Path Normalization for Traffic Surveillance Video Retrieval

Ehsan Lotfi[1], H.R. Pourreza [2]
1- Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran.
Email: elotfi@ieee.org (Corresponding author)
2- Departments of Computer Engineering, Ferdowsi University of Mashhad, Iran
Email: hpourreza@um.ac.ir

ABSTRACT:
In this paper we present a novel method based on path normalization for classification in the traffic surveillance videos. Extracting the low level feature vectors in various sizes and recording all spatial-temporal information without fixed sampling rate are the main reason in this normalization. The normalized feature vectors are used for unsupervised learning and since most people of society have legal traffic behaviors system can extract the necessary knowledge automatically to detect illegal behavior. In the proposed structure, decision making for these behaviors is based on spatial-temporal features. The experimental results show high accuracy in trajectories classification using path normalization.

KEYWORDS: Trajectories clustering, Behavior extraction, Spatial-temporal features

1. INTRODUCTION

In many surveillance videos applications, motion extraction is main part of object's behavior detection. Object's motion can be presented by trajectory and analyzing the object's behaviors, detecting the legal and illegal behaviors and retrieving the events, all depend on automatic trajectory extraction. Additionally the classification is a main task of trajectory extraction and event retrieval [1]. In our previous work [2], we presented a novel classification method to retrieve trajectory. Here we propose a path normalization method which can be used in classification step.

Recently various works in field of trajectories classification were done[3][4][5][7]. Vehicle tracking and trajectory categorization obtain the local routs [7] [8][9] like as a work which was done by Hu [10] who proposed a new feature named directional histogram trajectory and after smoothing the trajectories, it classified them using dominant-set clustering method. The proposed feature in this work was extracted based on certain number of sample points on trajectories. Hu [10] utilized Fuzzy-SOM and Hunter et. al [11] used Self-Organizing Maps for this purpose where motion path was modeled using polynomial and a measurement was proposed based on distance between polynomialsused the min, max and mean distance between trajectories as a measurement. Weiming [12] used the competitive networks to produce a statistic model of trajectories and proposed a hierarchical special-temporal classification based on equal and comparable number of sample points on trajectories. The measurement in that work was distance between trajectories. This distance was evaluated based on sample points. Among works which used spatial–temporal features[13], [14], [15], [16]. Junejo [16] used cut-graph method. Also SVM was used for anomalous trajectory detection [17] where the trajectories were first subsampled in order to obtain a fixed-size feature vector, then training trajectories were learned using single class SVM and hyper volumes which include all illegal path be obtained.

2. PROCEDURE STRUCTURE

Regarding the surveying, most of papers define the measurement of similarity between trajectories based on certain number of or comparable number of sample points on the trajectories. So they cannot be robust methods for trajectory extraction; Extracting same number of points from each trajectory may cause many temporal information be lost. To solve this problem and increase the performance in trajectories classification we should normalize the paths before classifying them. This is our main idea. In fact we must model the trajectories in away which converts trajectories with various points into certain number of parameters and trajectories classification be done by using this parameters. We define a scene for all frames. Proposed system track the motion cluster centers at scene from entering instant until leaving instant. This tracking includes temporary stay so that all special-temporal information is saved. In proposed structure we avoid object detection and trajectories are obtained using motional cluster tracking. In general structure first dynamic frontground is extracted from static background and motional features are determined. In this step the size of feature vectors is various. In the Next step the feature vectors of trajectories are normalized into fixed size. Finally unsupervised classification is done using these normalized vectors. Since most people of society have legal traffic behaviors system can learn the necessary knowledge to detect illegal behavior and the new behavior which is not near any learned classes is anomalous.

2.1. Feature extraction

For extracting dynamic frontground from background in videos we use difference of consecutive frames. Suppose that we want to process two consecutive frames, we divide first frame into various fixed-sized block which are not overlap with together. The blocks which have following 2 condition are selected and searched on second frame. The search area on second frame is a rectangular-shape around location of block. The conditions include, 1) whole of block should situate in the frontground. 2) the block must consist enough detail. Second condition is possible to obtain by defining a threshold on Standard Deviation of intensity vector of block. Figure 1 shows a sample of frontground extraction and Figure 2 shows some blocks which satisfy the above conditions.

Figure 1 Extracting frontground from background

Figure 2 Sample of blocks which satisfy the conditions and are worth for searching on the next frame and classification.

Extracting the start point of block in the first frame and finding it in second frame are basic part of low level feature extraction. So the feature vector of each moving block includes; start point of block, (x,y); the velocity of block's motion, (V_x , V_y) which is calculated by

$$V_x = d_x / d_t , \; V_y = d_y / d_t$$

The above vector is obtained by processing two consecutive frames. Whereas the d_x and d_y are calculated per time unit, we can omit the denominators. After extracting the features we cluster them using Kmeans. The learning phase causes some clusters that include same block in location and velocity. A cluster center has 4 features include;1) *x*, the mean of lengths of blocks which are located in the cluster; 2) *y*, the mean of widths of blocks which are located in the cluster. 3) the mean of blocks motion on axis *x*. 4) the mean of blocks motion on axis *y*.

To track the motion of cluster center, we use the following algorithm;

- Take cluster i of frame t
- Predict the center of cluster i in frame t+1:
  - $X = x_i + dx_i$
  - $Y = y_i + dy_i$
- Search the set of all cluster centres in frame *t+l* and find the closest cluster (*j*) to vector $\{X,Y, dx_i, dy_i\}$
- Add (x_i, y_i) to *Fs* set, and (dx_i, dy_i) to *Ft* set

- Let $i = j$ and $t = t+l$
- If the cluster centre has not exited the scene, proceed to the first

The result of cluster's center tracking is two sets. One of them (*Fs*) includes the points which crossed by trajectory and other (*Ft*) includes wee motion of cluster at that points.

2.2. Path Normalization

In the first we define a scene for all frames. Dimension of scene can be in proportion of frame size. Then we apply following condition; the trajectories which enter the scene and leave it were included by the algorithm.

According to the presented algorithm, two feature vectors are extracted from all trajectories;

Fs = $\{fs_1, fs_2, fs_{3,...}\}$
Ft = $\{ft_1, ft_2, ft_{3,...}\}$

Where $fs_i=(x_i,y_i)$ is i^{th} Point which is tracked and $ft_i=(dx_i,dy_i)$, $dx_i=x_{i+1} - x_i$, $dy_i=y_{i+1} - y_i$

Fs includes the points which crossed by trajectory and *Ft* includes wee motion of cluster at that points then *Fs* presents spatial feature and *Ft* presents temporal feature. The size of those vectors is similar and the size of feature vectors for various trajectories may be diversity. In the fact the trajectory is estimated by any points which present it in the scene and the trajectory is not subsampled with certain number of points. Since size of the low level feature vectors is variety, we cannot use them for classification directly and we must normalize them into fixed-size vectors. We normalize them in separate ways. For normalizing the *Fs* we fit a least squares polynomial of degree (M-1) for each trajectory then we obtain M parameters. Our goal is not to estimate trajectory with polynomial. We just convert various-size feature vectors to M representative feature vectors. For normalizing the *Ft*, we cluster the tracked points according to track into N segments. For example if number of tracked points is 16 and N =3, put first 5 points into first segment and second 5 points into second segment and put the reminder into third segment. Then calculate the mean of the motions which are included by each segment. So *Ft* as a feature vector which includes the motional information of various points converts to N representative feature vector. Let to survey the normalized feature vectors;

Fs_n, the normalized feature vector of *Fs* that includes M-coefficient of fitting polynomial;

$$Fs\_n = \{A_1, A_2, \ldots, A_m\}$$

Ft_n, the normalized feature vector of *Ft* that includes mean of motion per unit of time in N segment.

$$Ft\_n = \{(\bar{d}x_1, \bar{d}y_1), (\bar{d}x_1, \bar{d}y_1), \ldots, (\bar{d}x_n, \bar{d}y_n)\}$$

3. TRAJECTORIES CLUSTERING

Normalized feature vectors may be used in trajectories classification. There are two viewpoints for using these vectors and data fusion. In first viewpoint, new feature vector is composed by collection of the feature vectors. Size of the new vector will be (N+M). In Second viewpoint, two-step classification is performed. First-step classification is done using spatial information of trajectories and second-step clustering is performed using temporal information of trajectories. Actually the first is performed based on tracked points and overall schema of trajectories and the second is achieved based on quality of and condition of the motions. Supervised or unsupervised learning can be used in classification. Since most people of society have legal traffic behaviors, the result of machine learning by unsupervised method is some classes that show legal motions then in recognition phase if new trajectory doesn't fall to any learned class, warning for illegal behavior should report. In experimental results we found best results for the second view point in data fusion and two-step clustering. First classifier applies spatial feature vectors and the second classifier applies temporal feature vectors. Consider two following example as results of proposed hierarchical clustering;

Figure 3 shows three paths from C to A. One of them has a temporary stay at B and others move from C to A directly. First classifier should classify them into one class because their overall schemas are same. But the second should classify them into 2 classes Because of temporary stay at B, zero padding may be performed and may cause difference values for N segment in temporal feature vector.

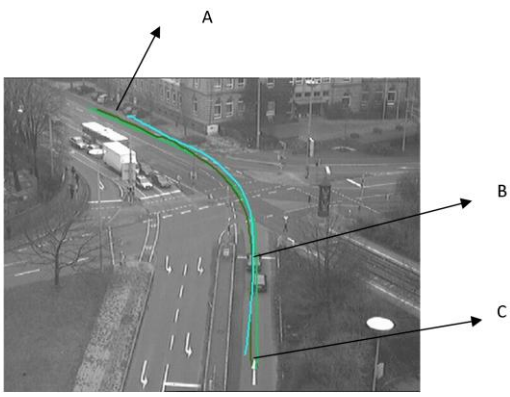


Figure 3 One of moving cluster has temporary stay at B and others move directly from C to A.

Figure 4 shows a cluster center by red and a new path by blue. In first clustering new path is assigned in red class. Blue moving cluster move rapidly then in second clustering it's rejected by red class because of difference in temporal information. The rejection and the acceptation are obtained by defining a threshold on distance of new path from class center.

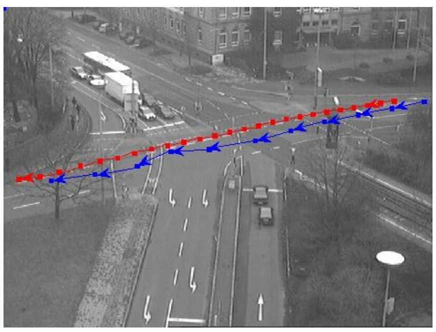


Figure 4 a cluster center presented by red and a new path presented by blue. They have same spatial feature vectors but different temporal vectors because of blue cluster's speed

4. EXPERIMENTAL RESULTS

We collect the datasets from two sources. Some videos obtained using published dataset by Institut für Algorithmen und Kognitive Systeme and others obtained at Traffic Surveillance and Control Center of Mashhad, Iran. There are more than 7,000 frame from 9 traffic surveillance cameras with different weather conditions, lighting, image quality and fields of view. See Figure 5, the trajectories were extracted from surveillance frames in various conditions. Process was performed with distance of 5 frames and based on blacks. The size of blocks will be 15*15 where the size of frames is 576*768 and will be 5*5 where the size of frames is 320*240. Sample results of cluster tracking are shown in Figure . The Trajectories presented by any points which are located at the scene. According to the proposed method for normalization we extracted 2 normalized vectors from each trajectory. One of them, *Fs_n*, includes 3-coefficient of fitting polynomial and another, *Ft_n* includes mean of vertical-horizontal motions in 4 segments. So the size of first vector is 3 and size of second vector is 8.

Figure 5 surveillance frames in various conditions

Figure 6 a view of crossroads which is in the data base and the extracted trajectories.

In experimental results we found best result for the second view point in data fusion and two-step clustering. We compare some method in order to find best classifier for first step. The methods applied the *Fs_n* to train. Table 1 shows the results. In this scenario, total number of trajectories is 96. The number of training trajectories is 37 and we use other 37 to test. In the table, the rate of training denotes the rate of correct class in training data.

Table 1 comparison of methods

	SOM	K-means	Fuzzy-Kmeans	Subtractive Fuzzy-Kmean	Multiclass-SVM
Topology, neuron#	[2,3], 6	-	-	-	-
Class#	-	6	6	6	6
Rate of training	79%	95%	98%	100%	36%
Hit ration (Test phase)	75%	84%	86%	98%	10%

According to the table best result is assigned by Subtractive Fuzzy-Kmeans but this method is supervised. Since we want to give the system least priori-information we choice Fuzzy-Kmeans which has proper result and we just determine maximum possible number of classes. This maximum number as an input of system can be fixed and it's not important that the number of classes is more than actual number because illegal behavior extraction is important not number of it. Figure shows the result of fuzzy-kmeans which used spatial feature vectors. According to the figure it's clear that the trajectories from G to F are classified in two classes which present a valid behavior. The trajectories from C to A and from E to D are classified into similar class. They are same in overall schema but are different in direction and temporal information and it's possible that they will be classified in second step which use temporal information. The 5 classes listed below are shown by Figure .

1-{(C-A),(E-D)}, 2-{G-F}, 3-{G-F}, 4-{(G-D),(E,F)}, 5-{G-A}

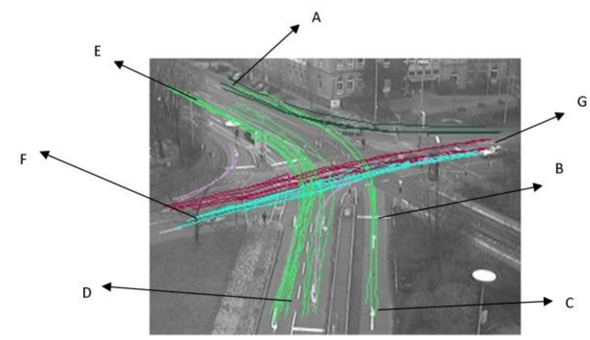


Figure ٧ classification result using spatial feature vectors and fuzzy-kmeans.

We choice 2-dimention SOM with 2 neurons as second classifier which just applies temporal vectors (*Ft_n*). Figure 1 shows the result of clustering by unsupervised learning. The SOM distinguish between C-A and E-D because they are difference in direction and temporal information. According to the Figure It's clear that new class which falls in C-A is extracted (Figure 8.i). The trajectories of this class have temporary stay at the point B. The 8 classes listed in Figure are extracted from the result of first classifier by second classifier. Finally

Table **2** shows the result of tracking and clustering with various numbers of classes as input of system.

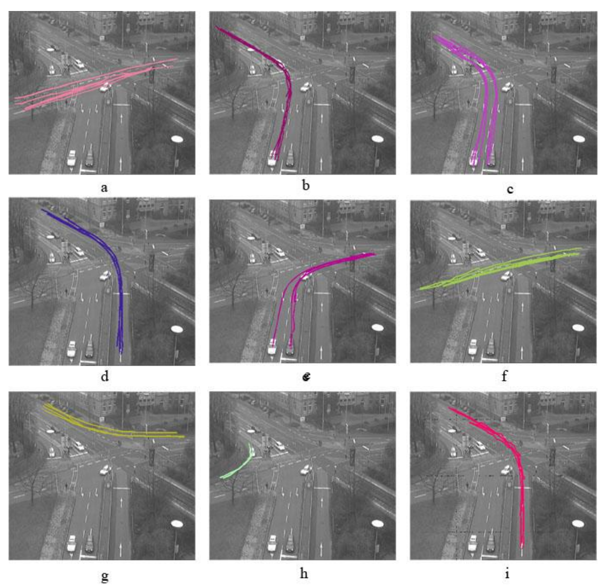
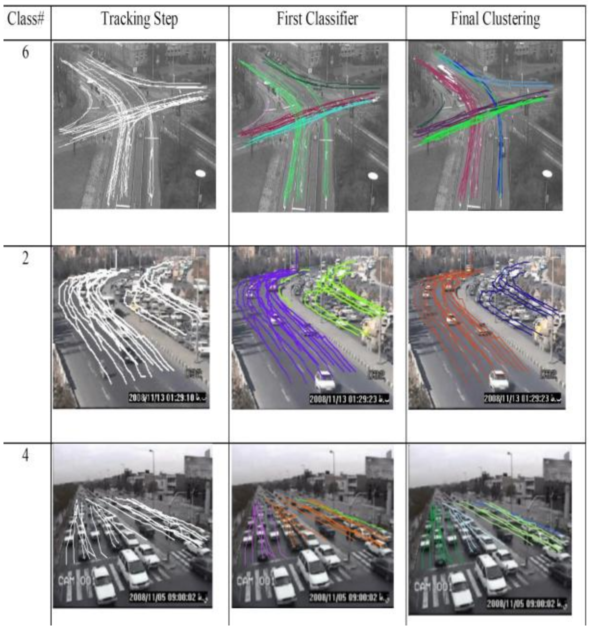


Figure ٨ final results of clustering

Table 2 sample result of tracking and clustering



5. CONCLUSION

In this paper the trajectories presented by 2 vectors include spatial and temporal features. We proposed a method in order to normalize path and proposed hierarchical clustering that applied normalized vectors. According to experimental results best classifier for first step is fuzzy kmeans. We used SOM as second classifier. Unsupervised learning provides the system that enables to extract legal behaviors automatically then enables to detect illegal behavior using least priori-knowledge. The experimental results show that trajectories clustering are performed in high accuracy.

REFERENCES

[1] Ehsan Lotfi and Mahdi Yaghoobi, Concept Retrieval based on a Combination of Fractal Coding, Fuzzy Rule Based System and SVM, International Journal of Fractals, Vol. 19, No. 2 (2011) 1–10, DOI: 10.1142

[2] Ehsan Lotfi, H. R. Pourreza, Trajectory Clustering and Behaviour Retrieval from Traffic Surveillance Videos, Majlesi Journal of Multimedia Processing, Vol. 1, No. 2, 2011

[3] Piciarelli, C., Micheloni, C. and Foresti, G.L., 2008. Support vector machines for robust trajectory clustering: IEEE Int. Conference on Image Processing (ICIP), San Diego, CA, USA, 2540 – 2543

[4] Fashandi H. and Eftekhari-Moghadam A.M., 2005. An image mining approach for clustering traffic behaviors based on knowledge discovery of image databases: IEEE Int. Conf. on Computational Intelligence for Measurement Systems and Applications, July , pp. 203-207

[5] Li, X., Weiming Hu, Wei Hu, A Coarse-to-Fine Strategy for Vehicle Motion Trajectory Clustering, Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)

[6] Zhou, Y. and et al., 2007. Detecting anomaly in videos from trajectory similarity analysis: IEEE Int. Conf. on Multimedia and Expo, Beijing, China., 1087-1090

[7] Melo, J., Naftel, A., Bernardino, A. and Santos-Victor, J., 2004. Retrieval of vehicle trajectory and estimation offline geometry using non-stationary traffic surveillance cameras: Advance Concepts for Intelligent Vision Systems, 104-111

[8] Naftel, A., Melo, J., Bernardin, A., Santos-Victor, J., 2004. Highway lane detection and classification using vehicle motion trajectories: Int. Conf. on Visualization, Imaging, and Image Processing, Marbella, Spain , 556-560

[9] Dimitrios Makris, Tim Ellis, .Path Detection in Video Surveillance,. Image and Vision Computing,20(12):895-903, 2002

[10] Hu, W., Dan Xie, Tieniu Tan, Steve Maybank,"Learning Activity Patterns Using Fuzzy Self-Organizing Neural Network"," IEEE Trans. Systems, Man and Cybernetics - Part B, 34(3):1618-1626, 2004.

[11] Owens J. and Hunter A., 2000. Application of the Self-Organizing Map to Trajectory Classification: IEEE Int. Workshop on Visual Surveillance, 77–83

[12] Hu, W., Xie, D., Fu, Z., Zeng, W. and Maybank, S., 2007. Semantic-Based surveillance video retrieval: IEEE Transaction on Image Processing, 16, (4), 1168-1181

[13] Kuettel, D., Breitenstein, M., Van Gool, L., and Ferrari, V., 2010, What's going on? Discovering spatio-temporal dependencies in dynamic scenes: IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2010), pp. 1951-1958.

[14] Fan Jiang, Junsong Yuan, Sotirios A. Tsaftaris, Aggelos K. Katsaggelos, 2011, Anomalous video event detection using spatiotemporal context: Computer Vision and Image Understanding, 115 323–333

[15] Snoek, C.G.M., Huurnink, B. and Hollink, L., 2007. Adding semantics to detectors for video retrieval: IEEE Transaction on Multimedia, 9, (5), 975-985

[16] Junejo, Imran N., Omar Javed and Mubarak Shah, Multi Feature Path Modeling for Video Surveillance, Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, DOI: 10.1109

[17] Piciarelli, C., Micheloni, C. and Foresti, G.L., 2008. Support vector machines for robust trajectory clustering: IEEE Int. Conference on Image Processing (ICIP), San Diego, CA, USA, 2540 - 2543