

Incidents Prediction in Road Junctions Using Artificial Neural Networks

Tarik HAJJI, Aicha ALAMI HASSANI and Mohammed OUZZANI JAMIL

Laboratoire Systèmes et Environnements Durables (SED), Faculté des Sciences de l'Ingénieur (FSI), Université Privée de Fès (UPF), Fez, Morocco

hajji@upf.ac.ma

Abstract. The implementation of an incident detection system (IDS) is an indispensable operation in the analysis of the road traffics. However the IDS may, in no case, represent an alternative to the classical monitoring system controlled by the human eye. The aim of this work is to increase detection and prediction probability of incidents in camera-monitored areas. Knowing that, these areas are monitored by multiple cameras and few supervisors. Our solution is to use Artificial Neural Networks (ANN) to analyze moving objects trajectories on captured images. We first propose a modelling of the trajectories and their characteristics, after we develop a learning database for valid and invalid trajectories, and then we carry out a comparative study to find the artificial neural network architecture that maximizes the rate of valid and invalid trajectories recognition.

Keywords. Incidents prediction, Images and videos processing, ANN, Artificial intelligence.

1. Introduction

The installation of surveillance cameras is an essential operation for the security, the detection of fraud, thefts, burglaries, assaults, vandalism. However, this operation is necessary but insufficient to meet the requirements of the monitoring. Consequently, the installation of an additional software layer like an IDS is highly recommended to increase the monitoring system performances.

The French Tunnel Research Engineering published in 2015 a report entitled «automatic incident detection using image analysis in the tunnel» [1]. The report contains general information and history of IDS, different types of incidents to detect the description of the operating principle such as the arrangement of cams and disruption management, the operating mode in the case of detection of an incident by the IDS valuation method.

1.1. Artificial neural networks

ANN are problem-solving tools inspired by the biological neural networks that constitute human brains, and it's a network of many very simple processors ("units"), each possibly having a local memory. It is based on a collection of connected artificial neurons organized in layers. Most ANN have three layers: the input layer, a hidden layer, and the output layer.



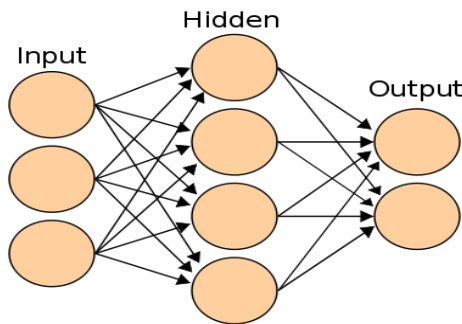


Figure 1. ANN[2]

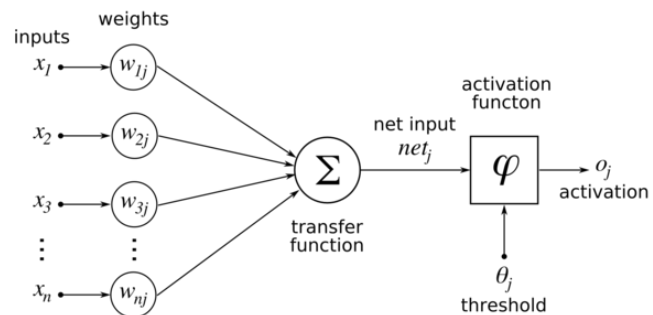


Figure 2. Artificial neural[2]

(Fig. 1) shows the structure of an ANN containing one hidden layer and one output layer with 4 neurons and 2 neurons, respectively, and three inputs. An artificial neural is a function able to realize a non-linear sum of its inputs (Fig. 2.). Most ANN have some sort of "training" rule whereby the weights of connections are adjusted on the basis of presented patterns. The Back-Propagation algorithm [2] which is one of the most popular learning algorithms is a method used in ANN to calculate the error contribution of each neuron after a batch of data is processed.

2. State-of-the-art

The aim of this part is to make an overview of the various methods of abnormal event detections in videos. The analysis of moving objects behavior in videos attracts a great number of researchers and therefore several approaches have been proposed. Most of them deem movement as low-level information. They then define intermediate level descriptors to model the studied behavior.

2.1. Foreground obtained using a background extraction process:

The simplest way to separate foreground from background is to acquire a background image that does not contain any moving object. A subtraction between each new image and initial background image is then performed with respect to a threshold. However a single image of background is often not available because it changes continuously due to various events. The background modeling is used in a variety of applications such as video surveillance [3, 4] or multimedia [5].

Methods can be classified into the following categories: background basic modeling [6], background statistical modeling [7], background fuzzy modeling and background estimation [8]. They can be also classified according to prediction [9], recursion [3], adaptation [10], or modality [11].

The Gaussians are then heuristically marked as follows: the darkest component is marked as a shadow and the other two components are differentiated according to the variance (the greatest variance is labelled as a vehicle and the other as the road). The foreground classification is made by associating each pixel with the corresponding Gaussian. An incremental algorithm is used for background maintenance to perform real-time processing. However, this process suffers from a lack of adaptation to the changes that appear in the scene through time. Stauffer and Crimson [12] have generalized this idea by modeling the color intensity of each pixel over the previous time interval $\{X1, X2, Xt\}$ by a mixture of Gaussians.

2.2. Points of interest:

Points of interest are pixels of the image that are likely to be tracked effectively in the following frames. The Harris algorithm for interest point's extraction [13] is known for its invariance.

3. Methodology

Our approach consists in starting with the modeling of a moving object (MO) by a motion vector composed of a collection of characteristics representing its behavior in the video. Vehicles or pedestrians or other objects captured by the cameras in the monitored area are referred to as movable objects. A Motion Vector (MV) of a MO also designates a mathematical object capable of

representing the state of a moving object at a given instant. Then we build a Knowledge Database (KDB) formed by MV for different MO and the score of each MV. The KDB is used as a data source for the system-learning algorithm. Afterwards, a comparative study is made to select the most suitable ANN architecture for this subject in terms of recognition rate and complexity. Finally, the method is evaluated and the results are compared with other SDIs.

3.1. Design of the motion vector

In order to characterize a MO at a given instant, a set of characteristics of the object's movement is extracted. In this study, we consider the following characteristics: the mask, the speed, the acceleration, the type of the trajectory, the direction, and the volume. These characteristics represent standardized mathematical quantities:

- The mask indicates the number of the area occupied by the moving object on the capture
- The type of the trajectory indicates the moving object type of movement: rectilinear uniform, rectilinear, sinusoidal, diagonal, circular ...
- The direction indicates the movement direction, it is a number between 0 and 386.
- Speed and acceleration are two classical physical quantities
- The volume indicates the area occupied by the moving object on the capture.

Thus the motion vector for an object is defined as follows $V_m = [mask, velocity, acceleration, trajectory, direction, volume]$. The figure below shows an example of a motion vector for a moving object:

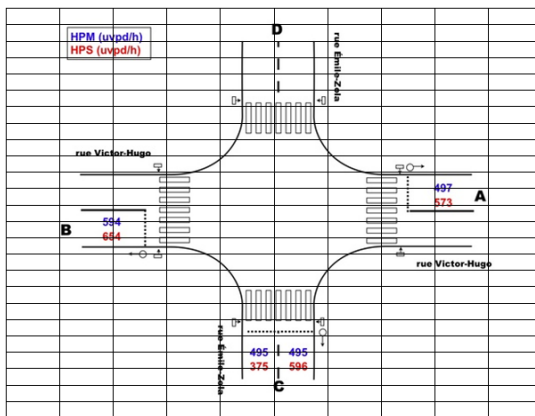


Figure 3. Moving objects: A, B, C and D

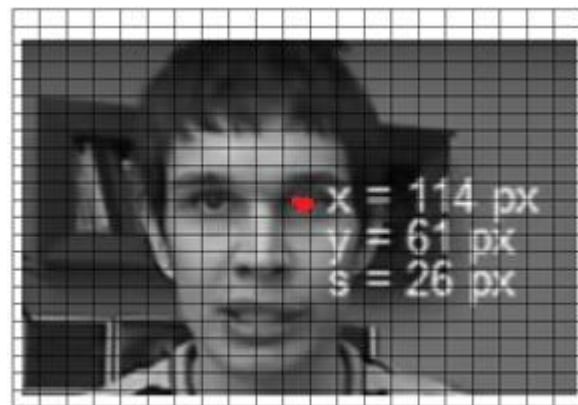


Figure 4. The mask deduction of the MO

The (Fig. 3) is composed of 4 moving objects A, B, C and D. The motion vector for each of them is respectively: $V_m(A)=[4,60, 0.5, 2, 180, 10]$; $V_m(B)=[4,60, 0.5, 2, 180, 10]$; $V_m(C)=[4,60, 0.5, 2, 180, 10]$; $V_m(D)=[4,60, 0.5, 2, 180, 10]$;

3.1.1. Mask of a moving object. The mask of a moving object is an identifier with a specific area on the capture. The zones are numbered from 1 to n, where n is the number of mask, in figure n equals to 200. In the case where the mobile object is distributed over more than one mask, then the most occupied mask is considered. We use *OpenTLD-master* (TLD) [14] which is an algorithm for tracking unknown objects in unconstrained video streams. The object of interest is defined by a bounding box in a single frame. TLD follows simultaneously the object, learns its appearance and detects it whenever it appears in the video. The algorithm is implemented by **Z. KALAL** in form of a Matlab library, and then we extract the coordinates of the MO on the scene captured by the camera and infer from it the mask. The (Fig. 4) explains the operating mechanism for the mask deduction of the MO (Red dot), which is the eye of the **Z. KALAL**.

3.1.2. Moving object trajectory. The trajectory of a moving object designates the motion type of an MO. We have considered a continuous notation of the moving object trajectories: 1 for uniform

rectilinear, 2 for rectilinear, 3 for sinusoidal, 4 for diagonal and 5 for circular. We used the notion of polynomial interpolation [15] for the classification of the MO trajectories in a mask.

The principle of the technique is to detect the trajectory curve according to the occupations of the moving object in a mask. The occupations are represented by the moving object coordinates in relation to a given reference. We compute these coordinates by *OpenTLD-master*.

For the recognition of the trajectory type represented by the polynomial, we used an ANN with supervised learning on a set of points of the predefined trajectories. The trajectories studied in this work, each one of them is represented by a Lagrange polynomial. These polynomials allow us to build a learning database on a set of points. For each polynomial we compute the images of a collection of n points $X_i = P_j[X_i]$ (P_j represents the polynomial of the trajectory j and X_i represents the point i) and a learning database is obtained for the ANN. The (Tab. 1) shows an extract of this database.

Table 1. Learning database for the trajectories classification

X_i	$P_1[X_i]$	$P_2[X_i]$	$P_3[X_i]$	$P_4[X_i]$	$P_5[X_i]$
0,90576579	0,52169653	0,10916584	0,33935151	0,3173368	0,37169586
0,61152489	0,45771693	0,46002726	0,55990782	0,92800427	0,26958747
0,72269234	0,93201435	0,83723601	0,72172925	0,69169074	0,24253564

We use an ANN with an input layer of 6 artificial neurons, a hidden layer of 3 artificial neurons and an output layer of 5 artificial neurons and the backpropagation algorithm for learning.

3.1.3. Moving object speed. It is a mathematical quantity, which designates the number of pixels traversed by a moving object in a time interval (pixels / time). We have experimentally fixed the time interval required to calculate the speed of a moving object on a mask. We exploited the coordinates of the mobile object returned instantaneously by *OpenTLD-master* to calculate the number of pixels travelled by the moving object during a fixed time interval. The principle of the algorithm for calculating the moving object speed is simple; we saved the couple (position, time) in a list. Then we considered that the speed of the object during a fixed duration is the distance between the position recorded at the initial moment and the position recorded at the final moment of the duration.

3.1.4. Moving object acceleration. This parameter represents the evolution of the mobile object movement in a mask. It is a mathematical quantity that measures the speed difference between two instants.

3.1.5. Moving object direction. We designate by this mathematical magnitude the angle between the motion direction of the moving object and the horizontal axis used as a reference on a segment on a mask.

3.1.6. Moving object volume. Several solutions are available and able to calculate the actual dimensions of an object on an image captured as *UPHOTOMEASURE* [16]. In this work we considered only the virtual volume of the moving object and not the actual volume occupied in the real space.

3.2. Design of the knowledge base

We have experimentally built this knowledge base (Tab. 2) in our laboratory. We created a prototype of a roundabout, used a camera to monitor and simulate moving objects by objects that we have prepared in our laboratory. Then we programmed valid and invalid configurations and finally computed the motion vectors for each moving object.

Table 2. Base of knowledge

MO	Mask	Trajectory	Speed	Acceleration	Direction	Surface	Score
----	------	------------	-------	--------------	-----------	---------	-------

1	625	9	145	5	195	97	3
2	47	6	129	-4	223	24	4
3	58	3	58	-3	136	4	1
4	417	7	35	-8	0	73	4
5	75	6	116	6	67	78	6

3.3. Implementation:

We have developed a JAVA application that generates an ANN with the postulated architecture based on a simple description in the form of an xml meta-model. The application also has the ability to start learning of the ANN by the back-propagation of the gradient. The learning data are loaded via input.txt and output.txt files and the learning results are generated in the results.txt file.

4. Results

The (Tab. 3) shows the results of a comparative study to choose the structure of the most adopted ANN for this subject in terms of convergence rate.

Table 3. Comparative study results

ANN	Quadratic Error	Number of iterations	Recognition Rate	
			Learning Database (%)	Test Database (%)
6 : 3 : 1	0.01	10^6	100	97
		10^9	100	99
		10^{10}	100	99,8
	0.0001	10^6	100	98
		10^9	100	99,6
		10^{10}	100	100
6 : 2 : 1	0.01	10^6	100	97
		10^9	100	99
		10^{10}	100	99,8
	0.0001	10^6	100	98
		10^9	100	99,6
		10^{10}	100	100
6 : 3 : 2 : 1	0.01	10^6	100	97
		10^9	100	99
		10^{10}	100	99,8
	0.0001	10^6	100	98
		10^9	100	99,6
		10^{10}	100	100

Based on these results, we can conclude that with the ANN characterized by: an input layer composed of 6 artificial neurons, a hidden layer composed of 3 artificial neurons, an output layer composed of a single artificial neuron, a quadratic error average of the order of 10^{-4} and with an iteration number equal to 10^{10} we have a perfect recognition rate equal to 100% tested on the learning database.

5. Conclusion

In this work, we deal with the subject of automatic detection of undesirable events in surveillance cameras in the roundabouts of cities in order to propose intelligent, rapid and effective intervention. We proposed a solution based on artificial neural networks with a supervised learning for the

classification of detected undesirable events. These events are detected via a vector value calculated in real time on the videos of the surveillance cameras. The experimental results are very encouraging and our perspective is to evaluate this approach on a real case.

References

- [1] Centre d'Etudes des Tunnels, détection automatique d'incident par analyse d'images en tunnel, France, 2015, http://www.cetu.developpement-durable.gouv.fr/IMG/pdf/CETU_DocInfo_DAI_2015.pdf
- [2] Syed Danish Ali and Rahul Ahuja, The Evolution and Core Concepts of Deep Learning & Neural Networks, AUGUST 3, 2016, <https://www.analyticsvidhya.com/blog/2016/08/evolution-core-concepts-deep-learning-neural-networks/>.
- [3] Cheung and . Kamath, Robust Background Subtraction with Foreground Validation for Urban Traffic Video, Special Issue on Advances in Intelligent Vision Systems : Methods and Applications, pp.2330-2340, 2005.DOI10.1155/ASP.2005.2330.
- [4] Jiyun Zhai, Xin Zhou, Congqing Wang, "A moving target detection algorithm based on combination of GMM and LBP texture pattern", Guidance Navigation and Control Conference (CGNCC) 2016 IEEE Chinese, pp. 1057-1060, 2016.
- [5] A.PANDE, A. VERMA et A. MITTAL : Network aware optimal ressource allocation for e-learning videos. In 6th International Conference on mobile Learning, Melbourne- Australia, 2007.
- [6] Lee, B.; Hedley, M. Background estimation for video surveillance. In: Image & Vision Computing New Zealand (IVCNZ '02); 2002; Auckland, NZ. 2002. 315-320.
- [7] Xuannan Ye, Wanggen Wan, "Fast background modeling using GMM on GPU", Audio Language and Image Processing (ICALIP) 2014 International Conference on, pp. 937-941, 2014.
- [8] S. Messelodi, C. Modena, N. Segata, and M. Zanin, A Kalman Filter Based Background Updating Algorithm Robust to Sharp Illumination Changes, 13th International Conference on Image Analysis and Processing (ICIAP), pp.163-170, 2005. DOI10.1007/11553595_20
- [9] Wang and . Suter, A Novel Robust Statistical Method for Background Initialization and Visual Surveillance, Asian Conference on Computer Vision (ACCV), pp.328-337, 2006. DOI10.1007/11612032_34
- [10] F. Por03 and . Porikli, Human body tracking by adaptive background models and meanshiftanalysis, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2003.
- [11] Z. Tao and N. Ram, Bayesian human segmentation in crowded situations, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), p.459, 2003.
- [12] Stauffer and . Grimson, Adaptive background mixture models for real-time tracking, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.246-252, 1999.
- [13] Mrinal Kanti Bhowmika, Sandip Shilb, Priya Sahac: Feature Points Extraction of Thermal Face using Harris Interest Point Detection: International Conference on Computational Intelligence: Modeling Techniques and Applications(CIMTA) 2013.
- [14] Z.Kalal, k.Mikolajczyk and J.Matas: Tracking-learning-detection, Journal Pattern Analysis and Machine Intelligence, 2012.
- [15] Endre Süli and David Mayers (2003). An Introduction to Numerical Analysis, Chapter 6. Cambridge University Press. ISBN 0-521-00794-1.
- [16] Digicontractor Corporation:http://www.uphotomeasure.com/UPM_a2/index.htm, 2016.