

Description and Recognition of Activity Patterns Using Sparse Vector Fields*

Ana Portêlo¹, Andrea Cavallaro², Catarina Barata³, and Jorge S. Marques³

¹ INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

² Centre for Intelligent Sensing, Queen Mary University of London, UK

³ Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract. Far-field activities represented as time series or trajectories can be summarized in compact representations of frequent patterns. Popular representations such as clustering or probabilistic modeling of trajectories often do not inform about both velocity and direction of motion, which are by definition visually and quantitatively embedded in vector fields. However, a common use of vector fields may dismiss information about forbidden areas, or regions with concurrent activity patterns. To address this problem we present a non-iterative layered vector field estimation process that yields sparse vector field abstractions of activity patterns from groups of trajectories. The key feature of our approach is the estimate of the probability density function (PDF) of targets positions: it automatically tunes the cost function parameter, and serves as weights in the sparse estimation problem. We also propose a trajectory labeling algorithm that labels trajectories according to their activity patterns using the vector field abstractions. Experiments in synthetic and real trajectory data show that the proposed estimation approach yields correctly sparse vector fields, which are similar to known generating vector fields, and 5-12% higher labeling accuracy on test trajectories when compared to other generative models. Outlier trajectories are also detected.

Keywords: Activity patterns, Vector fields, Sparse optimization, Trajectory Labeling, Oulier detection

1 Introduction

Far-field activities can be described by trajectories [10, 4], their spatial and angular features [9, 1], or by generative models [2, 12, 6, 11]. A high-level description of the spatial distribution of frequent activity patterns can support anomaly detection [8] and accessibility planning [5], and encode semantic regions of the scene [13].

* This work was partially supported by FCT under project SPARSIS - Sparse Modeling and Estimation of Vector Fields, contract PTDC/EEIPRO/0426/2014 and plurianual funding UID/EEA/50009/2019 and UID/CEC/50021/2019 (INESC-ID).

In the literature, both diffusion and probabilistic models have been used to describe frequent activity patterns and to label trajectories according to activity patterns. On the one hand, *diffusion* models such as optical flow models have been used to describe coherent motions and semantic regions [14], and heat maps based on thermal diffusion processes have been used to capture the temporal motion information of activities [7]. On the other hand, *probabilistic* models such as Hidden Markov Models have been used to detect the points of interest in a scene [11], and Dirichlet Process Mixture Models have been used to robustly label new trajectories according to their activity patterns [6]. These studies either focused on attributing a single semantic meaning to specific spatial regions [14, 11], or on labelling trajectories according to their activity patterns [7, 6]. However, neither of them provide straightforward information about velocity and direction of the activity patterns, and only the latter are able to describe different global activity patterns in the same region of the scene. Typically, this type of motion information is associated to vector fields which embed both the physical meaning of motion (*i.e.*, velocity and direction) and the semantic interpretation of different regions of the scene [12, 5].

We propose to (i) describe multiple, complex activity patterns using layered vector fields and (ii) label test trajectories according to activity patterns using the estimated vector fields. Our approach extends that of previous studies in that: (a) it imposes data-driven sparsity to the vector field abstractions to prevent erroneous extrapolations in regions with no target data (contrarily to [5, 12]); (b) it is sensitive to concurrent activity patterns (contrarily to [14, 11]); and (c) it provides information on the velocity and direction of activity patterns (contrarily to [7, 6]).

The proposed vector field estimation uses a cost function specifically designed to yield sparse estimates. Contrarily to other studies, which induce sparsity of the vector field estimates through the l_1 -norm [3], this work does so through statistical conditioning on available data – the estimated spatial Probability Density Function (PDF) of the targets positions restricts vector field estimates to the regions where targets are observed. The proposed cost function further benefits from automatic parameter tuning using targets positions and trajectory features. Layered vector field abstractions can be obtained if the proposed approach is applied on pre-clustered trajectories with similar activity patterns. We assess the accuracy of the vector field abstractions of synthetic trajectories by comparing the estimated and generating vector fields, and the correct sparsity by comparing the estimated vector fields in regions with no target data with the null vector field. Vector field comparisons focus on the mean vector length (RMSL) and the vector similarity coefficient (R) [16].

Moreover, we propose a trajectory labeling algorithm according to activity patterns. The displacement error between test trajectories and generated trajectories using the estimated vector fields is the measure for classification. This way, test trajectories are sorted according to activity patterns or detected as outliers. We assess the accuracy of the trajectory labeling algorithm by comparing the attributed and the observed activity pattern labels of test trajectories.

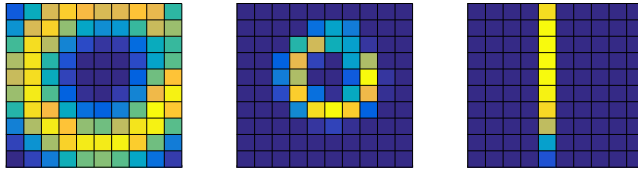


Fig. 1. Probability density functions of the targets positions for 3 different activity patterns from the synthetic data set (D_0). Gradient colours of ascending density of data points from dark blue to yellow.

2 Estimation of Multiple Vector Fields

We first aim to estimate the vector field, \mathbf{T} , that describes the activity patterns of a set of S trajectories, $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_S\}$. Let $t = 1, \dots, L_s$ and the target position, $\mathbf{x}_s(t)$, in the image plane of a camera be *driven* by \mathbf{T} according to

$$\mathbf{x}_s(t) = \mathbf{x}_s(t-1) + \mathbf{T}(\mathbf{x}_s(t-1)) + \mathbf{w}_s(t), \quad (1)$$

where $\mathbf{w}_s(t) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \forall t$, is a white random perturbation. Let the image plane be normalized, thus $\mathbf{x}_s(t) \in [0, 1]^2 \forall t$.

The vector field, $\mathbf{T} : [0, 1]^2 \rightarrow \mathbb{R}^2$, is defined only at the grid nodes of an over-imposed regular, uniform grid, $\mathcal{G} = \{\mathbf{g}_n \in [0, 1]^2, n = 1, 2, \dots, N\}$, on the image plane. As the target trajectories can be defined in any image coordinate, even if it does not correspond to a grid node ($\mathbf{x}_s(t) \notin \mathcal{G}$), we bilinearly interpolate to represent the vector field that drives the target position on any coordinate of the image plane, $\mathbf{x}_s(t) \notin \mathcal{G}$:

$$\mathbf{T}(\mathbf{x}_s(t)) = \sum_{n=1}^N \phi_n(\mathbf{x}_s(t)) \mathbf{t}_n, \quad (2)$$

where $\phi_n(\mathbf{x}_s(t))$ are the interpolation coefficients of the velocity vectors, \mathbf{t}_n , at the grid nodes. The matrix of interpolation coefficients for set \mathcal{X} is Φ .

Vector field estimation corresponds to an optimization problem where \mathbf{T} is the minimizer of a given cost function that has to induce data-driven sparsity. To impose *sparsity* of the vector field estimates in the regions where target data does not exist, the velocity vectors in \mathbf{T} are weighted by 1 minus the spatial probability density function (PDF) of targets positions, *i.e.*, $\mathcal{D} = \mathbb{1} - \Gamma_p$, $\Gamma_p \in \mathbb{R}^N$. Γ_p is the estimated PDF of the targets positions using the Parzen window algorithm over set \mathcal{X} . Then, to get its value at the grid nodes, we discretize at the desired image coordinates (Fig. 1).

The cost function is therefore defined as

$$f(\mathbf{T}) = \|\mathbf{V} - \mathbf{T} \Phi\|_2^2 + \alpha \|\mathbf{T} \circ \mathbb{1} \mathcal{D}^\top\|_2^2, \quad (3)$$

where $\mathbb{1}$ is of size $[2 \times 1]$, $\|\cdot\|_2$ defines the l_2 -norm of a vector, “ \circ ” represents the Hadamard product, and $\mathbf{T} \in \mathbb{R}^{2 \times N}$, $\mathbf{V} \in \mathbb{R}^{2 \times M}$, $\Phi \in \mathbb{R}^{N \times M}$, $M = \sum_{s=1}^S (L_s - 1)$,

are given by

$$\mathbf{T} = [\mathbf{t}_1 \dots \mathbf{t}_N], \quad (4)$$

$$\mathbf{V} = [\mathbf{v}_1(2) \dots \mathbf{v}_1(L_1) \mid \dots \mid \mathbf{v}_S(2) \dots \mathbf{v}_S(L_S)], \quad (5)$$

$$\Phi = \left[\begin{array}{ccc|ccc} \phi_1(\mathbf{x}_1(1)) \dots \phi_1(\mathbf{x}_1(L_1 - 1)) & & & \phi_1(\mathbf{x}_S(1)) \dots \phi_1(\mathbf{x}_S(L_S - 1)) & & \\ & \vdots & & & \vdots & \\ \phi_N(\mathbf{x}_1(1)) \dots \phi_N(\mathbf{x}_1(L_1 - 1)) & & & \phi_N(\mathbf{x}_S(1)) \dots \phi_N(\mathbf{x}_S(L_S - 1)) & & \end{array} \right], \quad (6)$$

with matrix \mathbf{V} composed of the velocity vectors between consecutive target positions, *i.e.*, $\mathbf{v}_s(t) = \mathbf{x}_s(t) - \mathbf{x}_s(t-1)$.

In Eq. (3), α is correlated with the grid resolution N . To avoid manual parameter input in every estimation procedure, we propose an *automatic tuning* based on the cardinality of the non-zero elements (*i.e.*, $|\cdot|_{\neq 0}$), expected value (*i.e.*, $\mathbb{E}[\cdot]$), and standard deviation [*i.e.*, $\sigma(\cdot)$] of the estimated PDFs of target and trajectory features,

$$\alpha = 1 - \frac{|\Gamma_p|_{\neq 0}}{N}, \quad (7)$$

$$N = \max \{N_{\min}, |\Gamma_c|_{\neq 0} > \mathbb{E}[\Gamma_c] + 1.5 \sigma(\Gamma_c)\}, \quad (8)$$

where Γ_p is the spatial PDF of the target positions as before; $\Gamma_c \in \mathbb{R}^N$ is the average curvature of the trajectories at the grid nodes, which is estimated using the velocity angles $\theta(t) = \tan^{-1}\left(\frac{y(t)-y(t-1)}{x(t)-x(t-1)}\right)$ [8, 1]; and N_{\min} is the minimum grid resolution selected by the user. In (8), very curly trajectories (*i.e.*, extreme values of the distribution of Γ_c) define the grid resolution.

Multiple vector fields can be estimated using (3) if it is applied to each set of pre-clustered trajectories (\mathcal{X}_k) with similar activity patterns, *e.g.* using multiple features [1]. In the following, we assume that the pre-clustering step has taken place and that we have access to the sets \mathcal{X}_k .

3 Activity Pattern Labeling

Our second aim is to label trajectories according to their activity patterns. To achieve this aim, we first estimate the \mathbf{T}_k following the above approach and using only trajectories from training sets \mathcal{X}_k . Then, we propose the following labeling algorithm:

1. Trajectory labeling:
 - (a) Generate trajectories from the starting point of a given test trajectory using the estimated \mathbf{T}_k and (1);
 - (b) Compute the displacement error as the euclidean distance between the generated and the test trajectories;
 - (c) Label each test trajectory with the activity pattern (vector field abstraction) that yields the smallest displacement error.

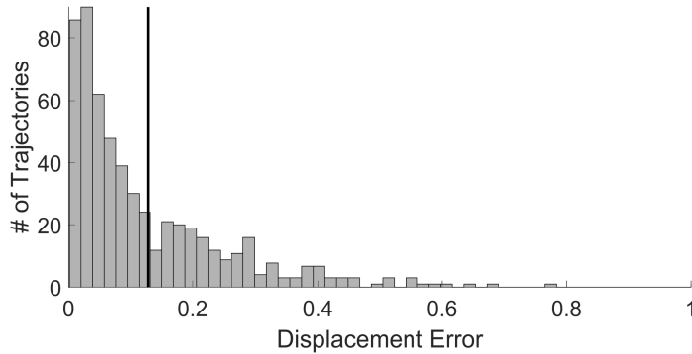


Fig. 2. Histogram of displacement errors between validation trajectories and generated trajectories using the known labels of activity patterns (vector fields). The cutoff threshold is shown in black. The validation trajectories used in this example come from the real data set D_3 .

2. Outlier detection using threshold:

- (a) Compute the cutoff threshold as the sum of the median and the median absolute deviation (MAD) of the displacement errors obtained from steps 1.(a) and 1.(b) applied on a set of validation trajectories (Fig. 2);
- (b) Label test trajectories as outliers of the labeled activity pattern from step 1.(c) if their displacement error is above the threshold.

In the Outlier detection step above, the cutoff threshold for outlier detection is the sum of the median and median absolute deviation of the displacement errors. We use the median and its absolute deviation instead of the mean and standard deviation given that the distribution of displacement errors is right skewed.

4 Experimental Results

4.1 Synthetic data

Assessment measures. Estimates of vector fields using synthetic data (\mathbf{T}_{est}) are assessed regarding both the *accuracy* when compared to the known generating vector field (\mathbf{T}_{ref}) and the correct *sparsity* compared to the null vector field (\mathbf{T}_0) in regions where no target data is observed. Let each node on the over-imposed grid be labeled according to its proximity to a given trajectory as an *active node*, if it belongs to a square of nodes containing part of a given trajectory, or a *non-active node*, if it does not belong to such a square of nodes. Thus, the region where no target data is observed is defined as the set of non-active nodes in the image plane, \mathcal{Z} , with respect to a given trajectory set.

The assessment measures compare pairs of vectors regarding the vector similarity coefficient (R), *i.e.*, the mean of the inner product of normalized vector

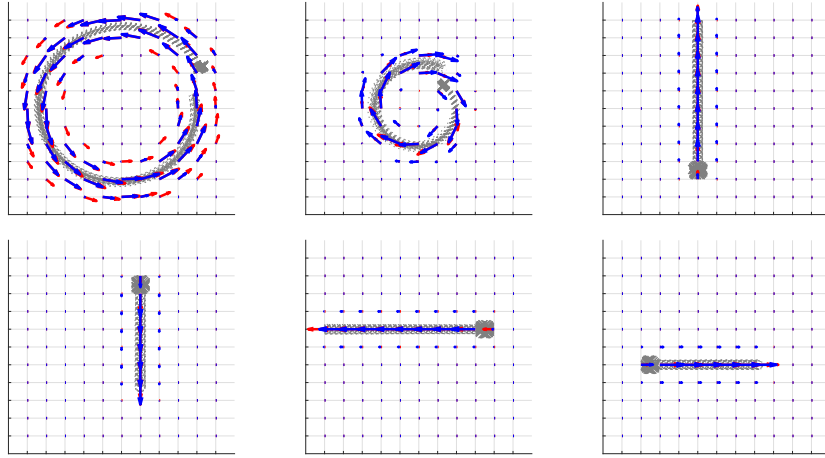


Fig. 3. Vector field estimates for the 6 activity patterns of the synthetic data set (D_0). Activity patterns are ordered across rows from left to right (top: 1, 2, 3; bottom: 4, 5, 6). Generating vector fields are shown in red, estimated vector fields in blue, generated trajectories in gray.

pairs from 2 vector fields A and B , defined as [16],

$$R = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \hat{\mathbf{t}}_i^A \cdot \hat{\mathbf{t}}_i^B; \quad (9)$$

and the vector root mean square length (RMSL), *i.e.*, the systematic difference in the mean vector length, defined as [16],

$$\text{RMSL} = L_V^2 = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \left\| \mathbf{t}_i^A - \mathbf{t}_i^B \right\|_2^2; \quad (10)$$

where “ \cdot ” is the inner product, $\hat{\mathbf{t}} = \frac{\mathbf{t}}{\sqrt{\|\mathbf{t}\|_2^2}}$, $\|\cdot\|_2$ represents the l_2 -norm of a vector, and $|\cdot|$ represents the cardinality of a set. In the case of *accuracy* assessment, $A = \mathbf{T}_{\text{ref}}$, $B = \mathbf{T}_{\text{est}}$, and $\mathcal{P} = \mathcal{G}$, the set of grid nodes. In the case of *correct sparsity* assessment, $A = \mathbf{T}_0$, $B = \mathbf{T}_{\text{est}}$, and $\mathcal{P} = \mathcal{Z}$, the set of non-active nodes. The optimal values for these measures are $(R, \text{RMSL}) = (1, 1)$ and $\text{RMSL} = 0$, respectively for accuracy and sparsity assessments.

Data set. The synthetic data set (D_0) has 300 trajectories generated using 6 different activity patterns. We use D_0 as a proof of concept for the assessment of accuracy and correct sparsity of the estimated vector fields.

Results. Fig. 3 shows that the vector field estimates are very similar to the generating vector fields not only in terms of magnitude (RMSL) and direction

Table 1. Overall accuracy of trajectory labeling for the proposed approach and comparison with literature results [6, 14, 11].

Data Set	activity patterns	traj. per pattern	Overall Accuracy (%)	
			Ours	Literature
D_0	6	50	100.00	–
D_1	15	18–278	87.81	[82.00, 86.70]
D_2	11	20	95.45	[86.30, 91.80]
D_3	19	100	99.47	[95.00, 98.00]

(R) but also regarding sparsity in the regions with no data, as expected. The accuracy of the estimated vector fields (activity patterns) is respectively (R , RMSL) **1:** (0.920, 0.663); **2:** (0.952, 2.305); **3:** (0.982, 0.979); **4:** (0.973, 0.723); **5:** (0.954, 9.787); **6:** (0.968, 0.763), and the RMSL of the vector field estimates corresponds to sparse vector fields, *i.e.*, all below $2.94\text{e-}04$.

4.2 Real data

Assessment measures. Activity pattern labeling accuracy is computed by comparing attributed and known trajectory labels taking into account the cutoff threshold as

$$Acc = \frac{\sum \text{diag}(\mathcal{M})}{\sum_{ij} \mathcal{M}_{ij}}, \quad (11)$$

where \mathcal{M} is the confusion matrix in a problem with multiple activity patterns.

Data set. The real data sets we used are: D_1 (Hu), containing 1500 trajectories with 15 activity patterns [6]; D_2 (Wang), containing 220 trajectories with 11 activity patterns [14]; D_3 (Morris), containing 1900 trajectories with 19 activity patterns [11]. We use these data sets to assess activity pattern labeling and outlier detection.

Results. Table 1 shows that overall the proposed algorithm correctly labels trajectories according to their activity patterns with an accuracy above that described in the literature. More specifically with higher accuracy than Heat-map, HMM, and DPMM, which are comparable generative models used to describe activity patterns [6, 14, 11, 15].

Regarding outlier detection, note that the proposed algorithm always assigns an activity pattern to a trajectory – the attributed activity pattern is the one that generates trajectories with the smallest displacement error relative to the test trajectory. However, if the displacement error is above the threshold the respective trajectory is plotted in a different colour than the others and tagged as an outlier. Fig. 4 shows examples of 2 activity patterns for each real data set, which have similar motion patterns but different semantic meanings. Concerning Activity Pattern I, only D_1 has outlier trajectories from two other activity patterns (shown in different colors, Fig. 4 middle row). Concerning Activity Pattern

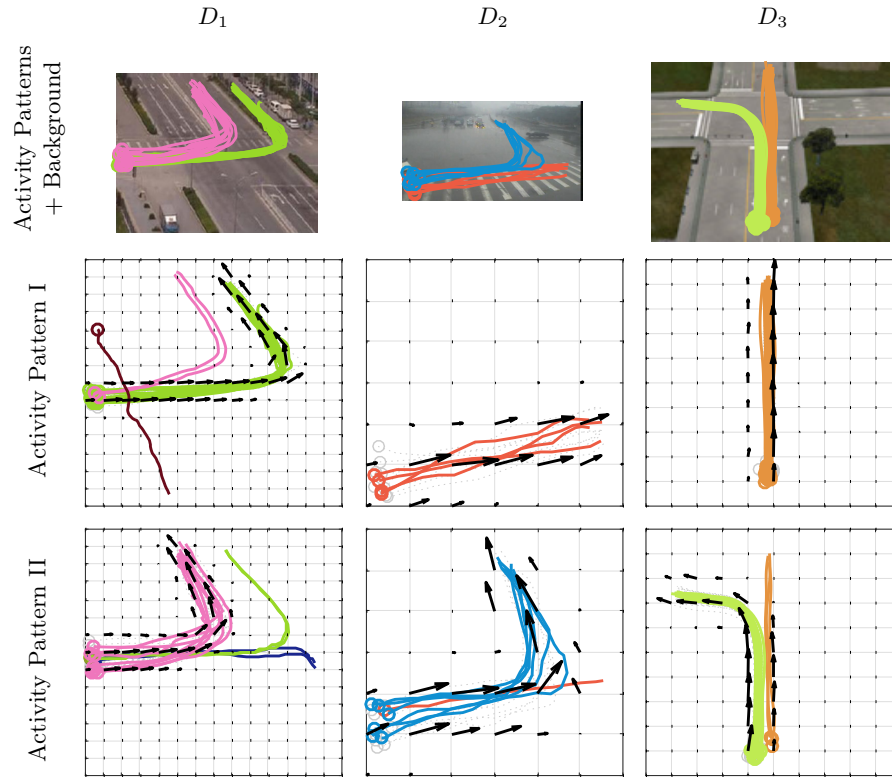


Fig. 4. Examples of trajectory labeling using the real data sets (D_1 , D_2 , and D_3), only two example activity patterns are shown for each data set. *Top row.* Overview of background image and test trajectories. *Middle row.* Correctly labelled trajectories and over-imposed estimated vector fields (black arrows) for Activity Pattern I. *Bottom row.* Correctly labelled trajectories, over-imposed estimated vector fields (black arrows), and outliers (shown in different colors) for Activity Pattern II.

II, all the data sets have outlier trajectories from Activity Pattern I, and D_1 also has outlier trajectories from one additional activity pattern (Fig. 4 bottom row).

The proposed approach yields vector field abstractions that can distinguish between similar activity patterns with different underlying semantics, given that for each data set, trajectories which were wrongly labelled as having one activity pattern were correctly detected as outliers of that activity pattern. For example, the green outlier trajectories from D_1 (Fig. 4 bottom left panel) are detected as outliers of that activity pattern – whereas the vector field of interest describes a left turn into the primary road, the outlier trajectories correspond to targets that instead performed a left turn into a secondary road. Similar examples for the other two data sets are shown in the bottom row of Fig. 4.

5 Conclusion

We proposed a vector field estimation approach that copes with dense trajectory data and yields compact abstractions of frequent activity patterns. The proposed approach abstracts frequently observed activity patterns and embeds data-driven sparsity, through the estimated spatial Probability Density Function (PDF) of the targets positions. Moreover, it informs about the physical and semantic meaning of the observed activity patterns. Finally, the estimated vector fields can be used to label new trajectories and detect outliers according to their activity pattern, with an improvement of about 5-12% on trajectory labeling accuracy when compared to other generative models.

References

1. Anjum, N., Cavallaro, A.: Multifeature object trajectory clustering for video analysis. *IEEE Transaction on Circuits and Systems for Video Technology* **18**(11), 1555–1564 (2008)
2. Barão, M., Marques, J.S.: Gaussian Random Vector Fields in Trajectory Modelling. In: *Proc. of the 19th Irish Machine Vision and Image Processing Conf. (IMVIP)*. pp. 211–216 (2017)
3. Barata, C., Nascimento, J.C., Marques, J.S.: A sparse approach to pedestrian trajectory modeling using multiple motion fields. In: *IEEE Int. Conf. on Image Processing (ICIP)*. pp. 2538–2542 (2017)
4. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(4), 743–761 (2012)
5. Ferreira, N., Klosowski, J.T., Scheidegger, C.E., Silva, C.T.: Vector Field k-Means: Clustering Trajectories by Fitting Multiple Vector Fields. In: *Eurographics Conf. on Visualization (EuroVis)*. vol. 32 (2013)
6. Hu, W., Li, X., Tian, G., Maybank, S., Zhang, Z.: An Incremental DPMM-Based Method for Trajectory Clustering, Modeling, and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(5), 1051–1065 (2013)
7. Lin, W., Chu, H., Wu, J., Sheng, B., Chen, Z.: A Heat-Map-based Algorithm for Recognizing Group Activities in Videos. *IEEE Trans. Circuits and Systems for Video Technology* **23**(11), 1980–1992 (2013)
8. Marques, J.S., Figueiredo, M.A.T.: Fast Estimation of Multiple Vector Fields: Application to Video Surveillance. In: *7th Int. Symposium on Image and Signal Processing and Analysis (ISPA)* (2011)
9. Mirge, V., Verma, K., Gupta, S.: Dense traffic flow patterns mining in bi-directional road networks using density based trajectory clustering. *Advances in Data Analysis and Classification* **11**(3), 547–561 (2017)
10. Morris, B.T., Trivedi, M.M.: A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* **18**(8), 1114–1127 (2008)
11. Morris, B.T., Trivedi, M.M.: Trajectory Learning for Activity Understanding: Unsupervised, Multilevel, and Long-Term Adaptive Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(11), 2287–2301 (2011)

12. Nascimento, J.C., Figueiredo, M.A.T., Marques, J.S.: Activity Recognition Using a Mixture of Vector Fields. *IEEE Transactions on Image Processing* **22**(5), 1712–1725 (2013)
13. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes. In: *European Conf. on Computer Vision (ECCV)* (2016)
14. Wang, W., Lin, W., Chen, Y., Wu, J., Wang, J., Sheng, B.: Finding Coherent Motions and Semantic Regions in Crowd Scenes: A Diffusion and Clustering Approach. In: *European Conference on Computer Vision (ECCV)*. pp. 756–771 (2014)
15. Xu, H., Zhou, Y., Lin, W., Zha, H.: Unsupervised trajectory clustering via adaptive multi-kernel-based shrinkage. In: *The IEEE International Conference on Computer Vision (ICCV)* (2015)
16. Xu, Z., Hou, Z., Han, Y., Guo, W.: A diagram for evaluating multiple aspects of model performance in simulating vector fields. *Geoscientific Model Development* **9**(12), 4365–4380 (2016)