information of an input trajectory. Due to the 3D feature of input trajectories, the constructed tube is extended to a combination of 3D equipotential surfaces instead of 2D equipotential lines, as in the left figure of Fig. 9. Similarly, the obtained droplet is also extended from a surface to a sphere which is represented by a center point and a set of boundary points located on a closed surface surrounding it (cf. right figure in Fig. 9).

Finally, a skeleton sequence can be described by concatenating droplet vectors from different body points and for different action classes, as

$$\mathbf{U}^m = [\omega_1 \mathbf{V}^m_{1,1}, \ldots, \omega_{N_B} \mathbf{V}^m_{N_B,1}; \omega_1 \mathbf{V}^m_{1,2}, \ldots, \omega_{N_B} \mathbf{V}^m_{N_B,2};$$
$$\omega_1 \mathbf{V}^m_{1,N_A}, \ldots, \omega_{N_B} \mathbf{V}^m_{N_B,N_A}], \quad (14)$$

where $\mathbf{U}^m$ is the droplet-based feature vector for the mth skeleton sequence. $\mathbf{V}^m_{\beta,\alpha}$ is the droplet vector derived for βth body point and αth action class. ω_β is the weighting factor balancing the relative importance of βth body point and it can be decided by cross-validation [36]. N_B and N_A are the total number of body points and action classes.

During 3D action recognition, we first derive a droplet feature vector $\mathbf{U}^m$ for an input skeleton sequence, and then utilize a classifier to recognize the action class of this skeleton sequence. In this paper, we use two different classifiers : 1) KNN, and 2) one-against-all linear SVM classifier [30].

6 EXPERIMENTAL RESULTS

We evaluate the performance of our trajectory representation approach on three trajectory analysis applications: trajectory clustering, trajectory classification & abnormality detection, and 3D action recognition. The experiments are performed on multiple benchmark trajectory datasets including Vehicle Motion Trajectory dataset (VMT) [4], Simulated Traffic Intersection dataset (CROSS) [7], our own constructed crossroad traffic dataset (TRAFFIC)[2], and MSR-Action3D Dataset (MSR) [8]. λ_1 and λ_2 in (11) are set as 2 and 0.1, which are decided from experimental statistics.

6.1 Trajectory Clustering

We perform trajectory clustering experiments on a benchmark Vehicle Motion Trajectory dataset (VMT) [4]. The VMT dataset includes 1,500 real-scene vehicle trajectories labeled in 15 clusters. Some example trajectories in VMT dataset is shown in Fig. 10. In this experiment, we cluster trajectories into the same number of clusters as the ground
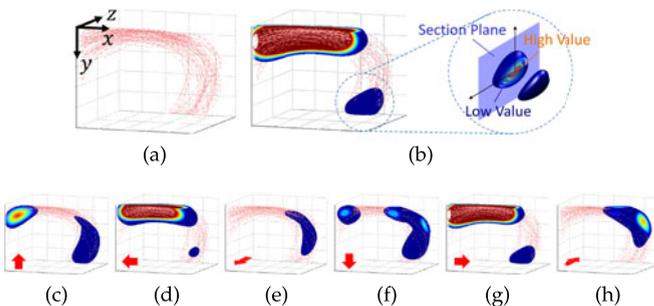

Fig. 8. An example of thermal transfer fields for 3D trajectories. (a) Given 3D trajectories. (c)-(h) Constructed thermal transfer fields for the given 3D trajectories in (a). Note that (c)-(h) correspond to the thermal transfer fields in directions y^-, x^-, z^-, y^+, x^+, z^+, respectively. (b) An illustration of details inside a 3D thermal transfer field when cutting a high-value volume in a transfer field.
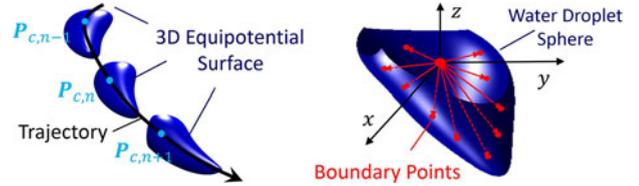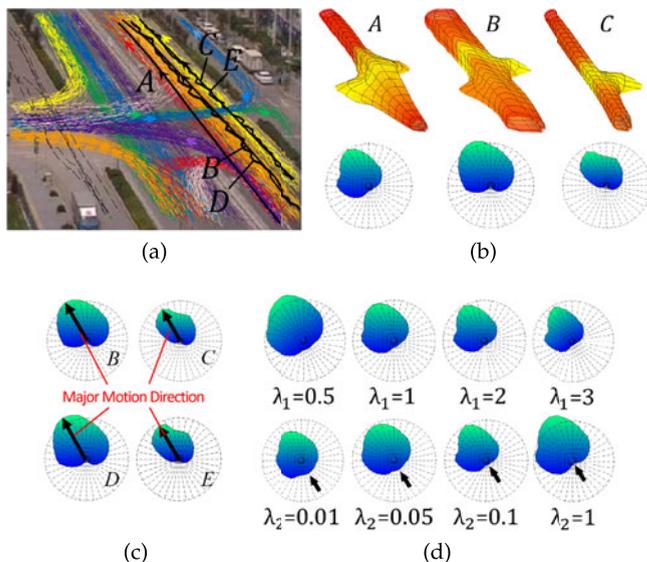
2. www.dropbox.com/s/ahyxw6vqypgb0uf/TRAFFIC.zip?dl=0

Fig. 10. (a) Trajectories from different groundtruth clusters in VMT dataset (color curves) and five input trajectories (black curves labeled by A-E). (b) Upper: 3D tubes for trajectories A, B, and C in (a); lower: Water droplets for A, B, and C. (c) Comparison of water droplet results between clean trajectories (B, C) and noisy trajectories (D, E) in (a) (note that B, D are from one cluster and C, E are from another cluster). (d) Upper: Droplets for trajectory A's 3D tube in Fig. 10b under different λ_1 where $\lambda_2 = 0.1$; lower: Droplets for trajectory A's 3D tube in Fig. 10b under different λ_2 where $\lambda_1 = 2$. (Best viewed in color.)

truth (i.e., 15 clusters), and evaluate the consistency between trajectory labels in the clustering results and those in the ground truth.

We compare our approach with seven methods: 1) Using Euclidean distance to measure the distance between trajectories plus using K-means clustering to cluster trajectories (**ED+Kmeans**); 2) Using euclidean distance plus spectral clustering [29] (**ED+SC**); 3) Using dynamic time warping to measure trajectory distances plus K-means [37] (**DTW+Kmeans**); 4) Using DTW plus spectral clustering [38] (**DTW+SC**); 5) Using a time-sensitive Dirichlet process mixture model [4] to represent and cluster trajectories (**tDPMM**); 6) Using a 3-stage hierarchical learning model [7] to represent and cluster trajectories (**3SHL**); 7) Using a heat-map model [12] to represent and cluster (**HM**).

Note that the 'tDPMM' and '3SHL' methods are state-of-the-art trajectory-modeling methods which increase the informativeness of trajectories by constructing probability models for each trajectory class. Besides, the 'HM' method encodes the temporal variation of a trajectory.

Moreover, in order to evaluate the effectiveness of our water droplet process, we further include the results of three additional methods: 1) Using our approach to construct 3D tubes for trajectory representation, plus using Hausdorff distance [24] to capture the high-dimensional information in these 3D tubes for trajectory clustering (**3D Tube+Hausdorff**); 2) Using 3D tubes plus using a state-of-the-art Grassmann manifold method [20] (**3D Tube+Manifold**); 3) Using our approach to achieve thermal diffusion maps for each trajectory point (cf. Fig. 3), and directly concatenating these thermal diffusion maps as the representation of a trajectory, finally using the Grassmann manifold method [20] to capture the high-dimensional information (**Thermal map+Manifold**).

TABLE 1
Cluster Learning Accuracy for Different Methods on VMT Dataset (%)

Method	Cluster Accuracy
ED+Kmeans	82.6
ED+SC[29]	85.0
DTW+Kmeans[37]	83.2
DTW+SC[38]	85.3
tDPMM [4]	86.7
3SHL [7]	84.4
HM [12]	82.0
3D Tube+Hausdorff	**91.5**
Thermal Map+Manifold	**92.2**
3D Tube+Manifold	**93.6**
Ours	**93.8**

Note that the major difference between the 'Thermal map +Manifold' method and '3D Tube+Manifold' method is that 'Thermal map+Manifold' skips the step of equipotential line extraction (cf. Section 3.2), and directly utilizes a thermal diffusion map to represent a trajectory point.

6.1.1 Comparison of Clustering Results

Table 1 compares the cluster learning accuracy [4] for different methods on VMT dataset, where the cluster learning accuracy measures the total percentage of trajectories being correctly clustered. From Table 1, approaches using our 3D tube representation (3D Tube+Hausdorff, 3D Tube+Manifold, Thermal map+Manifold, Ours) achieve obviously better clustering results than the compared methods. This demonstrates the usefulness of our 3D tube representation. Moreover, we can also observe from Table 1 that: 1) Our approach, which integrates both 3D tube representation and water droplet process, achieves the best clustering results. This demonstrates the effectiveness of our tube+droplet framework. 2) The '3D Tube+Manifold' method has slightly better results than the 'Thermal map+Manifold' method. This implies that equipotential lines (cf. Section 3.2) can not only capture the useful information in thermal diffusion maps, but also suitably avoid the disturbance from noise in thermal diffusion maps.

6.1.2 Robustness to Noises and Trajectory Breaks

We further demonstrate the effectiveness of our approach in dealing with noisy or broken trajectories. Following [4], we add Gaussian noise to all points in a trajectory to simulate a noisy trajectory. Three noise levels are used to derive three trajectory datasets with different noise strengths (i.e., Noise Level 1, 2, and 3 in Table 2). Similarly, we omit the initial or last G points in two of 10 trajectories in each cluster to simulate datasets with broken trajectories (i.e., omit $G = 10$ percent, $G = 20$ percent, $G = 30$ percent, and $G = 40$ percent in Table 2) [4].

Table 2 compares the clustering results of different methods on noisy or broken trajectories derived from VMT dataset. Table 2 shows that our approach achieves the best clustering results under different noise or trajectory break levels. Moreover, when noise or trajectory break level increases, the clustering performance decrease by our approach is relatively small among the

TABLE 2
Cluster Learning Accuracy with Different Noise or Trajectory Break Levels on VMT Dataset (%)

Datasets	ED+Kmeans	ED+SC [29]	DTW+Kmeans [37]	DTW+SC [38]	tDPMM [4]	3SHL [7]	HM [12]	3D Tube+ Hausdorff	3D Tube+ Manifold	**Ours**
Noise Level 1	80.6	83.3	81.2	83.5	84.3	83.2	80.3	88.7	91.3	**91.7**
Noise Level 2	78.8	81.2	80.0	81.9	83.3	80.1	78.9	85.3	89.1	**90.7**
Noise Level 3	77.0	81.2	78.8	74.6	81.5	79.2	71.8	79.5	86.4	**88.1**
Omit $G = 10\%$	80.8	84.4	81.1	85.1	86.1	84.0	82.0	90.6	92.9	**93.1**
Omit $G = 20\%$	78.9	82.7	76.5	83.5	85.7	80.4	81.8	89.7	91.3	**92.1**
Omit $G = 30\%$	76.5	79.3	74.4	81.1	81.8	76.7	80.3	87.2	88.7	**89.5**
Omit $G = 40\%$	72.3	76.1	71.8	77.5	78.1	74.1	78.1	84.3	83.2	**87.2**

compared methods, this further demonstrates the robustness of our tube-and-droplet approach when handling noises or trajectory breaks.

6.1.3 Effectiveness of 3D Tube Representation

Figs 10 illustrates a more detailed example about the effectiveness of our 3D tube representation. In Fig. 10a, the colored curves are trajectories to be clustered and the black trajectories labeled in A, B, and C are three trajectories from them. In order to ease the discussion, trajectories from different groundtruth clusters are displayed by different colors. Since trajectories A, B, and C come from three trajectory clusters which are located close to each other and have similar motion patterns (cf. the clusters in yellow, orange, and red in Fig. 10a), existing methods have limitations in differentiating them which only consider the pairwise correlation between trajectories (ED+Kmeans, ED+SC, DTW+Kmeans, DTW+SC, HM) or intra-cluster correlation among trajectories (tDPMM, 3SHL).

With our 3D tube representation, the differences among A, B, and C can be properly highlighted by embedding the complete motion information from all trajectories, as the upper figures in Fig. 10b. For example, the 3D tube of trajectory A shows an obvious leftward convex part since there is a large left-turn pattern provided by the purple cluster along A's route. Trajectory B's tube is thicker since it is located in the middle of a large upleft-ward motion pattern jointly provided by the yellow, orange, and red clusters. Besides, trajectory C's tube includes an obvious rightward convex part due to the rightward contextual patterns provided by the blue and green clusters next to C. Therefore, by suitably capturing the high-dimensional information in these 3D tubes, the difference among trajectories can be effectively reflected in the resulting droplet vectors (cf. the lower figures in Fig. 10b).

6.1.4 Effectiveness of Water Droplet Process

In order to evaluate our water droplet process, we compare our approach with '3D Tube+Hausdorff', '3D Tube+Manifold', and 'Thermal Map+Manifold'. These methods use the same 3D tube representation to depict a trajectory, but use different schemes to capture the high-dimensional information in a 3D tube. The clustering results of these methods under different noise or trajectory break levels are shown in Table 2. Furthermore, Table 3 compares the running time of the 3D tube information handling & clustering steps in these methods. We observe that:

- From Table 2, the clustering-accuracy difference between our approach and '3D Tube+Hausdorff' becomes larger for higher noise levels. This implies that satisfactory results cannot be easily achieved without suitably capturing the high-dimensional information in 3D tubes. More specifically, since Hausdorff distance is easily affected by noise, its results decrease more rapidly for large noises. In contrast, since our water droplet process characterizes a 3D tube by accumulating information at different parts in a tube (cf. (10)), the disturbance from noise can be effectively reduced.

- Although '3D Tube+Manifold' and 'Thermal Map+Manifold' can achieve relatively better results than '3D Tube+Hausdorff', their computation complexities are considerably high (cf. Table 3). Comparatively, our water droplet process handles information of a 3D tube in a simple but effective way, which is able to perform clustering in less than 1 minute while achieving better results than '3D Tube+Manifold' and 'Thermal Map+Manifold'.

Moreover, Fig. 10c shows the droplet results for four trajectories where trajectories B and D belong to the orange cluster and trajectories C and E belong to the yellow cluster. Since trajectories D and E are interfered by noise, the ambiguity among trajectories increases. However, with our water droplet process, these noisy effects can be properly reduced and the major characteristics of a 3D tube can be properly obtained. For example, in Fig. 10c, the common characteristics of trajectories B and D are captured in their droplets which include larger sectors on both sides of the trajectories' major motion direction.

6.1.5 Effect of Different Parameter Values

Fig. 10d shows the droplet results for trajectory A's 3D tube in Fig. 10b under different λ_1 and λ_2 values (cf. (11)).

TABLE 3
Running Time of Different Steps for Table 2

Method	Transfer field +3D tube	Handling 3D tube +clustering
3D Tube+Hausdorff		38.4 min
3D Tube+Manifold		12.0 hr
Thermal Map+Manifold	11.2 min	12.0 hr
Ours		**47.5 sec**

(Note: 'Transfer field+3D tube' refers to the steps of constructing thermal transfer fields and 3D tubes 'Handling 3D tube+clustering' refers to the steps of handling the high-dimensional information in 3D tubes and performing clustering. Implemented by Matlab and evaluated on a PC with 4 core CPU and 8G RAM.)
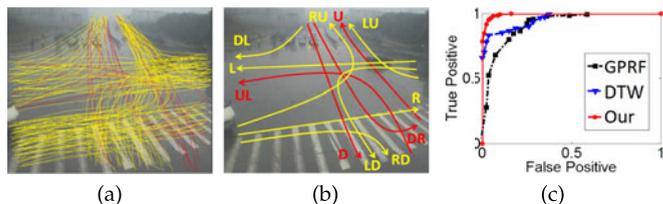
(a)   (b)   (c)

Fig. 11. (a) Example trajectories of the TRAFFIC dataset (yellow curves: normal trajectories, red curves: abnormal trajectories). (b) Major motion routes of different trajectory classes. The yellow and red curves indicate routes for normal and abnormal trajectory classes ('U','D','L','R' refers to 'upward','downward','leftward','rightward' respectively. For example, 'LU' means moving leftward and then upward). (c) ROC curves of different methods when discriminating normal/abnormal trajectories under different abnormality detection thresholds. (Best viewed in color.)
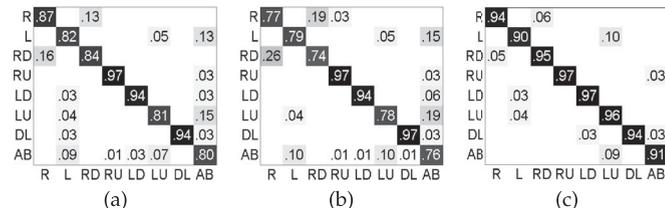

(a)   (b)   (c)

Fig. 12. Confusion matrices of different methods when classifying normal trajectories and detecting abnormal ones (denoted as 'AB'). (a) GPRF [6]. (b) DTW [37]. (c) Ours.

From Fig. 10d, we can see that λ_1 mainly controls the effect from a 3D tube's shape. More specifically, when λ_1 increases, a droplet becomes more concentrated on a tube's true motion directions and convex parts, while sectors not along a tube's direction or convex part will shrink. At the same time, with the increase of λ_1, the thickness of a tube is also reflected more obviously in a droplet. For example, since trajectory A's 3D tube in Fig. 10b is relatively narrow, increasing λ_1 will decrease the size of its droplet and make the droplet more coherent with the tube's thickness.

Besides, Fig. 10d reveals that λ_2 controls the impact of a 3D tube's route. When λ_2 is extremely small, the route information of a 3D tube cannot be fully included in a droplet (i.e., a droplet cannot tell whether a trajectory is moving upward or downward along a 3D tube, cf. Fig. 10b). This will create obvious sectors in the opposite direction to a trajectory's motion route (cf. the leftmost figure in Fig. 10d). When λ_2 increases, the route information is more clearly included in its corresponding droplet, and sectors opposite to a trajectory's route will shrink. We set λ_1 and λ_2 as 2 and 0.1. Experiments show that these values can properly embed both the shape and route information of a 3D tube and create satisfactory results.

6.2 Trajectory Classification and Abnormality Detection

6.2.1 Results on TRAFFIC Dataset

We perform experiments of trajectory classification & abnormality detection on our own constructed TRAFFIC dataset. This dataset includes 300 real-scene trajectories where 200 trajectories are for normal activities and the other 100 trajectories are abnormal ones. The normal trajectories includes seven classes, with about 30 trajectories for each class. The major motion routes of different trajectory classes are indicated in Fig. 11b. Some example normal and abnormal trajectories are shown in Fig. 11a.

Note that this is a challenging dataset in that: 1) The total number of trajectories in the dataset is small, making it difficult to construct reliable models; 2) The motion trajectories within the same class have large variations due to the large width of roads; 3) Due to the visual angle of the surveillance camera, trajectories from different class are easily confused and are difficult to differentiate.

We compare our approach (cf. Section 5.2) with two methods: 1) The GPRF method [6] which introduces Gaussian process regression flows to model the location and velocity probability for each trajectory class, and utilizes

them to classify trajectories (**GPRF**); 2) The DTW method [37] which classifies a test trajectory by measuring its dynamic-time-warping distance with the center of different trajectory classes (**DTW**).

We split the dataset into 50 percent training-50 percent testing parts. Note that in our experiments, only normal trajectories are used for training. Four independent runs are performed where the training and testing sets are randomly selected in each run, and the final results are averaged. Fig. 12 compares the confusion matrices of different methods for classifying normal trajectories and detecting abnormal ones. Fig. 11c further compares the ROC curves of different methods when discriminating normal/abnormal trajectories under different abnormality detection thresholds.

From Figs. 12 and 11c, we can observe that:

- Our approach can achieve obviously better results than the compared methods. More specifically, the compared methods have low effectiveness in discriminating trajectory classes such as *RD*, *L*, and *LU*. This is because trajectories in these classes are easily confused with similar trajectories from other classes such as *R* and *AB* (e.g., *L*, *LU* are similar to patterns *UL* and *U* in the abnormal class *AB*, cf. Fig. 11b). Comparatively, since our approach introduces informative tube-and-droplet representation to capture the subtle difference between trajectory groups, more satisfactory results can be achieved.

- Although the GPRF method constructs probability models to encode the spatial-temporal variation of trajectory classes, it still creates less satisfactory results. This is because: (a) The GPRF method only focuses on modeling the contextual information inside each trajectory class, which has limitations in differentiating trajectories from similar classes; (b) The number of training trajectories is small, which makes it difficult for the GPRF method to construct reliable probability models. Comparatively, our approach is able to work reliably under relatively small trajectories. Besides, by leveraging the complete contextual motion information, the subtle difference between different classes are also properly discriminated by our approach, thus obtaining more improved results.

6.2.2 Results on CROSS Dataset

We also perform experiments of trajectory classification and abnormality detection on CROSS dataset [7]. It includes 11,400 normal trajectories labeled in 19 clusters and 200 abnormal trajectories (one example in Fig. 2).

Following [7], we utilize 1,900 normal trajectories in training. The constructed thermal transfer fields or

TABLE 4
Trajectory Classification and Abnormality Detection
Results on CROSS (%)

Method	Classification (CA)	Abnormality Detection (DR/FPR)
DTW [37]	95.6	71.3/55.5
GPRF [6]	98.0	80.5/23.5
tDPMM [4]	98.0	91.0/**23.3**
3SHL [7]	96.8	85.5/23.5
Ours	**98.6**	**91.3**/23.5

trajectory class models are then utilized to classify the rest 9,500 normal and 200 abnormal trajectories. We use (13) to detect abnormal trajectories and a simple K-nearest neighbor (KNN)[39] strategy to classify normal trajectories.

Table 4 compares the trajectory classification & abnormality detection results. Classification results are evaluated by classification accuracy (CA) which is the ratio between total number of correctly classified normal trajectories and total number of normal trajectories. Abnormality detection results are evaluated by abnormality detection rate (DR) and abnormality false positive rate (FPR) [7]. Note that besides the DTW and GPRF methods, we also include the results of two state-of-the-art methods on CROSS dataset in Table 4 (i.e., 3SHL [7] and tDPMM [4]).

Our approach achieves the best performance in classification. When detecting abnormal trajectories, our approach can also achieve obviously improved results than the compared methods (DTW [37], 3SHL [7], and GPRF [6]) and similar results to a state-of-the-art tDPMM method [4].

6.3 3D Action Recognition

Finally, we evaluate the performance of our approach in 3D action recognition. We perform experiments on a benchmark MSR-Action3D Dataset [8] which includes 557 3D skeleton sequences for 20 human actions performed by 10 different subjects. One example skeleton sequence is shown in Fig. 7.

Following the previous works on MSR-Action3D dataset [8], [40], we evaluate the recognition accuracy over all 20 actions where actions of half of the subjects are used for training and the rest actions are used for testing. Besides, a trajectory alignment process similar to [31] is applied as a pre-processing step to reduce 3D trajectory variations.

We utilize the process in Section 5.3 to implement our approach for 3D action recognition, where 'Droplet+KNN' and 'Droplet+SVM' in Table 5 refer to using KNN and SVM classifiers to recognize our droplet feature vectors (cf. (14)), respectively. Moreover, we also include the results by combining our droplet feature vector with a state-of-the-art 'Moving Poselets' method [41] which introduces sophisticated mid-level classifiers to improve recognition accuracy (cf. 'Droplet+Moving Poselets' in Table 5). Specifically, we concatenate our droplet feature vectors with the body point velocity & acceleration features used in [41], and follow the 'Moving Poselets' classification process [41] to recognize the action class of the concatenated feature vectors.

We compare our approach with the state-of-the-art 3D action recognition methods using skeleton sequences [32], [33], [34], [35], [40], [41], [42], [43]. Table 5 shows the recognition accuracy. According to Table 5, our 'Droplet+SVM'

TABLE 5
Recognition Accuracy Comparison on
MSR-Action3D Dataset (%)

Method	Recognition Accuracy
Eigen joints [42]	82.3
HON4D+D_{disc} [40]	88.9
Actionlet Ensemble [32]	88.2
Skeleton Quads [34]	89.9
Pose Set [43]	90.2
Manifold Learning [35]	91.2
Moving Pose [33]	91.7
Moving Poselets [41]	93.6
Droplet+KNN	91.2
Droplet+SVM	92.1
Droplet+Moving Poselets	**93.9**

approach outperforms all the existing techniques except [41]. This demonstrates that our tube-and-droplet framework can be reliably applied to handle sequence analysis with multiple trajectories. Besides, our 'Droplet+KNN' approach also achieves satisfactory results. It implies that our droplet features can effectively capture the discriminative characteristics of trajectories, such that good results can be achieved with simple recognition strategies such as KNN. Moreover, the 'Droplet+Moving Poselets' approach achieves the best performance. It further indicates that our droplet features can be effectively combined with more sophisticated recognition strategies to achieve further improved performances.

7 CONCLUSION

In this paper, we study the problem of informative trajectory representation and introduce a novel tube-and-droplet framework. The framework consists of three key ingredients: 1) introducing the idea of constructing thermal transfer fields to embed the global motion patterns in a scene; 2) deriving equipotential lines and concatenating them into a 3D tube to establish a highly informative representation, which properly embeds both the motion route and the contextual motion pattern for a trajectory; 3) introducing a simple but effective droplet-based process to effectively capture the rich information in 3D tube representation. We apply our tube-and-droplet approach to various trajectory analysis applications including clustering, abnormality detection, and 3D action recognition. Extensive experiments on benchmark demonstrate the effectiveness of our approach.

### REFERENCES

[1] X. Wang, K. Tieu, and E. Grimson, "Learning semantic scene models by trajectory analysis," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 110–123.
[2] J. J. Little and Z. Gu, "Video retrieval by spatial and temporal structure of trajectories," in *Proc. SPIE*, 2001, pp. 545–552.
[3] Y.-K. Jung, K.-W. Lee, and Y.-S. Ho, "Content-based event retrieval using semantic scene interpretation for automated traffic surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 2, no. 3, pp. 151–163, Sep. 2001.

[4] W. Hu, X. Li, G. Tian, S. Maybank, and Z. Zhang, "An incremental DPMM-based method for trajectory clustering, modeling, and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1051–1065, May 2013.

[5] H. Xu, Y. Zhou, W. Lin, and H. Zha, "Unsupervised trajectory clustering via adaptive multi-kernel-based shrinkage," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4328–4336.

[6] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1164–1171.

[7] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2287–2301, Nov. 2011.

[8] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. IEEE Conf. Comput. Vis. Patt. Recog. Workshops*, 2010, pp. 9–14.

[9] W. Hu, D. Xie, T. Tan, and S. Maybank, "Learning activity patterns using fuzzy self-organizing neural network," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 3, pp. 1618–1626, Jun. 2004.

[10] A. Naftel and S. Khalid, "Motion trajectory learning in the DFT-coefficient feature space," in *Proc. IEEE Int. Conf. Comput. Vis. Syst.*, 2006, pp. 47–47.

[11] T. Palpanas, M. Vlachos, E. Keogh, D. Gunopulos, and W. Truppel, "Online amnesic approximation of streaming time series," in *Proc. Int. Conf. Data Eng.*, 2004, pp. 339–349.

[12] W. Lin, H. Chu, J. Wu, B. Sheng, and Z. Chen, "A heat-map-based algorithm for recognizing group activities in videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1980–1992, Nov. 2013.

[13] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson, "Trajectory analysis and semantic region modeling using nonparametric hierarchical Bayesian models," *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 287–312, 2011.

[14] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *Int. J. Comput. Vis.*, vol. 50, no. 2, pp. 203–226, 2002.

[15] J. C. Nascimento, M. A. Figueiredo, and J. S. Marques, "Trajectory classification using switched dynamical hidden Markov models," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1338–1348, May 2010.

[16] M. Müller, "Dynamic time warping," *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007, pp. 69–84.

[17] M. Vlachos, D. Gunopulos, and G. Kollios, "Discovering similar multidimensional trajectories," in *Proc. Int. Conf. Data Eng.*, 2002, pp. 673–684.

[18] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, no. 1, pp. 37–52, 1987.

[19] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, 2000.

[20] Y. M. Lui, "Tangent bundles on special manifolds for action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 930–942, Jun. 2012.

[21] D. Gong, X. Zhao, and G. Medioni, "Robust multiple manifolds structure learning," in *Proc. Int. Conf. Mach. Learn.*, 2012, pp. 321–328.

[22] A. Goh and R. Vidal, "Segmenting motions of different types by unsupervised manifold clustering," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2007, pp. 1–6.

[23] E. Sangineto, "Pose and expression independent facial landmark localization using dense-surf and the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 624–638, Mar. 2013.

[24] Y. Gao, M. Wang, R. Ji, X. Wu, and Q. Dai, "3-D object retrieval with Hausdorff distance learning," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 2088–2098, Apr. 2014.

[25] S. Patankar, *Numerical Heat Transfer and Fluid Flow*. Boca Raton, FL, USA: CRC Press, 1980.

[26] R. H. Pletcher, J. C. Tannehill, and D. Anderson, *Computational Fluid Mechanics and Heat Transfer*. Boca Raton, FL, USA: CRC Press, 2012.

[27] J. N. Reddy, *An Introduction to Nonlinear Finite Element Analysis: With Applications to Heat Transfer, Fluid Mechanics, and Solid Mechanics*. Oxford, U. K.: Oxford Univ. Press, 2014.

[28] B. R. Munson, D. F. Young, and T. H. Okiishi, *Fundamentals of Fluid Mechanics*. New York, NY, USA: Wiley, 1990.

[29] U. Von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[30] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Tech.*, vol. 2, pp. 1–27, 2011.

[31] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3D skeletons as points in a Lie group," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 588–595.

[32] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning actionlet ensemble for 3D human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 914–927, May 2014.

[33] M. Zanfir, M. Leordeanu, and C. Sminchisescu, "The moving pose: An efficient 3D kinematics descriptor for low-latency action recognition and detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2752–2759.

[34] G. D. Evangelidis, G. Singh, and R. Horaud, "Continuous gesture recognition from articulated poses," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2014, pp. 595–607.

[35] R. Slama, H. Wannous, M. Daoudi, and A. Srivastava,, "Accurate 3D action recognition using learning on the Grassmann manifold," *Pattern Recognit.*, vol. 48, no. 2, pp. 556–567, 2015.

[36] M. W. Browne, "Cross-validation methods," *J. Math. Psychology*, vol. 44, no. 1, pp. 108–132, 2000.

[37] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDD Workshop*, 1994, vol. 10, no. 16, pp. 359–370.

[38] B. Morris and M. Trivedi, "Learning trajectory patterns by clustering: Experimental studies and comparative evaluation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 312–319.

[39] K. Fukunaga and L. Hostetler, "K-nearest-neighbor Bayes-risk estimation," *IEEE Trans. Info. Theory*, vol. 21, pp. 285–293, 1975.

[40] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4D normals for activity recognition from depth sequences," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2013, pp. 716–723.

[41] L. Tao and R. Vidal, "Moving poselets: A discriminative and interpretable skeletal motion representation for action recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshop*, 2015, pp. 61–69.

[42] X. Yang and Y. Tian, "Eigenjoints-based action recognition using naive-bayes-nearest-neighbor," in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshops*, 2012, pp. 14–19.

[43] C. Wang, Y. Wang, and A. Yuille, "An approach to pose-based action recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2013, pp. 915–922.
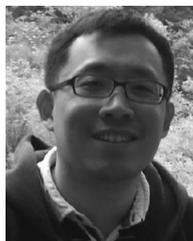
Weiyao Lin received the BE and ME degrees from Shanghai Jiao Tong University, China, in 2003 and 2005, and the PhD degree from the University of Washington, Seattle, in 2010. He is currently an associate professor in the Department of Electronic Engineering, Shanghai Jiao Tong University, China. His research interests include image/video processing, video surveillance, computer vision. He is a senior member of IEEE.
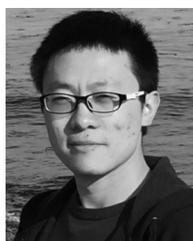
Yang Zhou received the BS degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013. He is currently working toward the ME degree in electrical engineering at Shanghai Jiao Tong University, China and Georgia Institute of Technology. His research interests include image processing & computer vision.

Hongteng Xu received the BS degree from Tianjin University, Tianjin, China, in 2010. In fall 2010, he joined the dual-master program in Georgia Institute of Technology and Shanghai Jiao Tong University, and graduated in Spring 2013. Currently, he is working toward the PhD degree in the School of Electrical and Computer Engineering, Georgia Tech. His research interests include computer vision, data mining and machine learning.
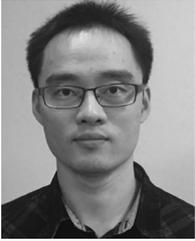
Junchi Yan received the MS and PhD degrees from Shanghai Jiao Tong University, China. He is currently a research staff member and master Inventor with IBM. He received the IBM Research Accomplishment and Outstanding Accomplishment Award in 2013, 2014. His research interests include computer vision and machine learning applications.
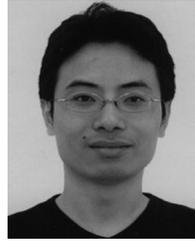
Jianxin Wu received the BS and MS degrees in computer science from Nanjing University, and the PhD degree in computer science from Georgia Institute of Technology. He is currently a professor with Nanjing University, China and was an assistant professor with Nanyang Technological University, Singapore. His research interests are computer vision and machine learning.

Mingliang Xu received the BE and ME degree from Zhengzhou University, China, in 2005 and 2008, respectively, and the PhD degree from State Key Lab of CAD&CG, Zhejiang University, China, in 2012, all in computer science. Currently, he is an associate professor in the School of Information Engineering, Zhengzhou University, China. His research interests include computer graphics and computer vision.

Zicheng Liu received the BS degree from Huazhong Normal University, China, in 1984, the MS degree from Institute of Applied Mathematics, Chinese Academy of Sciences, in 1989, and the PhD degree in computer science from Princeton University, in 1996, He is a principal researcher with Microsoft Research Redmond. Before joining Microsoft Research, he worked with Silicon Graphics Inc. His current research interests include human activity recognition, 3D face modelling, and multimedia signal processing. He is a fellow of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.