Abnormal High-Level Event Recognition in Parking lot

Najla Bouarada Ghrab[1(✉)], Rania Rebai Boukhriss[1], Emna Fendri[2], and Mohamed Hammami[2]

[1] MIRACL Laboratory, ISIMS, University of Sfax, BP 242, 3021 Sakiet Ezzeit Sfax, Tunisia
najla.bouarada@yahoo.fr, rania.rebai@hotmail.fr

[2] MIRACL Laboratory, FSS, University of Sfax, Road Sokra km 4, BP 802, 3038 Sfax, Tunisia
fendri.msf@gnet.tn, mohamed.hammami@fss.rnu.tn

Abstract. In this paper, we presented an approach to automatically detect abnormal high-level events in a parking lot. A high-level event or a scenario is a combination of simple events with spatial, temporal and logical relations. We proposed to define the simple events through a spatio-temporal analysis of features extracted from a low-level processing. The low level processing involves detecting, tracking and classifying moving objects. To naturally model the relations between simpler events, a Petri Nets model was used. The experimental results based on recorded parking video data sets and public data sets illustrate the performance of our approach.

Keywords: Object classification · Simple event · Scenario Abnormal event

1 Introduction

High-level video event recognition has received a lot of attention recently due to its diverse application areas in video surveillance. The goal of high-level video event recognition is to describe and infer a variety of abnormal behaviors that occur in the scene. Unlike normal events, the abnormal behaviors in a scene are rare. The focus of this study was to detect abnormal high-level events that may occur in a parking lot monitored by a surveillance camera through an explicit modeling. The high-level events recognition problem is hierarchical since the high-level event or scenario consists of multiple simple events with certain spatial, temporal, and logical relations. A simple event can be an action performed by a single object or an interaction between multiple objects. For this reason, we were motivated to address the problem at different levels. Our method is based on the extraction of features from a low level processing enhanced with spatio-temporal analysis. To describe and recognize high-level events or scenarios that incorporate simple events with temporal and spatial relations, a Petri Nets model

A. Abraham et al. (Eds.): ISDA 2017, AISC 736, pp. 389–398, 2018.
https://doi.org/10.1007/978-3-319-76348-4_38

was used. Our contributions in this paper are the following: First, the objects are automatically and accurately detected and classified. Second, the simple events are defined through a spatio-temporal analysis of the features related to the detected objects in the scene. Moreover, the abnormal scenarios are represented using the Petri Nets model.

The remainder of this paper was organized as follows. Section 2 reviewed the related work to the problems addressed by the proposed approach. Section 3 described our proposed method of high-level event recognition for an abnormal event detection. Section 4 detailed the experimental results. Finally, we conclude this work in Sect. 5.

2 Related Work

The generic process for surveillance system [1,2] consists of two processing levels: low-level and high-level. In what follows, the issues related to this system were discussed.

The low-level processing consists of three principle steps: moving object detection, tracking and classification. The moving object detection aims to isolate the foreground pixels that participate into any kind of motion observed in a given scene. The contributions reported in the literature can be classified in four main categories based either on inter-frames differences [3], background modeling [4], optical flow [5] or on a combination of two or all of these methods [6]. The next step is the objects tracking to locate the objects in time and extract their trajectories. There are two categories of tracking methods: Points based approach [7] and Model based approach [8] (silhouettes or kernel). Objects classification aims to improve the reliability of the high level processing results by classifying detected objects appearing in a given scene into classes of humans or vehicles. In the literature, object classification methods can be categorized primarily into four main approaches: Shape-based methods [9], Texture-based methods [10], Motion-based methods [11] and Hybrid methods [12]. These methods generally use a machine learning technique for the moving object classification.

The simple events can be a human action performed in a short time or an interaction between two objects. The high level processing takes into account the information obtained from the low-level processing in order to represent such simple events. These are combined afterwards to include the high-level events. In the literature, various techniques have been proposed for action recognition. These techniques can be classified as follows: (i) Human body model based methods [13], (ii) Holistic methods [14], and (iii) Local feature methods [15]. To represent the high-level events with a formal model and recognize these events as they occur in the video sequence, a great deal of work has also been proposed. Some researchers deal with state models that model the state of the video event in space and time using semantic knowledge, e.g. Finite-State Machines (FSM) [16], Bayesian Networks (BNs) [17], Hidden Markov Models (HMM) [18]... Other works model the high-level event by defining a set of semantic rules, constraints, and relations at the symbol level. Grammar based methods [19] and Petri Nets [20] are the most used in most of the studies dealing with this issue.

3 Proposed Approach

In this section we introduce our approach of high-level events recognition in a parking lot. As presented in Fig. 1, our approach consists of two levels with an increasing level of abstraction. In a low-level processing, we are interested in detecting, tracking and classifying moving objects to obtain the position and class of each object. In a high-level processing, we rely on a spatio-temporal analysis to recognize the simple events; as for the high-level events, they are identified with a Petri Nets model.

3.1 Low-Level Processing

The process of our low-level processing is composed of two phases: (1) an off-line phase adopting a data mining process to construct the appropriate prediction model for the moving object classification and (2) an on-line phase to detect, track and classify the moving objects in order to obtain the position and class of each detected moving object in the scene.

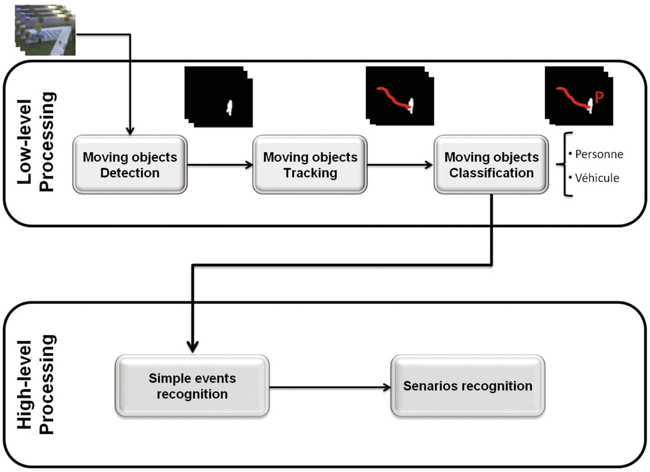

Fig. 1. The process of our approach.

Off-line Phase: The proposed method for moving objects classification operates in two steps: (a) the first step is devoted to the preparation of data for the learning phase; (b) the second step consists in finding the most appropriate prediction model using several data-mining algorithms in order to best discriminate the two classes of moving objects: Pedestrian (P) and Car (C).

(a) *Data Preparation:* In this step, we start by the construction of a large and representative learning dataset and a test dataset. Our training data is composed of 4583 moving objects of class P and 849 moving objects of class C. As for the test data set, it contains 4682 moving objects of class P and 108 moving objects of class C. Thereafter, a selection of the best features is required to ensure the best performance of the moving objects classification. In our work, we have selected seven shape based features that are Height, Width, Aspect Ratio, Area, Perimeter, Compactness and Anthropomethry; two famous texture based features that are the Histogram of Oriented Gradients (HOG) and the Local Binary Pattern (LBP) descriptors and one motion based feature that is the velocity of the object [21].

(b) *Data Mining:* The aim of this step is the construction of the appropriate prediction model. To this end, we introduce a new field of application of data mining techniques. Thus, we have selected 4 different learning algorithms: the SVM with polynomial kernel (Poly Kernel), SVM with Radial Basis Function Kernel (RBF Kernel), Multilayer Perceptron Neural Network (MLP) and the algorithm C4.5 of decision trees. The quality and the stability of the prediction models constructed by these four data mining algorithms are evaluated intensively in our experimental study to determine the most suitable prediction model for our classification task. This classification is relevant for the performance of simple event recognition.

On-line Phase: The online phase of our low-level processing is composed of three principle steps:

– *Moving Objects Detection:* To ensure better results of moving objects detection, we adopted a background modeling based method with a dynamic matrix and spatio-temporal analyses of scenes [22]. In addition, the adopted method originality lies in the integration of the principle of the methods based on inter-frame differences in the background modeling step and in the way this integration was carried out, making it more adaptable and independent of the moving objects speeds and sizes.
– *Moving Objects Tracking:* For the moving objects tracking, we adopted the method presented in [23] based on merging and splitting detection, and feature correspondence. This method can track multiple objects with long-duration even if there are partially or completely occluded.
– *Moving Object Classification:* Our moving object classification method is based on a hybrid approach combining the best features based on shape, texture and motion using a data-mining process to ensure a better classification of the moving objects. This classification is based on the prediction model constructed in the off-line phase of our approach.

3.2 High-Level Processing

Our high-level processing consists of simple events and scenarios recognition.

Simple Events Recognition. We were interested in recognizing a set of simple events based on the information extracted from the first level processing, namely the object Bounding Box, the object blob, the object center position and the object class. Thus, each moving object is characterized by the spatial features presented in Eq. 1:

$$MO = [x, y, id, R, x_{top}, y_{left}, x_{height}, y_{width}] \tag{1}$$

where (x, y) are the center coordinates, id is the class, R is the blob, x_{top}, y_{left}, x_{height}, y_{width} are the coordinates of the Bounding Box.

In our work, the simple events were defined with respect to persons and cars. A simple event can be the action performed by one object (car enters, car moves, car stops, person enters) and can also be performed between two or more objects, which is commonly referred to an interaction (person walks-towards, person joins car, person joins person, person turns car, person splits from car, person splits from person, person walks away, person runs away). We have two types of interactions: person-to-person and person-to-car. An important aspect in the abnormal event detection is to differentiate between the simple events of the person. A person running away is different from a person walking. So, we enhanced our spatio-temporal analysis with a motion based descriptor defined by Histograms of Optical Flow (HOF) [24]. This descriptor allows differentiating between walks-towards and runs-towards and between walks-away and runs-away. The HOF descriptor was computed by blocks. The number of blocks used is 3∗3 blocks in the spatial domain and 2 in the temporal domain. The responses per block were computed using Lucas-Kanade method [25]. Then for each response the magnitude was quantized in o orientations, usually $o = 8$. Finally, adjacent blocks were concatenated to form the descriptors. To determinate the action of a person, the Bounding Box extracted during the detection and classification steps were firstly resized to the maximum dimension of the Bounding Box. Then they were organized in a figure centric spatio-temporal volume for each person. The Bounding Box were segmented to cycles where a cycle begins with an initial swing of the two feet. Finally, the HOF descriptors were calculated for each cycle.

For the recognition of interactions between objects, a vector of spatio-temporal features was extracted for each moving object at each time t, as presented in Eq. 2:

$$f^t = [d_{ij}, op_{ij}, th_{ij}, se_{ij}] \tag{2}$$

where d_{ij} is the distance between the two objects, op_{ij} is the overlap between the two objects, th_{ij} is the angle between the vector that forms the objects centers and the original axis and se_{ij} is the closest segment between the objects. The extracted interactions between two objects were detailed in Table 1.

Scenarios Recognition. Scenarios were defined by a Petri Nets model which allows benefiting from the following advantages:

- It naturally models the relations between the simpler events including non sequential (and sequential) temporal relations, spatial and logical composition, hierarchy, concurrency, and partial ordering;

Table 1. Determination of interactions between two objects.

Interaction	Determination
MoveAway (O_1, O_2)	$d^{t+1}(O_1, O_2) < d^t(O_1, O_2)$
MoveFrom (O_1, O_2)	$d^{t+1}(O_1, O_2) > d^t(O_1, O_2)$
Merge (O_1, O_2)	$op^t(O_1, O_2) = 0$ and $op^{t+1}(O_1, O_2) = 1$
Split (O_1, O_2)	$op^t(O_1, O_2) = 1$ and $op^{t+1}(O_1, O_2) = 0$
TurnB (O_1, O_2)	$op^t(O_1, O_2) = 1$ and $SE =$ 'b' and $th \in [-90, 90]$
TurnL (O_1, O_2)	$op^t(O_1, O_2) = 1$ and $SE =$ 'l' and $th \in [-45, 135]$
TurnR (O_1, O_2)	$op^t(O_1, O_2) = 1$ and $SE =$ 'r' and $th \in [-90, 0]$
TurnF (O_1, O_2)	$op^t(O_1, O_2) = 1$ and $SE =$ 'f' and $th \in [90, 180]$

– It provides a nice graphical representation of the event model;
– PN event models are specified manually without any need of long video for training.

A Petri Net model is made up of set of Places, a set of Transitions, a set of Arcs between a Place and a Transition or between a Transition and a Place, and the tokens. In our work, the set of simple events are represented by the transitions and the places are the state describing a situation of the objects. Each token in a place represents a distinguished moving object which can be cars or persons. As Petri Nets model the scenarios as a combination of simple events with the spatial, temporal and logical relations, we start by presenting these relations:

Temporal Relations: Temporal relations are heavily important when describing scenarios. We adopt the Allen's relations [26] between two intervals: 'before', 'meets', 'overlaps', 'stars', 'during' and 'finishes'.

Spatial Relations: A spatial relation is a relation between spatial objects. Thus, they are only defined in their interactions.

Logical Relations: The logical relations include and, or and not. It is designed for concatenating the spatial and temporal relations.

This research study addressed two interesting scenarios of theft in a parking: theft from a car and theft from a person. The two Petri Net models are defined respectively in Fig. 2(a) and (b).

In the first scenario, we have two objects: a Car C_0 and a Person P_1. The Petri Net model combines the simple events with a sequential order. Whenever a car enters the scene, a new token is inserted in the place p_1. If the car moves, its token is moved from p_1 to p_2 and so on. Whenever a Person P_1 enters the scene, a new token is inserted in the place p_4. If the person P_1 walks towards C_0, the token is relocated to p_4. At the end, the number of token in p_7 denotes the number of objects in the scene. In the second scenario, the interactions are between two persons P_1 and P_2. Whenever a person P_1 enters and moves in the

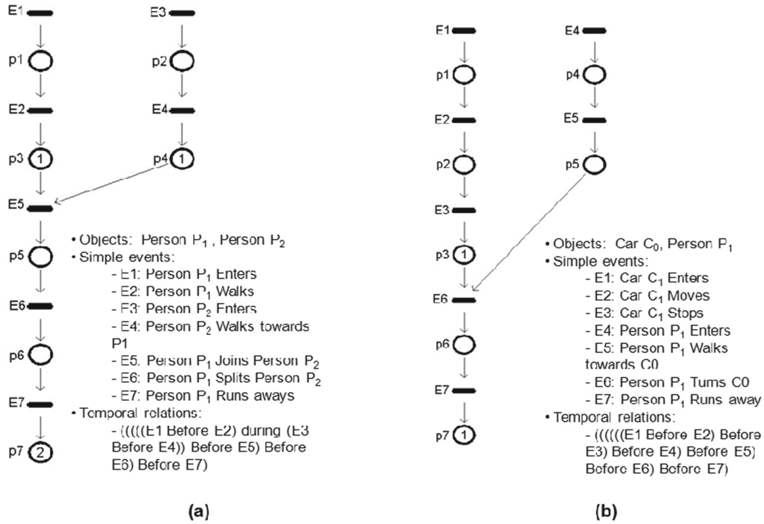

Fig. 2. Petri Net representation.

scene, a new token is inserted in the place p_1 then replaced to p_3. Whenever a person P_2 enters and moves in the scene, a new token is inserted in the place p_2 then relocated to p_4. The tokens from p_3 and p_4 are not matched in p_5 until the simple event 'joint' is detected. The same process is performed until the token is inserted into p_7.

4 Experimental Results

In this section, we presented two series of experiments: The first series detailed the experiment concerning the moving objects classification of the low level processing, whereas the second series gave the results of our high-level event recognition.

4.1 First Series of Experiments

In this series of experiments, we aimed to show the performance of the moving objects classification. The experiments were performed to select the most adequate learning technique. We used a corpus[1,2] of data recorded in different scenes and in various weather conditions. The data contain moving objects of different classes. In this study, we proceeded to select the most adequate learning technique. We used such different classifiers as the Multilayer Perceptrons (MLP), Support Vector Machines (SVM) with polynomial kernel and RBF kernel and C4.5 based decision trees. We compared the cited data mining algorithms using

[1] http://vcipl-okstate.org/pbvs/bench/.
[2] http://www.ino.ca/Video-Analytics-Dataset.

the Total Correct Classification (TCC) rate. The results of this experiment are shown in Fig. 3 in which the best TCC rates are achieved by the SVM with polynomial kernel.

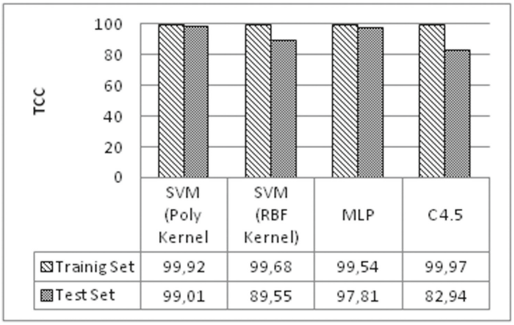

Fig. 3. Moving object classification results with different machine learning techniques.

4.2 Second Series of Experiments

Given the lack of public datasets representing abnormal scenarios in a parking lot, we applied our framework on our video stream captured from a static camera monitoring a university parking lot. In this dataset, 28 sequences are recorded where 14 participants. The dataset involves two types of thefts: theft from a car and another from a person. Our system is used to model and detect these abnormal scenarios. Once the abnormal scenario is detected, an alert is trigged by the system. For the two thefts scenarios, we have eight simple events: Enter, Join, Split, Walk-towards, Run-towards, Walk-away, Run-away, and Turn around.

Since the high-level event recognition is related to the simple events recognition, the final results are strictly bounded by the results of the simple event recognition. To validate our method in the case of a simple event performed by a single object, we extracted a model of walk and run using a popular benchmark dataset WEIZMANN [27]. The WEIZMANN dataset contains 90 videos of 10 actions performed by 9 persons. We prepared a learning data and used the SVM

Table 2. Performance of simple events recognition.

	Enters	Joins	Splits	Moves-toward	Moves-away	Turns	Walks	Runs
TP	6	5	4	134	172	601	20	6
FP	0	0	0	0	0	0	0	0
FN	0	0	0	7	0	0	6	0
Precision	1	1	1	1	1	1	1	1
Rappel	1	1	1	0.95	1	1	0.89	1

with Poly kernel to recognize the simple event. Based on this model, we achieved the results presented in Table 2 that shows the performance of the simple events defined with spatio-temporal analysis.

5 Conclusion

We have presented a new approach of high-level video recognition for an abnormal event detection in a parking lot. Our approach proceeds by detecting, tracking and classifying the moving objects in the scene first. Based on the position and the class of the detected moving object, a module of simple events recognition is then performed. The defined simple events are combined with spatial, temporal and logical relations to recognize the abnormal scenarios through a petri Nets model. The achieved results prove the efficiency of our approach.

References

1. Kim, I.S., Choi, H.S., Yi, K.M., et al.: Intelligent visual surveillance a survey. Int. J. Control Autom. Syst. **8**(5), 926–939 (2010)
2. Vishwakarma, S., Agrawal, A.: A survey on activity recognition and behavior understanding in video surveillance. Vis. Comput. **29**(10), 983–1009 (2013)
3. Cheng, Y.H., Wang, J.: A motion image detection method based on the inter-frame difference method. Appl. Mech. Mater. **490–491**, 1283–1286 (2014)
4. Jian-Ping, T., Xiao-lan, L., Jun, L.: Moving object detection and identification method based on vision. Int. J. Secur. Appl. **10**(3), 101–110 (2016)
5. Hariyono, J., Hoang, V.D., Jo, K.H.: Motion segmentation using optical flow for pedestrian detection from moving vehicle. In: Hwang, D., Jung, J.J., Nguyen, N.T. (eds.) Computational Collective Intelligence. Technologies and Applications. ICCCI 2014. LNCS, vol. 8733. Springer, Cham (2014)
6. Kushwaha, A.K.S., Srivastava, S., Srivastava, R.: Multi-view human activity recognition based on silhouette and uniform rotation invariant local binary patterns. Multimed. Syst. J. **32**(4), 451–467 (2017)
7. Gargi, P., Rajbabu, V.: Mean LBP and modified fuzzy C-means weighted hybrid feature for illumination invariant meanshift tracking. Signal Image Video Process. **11**(4), 665–672 (2017)
8. Yang, Y., Cao, Q.: A fast feature points-based object tracking method for robot grasp. Int. J. Adv. Robot. Syst. **10**(3) (2013)
9. Dedeoğlu, Y., Töreyin, B.U., Güdükbay, U., Çetin, A.E.: Silhouette-based method for object classification and human action recognition in video. In: Huang, T.S., et al. (eds.) Computer Vision in Human-Computer Interaction. ECCV 2006. LNCS, vol. 3979, pp. 64–77. Springer, Heidelberg (2006)
10. Moctezuma, D., Conde, C., Diego, I.M., Cabello, E.: Person detection in surveillance environment with HoGG: Gabor filters and histogram of oriented gradient. In: ICCV Workshops, pp. 1793–1800 (2011)
11. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. Pattern Recognit. **36**(3), 585–601 (2003)
12. Zhang, Z., Cai, Y., Huang, K., Tan, T.: Real-time moving object classification with automatic scene division. In: IEEE International Conference on Image Processing, pp. 149–152 (2007)

13. Chakraborty, B., Bagdanov, A.D., Gonzàlez, J., Roca, X.: Human action recognition using an ensemble of body-part detectors. Expert Syst. **30**(2), 101–114 (2012)
14. Bobick, F., Davis, W.: The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Mach. Intell. **23**(3), 257–267 (2001)
15. Wang, H., Kläser, A., Schmid, C., Cheng-Lin, L.: Dense trajectories and motion boundary descriptors for action recognition. Int. J. Comput. Vis. **103**(1), 60–79 (2013)
16. Lv, F., Nevatia, R.: Single view human action recognition using key pose matching and Viterbi path searching. In: IEEE Conference on Computer Vision and Pattern Recognition, p. 18 (2007)
17. Wang, Y., Cao, K: A proactive complex event processing method for large-scale transportation Internet of Things. Int. J. Distrib. Sens. Netw. **10**(3) (2014)
18. Bansal, N.K., Feng, X., Zhang, W., Wei, W., Zhao, Y.: Modeling temporal pattern and event detection using Hidden Markov Model with application to a sludge bulking data. Procedia Comput. Sci. **12**, 218–223 (2012)
19. Zhang, Z., Tan, T., Huang, K.: An extended grammar system for learning and recognizing complex visual events. IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 240–255 (2011)
20. Ghanem, N., DeMenthon, D., Doermann, D., Davis, L.: Representation and recognition of events in surveillance video using Petri nets. In: Conference on Computer Vision and Pattern Recognition Workshop, pp. 112–121 (2004)
21. Boukhriss, R.R., Fendri, E., Hammami, M.: Moving object classification in infrared and visible spectra. In: International Conference on Machine Vision (2016)
22. Hammami, M., Jarraya, S.K., Ben-Abdallah, H.: On line background modeling for moving object segmentation in dynamic scene. Multim. Tools Appl. **63**(3), 899–926 (2013)
23. Yang, T., Li, S.Z., Pan, Q., Li, J.: Real-time multiple objects tracking with occlusion handling in dynamic scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2005)
24. Uijlings, J., Duta, I.C., Sangineto, E., Sebe, N.: Video classification with densely extracted HOG/HOF/MBH features: an evaluation of the accuracy/computational efficiency trade-off. Int. J. Multimed. Inf. Retr. **4**(1), 33–44 (2014)
25. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: International Joint Conference on Artificial Intelligence (1981)
26. Allen, J.F., Ferguson, F.: Actions and events in interval temporal logical. J. Log. Comput. **4**(5), 531–579 (1994)
27. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: IEEE International Conference on Computer Vision (2005)