8-26

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

The Agent WiSARD Approach to Intelligent Active Video Surveillance Systems

Massimo De Gregorio
Istituto di Cibernetica "E. Caianiello" – CNR
Via Campi Flegrei 34 – 80078 Pozzuoli (NA) – ITALY
m.degregorio@cib.na.cnr.it

Abstract

The Agent WiSARD methodology for intelligent active video surveillance systems is proposed in this paper. The hybrid neurosymbolic system (called ISIDIS) is based on the integration of virtual neural sensors and BDI agents. The use of virtual neural sensors coupled with symbolic reasoning allow the system to work on different scenarios and in any light condition.

1. Introduction

The Agent WiSARD methodology, here presented in the domain of intelligent active video surveillance systems, has already been adopted in different domains. The very first application was devoted to the classification and reconstruction of architectural portal shapes [1], then it was used for understanding and explaining geometrical figures ([2][3] for 2D images and [4] for 3D images).

Taking advantage of the characteristics of Agent WiSARD (modularity, flexibility, etc.), we approach the problem of active video surveillance from a different point of view. Most of the video surveillance systems are based on algorithms of motion detection [5][6] or motion tracking [7]. Moreover, in order to get good performances by these systems in outdoor applications one has to deal with light and weather conditions [8]. Eventually, one has even to deal with high computational and hardware costs. The advantages of using virtual neural sensors and symbolic reasoning for interpreting their outputs, makes the system ISIDIS both very light from the computational and hardware point of view and quite robust in performances.

ISIDIS[1] is an intelligent active video surveillance system of the latest generation. It is able of alerting the control room only in case of particular events. Taking advantages of neural network and artificial intelligence methodologies, the system can be applied to different scenarios (in particular risk danger places). It can be adapted, by the user, very easily to a variety of sites, like tunnels, bridges, overpasses, railroad crossings, etc. The system works in real time on images coming from video cameras with different light conditions (it works 24 hours a day) and weather changes. The use of ISIDIS allows security officers to not observe constantly the control room monitors. In order to attract the controllers attention in real time, if the system recognizes an abnormal situation it switches on the monitor connected to that site.

[1] The system has been designed and developed by Neatec SpA, Istituto di Cibernetica – CNR, and Università di Napoli "Federico II". The patent pending on ISIDIS is n. RM 2006A000153 – 20/03/2006.

2. System architecture

The system is formed by three different modules: image pre-processing module, neural module and symbolic reasoning module.

The first module takes the image from the frame grabber (the system works on 10 images per second) and convert it in a binary (black and white) format [9][10]. In order to speed up the system performances, the conversion is applied only to those parts of the image covered by the virtual neural sensors (see figure 1).

Fig. 1: image before and after conversion.

The second module, that is the sensory part of the system, is implemented via a set of RAM-discriminators (see §3): one for each virtual sensor. There are three different sensors to be placed on the image: green, red and blue. The green sensors are going to cover those parts of the scene where certain movements are allowed (for instance, train transits); the red ones, those parts where some movements could be very dangerous or risky. In order to allow the system to get accustomed to light changes, the values reported by blue sensors are used to normalize the other sensor outputs. This values normalization, in addiction to the noise resistant property typical of neural networks, allow the system to have the same performances during the entire day.

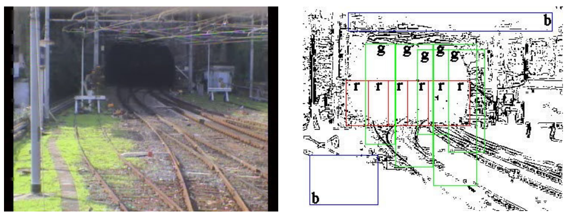

Fig. 2: an example of virtual neural sensor arrangement.

In figure 2 an example of virtual sensor placing is reported. The green sensors (**g**) follow the shapes of the trains entering and leaving the tunnel, the red ones (**r**) are placed at the entrance of the tunnel, and the two blue sensors (**b**) in places where no movements should be present.

The symbolic module, formed by BDI agents, is devoted to the interpretation and evaluation of the sensor outputs. In fact, the alert level depends on the combined information coming from the whole set of virtual neural sensors (see §4). On the basis of their pattern of activation during time, the BDI agents can distinguish many different situations: from normal situation (for instance, train transit) to what is considered a dangerous events for the specific domain of application (one has to remind that the examples here reported are related to scene representing entrance of railway tunnels, but the system can be applied to many different domains).

3. Neural Module

The neural module of this hybrid system is formed by as many RAM-discriminators (WiSARD) [11] as the number of sensors is. To each virtual sensor a RAM-discriminator is associated.

WiSARD is an adaptive pattern recognition machine which is based on neural principles. It is a weightless system whose basic components are RAM-discriminators. RAM-discriminators were selected as digital neural components for the system on the basis of the following considerations:

- they can be trained on-line;
- RAM-discriminators are tailored for efficient implementation on conventional computers;
- the use of artificial neurons more closely reflecting biological neurons would not make a difference.

The part of the image (previously filtered) covered by a sensor forms the input of the corresponding RAM-discriminator (the kind of RAM-discriminator adopted for this system accepts only black and white images as input).

The neural networks are trained only if the symbolic module agrees (see §4) and they are trained with a set of images on which the alert level is "normal". So doing, in the classification phase, the neural module reacts each time an abnormal situation is present in that part of the image covered by the corresponding virtual sensors.

The output of the neural module is formed by the pair *response-colour* of each sensor, and it is passed to the symbolic module that already knows the sensor strategic positioning.

In figure 3 the sensor outputs for two different situations are reported.

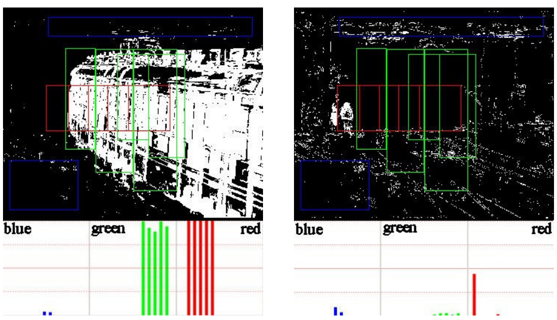

Fig. 3: sensor outputs: train on the left; intruder on the right.

In the system two different time intervals can be fixed: δt (training interval) and Δt (classification interval). During δt the neural networks are trained with a set of frames in which the alert level is "normal" (this avoids the neural networks to be trained during abnormal situations – see §4). Δt is the classification time length. For indoor applications (where light changes are very rare) Δt can be as long as the user prefers (depending on the kind of application the system is adopted for); while, for outdoor applications a good length for Δt is about 10 seconds (this is the value currently used in ISIDIS). With these settings, the system can easily face light and weather changes.

4. Symbolic Module

Although there is not a general agreement on agent definition, social ability is undoubtedly one of the main quality that an agent must hold [12]. This ability to interact – with either the environment, or a user, or other agents – leads naturally to candidate agents as module of a hybrid system. Among the agent architectures proposed during the last years, BDI (Belief-Desire-Intention) model is surely one of the best known and most appealing [13].

To design and implement a BDI agent, we need to set its starting beliefs, plans and goals. Beliefs are the knowledge the agent has about the environment and itself; they can be represented by first order logic predicates. Plans are the means an agent uses to operate: they shape its behaviour [14]; goals are system states the agent wants to reach. Either when a goal is assigned to an agent or when the agent itself has an own goal, it selects a plan to achieve it. Once the agent tries to execute the plan, it gets an intention (intentions are plan instances).

The environment in which the agent acts in ISIDIS is formed by the set of virtual neural sensors and the control room. Its beliefs are formed by the set of *responses-colour* both on the current frame and on a certain number of previous frames (this helps the system in avoiding false alarms due to sudden light or image resolution changes). The main plans of the agents are those devoted to alarm the control room if something of "unusual" is happening. (One has to remind that the system is not taking any decision but it has just to alert who is in charge for that.)

The agent is capable of distinguishing between allowed and not allowed movements, it takes into account all the maintenance time schedule, and it decides whether the neural networks have to be trained (in cases where the agent evaluates the situation different from "normal", the neural networks must keep classifying).

Four different levels of alert are considered by the agent: normal, warning, pre-alarm, alarm.

To set the alarm level, the agent takes into account:

- the responses about the current frame and a set of previous frames;
- the sequence of activation (both for green and for red sensors);
- the single activation (for red sensors);
- the noise level detected by the blue sensors.

With these information, the agent is capable of:

- distinguishing between allowed (for example, a train entering the tunnel) and not allowed movements (human beings or big animals nearby the entrance);
- understanding whether an abnormal situation is

occurring (for instance, a tree on the railway or someone entering the tunnel);
- inhibiting the neural network training phase;
- alerting the control room;
- sending images and messages to the controllers.

The alert is given switching on the related monitor in the control room. So doing, people in the control room are not obliged to watch the monitor constantly. For this reason, the system is considered an active video surveillance system.

5. System performances

The system has been tested for one year at the entrance of a railway tunnel. It worked 24 hours a day analyzing 10 frames per second. The results are very encouraging; in fact, sometimes it gives a couple of alarms in a day (usually during the night). Even in very extreme situations (heavy rain, night filter on during daylight, very low image resolution, …) the system performed very well. In many cases, people working in the control room tried to enter the tunnel just to check whether the system was working and whether the system could have been cheated: but they have been detected by ISIDIS.

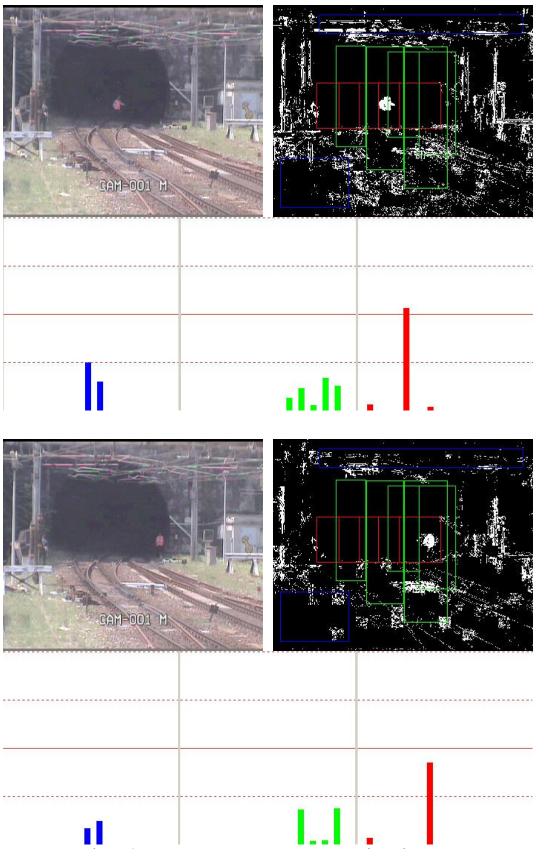

Fig. 4: system outputs on noisy images.

In figure 4, one can notice how the system reacts in those cases considered dangerous (with respect to the sensor response patterns, the symbolic module labels both situations as “alarm”). These outputs were elaborated by an early version of the system and on very noisy images taken by a webcam.

The system elaboration time, from the real image (frame) to the alert level evaluation, is less than 0.1 sec.

In figure 5 two different outputs of ISIDIS (first release) are shown. The first one on the left image is considered by the system an allowed movement (even though the system labels this events as “warning”); on the right image an intrusion has been detected and the alert level is the highest.

Fig. 5: left: allowed movement; right: alarm.

6. Conclusions and future works

Being a low-cost system, ISIDIS can be adopted for controlling the big quantities of tunnels (entrances and exists) on railway and underground networks, archaeological sites, museums and all those places were the presence of a human being could cause serious damages.

The whole system behavior does not depend on the kind of hardware used to catch images (webcams, thermal cameras, …).

The system has been designed and implemented in the view of its application to those domains where intrusions must be detected.

In order to implement WiSARD systems one needs only a certain amount of RAM and some little control. Moreover, most of the symbolic reasoning carried out by the BDI agent can be easily implemented on a NSP (Neuro-Symbolic Processor – [15]). These characteristics led us to work for implementing the whole video surveillance system on hardware, where neural virtual sensors will be implemented on microcontrollers and NSP on FPGA. The cost of the resulting board will be very low with respect to other actual systems.

7. Acknowledgments

The author gratefully acknowledge support from Neatec SpA and collaboration from prof. E. Burattini and his research group at the Università di Napoli “Federico II”.

References

[1] De Gregorio M., “Is that portal ghotic? A hybrid system for recognising architectural portal shapes”, *MVA '96: IAPR Workshop on Machine Vision Applications*, 12-14 November 1996, Tokyo, Japan, pp. 389-392.

[2] Burattini E., Coraggio P., De Gregorio M., “Agent WiSARD: a hybrid system for reconstructing and understanding two-dimensional geometrical figures”, *3rd International Conference on Hybrid Intelligent System HIS2003*, 14-17 December 2003, Melbourne, Australia, pp. 887-896.

[3] Burattini E., Coraggio P., De Gregorio M., Ripa B., "Agent WiSARD: go and catch that image", *IAPR - TC3 International Workshop on Artificial Neural Networks in Pattern Recognition*, 12-13 September 2003, Florence, Italy, pp. 89-95.

[4] Burattini E., De Gregorio M., Coraggio P., Staffa M., "Agent WiSARD in a 3D world", in *Artificial Intelligence and Knowledge Engineering Applications: a Bioinspired Approach – Part II*, ed. J. Mira & J.R. Alvarez, LNCS 3562 Springer, 2005, pp.272-280.

[5] Kim J. B., Kim H. J., "Efficient region-based motion segmentation for a video monitoring system", *Pattern Recognition Letters*, 24 (1-3), 2003, pp. 113-128.

[6] Migliore D. A., Matteucci M., Naccari M., "A Revaluation of frame Difference in Fast and Robust Motion Detection", *VSSN06*, publisher ACM, (EDT) Sheridan Printing, 2006.

[7] Sánchez-Nielsen E., Hernández-Tejera M., "Heuristic algorithms for fast and accurate tracking of moving objects in unrestricted environments", in *Brain, Vision, and Artificial Intelligence*, eds. M. De Gregorio *et al.*, LNCS 3704 Springer, 2005, pp. 507-516.

[8] Matsushita Y., Nishino K., Ikeuchi K., Sakauchi M., "Illumination normalization with time-dependent intrinsic images for video surveillance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, 10, 2004, pp. 1336- 1347.

[9] Canny J., "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, Nov 1986.

[10] Prati A., Mikic I., Trivedi M., Cucchiara R.,. "Detecting moving shadows: algorithms and evaluation", *IEEE Trans. PAMI*, 25, 2003, pp. 918-923.

[11] Aleksander I., Thomas W., Bowden P., "WiSARD, a radical new step forward in image recognition", *Sensor Rev.*, 1984, pp. 120-124.

[12] Wooldridge M., Jennings N. R., "Intelligent agents: Theory and Practice", *The Knowledge Engineering Review*, 10, 2 pp. 115-152, 1995.

[13] Wooldridge M., *Reasoning about rational agents*, The MITT Press: Cambridge, MA, 2000.

[14] Bratman M. E., *Intentions, Plans and Practical Reasoning*, Harvard Univ. Press: Cambridge, MA, 1987.

[15] Burattini E., De Gregorio M., Ferreira V.M.G., França F.M.G., "NSP: a Neuro-Symbolic Processor", in *Artificial Neural Nets: Problem Solving Methods – Part II*, ed. J. Mira & J.R. Alvarez, LNCS 2687 Springer, 2003, pp.9-16.