Trajectory Clustering and Behaviour Retrieval from Traffic Surveillance Videos

Ehsan Lotfi
Department of Computer Engineering, Islamic Azad University, Mashhad Branch, Iran

ABSTRACT:
The system presented here uses a minimum of prior knowledge to retrieve trajectories and abnormal behaviour. This system is based on the premise that most members of society choose to comply with traffic laws. The system input is only two parameters, namely (1) the maximum number of junctions available and (2) the rate of infractions in the city. In an unsupervised manner, the new system is capable of learning all legal behaviour and, on its own, extracting the knowledge required for detection of illegal behaviour. The query model used in this system is based on keywords and user sketches. Keywords are labels applied automatically by the system to express the activity models. The proposed system also optimizes user sketches before implementation. Practical implementations have demonstrated the high efficiency of this system in learning legal behaviour and detecting illegal practices.

KEYWORDS: trajectory retrieval; behaviour extraction; traffic surveillance video; query optimization; SOM; SVM.

1. INTRODUCTION

In surveillance applications, a video conveys meaningful messages to its audience. For example, by watching a traffic surveillance video, the audience understands concepts such as turning left, speeding driving, going the wrong way, and eventually the observance and infractions of traffic laws. The fact that a human understands the concepts of a video while a machine cannot is normally referred to as the semantic gap between man and machine (Snoek 2007). If computers were capable of extracting the meanings contained in a video, such meanings could then be used for automatic indexing and the retrieval of traffic concepts from a given database. Several query models may be addressed along the retrieval system. Hu (2007) presented a structure that uses sketch and keyword queries to retrieve traffic concepts. Several approaches are proposed for retrieval, including content-based information retrieval in this method, the system models the contents of the video in a manner effective for retrieval (Doulamis 2000, Piriou 2006). Approaches introduced in recent years are mostly semantic-based methods. A group of such methods use the semantic description of object motion to retrieve concepts from traffic videos (Haag 2000, Liu 2001, Fashandi 2005, Hu 2007, Safara 2008, Kuettel 2010, Fan Jiang 2011). In Hu's (2007) work, first the objects are extracted from video frames and their trajectories tracked. Then the trajectories are classified for the training of object motion models. Eventually, the semantic description is added to the motion models. It must be noted that this semantic description is added manually. It is through these semantic descriptions that the user retrieves preferred concepts. In the structures which are proposed by Safara (2008) and Fashandi (2005) information is extracted from consequential images taken from a crossroads in natural language. The process involves stages such as object detection, trajectory extraction, object activity class definition, and the assignment of assigning each extracted trajectory to one of the defined classes. The inputs of the designed fuzzy system are these same motion classes and its outputs are the legal or illegal nature of the motion. In fact, all traffic regulations have been modelled by the means of a fuzzy system. It must be noted here that the fuzzy system is made and calibrated by an expert. Actually in semantic-based methods, the system extracts several visual classes automatically. The semantic descriptions and knowledge are added manually, through tables (Hu 2007) or a fuzzy system (Fasahndi 2005, Safara 2008). Classification of trajectories and the addition of semantic descriptions is the foundation of semantic-based methods. The addition of knowledge to previous visual models is done manually. The system presented here acquires the knowledge of behaviour in an unsupervised manner and with a minimum of prior knowledge, therefore eliminating the need for the manual addition of semantic descriptions. Thus, descriptions, such as normal or abnormal behaviour and the type of motion in terms of speed and space, are automatically defined in the Metadata.

Since object motion extraction makes up a significant part of object motion behaviour detection, various methods of trajectory classification are surveyed here. Object motions are mostly displayed with their trajectories. The analysis of object motion behaviour, detection of legal and illegal behaviour, and, finally the retrieval of events all depend on automatic trajectory detection. On other hand, the classification of trajectories may result in overall regional motion trajectories (Hu 2004). Previous papers on trajectory classification have used the Self-Organizing Map (SOM; Owens 2000) and fuzzy self-organizing map (FSOM; Hu 2004) for this purpose. In the structure proposed by Hu (2007) a hierarchical structure was utilized for trajectory classification. The spatial-temporal classification offered in this paper is based on the number of equal and comparable points on the trajectories. The classification criterion is the distance between trajectories that is the distance calculated according to the same points. Another work using spatial-temporal features is (Junejo 2004), which employs the cut-graph method. Makris (2002) worked with agglomerative clustering and Piciarelli (2008) utilized support vector machines (SVM) to detect abnormal trajectories. In (Piciarelli 2008), fixed-length property vectors are obtained from trajectories, and training trajectories are classified by single class SVM. Thus, hyper volumes are obtained, including those of all legal trajectories. If a new trajectory is outside of the calculated hyper volume, it is anomalous. In this framework, the only training data is illegal behaviours. In (Hu 2006), once the trajectories are soothed, a new feature, called a trajectory directional histogram and a dominant-set clustering method are offered through which the trajectories are classified. In (Melo 2004), the trajectory of each object is modelled with a multinomial expression, providing a criterion for the classification of trajectories based on the distance between multinomial. In (Naftel 2004), minimum, maximum, and average distances between lines are used for classification.

2. PROPOSED STRUCTURE

All structures presented for semantic retrieval from traffic surveillance videos are based on previously defined knowledge which is somehow given to the system. Examples include the works of Piciarelli (2008), in which the SVM classifier is trained with legal behaviour in a supervised manner in order to make it capable of detecting illegal behaviour. Another example is the work of Hu et al. (2007) that uses descriptive tables to retrieve the activities. These tables are manually given to the system. The system offered by Hu et al. (2007) detects abnormal behaviour based on its distance with the centre of trained models therefore it is not capable of indexing in a totally automatic manner. Hu et al. (2007) failed to clarify two key issues: (1) the automatic definition of threshold for the distance from the centre of the trained models for the purpose of detecting of illegal behaviour according to the possibility of crossroad diversification inside the database and (2) the automatic definition of the number of classes for trajectory classification in different crossroads. The system offered here solves those two key issues and uses a minimum of prior knowledge to retrieve activities and abnormal behaviour and it also files metadata table fields in a totally automatic manner. This system works unsupervised and is based on the premise that most members of society choose to comply with traffic laws and that, after learning legal behaviour; they proceed to define abnormal practices. The system works with a minimum of prior knowledge and only two parameters, namely (1) the maximum number of junctions available and (2) the rate of infractions in the city. Indexing of the concepts is handled in a totally automatic manner. Once the spatial-temporal vectors of motion are extracted, a hierarchical clustering is used to extract the manner of motion. In parallel and by using SOM, spatial-temporal vectors are classified according to infraction rates, possible abnormal data ruled out, and the remaining data used for the training of several SVM single classes. In an unsupervised manner, the new system is capable of learning all legal behaviour and, on its own, extracting the knowledge required for detection of illegal behaviour. The query model used in this system is based on keywords and user sketches. Keywords are labels applied automatically by the system to express the activity models. The proposed system also optimizes user sketches before implementation. The overall proposed structure is presented in Figure 1.

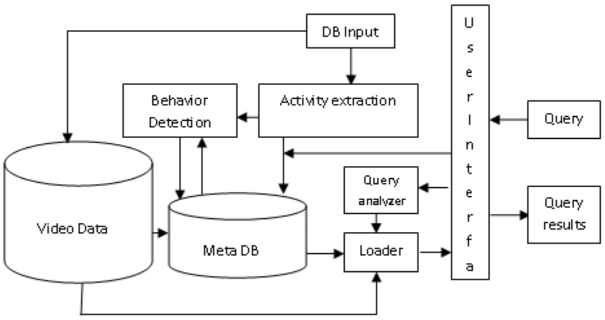


Figure 1 Overall structure of proposed system

2.1. Trajectory Classification

Extraction of trajectories is a very critical part of the traffic behaviour retrieval process. The overall structure is such that first the dynamic background is separated from the static one and all motion features are extracted. The feature vector of each trajectory may be of any size in this stage. In the next stage, the feature vectors of all trajectories are normalized and turned into vectors with known lengths. Then, trajectory classification is performed according to normalized features. Figure 2 demonstrates the stages of motion extraction.

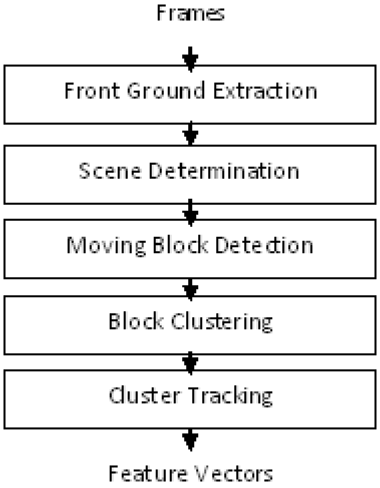


Figure 2 Trajectory extraction

We used consequential frame differences to separate the background and dynamic foreground of a given video. The scene may have dimensions proportionate to the frame or smaller. The processing of two consecutive frames gives the following feature vector for each moving block.

Fv (i)={x,y,r,g,b,Vx,Vy}

(x,y) is the block beginning coordinate; r, g, b are average block pixel colour values; Vy, Vx are the velocity (speed) resulting from block motion calculated by equation V = dx/dt. Once the features are extracted, clustering occurs through Fuzzy-K means. After training, the cluster centre is formed by the blocks pointing to the displacement of a group of close blocks with similar displacement in the same direction. The following algorithm is used to follow the cluster centre motion:

- Take cluster *i* of frame *t*
- Use the following equations to estimate the cluster centre point in frame *t+l:*
  - $X = x_i + dx_i$
  - $Y = y_i + dy_i$
- 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

To normalize the resulting vectors, we take n samples from each spatial and temporal vector and extract them by finding the means on different parts of the trajectory. Normalized vectors may be used to classify the trajectories. There are two possible approaches for using the feature vectors. First, there should be one classification stage using the spatial feature vector, thus leaving the second classification for the results of the first one by using the temporal feature vector. Second, a collection of the two vectors and a classifier must be used for trajectory classification for the sake of data fusion. The proposed structure for the classification of behaviour utilizes both approaches. The first approach is employed in two stages to extract the type of motion and the second one to extract the normality or abnormality of the trajectory. The proposed method for clustering behaviour for the automatic detection of abnormal behaviour is such that, first, the trajectories are classified in terms of spatial-temporal features. Then for each class, there is another classifier to define the normal or abnormal nature of the trajectory. See the following Figure 3, a model is presented for the extraction of traffic behaviour. Using a spatial vector, SOM1 classifies each trajectory into a motion class at the crossroads and assigns a label to it. SOM11, SOM12... and SOM1n divide continuous trajectories into two classes via their temporal vectors. In the training phase, SOM3 classifies training data, including spatial-temporal vectors, so that each cluster contains a number of trajectories. Then, the furthermost trajectories from the cluster centre are ruled out so that their number is proportionate to the infraction rate, which is given to the system as a fixed parameter. The rest of the trajectories in each class are used for Single Class SVMs training in terms of normal behaviour. In the test phase, SOM3 guides each trajectory to one of the trained SVMs model to detect its abnormality. It must be noted that training data include both legal and illegal behaviour.

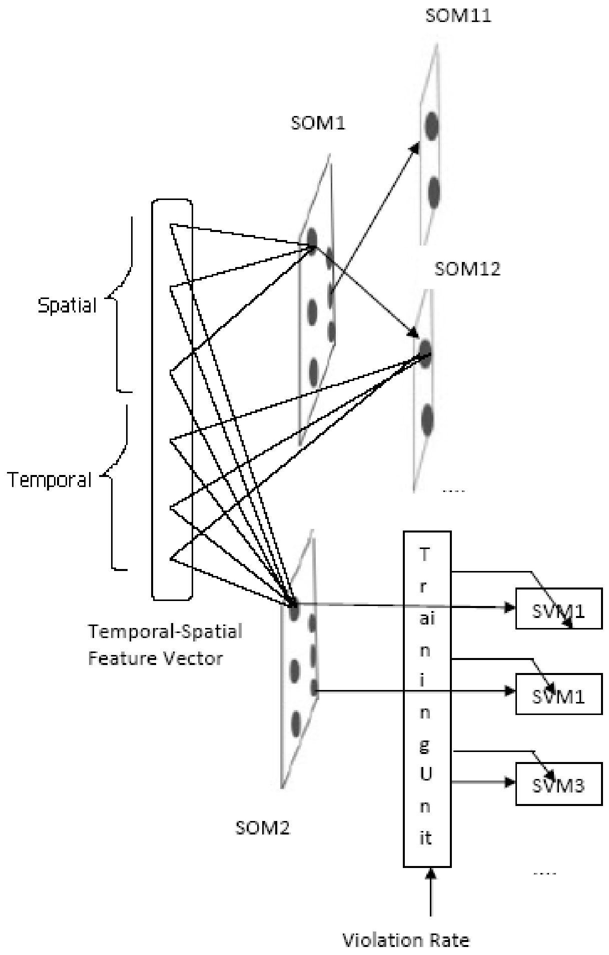


Figure 3 Proposed structure for behaviour clustering

2.2. Cluster Estimation

Except for second level SOMs, for which there are a fixed number of two neurons, the number of neurons for other classifiers differs in various videos. It is known that the maximum number of classes in this problem may be defined as the fixed parameter of the system. For example, the maximum number of classes was 8 in DB. We seize the point that optimal clustering minimizes the intra-class variance while maximizing inter-class variance, and so we the following heuristic equation may be defined.

$$cf = \frac{|\sum_{class\#}(nubmer\ of\ vectores\ in\ class)\times(variance\ of\ vectores\ in\ class)|}{|\ variance\ of\ class's\ centers|+1} \tag{1}$$

The numerator is made of the total result of the number of vectors of each class multiplied by the class variance, while the denominator is made up of inter-class variance. In the training phase, clustering may be done for a number of clusters from 2 to a maximum number and so the above equation may be applied to any result. Clustering with the number of clusters minimized by the above equation will be the final result. SOM1 classifies each trajectory into a motion class at the crossroads through the spatial vector and gives it a label. SOMs of the second level divide all continuous trajectories into two classes, namely Low Speed and High Speed, through their temporal vectors. SOMs also define the output of SVM Single Classes for each trajectory, i.e. their normal or abnormal nature. The extracted information and knowledge may be stored inside a metadata such as in the following figure. Each field is devoted to one of the different parts of the system, such as data collection, trajectory extraction, and classifier, and these are filled automatically. Retrieval may be done based on the formed metadata. Figure 4 shows the formed metadata.

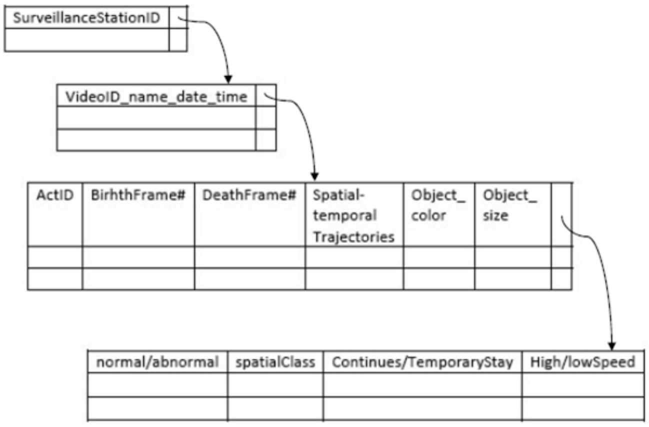


Figure 4 Proposed structure of meta data

2.3. Query Model and Behaviour Retrieval

Up to the previous section, different parts of the system filled the metadata fields in an automatic manner. Retrieval, based either on either keywords or sketches, is performed according to these fields. Before anything else the user must define the Station ID (root table) and date-time (as the video ID). In the proposed system, queries may be based on user sketches. Since some queries can not be expressed with keywords, a sketch comes in handy to be. Queries such as “a vehicle on a given trajectory” are of this nature. The user interface is prepared for the input of user sketches. Once the trajectory is given, normal vectors are extracted and fields with a minimum Euclidian distance from the given trajectory are extracted as answers, via the fourth field of the third table.

As seen in the trajectory classification section, the system automatically classifies the trajectories in an unsupervised manner. After classification, a label to each class is assigned. Labels are then employed in queries through keywords and the settings of Station ID and Video ID by the user. Thus, the user is able to make the following queries:

“ all {abnormal} behaviours”
“a {blue} car with abnormal behaviour”
“a {white} car in class{..} with {high/Low} speed ”

Since the tracker has already extracted the average cluster colours, the first field is extractable. The second and third fields are devoted to labels class defined by the system after completion of the training phase.

In sketch-based retrieval, it has been observed that different user sketches for one query present different results. This is because the sketch is not optimized among the available trajectories. To optimize the retrieval process, one stage of retrieval is performed by the user sketch method. The retrieval criterion is the Euclidian distance from the user query. Retrieved trajectories are considered as the primary population.
Figure 5 presents the process of decoding and bit string display. Assessing a trajectory needs sampling of n (number of query points) points on it. We then have:

$$Traj_i = \{(x_1, y_1), (x_2, y_2), \ldots., (x_n, y_n)\}$$

$$X_i = \{x_1, x_2, \ldots., x_n\}, Y_i = \{y, y_2, \ldots., y_n\}$$

The following assessment function is now used to find the best answer:

$$\mathbf{Fitness}(\mathbf{Traj_i}) = \frac{|X_1 - X_{query}| + |Y_1 - Y_{query}|}{2} \quad (2)$$

The optimized result may be achieved through reproducing and minimizing the said function. The only point to consider when retrieving abnormal behaviour is that the size of the primary population must be small.

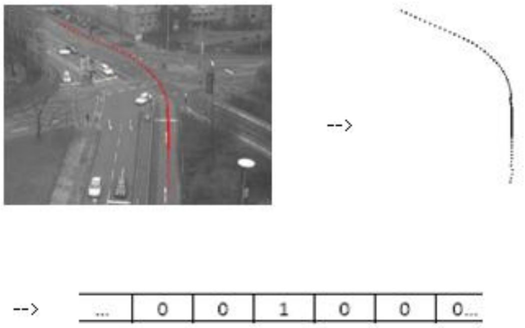


Figure 5 Coding process and trajectory representation

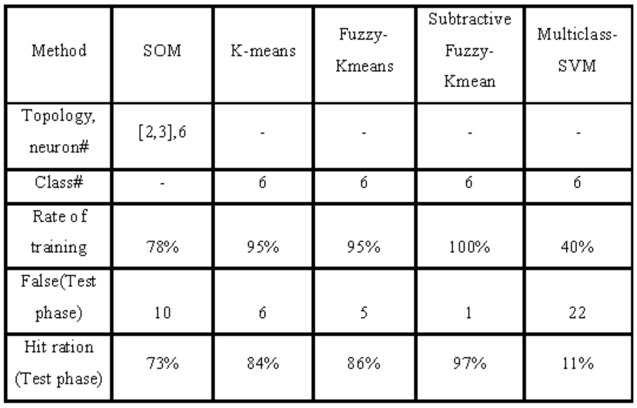
3. EVALUATION

To test and assess the offered algorithms, video data was collected from two sources. One group of data was provided by the Mashhad Traffic Surveillance and Control Center, Mahhsad, Iran and the other group was acquired from “the Institute fur Algorithmen und Kognitive Systeme.” Overall, more than 70,000 frames from at least six surveillance stations were used in the database. Processing was entirely done on fixed views of crossroads. Processing occurred on each five frames (l = 5), when the frame size was 768 × 576 and the block size was 15×15 and when the frame size was 320 × 240 and block size 5 × 5.

3.1. Trajectory Clustering

We compared several methods, all of which use the spatial vector for training, and aimed at finding the best classifier in the first phase. Results are provided in Table 1. The scenario has a total of 96 trajectories, of which 37 were dedicated to training and another 37 for testing.

Table 1 Comparing several methods for clustering



Method	SOM	K-means	Fuzzy-Kmeans	Subtractive Fuzzy-Kmean	Multiclass-SVM
Topology, neuron#	[2,3],6	-	-	-	-
Class#	-	6	6	6	6
Rate of training	78%	95%	95%	100%	40%
False(Test phase)	10	6	5	1	22
Hit ration (Test phase)	73%	84%	86%	97%	11%

According to the table, the best results belonged to Subtractive FCM, although the method is supervised. Since we want that the system requiring a minimum of a priori knowledge, we can choice FCM or SOM. They provide satisfactory results and it is only necessary to define the possible maximum of classes. As a system input, this maximum value may be fixed. The only problem in the implementation of the FCM was its high reliance on primary weight values and various results obtained in different executions. SOM poses no such problem and, in fact, is capable of keeping with the precision of FCM through more training data. Figure 6 provides the results of SOM1.

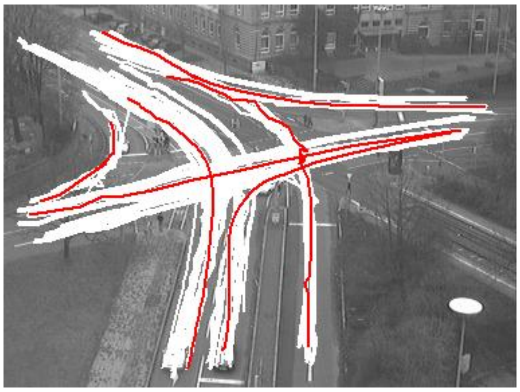
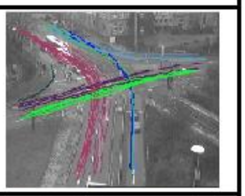
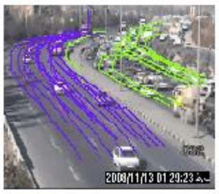
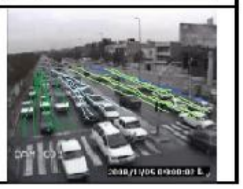


Figure 6 Result of SOM1

We chose a 2D SOM with two neurons using only temporal vectors as classifier of the second category. Once having performed class training, the system labelled slower classes as Low-Speed and faster ones as High-Speed, thus making them ready for querying. Zero was inserted into several rows of their temporal vectors to prevent them from entering the second clustering phase as momentary stop trajectories. Table 2 provides the results of tracking and classifying using different numbers of classes.

Table 2 Result of SOM2 by using Spatial-Temporal vectors

Class#	Tracking Step	SOM2 result
6		
2		
4		

To observe the elimination of abnormal trajectories from SVM training data, consider the following Figure 7: Two abnormal behaviours have been included presented in Figure 7.a which is present the trajectories of 6th class as SOM3 result. With an infraction rate of 5%, trajectories eliminated from SVM6 training data can be viewed in (b). As seen, these are only trajectories suitable for SVM6 training. In the test phase, the system successfully detected all sixteen cases of infractions available in the DB as abnormal behaviour such as zigzag driving, illegal U turn and so on, although it should be noted that four normal trajectories were mistakenly detected as abnormal.

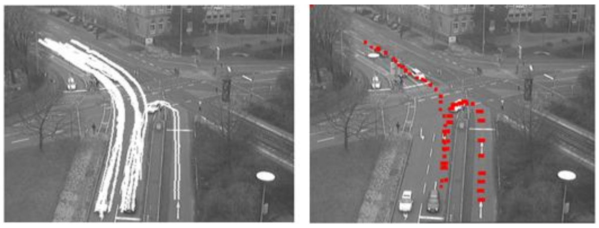


Figure 7 Rejecting illegal trajectories from SVM training data

3.2. Cluster Estimation

To assess the proposed cost function (1), the estimates of cluster numbers in Figure 8.a, b and c are displayed with their related crossroads in rows 1, 2, and 3 of Table 2. In each curve the minimum value of the cost function is considered as the suitable number of classes.

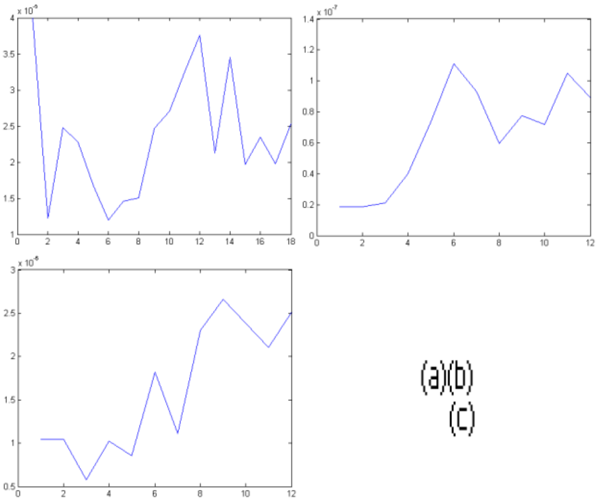


Figure 8 The minimum value of the cost function obtained in a) 6 for first row, b) 2 for second row and c) 3 for last row of table 2, (1 is not valid for class number)

3.3. Optimizing result

Figure 9 shows the result of the user query optimization by genetic algorithm and includes the main query, its optimized version, and its fitness information for different generations. The figure illustrates that the system can produce the proper query from the user query. Table 3 provides the retrieval results by optimized query. In the tables Recall formula is:

Recall = 100* (Correct Trajectories)/(Total + (False Trajectories))

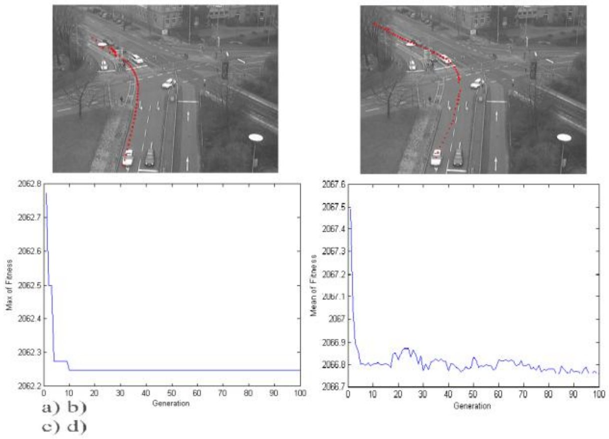
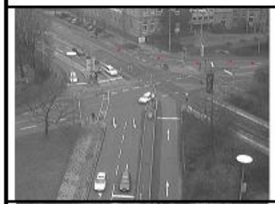
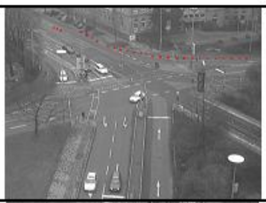
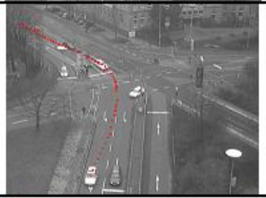
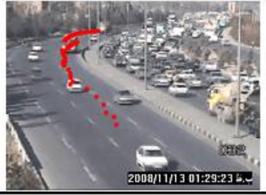


Figure 9 Optimizing the user query by using GA; a) user query b) optimized query c) maximum fitness d) minimum fitness

Table 3 Retrieval result by using the optimized query

User Query	Optimized Query	Recall
		75%
		93%
		73%

4. CONCLUSION

The system we have proposed here uses a minimum of a priori knowledge to retrieve activities and abnormal behaviour from traffic surveillance videos. For setup this system requires just two parameters in which knowledge is optimized in such a way that the system does not demand more information for the retrieval process. The parameters include: (1) the maximum number of existing junctions and (2) the rate of infractions in the city. The first parameter is used to make up the system's topology and the second to extract training patterns. Actually, this system is trained-based on Active Learning and so automatically indexes normal behaviour with its spatial temporal information and indexes illegal activities as abnormal behaviour. The proposed system uses hierarchical hybrid clustering to extract the manner of motion. In parallel and by using SOM, spatial-temporal vectors are classified according to infraction rates, possible abnormal data ruled out, and the remaining data used for the training of several SVM single classes. Experimental results show a high accuracy in hybrid classification, class number estimation, and the retrieval process. Also, a comparative improvement in retrieval using optimized user sketches was observed.

REFERENCES

[1] Doulamis, A.D., Doulamis, N.D. and Stefanos, 2000. A fuzzy video content representation for video summarization and content-based retrieval: Signal Processing, , 80, (6), 1049-1067

[2] 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

[3] 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

[4] Haag, M. and Nagel, H.-H. 2000. Incremental recognition of traffic situations from video image sequences: Image Vis. Comput., , 18, (2), 137–153

[5] Hu, W., Xie, D., Tan, T. and Maybank, S., 2004. Learning activity patterns using fuzzy self-organizing neural network: IEEE Trans. Systems, Man and Cybernetics - Part B, 34, (3), 1618-1626

[6] 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

[7] Junejo, I.N., Javed , O. and Mubarak Shah, 2004. Multi feature path modeling for video surveillance: Proc. Int. Conf. on Pattern Recognition, 716-719

[8] 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.

[9] Li, X., Hu, W., 2006. A coarse-to-fine strategy for vehicle motion trajectory clustering: Proc. Int. Conf. on Pattern Recognition, 591–594

[10] Liu, Z.Q., Bruton, et al., 2001. Dynamic image sequence analysis using fuzzy measures: IEEE Tran. on Syst., Man and Cyber., , 31, (4), 557–571

[11] Makris D. and Ellis. T., 2002. Path detection in video surveillance: Image and Vision Comp., 20, (12), 895-903

[12] 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

[13] 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

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

[15] Piriou, G., Bouthemy, P., and Yao J.F., 2006. Recognition of dynamic video contents with global probabilistic models of visual motion: IEEE Tran. on Image Processing, 15, (11), 3418-3431

[16] 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

[17] Safara F. and Naderi S., 2008. Fuzzy decision maker for knowledge discovery from image archives: 3rd International Conference on Geometric Modelling and Imaging, pp. 126-129

[18] 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

[19] 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