Representing and Recognizing Motion Trajectories: A Tube and Droplet Approach

Yang Zhou[1], Weiyao Lin[1*], Hang Su[1], Jianxin Wu[2], Jinjun Wang[3], Yu Zhou[4]

[1] Dept. of Electronic Engineering, Shanghai Jiao Tong University, China (*Corresponding Author)
[2] National Key Laboratory for Novel Software Technology, Nanjing University, China
[3] Institute of Artificial Intelligence and Robotics, Xi'an Jiao Tong University, China
[4] Institute of Information Engineering, Chinese Academy of Sciences, China

ABSTRACT

This paper addresses the problem of representing and recognizing motion trajectories. We first propose to derive scene-related equipotential lines for points in a motion trajectory and concatenate them to construct a 3D tube for representing the trajectory. Based on this 3D tube, a droplet-based method is further proposed which derives a "water droplet" from the 3D tube and recognizes trajectory activities accordingly. Our proposed 3D tube can effectively embed both motion and scene-related information of a motion trajectory while the proposed droplet-based method can suitably catch the characteristics of the 3D tube for activity recognition. Experimental results demonstrate the effectiveness of our approach.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: Video analysis

Keywords

Motion trajectory representation; Activity recognition; Droplet

1. INTRODUCTION

Motion trajectories are essential in describing object movement patterns over time. Visual analysis and recognition of motion trajectories are of considerable importance in many applications including video surveillance and video retrieval [1-6]. In this paper, we focus on the effective representation and recognition of motion trajectories.

First, properly representing motion trajectories is crucial in trajectory recognition. Many algorithms [4] directly utilized object locations at different time to represent object motion trajectories. However, this representation method is easily affected by the large variation of motion trajectories. To address this problem, Nascimento et al. [1] proposed low-level dynamic models to decompose trajectories into basic displacements. Chu et al. [2] created heat maps from trajectories to include temporal information while reducing the variances among trajectories. However, most of the existing methods only focus on modeling the absolute movements of objects while the scene-related information (e.g., whether the object is located in an unusual place in a scene) is not embedded. Thus, they will have limitations when recognizing scene-related abnormal activities such as entering an unusual place. Although some methods [3] utilized Gaussian process flows to model the location and velocity probability for each point in a trajectory, these flows are only used to measure the likelihood of a trajectory belonging to a specific trajectory type. Thus, they still do not actually embed information about the scene.

Second, accurately recognizing motion trajectories is another important issue. Some methods applied alignment or dynamic time warping to reduce trajectory variances and performed recognition by finding the most similar trajectory type [2, 4]. However, due to the large uncertainty of object motions, they still cannot completely avoid recognition errors from trajectory variances. Other methods constructed trajectory probability densities or graphical models for different activity patterns and recognized activities accordingly [1, 3, 10, 11]. However, these methods also have disadvantages in: (a) requiring large numbers of training data to achieve reliable probability models, and (b) have limitations in differentiating abnormal activities which only deviate slightly from normal ones [3]. Besides, Lin et al. [6] proposed to calculate transmission energies from trajectories for abnormality detection. Although this method can effectively detect abnormal activities, it still has limitations in differentiating normal activity patterns.

In this paper, we propose a new tube and droplet approach to represent and recognize motion trajectories. The contributions of our approach can be summarized in the following:

(1) We propose to derive scene-related equipotential lines for points in a motion trajectory and concatenate them together to construct a 3D tube for representing the trajectory. The proposed 3D tube can effectively embed both motion and scene-related information of a motion trajectory.

(2) We propose a droplet-based method which derives "water droplets" from 3D tubes and recognizes trajectory activities accordingly. The proposed droplet-based method can i) suitably catch the characteristics of 3D tubes, ii) effectively differentiate subtle normal/abnormal activities, iii) perform recognition with small number of training data.

The remainder of this paper is organized as follows. Section 2 describes the basic ideas of our approach. Section 3 presents the details of our 3D tube and droplet construction steps. Experimental results are shown in Section 4. Section 5 concludes the paper.

2. BASIC IDEAS

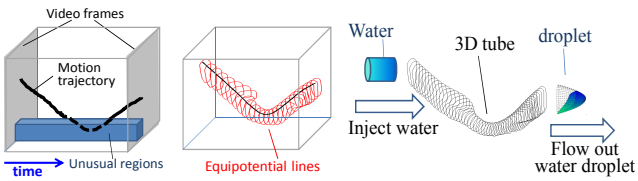
As mentioned, in order to properly handle scene-related activities, it is desirable to embed both object motion and scene-related information in trajectory representations. Thus, we propose to utilize 3D tubes to represent trajectories. As shown in Figure 1, (a) is an object motion trajectory, (b) is the equipotential lines for different points in the trajectory in (a), and (c) is a 3D tube representing the trajectory in (a). From Figure 1, we can see that a 3D tube is composed of equipotential lines where each equipotential line describes a point in a trajectory. More specifically, each equipotential line is located at its corresponding trajectory point while its shape is decided by the neighborhood

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
MM '14, November 03 - 07 2014, Orlando, FL, USA
Copyright 2014 ACM 978-1-4503-3063-3/14/11…$15.00.
http://dx.doi.org/10.1145/2647868.2654955

scene around the trajectory point (e.g., a trajectory point located in an unusual region will make its equipotential line shrink, as in Figure 1 and will be described in detail later). In this way, we can effectively embed both motion and scene-related information in a 3D tube with the route of the tube representing object motions and the shape of the tube representing information of the neighborhood scene.

Since 3D tubes include rich and high-dimensional information, the problem then comes to the selection of a suitable method for performing recognition based on these tube features. In this paper, we further propose a droplet-based method for activity recognition. In this method, we first "inject" water in one end of a 3D tube and then achieve a water droplet flowed out from the other end, as in Figure 1 (c). Since different activities are represented by 3D tubes with different shapes, by suitably modeling the water flow process, the flowed droplets can precisely catch the characteristics of 3D tubes. Thus, accurate recognition results can be achieved by parsing the shape of these droplets.

With the basic ideas of 3D tubes and water droplets, we can propose our motion trajectory representation and recognition approach. It is described in detail in the following section.



(a) (b) (c)

Figure 1. (a) Object trajectory; (b) Equipotential lines for different points in the trajectory in (a); (c) The final 3D tube representation and the process of droplet-based method.

3. THE APPROACH

The framework of our approach can be described in the following. First, equipotential lines are achieved for all points in the input trajectory. These equipotential lines are derived according to a pre-trained network which encodes the correlations among patches in the scene. Then, equipotential lines for different points are concatenated chronologically to construct a 3D tube for representing the input trajectory. After deriving a water drop by flowing water through the 3D tube, the activity of the input trajectory can be recognized by comparing this water drop with the trained drop shape patterns for different activities. In the following, we will describe details of the proposed approach.

3.1 3D Tube Construction

3.1.1 Network construction

In order to embed scene-related information into 3D tubes, we first need to construct a directed network to describe a scene. In this paper, we divide a scene into non-overlapping patches such that each patch can be viewed as a node in a network while the directed links between neighboring patches can be viewed as the directed edges in a network, as in Figure 2. With this network model, a trajectory moving in a scene can be modeled as a package transmitting through different nodes in its corresponding network. Thus, by properly constructing the edge weights in this network, the information of the scene can be effectively modeled.

In this paper, the directed edge weights between neighboring patches are calculated by:

$$W_{i\to j} = \sum_{k=1}^{N} R_{k,i\to j} \qquad (1)$$

where $W_{i\to j}$ is the weight for the directed edge from patch P_i to

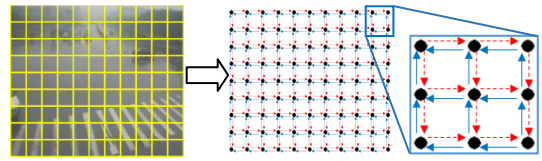


Figure 2. Separate a scene into patches and construct a directed network accordingly.

patch P_j. N is the total number of trajectories in the training set. $R_{k,i\to j}$ is the impact of the k-th training trajectory to edge $P_i \to P_j$. $R_{k,i\to j}$ can be calculated by:

$$R_{k,i\to j} = \sum_{r=1}^{L_k} \max\left(v_{k,r}(i\to j), 0\right) \cdot e^{-\|P_r - P_i\|^2} \qquad (2)$$

where L_k is the total number of points in trajectory k. $v_{k,r}(i \to j)$ is the displacement of the r-th point in trajectory k in the direction of $P_i \to P_j$, as in Figure 3. P_r is the patch where point r is located. $\|P_r - P_i\|^2$ is the distance between patches P_i and P_r.

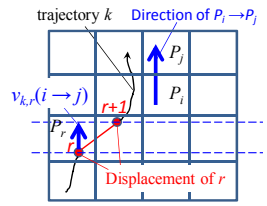


Figure 3. The way to calculate $v_{k,r}(i \to j)$.

From Eqs (1) and (2), we can see that the weight for a directed edge $P_i \to P_j$ is proportional to the total displacement strength of all training trajectory points in the direction of $P_i \to P_j$. Besides, a distance term $\|P_r - P_i\|^2$ is also included such that trajectory points closer to the directed edge will have more impact to the weight $W_{i\to j}$. In this way, if there are large numbers of training trajectories passing through P_i and following the direction of $P_i \to P_j$, a large weight will be assigned to $W_{i\to j}$ meaning that moving from P_i to P_j is normal. On the contrary, a small $W_{i\to j}$ will be assigned to indicate that following $P_i \to P_j$ is abnormal.

3.1.2 Achieving equipotential lines

With the constructed network, we can derive an equipotential line for each point in an input testing trajectory. More specifically, we first achieve an energy map for a trajectory point based on the constructed network, and then derive a constant-energy line from this energy map as the resulting equipotential line.

The energy map for a trajectory point r can be achieved by iteratively propagating energies outwards from the patch of r (i.e. P_r) to other patches in the scene. The trajectory point patch P_r is first allocated with an initial energy E_0=100 before propagation. And during each iteration, the energy propagated from a patch P_i to its outside neighboring patch P_j can be calculated by:

$$E_{i\to j} = E_i \cdot e^{-\frac{\gamma}{W_{i\to j}}} \qquad (3)$$

where $E_{i\to j}$ is the energy transmitted from P_i to P_j. E_i is the energy in patch P_i. $W_{i\to j}$ is the weight for the directed edge $P_i \to P_j$ achieved by Eq. (1). And γ is a constant. From Eq. (3), it is clear that fewer energies will be propagated if the motion from P_i to P_j is abnormal (i.e., small $W_{i\to j}$).

Figure 4 (b) and (d) show the energy maps of two points r_a and r_b in a trajectory in (a). In (b), since moving rightward from r_a is normal (because there are lots of dashed blue training trajectories moving rightward around r_a), more energies can be propagated to this direction, thus leading to a long rightward tail

in r_a's energy map. Comparatively, in (d), since r_b is located in an unusual/abnormal region (because there is no training trajectories moving around r_b), few energies can be propagated out, thus making r_b's energy map decay quickly around r_b.

With the energy map, the equipotential line can be easily achieved by finding a constant-energy line in the map. In this paper, we find the line whose energy is half of the peak energy E_0, as in Figure 4.

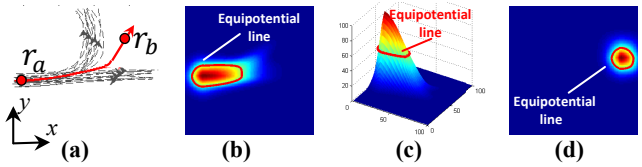


Figure 4. (a) The input trajectory (solid red line) and training trajectories (dashed black lines); (b) and (d): The energy maps and equipotential lines for trajectory points r_a and r_b in (a); (c): The energy map surfaces of (b) in 3D.

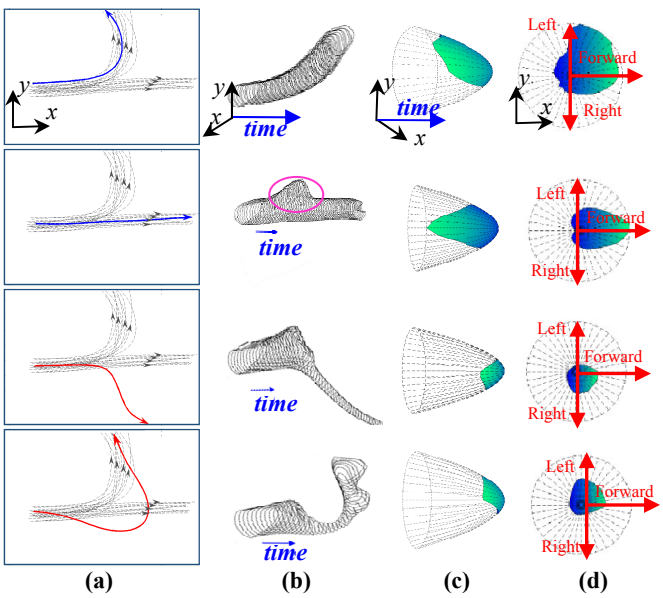


Figure 5. (a) Input trajectories (red/blue solid lines) and training trajectories (black dashed lines); (b) 3D tubes for the input trajectories in (a); (c) Droplets derived from the 3D tubes in (b); (d) Viewing the droplets in (c) from the left side.

3.1.3 Constructing 3D tubes

After deriving equipotential lines for each point in an input trajectory, a 3D tube can be constructed to represent this trajectory by concatenating all equipotential lines according to their temporal order in the trajectory. Figure 5 (b) shows the 3D tubes of the trajectories in (a). From Figure 5, we can see that:

(1) Our proposed 3D tubes include rich information about both trajectories' motion and their relationships with the surrounding scene. For example, the routes of 3D tubes represent actual motions of trajectories. The thicknesses of 3D tubes indicate whether trajectories are normal/abnormal (e.g., tubes will become narrow for abnormal trajectories). Moreover, the shapes of equipotential lines inside 3D tubes also indicate possible normal motion directions (e.g., the convex part in the pink circle in (b) indicates that moving upward (i.e., turn left) is another possible normal motion direction). Therefore, various trajectory activities can be effectively recognized with our 3D tubes.

(2) Compared with the 2D trajectory representation in (a), our 3D tube representation includes an additional time dimension. And this also has the advantage of properly displaying the object's motion speed information such as "stop" and "speed up".

Note that we also normalize 3D tubes to have the same length in order to avoid the influence of different trajectory lengths.

3.2 The Water-Droplet Method

After constructing 3D tubes for input trajectories, we need to find suitable methods to effectively utilize these high-dimensional tube features. In this paper, we propose to inject water into 3D tubes and derive "water droplets" to detect activities, as Figure 1.

Basically, a droplet can be described by a center point r and a set of boundary points h_m, as in Figure 6 (a). We define r as the point which follows the route of the input trajectory when passing through a 3D tube. And h_m is the point following the route constructed by the points on the m-th direction of r on all equipotential lines, as in Figure 6 (b). Assuming that r passes a 3D tube with a constant velocity while the velocity of h_m is changed according to the thickness and shape of the 3D tube, by calculating the relative distances between h_m and r at the tube output, the features of the 3D tube can be effectively captured.

In this paper, we define two distances at a tube output: H_m which is the distance between h_m and r in the time dimension, and D_m which is the distance between h_m and r in the x-y plane, as in Figure 6 (a). In order to simplify calculation, we further assume $D_m = H_m$ such that a 3D droplet can be simply described by H_m in a 2D plane, as in Figure 5 (d). According to the fluid viscosity theory [7], H_m can be calculated by:

$$H_m = \frac{1}{L}\sum_{t=1}^{L} v_{m-r,t} = \frac{1}{L}\sum_{t=1}^{L} \frac{d_{m,t}^2}{\mu(d_{m,t}, \theta_{m,t})} \qquad (4)$$

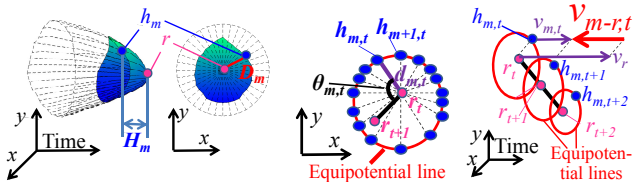
where L is the length of a 3D tube. $d_{m,t}$ is the distance between r_t and $h_{m,t}$, and $\theta_{m,t}$ is the angle between lines $r_t \rightarrow h_{m,t}$ and $r_t \rightarrow r_{t+1}$, as in Figure 6 (b) and (c). Note that $h_{m,t}$ and r_t are the locations of h_m and r at time t when going through a 3D tube (i.e., r_t is the t-th point in a trajectory and h_m is the m-th point on the equipotential line of r_t). $v_{m-r,t} = \frac{d_{m,t}^2}{\mu(d_{m,t},\theta_{m,t})}$ is the relative time-dimensional velocity between h_m and r at time t. $\mu(d_{m,t}, \theta_{m,t})$ is the viscosity coefficient of the tube calculated by:

$$\mu(d_{m,t}, \theta_{m,t}) = \frac{\beta}{d_{m,t} \cdot (\cos(\theta_{m,t}) + \lambda)} \qquad (5)$$

where λ and β are constant values.

From Eqs (4) and (5), we can see that the time-dimensional distance H_m is calculated by accumulating the time-dimensional velocity difference between h_m and r when passing through a 3D tube. The velocity difference $v_{m-r,t}$ is decided by the thickness and shape variations in a 3D tube. When a tube becomes thick (i.e., large $d_{m,t}$ and correspond to normal motions), $v_{m-r,t}$ will be enlarged and create large H_m. Besides, when h_m is located along the motion direction of a trajectory (i.e., the angle between $r_t \rightarrow h_{m,t}$ and $r_t \rightarrow r_{t+1}$ is small), the viscosity from a tube will be small, leading to large H_m.

Figure 5 (c) and (d) show the resulting droplets of the 3D tubes in (b). From Figure 5 (c) and (d), we can observe the effectiveness of our water-droplet method. For example, the droplets of normal tubes (the first and second rows in Figure 5) have large sizes than the droplets of abnormal tubes (the last two rows). Also, the droplets have large sectors in the trajectories' major motion directions. Thus, we can effectively use these droplet features to recognize both normal/abnormal activities and activities of different motion patterns. In our experiments, we construct feature vectors from droplets by concatenating H_m and utilize linear SVMs [8] to recognize activities.



(a) (b) (c)

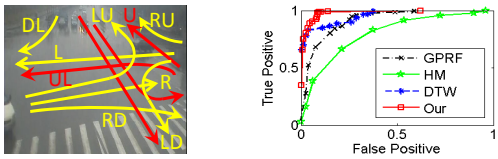
Figure 6. (a) Center point r, boundary point h_m, and their distances; (b) Illustration of $d_{m,t}$ and $\theta_{m,t}$ at time t; (c) $h_{m,t}$ and r_t at different t and relative velocity $v_{m-r,t}$.

4. EXPERIMENTAL RESULTS

In this section, we show experimental results for our proposed tube and droplet approach. In our experiments, the patch size is set to be 10×10, and γ, λ, and β in Eqs (3) and (5) are set to be 1, 5, and 1, respectively. We first perform experiments on a road dataset that we constructed. The dataset includes 300 trajectories obtained by a tracking method [9] 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), as in Figure 7 (a). Besides, some example abnormal trajectories are also displayed in Figure 7 (a). Note that this is a challenging dataset in that: (1) the total number of trajectories in the dataset is small; (2) The motion trajectories within the same class have large variance; (3) Many trajectories from different class are confusing and are difficult to differentiate.

We compare our approach with three methods: the GPRF method [3], the DTW method [4], and the heat-map (HM) method [2]. We split the dataset into 75% training-25% testing parts and perform recognition on the testing part [6]. Four independent experiments are performed where the training and testing sets are randomly selected in each experiment. And the results are averaged. Figure 7 (b) compares the ROC curves for different methods in recognizing normal/abnormal activities. Besides, Table 1 further compares the Miss, False Alarm (FA), and Total Error Rates (TERs) [6] for different methods in differentiating the seven normal activity patterns and the abnormal activity pattern.

From Table 1 and Figure 7 (b), we can see that the compared methods, which mainly perform recognition according to the similarity among trajectories' motion patterns, have low effectiveness when recognizing activities such as "L", "RU", and "RD" since their trajectories are easily confused with other similar activities such as "UL", "U", "R". Comparatively, our proposed approach can achieve obviously better results where the confusing activities are properly differentiated. This is because: (1) our 3D tube representation embeds scene-related information by using narrow tube sections to represent local abnormal displacements. Thus, abnormal activities can be precisely differentiated even if their overall trajectory shapes are similar to normal activities; (2) Our droplet features can catch the subtle motion direction by accumulating velocity differences between boundary and center points in 3D tubes. Thus, the confusing activities such that "R" and "RD" can also be effectively differentiated by our approach.



(a) (b)

Figure 7. (a) Examples of normal activities (yellow) and abnormal activities (red); (b) ROC curves of different methods in abnormality detection.

Table 1 Miss, FA, and TER rates in recognizing different normal patterns together with the abnormal activity pattern

		Our	GPRF [3]	DTW [4]	HM [2]
R	Miss	**4.2%**	9.7%	15.3%	8.2%
	FA	**0.5%**	1.4%	1.7%	0.7%
L	Miss	**3.8%**	7.1%	9.3%	11.2%
	FA	**0.2%**	0.5%	1.0%	1.8%
RD	Miss	**3.5%**	10.2%	18.5%	12.4%
	FA	**0.3%**	0.7%	1.4%	1.0%
LU	Miss	**2.6%**	0.3%	0.5%	2.6%
	FA	**0.5%**	0.2%	0.0%	0.4%
LD	Miss	**1.2%**	3.3%	14.1%	8.0%
	FA	**0.2%**	0.2%	0.6%	0.2%
RU	Miss	**2.4%**	14.7%	13.1%	17.9%
	FA	**0.5%**	1.8%	1.5%	2.1%
DL	Miss	**1.3%**	2.5%	1.4%	2.4%
	FA	**0.0%**	0.2%	0.0%	0.4%
Abnormal	Miss	**8.8%**	17.4%	24.3%	23.1%
	FA	**7.9%**	17.0%	26.7%	21.2%
TER		**5.3%**	10.7%	15.4%	14.1%

5. CONCLUSION AND FUTURE WORKS

In this paper, we propose to construct a 3D tube for representing a motion trajectory and then derive a "water droplet" from the 3D tube to recognize the trajectory. Experimental results demonstrate the effectiveness of our approach. Future work includes more experiment on public datasets and comparison with other graphical model-based methods [10, 11].

6. ACKNOWLEDGMENTS

This paper is supported in part by grants: Shanghai Pujiang Program (12PJ1404300), National Science Foundation of China (61471235, 61001146, 61025005, 61303170), Chinese national 973 grants (2010CB7314016), and SMC scholarship of SJTU.

REFERENCES

[1] J. Nascimento, M. Figueiredo, and J. Marques, "Trajectory classification using switched dynamical hidden Markov models," *IEEE Trans. Image Proc.*, pp. 1338-1348, 2010.

[2] H. Chu, W. Lin, J. Wu, X. Zhou, Y. Chen, and H. Li, "A new heat-map-based algorithm for human group activity recognition," *ACM Multimedia* (*MM*), pp. 1069-1072, 2012.

[3] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," *ICCV*, 2011.

[4] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *IJCV*, 2002.

[5] D. Ellis, E. Sommerlade, and I. Reid, "Modelling pedestrian trajectory patterns with gaussian processes," *Int. Conf. Computer Vision Workshops*, pp. 1229-1234, 2009.

[6] W. Lin, Y. Chen, J. Wu, H. Wang, B. Sheng, and H. Li, "A new network-based algorithm for human activity recognition in videos," *IEEE Trans. CSVT*, 24(5), pp. 826-841, 2014.

[7] V. Streeter, E. Wylie, and K. Bedford, "Fluid Mechanics," *McGraw-Hill press*, 1987.

[8] C. Chang and C. Lin. "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Tech.*, pp. 1-27, 2011.

[9] R. Hess and A. Fern, "Discriminatively trained particle filters for complex multi-object tracking," *CVPR*, 2009.

[10] X. Wang, K. T. Ma, G. Ng, and E. Grimson, "Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models, " *IJCV*, 2011.

[11] B. Morris and M. Trivedi, "Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach," *IEEE Trans. PAMI*, 2011.