Locality-constrained Multi-Instance Learning for Abnormal Trajectory Detection

Ruoyao Li

School of Information, Central University of Finance and Economics, Beijing 100081, China

[a]gemini4825@yahoo.com

Keywords: Abnormal trajectory detection, locality-constrained trajectory partition, Hierarchical Dirichlet Process-Hidden Markov model (HDP-HMM), multi-instance learning.

Abstract. Abnormal event detection based on trajectory has been extensively investigated in recent years; however, problems remain when processing an incomplete trajectory that usually has abnormality in some parts of the whole trajectory and the rest are normal. In this paper, we propose a locality-constrained multi-instance learning framework for abnormal trajectory detection. We explore local adaptability for robust trajectory classification, and partition each trajectory into tracklets by control points of cubic B-spline curves. Then, the tracklets are modeled by Hierarchical Dirichlet Process-Hidden Markov Model (HDP-HMM). Finally, the whole trajectory is considered within the multi-instance learning framework as bags, when abnormal ones are positive bags consist of tracklets, normal trajectories are negative bags with tracklets. With experimental results on the CAVIAR dataset, it shows that the proposed method achieves better performance than several recent approaches.

1. Introduction

Visual trajectory classification has a significant role in computer vision [1,2], because trajectories carry rich semantic information that is valuable for identifying crowdedness [3], behaviors [1], activities and events [4] of video scenes. The other is that recent advance about visual object tracking makes it possible to obtain long trajectories and reliable trajectory feature representations. That is why abnormal trajectory detection has attracted a lot of research attentions for years.

Literatures about trajectory classification mainly involve trajectory representation and trajectory modeling. Trajectory representation could be realized using methods including polynomial based curve fitting, Haar wavelet transform, minimum error-based polygonal approximation, B-spline curves, and Discrete Fourier Transform (DFT) coefficients [5, 6, 7]. In [6], trajectories are represented according to the spatiotemporal movements of objects using function approximation algorithms of least square polynomial, Cheybyshev polynomial and DFT. In [7], Haar wavelet coefficients and Least-squares Cubic Spline Curves Approximation (LCSCA) are adopted as parametric vectors to represent a trajectory. Compared with point-based trajectory presentation, parameterized trajectory representations substantially compress the trajectories, and are therefore more effective for similarity measure and clustering. In [7], it is demonstrated that trajectory representation with LCSCA outperforms other parameterized representations, as the least-square fitting procedure holds better fidelity to trajectories and insensibility to the variation of trajectory length.

In recent works, on aligned trajectory sequences, unsupervised or supervised learning methods are wildly used for abnormal trajectory detection. Trajectory clustering assigns similar trajectories to the same cluster using methods such as Self-Organizing Map (SOM) [8], hierarchical fuzzy K-means[9], while trajectory modeling constructs a parametric or nonparametric model to represent and index trajectories, i.e. Gaussian Mixture Models (GMMs) [10], hierarchical Bayesian Model [11], and hierarchical Hidden Markov Model [12]. Existing trajectory clustering and modeling methods solve a variety of problems in trajectory analysis, however, they still have the limitation to model trajectories incrementally. Therefore, when new trajectories reach, the model has to be retrained based on the previous and the new trajectories that would lead to the high computational complexity.

Recently, Wu [13] simultaneously trains a representation and detectors for categories with either weak or strong labels present. Further, multiple-instance learning approach is frequently used to train the detector and infer the object location in positive samples of some supervised approaches [14]. Inspired by the attraction and challenge of supervised multi-instance learning, it is introduced into abnormal trajectory detection. The discriminative trajectories frequently occur in the positive samples but rarely in the negative ones. However, the global learning strategy is challenged when there are noises or local variations in whole trajectories.

In this paper, we propose a locality-constrained multi-instance learning framework for abnormal trajectory detection. Comparing with traditional detection methods, the contribution of our paper is achieving good performance when processing incomplete trajectories. We use local adaptability for robust trajectory classification, and partition each trajectory into tracklets by control points of cubic B-spline curves [15]. Then, we model the tracklets by Hierarchical Dirichlet Process-Hidden Markov Model (HDP-HMM) [16]. Finally, we combine multiple-instance learning and trajectory partition as positive trajectory sample mining strategy and optimize model iteratively. The whole trajectory is considered within the multi-instance learning framework as bags, when abnormal ones are positive bags consist of tracklets, normal trajectories are negative bags with their tracklets. The experiments demonstrate that the proposed approach achieves comparable performance with existing supervised approaches.

The remainder of the paper is organized as follows. In section II, the framework is introduced in detail. In section III, we describe the locality-constrained trajectory partition, HDP-HMM, and multi-instance learning. Experimental results are presented in section IV and we conclude the paper in section V.

2. Locality-constrained Multi-Instance Learning Framework

In this section, we first present the locality-constrained multi-instance learning framework for abnormal trajectory detection in surveillance video, as in Fig. 1. The framework includes three stages: trajectory representation and partition, modeling and training, multi-instance learning.

Given a video scenario, firstly we do trajectory representation and partition. The objective of trajectory partition is to divide a long trajectory into tracklets based on the LCSCA feature vectors [15]. Accordingly, a long feature vector is divided into a set of short independent sub-vectors, each of which will be better represented by a model. Then, the HDP-HMM model is incorporated to train the detection model for each segment of tracklets, without the limits of numbers of states in HMM. Finally, we apply KL divergence [17] to measure the distance between each two trakclets, and apply Hausdorff [9] to measure the distance between two trajectories. In the trajectory detection stage, Citation-KNN [18] is used to learn the trajectory with many instances.

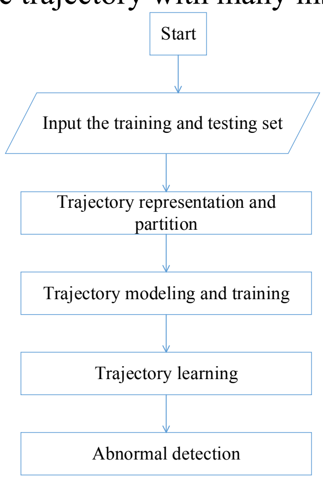

Fig.1. Illustration of the locality-constrained multi-instance learning framework.

3. Methodology

In the following, we firstly describe three parts of approaches as LCSCA-based trajectory representation and partition, HDP-HMM-based modeling and training, multi-instance learning with Citation-KNN, to explain how to partition, model and learn the patterns of trajectories.

3.1 Trajectory Representation and Partition.

Sillito and Fisher [7] proposed the LCSCA-based trajectory feature representation method, which is achieved by approximating each spatial-temporal trajectory with a uniform cubic B-spline curve parameterized by time. They showed that the LCSCA control point-based distance between trajectories with variable length is more robust than the original point-based distance. Consequently, we use the control points of cubic B-spline curves to extract a fixed-length parametric vector as trajectory representation.

Given a trajectory $t_i = \{(P_1^x, P_1^y), \cdots, (P_n^x, P_n^y)\}$, where n is the length of trajectory and (P_n^x, P_n^y) represents n-th position point of the trajectory [xv], the corresponding control point-based feature representation of t_i is $y_i = \{(C_1^x, C_1^y), \cdots, (C_p^x, C_p^y)\}$, p is the number of control points, and (C_p^x, C_p^y) represents the p-th control point, where C_p^x and C_p^y represent its normalized x-coordinate and y-coordinate respectively. With the predefined B-spline basis function (1), we can obtain $y_i = \Phi^{\dagger} t_i$, where $\Phi_{n,p} = B_{p,m}(s_n)$.

$$B_{p,1}(s_n) = \begin{cases} 1 & if\, \tau_p \le s_n < \tau_{p+1} \\ 0 & otherwise \end{cases}, \quad B_{p,m}(s_n) = \frac{s_n - \tau_p}{\tau_{p+\mathrm{m}-1} - \tau_p} B_{p,m-1}(s_n) + \frac{\tau_{p+\mathrm{m}} - s_n}{\tau_{p+\mathrm{m}} - \tau_{p+1}} B_{p+1,m-1}(s_n), \tag{1}$$

Then, the objective of trajectory partition is to divide a long trajectory into tracklets. The trajectory partition is based on the fact that trajectories of the same category should often have similar shapes and share control points on the cubic B-spline curves. We partition the trajectories into local tracklets (local shapes) based on the control points and then align the tracklets via the DTW algorithm [9]. After the trajectory partition, a trajectory T is represented as

$$T = \begin{bmatrix} t_1 \\ \vdots \\ t_i \\ \vdots \\ t_{p-1} \end{bmatrix}, \quad i = 1, \ldots, p-1 \tag{2}$$

Where t_i is the feature vector of the i-th tracklet based on trajectory partition and DTW alignment. We will model a trajectory with some tracklets of local similarity, called locality-constrained representation and partition.

3.2 Trajectory Modeling and Training.

With the the proposed LCSCA-based trajectory representation and partition, we further propose a discriminate model with HDP-HMM training procedure.

Considering the complexity of the movement patterns of objects in surveillance video, it is hard to get the prior knowledge for the model. When there is a difference between the numbers of assigned states and actual states in HMM, we involve HDP-HMM approach which can assign the original prior arguments with random values. Then, we update the state numbers and arguments in HMM instead of remaining the same state numbers. The key procedure is how to construct the HDP-HMM [19] for each tracklet in Fig. 2. The left part of the drawing is HDP, and the right is HMM. π_k is sampled from the Dirichlet process to describe the distribution of the k-th state transition matrix. θ_k is sampled from the prior distribution, it illustrates the observation distribution from the k-th state. z_t is the state of HMM, and y_t is the observation of HMM. Moreover, the Beam sampling [19] method is chosen to slice sample with dynamic programming quickly.

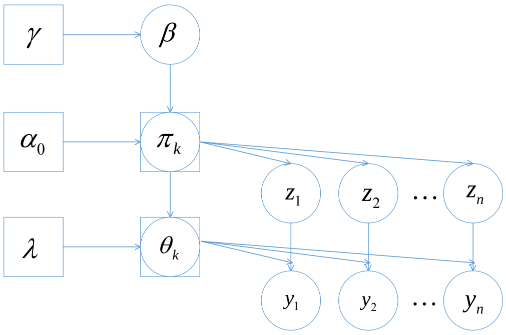

Fig.2. Structure of HDP-HMM

3.3 Trajectory Learning.

The existing approaches get the sample's label from the k nearest neighbors, so they usually will fail when the k nearest neighbors have different labels. To solve this problem, we use an algorithm named Citation-KNN [19]. This method is calculated not only by the nearest neighbors, but also by those trajectories that keep the new trajectory as their neighbor, described in Algorithm 1.

Algorithm 1. Citation-KNN Trajectory Learning

Input: X is the training set, Y is the testing set, k is the number of nearest neighbors.
Output: the label of detection for testing set
1. Compute the distances among tracklets with KL divergence,
2. Calculate the distances among trajectories with Hausdorff,
3. Sort all distances between training and testing sets by step 1 and step 2 on each trajectory,
4. Select k nearest neighbors from the testing set,
5. Count the labels of all nearest neighbors,
6. Recognize the testing trajectory as abnormal if the number of abnormal labels is more than normal ones.

Because the tracklets are described as a HMM instead of a feature vector, the traditional Euclidean distance is not suitable to measure the segments of trajectories. In the above-mentioned Citation-KNN method, we use Hausdorff distance to measure the difference among tracklets, and apply KL divergence to measure the similarity among trajectory model.

4. Experiments

We evaluate our approach on a public trajectory dataset: the CAVIAR ("INRIA") dataset [20].The CAVIAR dataset contains a series of trajectories in an entrance lobby. There are 11 entry-exit routes appearing in the dataset. Considering the traversal orientations, we have 22 categories of normal trajectories for dictionary construction. Each category has 100 simulated trajectories. In addition, lots of trajectories are of noises and local variations. There are 21 normal trajectories in the CAVIAR dataset for testing. These trajectories correspond to people walking directly from one exit to another. The Fig.3 illustrates the normal and abnormal trajectory examples.

Fig.3. Normal (left, blue) and abnormal (right, yellow) trajectory examples

In experiments, we follow [15] to use seven control points (p=7) for trajectory partition. Anomalous Correction Accuracy (ACC) is used to measure the proportion of corrected classified normal and abnormal trajectories, and recall value is calculated to measure the proportion of corrected labeled abnormal in testing samples. The recall value is very important in the field of abnormal event detection, for recognizing a normal trajectory as abnormal one is more serious than recognizing an abnormal one as normal.

There are some experiments to evaluate the effectiveness of the proposed approach in comparison with three groups of approaches: LCSCA + HDP-HMM + Multi-instance, LCSCA + HMM + Multi-instance, Whole Trajectory + HDP-HMM + Multi-instance. In Table I, the results of abnormal detection are very impressive. On the CAVIAR dataset, the proposed LCSCA + HDP-HMM + Multi-instance approach has a very high performance (ACC=87.33%, Recall=100%) when using all training samples of trajectory. It also gets good performance (ACC=79.30%) on the whole trajectories with noises and local variations. Compared with the HMM approach, the proposed approach has significant performance improvement. HMM has fixed number of states, beyond expressing the true distribution in practical application. That is why the second approach has lower accuracy. The third approach based on the whole trajectory which has noises and abnormality in some parts, ignoring the abnormal on some local tracklets, so this method has lowest accuracy and recall rate. However, the first LCSCA + HDP-HMM + Multi-instance method exactly solves these problems, so has highest values of accuracy and recall.

Table 1. Comparisons of abnormal detection

Dataset	Method	Accurac y (%)	Recall (%)
CAVIAR	LCSCA + HDP-HMM + Multi-instance	87.33	100.00
	LCSCA + HMM + Multi-instance	83.42	80.21
	Whole Trajectory + HDP-HMM + Multi-instance	79.30	71.40

5. Summary

In this paper, we propose a locality-constrained multi-instance learning strategy for abnormal trajectory detection in surveillance video scenes. The novel techniques introduced in this paper include trajectory representation and partition, trajectory modeling and training, trajectory learning. The proposed approach utilizes locality to represent and partition trajectories into independent tracklets, then a robust discriminative HDP-HMM is trained on tracklets. Finally, in the multi-instance learning framework, whether a testing trajectory is abnormal or not is determined by the classification from some bags. It means that the whole trajectory is considered as bags, when abnormal ones are positive bags consist of tracklets, normal trajectories are negative bags with tracklets. Thanks to the flexibility of the LCSCA representation and partition and the employed Citation-KNN method, our approach reports an efficient and robust detection. Experimental results on the CAVIAR dataset show the good performance of the proposed approach. The comparison to the recent approaches is also provided, which indicates that the approach improves the state of the art. Our method can also apply to other applications, such as event or action recognition.

References

[1]. Weinland, D., Ronfard, R., Boyer, E.: A survey of vision-based methods for action representation, segmentation and recognition. CVIU 115 (2011) 224–241

[2]. Nascimento, J.C., Marques, J.S., Lemos, J.M.: Modeling and classifying human activities from trajectories using a class of space-varying parametric motion fields. IEEE TIP 22 (2013) 2066–2080

[3]. Amit Adam, Ehud Rivlin, I.S., Reinitz, D.: Robust real-time unusual event detection using multiple fixed-location monitors. IEEE TPAMI 30 (2008) 555–560

[4]. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: IEEE CVPR. (2010) 1975–1981
[5]. Hu, W., Li, X., Tian, G., Maybank, S., Zhang, Z.: An incremental dpmm-based method for trajectory clustering, modeling, and retrieval. IEEE TPAMI 35 (2013) 1051–1065
[6]. Naftel, A., Khalid, S.: Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space. Multimedia Systems 12 (2006) 227–238
[7]. Sillito, R.R., Fisher, R.B.: Parametric trajectory representations for behavior classification. Multimedia Systems 12 (2006) 227–238
[8]. 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 (2008) 1114–1127
[9]. Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T., Maybank, S.: A system for learning statistical motion patterns. IEEE TPAMI 28 (2006) 1450–1464
[10]. Sillito, R.R., Fisher, R.B.: Semi-supervised learning for anomalous trajectory detection. In: BMVC. (2008) 1–10
[11]. Wang, X., Ma, K.T., Ng, G.W., Grimson,W.E.L.: Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. IJCV 95 (2011) 287–312
[12]. Vasquez, D., Fraichard, T., Laugier, C.: Growing hidden markov models: An incremental tool for learning and predicting human and vehicle motion. The International Journal of Robotics Research (2009)
[13]. Wu Q, Ng M K, Ye Y, et al. Multi-label collective classification via Markov chain based learning method[J]. Knowledge-Based Systems, 2014, 63(3):1–14.
[14]. Zhang M L, Wu L. LIFT: Multi-Label Learning with Label-Specific Features.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(1):1609-1614.
[15]. Li C, Han Z, Ye Q, et al. Locality-Constrained Sparse Reconstruction for Trajectory Classification[C]// 2014 22nd International Conference on Pattern Recognition (ICPR)IEEE Computer Society, 2014:2602-2606.
[16]. Fox E，Sudderth E，Jordan M I，et al．Developing a Tempered HDP-HMM for Systems with State Persistence． Technical Report，P-2777． Cambridge，USA: MIT，2007
[17]. Goldberger J, Gordon S, Greenspan H. An Efficient Image Similarity Measure Based on Approximations of KL-Divergence Between Two Gaussian Mixtures[C]// Computer Vision, IEEE International Conference onIEEE Computer Society, 2003:487-487.
[18]. Cui Y Y, Gao Y. Abnormal Event Detection Based on the Multi-Instance Learning[J]. Pattern Recognition & Artificial Intelligence, 2011, 24(6):862-868.
[19]. [Wang Jun，Zucker J D． Solving the Multiple-Instance Problem: A Lazy Learning Approach / / Proc of the 18th International Conference on Machine Learning． Stanford，USA，2000: 1119 － 1126
[20]. CAVIAR. URL http://homepages.inf.ed.ac.uk/rbf/CAVIAR/