Automatic security system for recognizing unexpected motions through video surveillance

Flavia Demarinis, Antonella Accettura, Fabio Garzia, Roberto Cusani
Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni, DIET
University of Rome La Sapienza
Rome, Italy
flaviademarinis.87@gmail.com, {antonella.accettura, fabio.garzia, roberto.cusani}@uniroma1.it

Abstract— **This paper deals with a study concerning the so called "smart video surveillance" system, starting from the consideration of unexpected motions. It is known that security staff, whose aim is to watch monitors and approach in case something bad and unlawful happens, in any kind of location, can keep the attention up for no more than twenty minutes, then the concentration falls down severely. Since this decrease of efficacy, it may be helpful a support system in watching and analysing the real-time or recorded scenes. According to the state of the art, unlike several other examples of smart video surveillance techniques (3,4,5,6), this one described in this article does not focus on the images, it does rather on the velocity parameter. Given a certain scenario, certain behaviours, thereby velocities, are expected, and it is supposed they might happen. The anomalies recognition is done using artificial neural networks, which are built and trained in order to compare an array (target) of expected velocities in normal conditions of life, to several arrays (input) of other velocities extracted from unexpected situations properly chosen. Once decided the tool for tracking the motions in the videos, then obtained the arrays of x and y velocities components, it is the time to build the artificial neural networks through iterations, which have been done changing the number of hidden neurons. The results are interesting enough and go right to the purpose, which consists in the choice of the best neural network.**

Keywords—smart video surveillance, intelligence, unexpected motions, security, automatic security system

I. INTRODUCTION

Unexpected motions are those motions that, given certain surrounding circumstances, (such as time, place of the world, society, location, and so on) might be considered unexpected as unusual if related to the motions already spotted in the same circumstances, for instance, or those ones which are commonly thought as reasonable according to the common sense. It could look much easier to recognize motions as unexpected or not through some very specific characteristics and parameters that should be taken into account and that might help to define a certain behavior. This is what in 2001 two psychologists Davison and Neale (1) tried to do through the following list in the book they wrote together "Abnormal Psychology": 1) statistical infrequency, 2) violation of norms, 3) personal distress, 4) disability or dysfunction, 5) unexpectedness, 6) time, 7) place, 8) culture, 9) time of the year, 10) season.

Then, if on one hand it is important to recognize the unexpected motions, on the other hand it is important to try a kind of mathematical modelling of the motions themselves [20]. The transition from a rational behavior to an irrational one is controlled by nervousness that influences the strength of the subject and the velocity of their movement. Le Roy F. Henderson tried to model people movements supposing that in normal conditions, crowd does behave as gas or fluid, even if with some revision. In large spaces, it should be possible to assume that people move according to the equation $Q(\rho)=\rho V(\rho)$ (1). Where Q represents people flow, ρ represents their density and V their velocity.

In panic crowded space situations, psychology shifts from individual to crowded psychology, it is said herding behavior, which is irrational and takes bad results, such as dangerous overcrowding, getaways get longer and sometimes fatalities increase (7,8,9,10) . There are specific behaviors and facts that can be distinguished in panic situations: 1) pedestrians become very nervous; 2) pedestrians try to walk faster for escaping from the source of the fear; 3) pedestrians push each other, so they interact physically; 4) movements become mismatched when pedestrians must go through a bottleneck; 5) physical interactions increase in crowded escape situations, and they might reach dangerous pressure, which could bend steel barriers and tear down brick in the walls; 6) arching and clogging could be observed at exits; 7) extra exits, such as side exits, are usually ignored and then not used as escape exit; 8) escapes are obstructed by fallen or injured people that become real obstacles; 9) people show herding behavior.

After the tragic Nine Eleven, there has been a great increase and development of video surveillance and video analysis (12,13,14). These instruments are very useful for controlling high-risk areas; they support human activities of control. Moreover, the terrorism threat has created a shift in the security paradigm from "investigation of incident" to "prevention of potentially catastrophic incidents", this is the main meaning video surveillance acquired.

Existent video surveillance systems allow to record information; meanwhile threat detection is left to human operators. It is an exhausting work, and many errors may occur. There were made many studies about effectiveness of human monitoring of surveillance videos, and a study made by US national Institute of Justice concluded as follows: "Such a task (manually detecting events in surveillance video), even

978-1-4799-3532-1/14/$31.00 ©2014 IEEE

when assigned to a person who is dedicated and well-intentioned, will not support an effective security system. After only 20 minutes of watching and evaluating monitor screens, the attention of most individuals has degenerated to well below acceptable levels. Monitoring video screens is both boring and mesmerizing. There are no intellectually engaging stimuli, such as when watching a television program." Smart video surveillance (21,22,23) has at least two advantages, even if computer programming to detect, classify and spatial and temporal coding pixels is a very complex issue: 1) real-time threat alert: there is an alert when a certain event, defined by users, happens in the camera field of view (for example automatic detection of missed objects); 2) rapid video search: this system allows to look for a specific object or event (for example it could be required to find all the red vehicles in the traffic flow).

II. THE CASE STUDY

The purpose of the study was to develop a system able to recognize unexpected motions in surveillance video thanks to the use of Artificial Neural Network. Researches demonstrated that security staff attention decreases after 20 minutes, for this it would be taken into account the possibility to create a system able to recognize unexpected motions analyzing velocities (16,17), rather than images like many others already did. In fact velocity changes when anything in the scene happens. For example, in case of shooting, people will run away from the source of danger increasing their speed. The same kind of behavior should be expected in case of fire or explosion. Therefore, an increasing of velocity in many subjects in the scene is a marker that something probably unexpected is happening.

Two codes of the same computation environment are recognised as useful, and used to get the purpose of this study (15). The first of them, known as "optical flow" (figure 1) allows to gain the spatial distribution of the apparent velocity by changing the brightness of the images. Horn-Schunck method is to obtain the velocity value. The results of video processing is an n*m*l matrix, where: 1) n and m are horizontal and vertical resolution pixels; 2) l represents the number of frames each video is composed by. For each frame it

Fig. 1. Example of a figure caption

is obtained a matrix of velocity of each pixel of the image.

The second code is known as "multi-object tracking" (18,19) that allows to recognize and track objects in the video recorded by a stable camera (figures 2 and 3). This is a more complex code than the previous, and it is able to calculate objects's velocity as the difference between the centroid's positions in two consecutive frames. Then there are two kinds of results: 1) a matrix of velocities: n*m matrix where n represents the number of tracks and m the number of frames of the analyzed video; 2) the array of tracks: made by all position changes for every single track, they are represented by a couple of values x, y. For every track, it has been obtained a couple of values for every frame where the track is. In this way it is gained the change of position for every track, even if it is not an absolute value. The purpose is to create a matrix of velocities for each video, made up by the modulus of the position change for every track. Unassigned tracks are found and initialized.

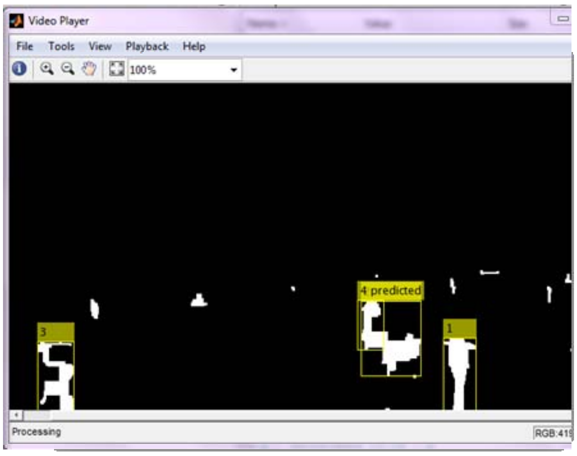

Fig. 2. A frame from the video with normal conditions and behaviors, through multi-object tracking code

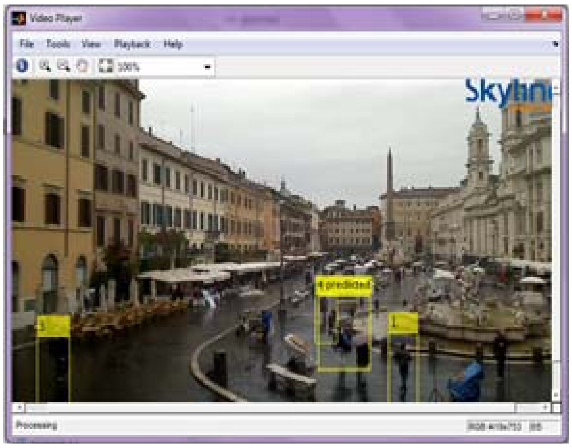

Fig. 3. A frame from the video with normal conditions and behaviors, through multi-object tracking code

A. The creation of the artificial neural network

The next step has been the creation of artificial neural networks (11) to analyse the velocities changing of the objects. For doing this they have been taken as term for comparison the velocities which are typical of normal conditions, these are extracted from the nineteenth video, that recorded in Piazza

Navona. It is now the time of the turning point: to detect unexpected movements.

An artificial neural network is a computational model inspired by central nervous system that is capable of machine learning and pattern recognition. It is an interconnected system of "neurons" that can compute values from inputs by feeding information through the network. Multi-object tracking code is that one used to create the network input. The input consists of two arrays with all the videos velocity values. These velocities do not belong to the number of tracks on the scene and the frame rate. The network is done by iterations, changing the number of neurons in the hidden layer. The final goal has been to find the best network.

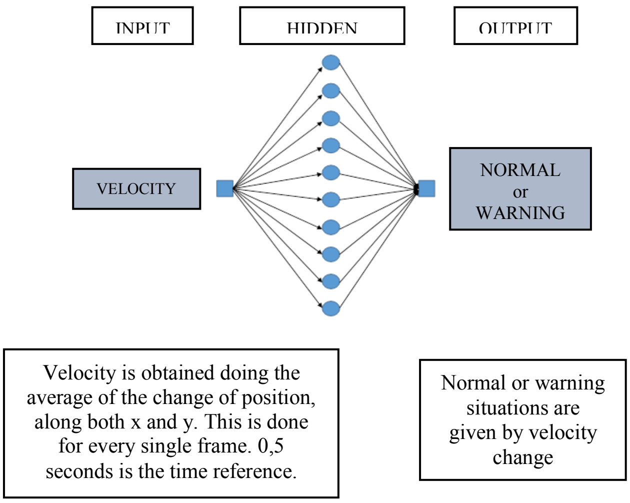

Fig. 4. Scheme of the artificial neural network

Before describing step by step what has been done, it is important to underline the fact that every video has got different duration, different number of tracks, different number of frames and different frame rate. The videos were chosen according to specific needs of urgency representations. The first thing has been done is the splitting of the couple of values, so that there is, for each track, distinguished x and y component arrays. Then, it has been calculated the velocity average, based on the number of tracks in every frame. The input has been established as independent from the number of tracks. The nineteen videos have three different frame rates: 1) two videos present a frame rate composed by 30 frames per second; 2) one video presents a frame rate composed by twenty frames per second; 3) seventeen videos present one frame rate composed by twenty-five frames per second. Further things to be known are that the time sample lasts 0.5 seconds, and the velocity is obtained doing the moving average on the frames. The input is a one unique array done by the velocities averages, frame after frame, for every single video. The target has been done by hands, simply watching every video, and putting, for each frame, 0 in case of normal situation, and 1 if danger. Lastly the output, it consists in an array composed by a sequence of 0, 1 (1 in case of warning situation, 0 in case of normal situation due to changing of velocity). It is the time now to create the Artificial Neural Network importing both input and target arrays, and choosing the number of neurons in the hidden layer.

Then the network might be trained, going right to the graphic results where performance and regression are represented. On a hand the performance allows to evaluate the network performance, on the other hand the regression is used to estimate the expected value influenced by a dependent variable (it estimates how well the network learns).

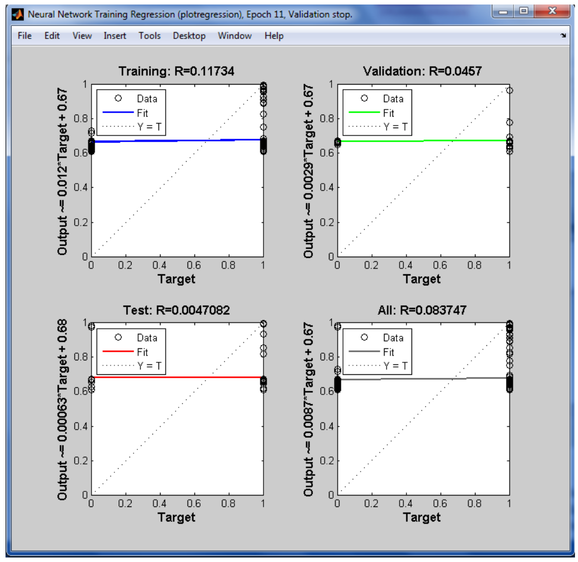

Fig. 5. The goodness of the neural network

Along x axis there is the number of epochs (iterations); along y axis the mean squared error between the real output and the artificial network output. The blue curve represents the training sets used to training the network using the target data, the green curve represents the validation set that is used to test the network and visualize if there is an overfitting problem, and the red curve represents the test set that is used to verify function and performance of the network. Dotted line represents the learning function, continuous line represents, instead, the function learned by the network. The circles are inputs and targets.

III. RESULTS

Iteration is used in order to obtain the best results, it is done changing the number of the hidden neurons time after time, and for both v_x and v_y. Performance and regression graphics have been analyzed to choose the best network.

The figure on the left represents v_x performance, varying the number of the hidden neurons from 10 to 100 the trend is included in a very close range. It is observed that from a certain value of hidden neurons, the network processing goes slower, this is why 55 hidden neurons network is chosen as the best network.

Fig. 6. Performance of x velocity components

At this point it has been important to estimate, through three different tests, the changes obtained modifying the training set, validation set and test set percentage for the chosen network. The tests have been conducted as follows: 1) 70% samples in training set, 15% samples in validation set and 15% samples in test set. Network stops after 10 iterations. and the best performance is at epoch 4 and 0.22684 as mean squared error. Regression is 0.19982 for training set; 2) 65% samples in training set, 20% samples in validation set and 15% samples in test set. Network stops after 8 iterations, and we obtain the best performance at epoch 2 and 0.22439 as mean squared error; that is very similar to the previous result. Regression is 0.24028 for training set; 3)45% samples in training set, 30% samples in validation set and 25% samples in test set. Network stops after 9 iterations and we obtain the best performance at epoch 3 and 0.22027 as mean squared error. Regression is 0.21345 for training set. The network number 2 has been chosen.

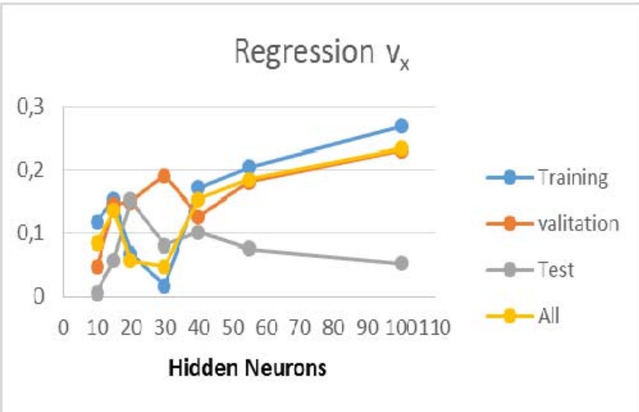

Fig. 7. Regression of x velocity components

Along these same lines it has been conducted an analysis for v_y, and as well in this case it has been observed that the performance values are included in a very close range. Doing the same considerations done above, 60 hidden neurons network has been chosen.

Training set, validation set and test set have been conducted and the results are as follows: 1) 70% samples in training set, 15% samples in validation set and 15% samples in test set.

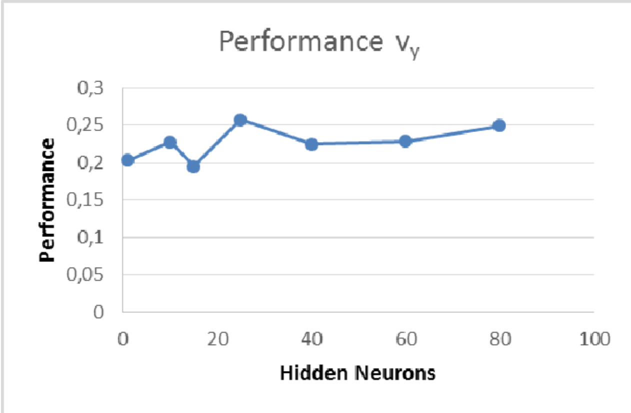

Fig. 8. Performance of y velocity components

Network stops after 10 iterations, the best performance is at epoch 4 and 0.21931 as mean squared error of. Regression is 0.19294 for training set; 2) 65% samples in training set, 20% samples in validation set and 15% samples in test set. Network stops after 9 iterations, the best performance is at epoch 3 and 0.22168 as mean squared error; that is very similar to the previous result. Regression is 0.20476 for training set; 3) 45% samples in training set, 30% samples in validation set and 25% samples in test set. Network stops after 28 iterations and the best performance is at epoch 22 and 0.21718 as mean squared error. Regression is 0.23221 for training set. In this case too, no consistent differences among the network, this is why the third network has been chosen.

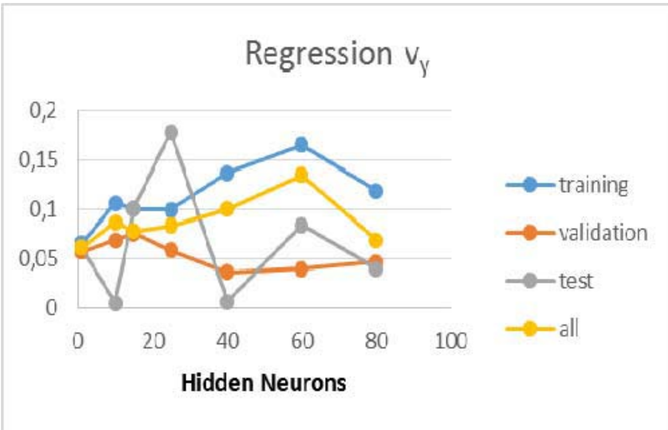

Fig. 9. Regression of y velocity components

IV. CONCLUSIONS

The purpose of this paper has been to find a system able to recognize unexpected motions in surveillance videos using Artificial Neural Network. The attempt has been the creation of a system able to recognize unexpected motions analyzing velocities and not images, and then one of the main issue is the definition of unexpected versus normal behaviour. Human tracking is the first step of the behaviour modelling, and two codes are chosen as able to recognize people velocities in video. The first one is the optical flow code, which allows gaining the spatial distribution of apparent velocities; these values are obtained thanks to the change of brightness in the image (Horn-Schunck method). The second one is the multi-object tracking that allows recognizing and tracking objects in a video recorded by a stable camera, this is a more complex code than the previous one and it is able to calculate objects' velocity too simply doing the difference between the centroid's positions in two subsequent frames. It consists in a changing of position for each track in each frame.

They have been created two artificial neural networks (according to x and y components) to analyze the velocity change of moving objects, to detect unexpected movements. Multi-object tracking is the code chosen, and its results are used as artificial neural network input, which is composed by 19 videos velocities arrays. The network has been created through an iterative method, changing the number of neurons in the hidden layer. The final goal was to find the best network to optimize the problem, starting from input and target.

Unlike what expected, even changing several times the number of neurons in the hidden layer, the results were involved in a close range for both v_x and v_y networks. In particular, there is not overlapping between the two lines of the learning function and the line that represents the function really learned. This may happen because of the use of arrays instead of matrix. However, this choice reason is related to the fact the videos have all different durations.

REFERENCES

[1] Ann Kring, Gerald C. Davison, John M. Neale, Sheri Johnson, "Abnormal Psychology 12th ed.", Wiley & Sons, Incorporated, John, 2012

[2] http://www.eoht.info/

[3] Zhengya Xu, Hong Ren, Wu, Smart Video Surveillance System, School of Electrical and Computer Engineering, Platform Technolology Institute, RMIT University, VIC 3001. Australia

[4] Yigithan Dedeoglu, Moving object detection, tracking and classification for smart video surveillance, August 2004

[5] Teddy Ko, A Survey on Behaviour Analysis in Video Surveillance applications

[6] Arun Hampapur, Smart Video Surveillance for Proactive Security, in IEEE Signal Processing Magazine, July 2008

[7] Mehdi Moussaïd, Dirk Helbing, Guy Theraulaz, How simple rules determine pedestrian behavior ancd crowd disaster, November 2010

[8] R.F. fahy, G. Proulx, 'Panic' and human behavior in fire, procreeding of the 4th International Symposium on Human Behavior, pp. 387-398, July 2009

[9] Anthony C. Davies, Jia hong Yin, Sergio A. Velastin. Crowded Monitoring using Image Processing, IEE Electronic and Communications Engineering journal, Vol. 7, No. 1, pp. 37-47, 1995

[10] Anthony Dearden, Yiannis Demiris, Oliver Grau, Tracking football Player Movement from a Single Moving Camera Using Particle filters, Preceedings of CVMP, pp. 29-37, IET Press, 2006

[11] Moumi Pandit, Mousumi Gupta, Image Recognition with the Help of Auto-Associative Neural Network, International journal of Computer Science and Security, Volume 5, 2011

[12] M. Elarbi-Boudihir, Khalid A. Al-Shlfan, Intelligent Video Surveillance System Architecture for Abnormal Activity Detection

[13] Yunyoung Nam, Seungmin Rho, Jong Hyuk Park, Intelligent Video Surveillance System: 3-tier context-aware surveillance system metadata, Springer Science, 2010

[14] Demetri Terzopoulos, A Reality Emulator Featuring Autonomous Virtual Pedestrians and its Application to Distributed Visual Surveillance

[15] Siamak Khatibi, Anders Hultgren, Introduction to Implementing Matlab in Computer Vision

[16] H. A. Rahim, U. U. Sheikh, R. B. Ahmad, A. S. M. Zain, Vehicle Velocity Estimation for Traffic Surveillance System, World Academy of Science, Engineering and Technology 45, 2010

[17] Satya Kalyan A., Divakar T., K N Rao, Mani Chakravarthi A., Vehicle Velocity Prediction & Estimation in 2d Video for night condition, International Journal of signal Processing, image Precessing and Pattern Recognition, Vol. 4, No. 2, Giugno 2011

[18] Bhavana C. Bendale, Prof. Anil R. Karwankar, Moving Object Tracking in Video Using Matlab

[19] A. Sai Suneel, Person or object Tracking and Velocity estimation in Real Time Videos, Publications of Problem & Application in Engineering Research, Vol. 4, 2013

[20] Ramin Mehran, Alexis Oyama, Mubarak Shah, Abnormal Crowd behavior Detection using Social Force Model

[21] Anthony C. Davies, Jia hong Yin, Sergio A. Velastin. Crowded Monitoring using Image Processing, IEE Electronic and Communications Engineering journal, Vol. 7, No. 1, pp. 37-47, 1995

[22] Duarte Duque, Henrique Santos, Paulo Cortez, Prediction of Abnormal Behaviors for Intelligent Video Surveillance Systems, IEEE, 2007

[23] Nahum Kiryati, Tammy Riklin Raviv, Yan Ivanchenko, Shay Rochel, Real-time Abnormal Motion Detection in Surveillance Video