Expert Systems With Applications 91 (2018) 138–149

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Intelligent video surveillance beyond robust background modeling

Eduardo Cermeño[a,*], Ana Pérez[a], Juan Alberto Sigüenza[b]

[a] Research Department, Vaelsys, Madrid 28043, Spain
[b] Departamento de Ingeniería Informática, UAM, Madrid 28049, Spain

ARTICLE INFO

Article history:
Received 10 April 2017
Revised 10 August 2017
Accepted 31 August 2017
Available online 1 September 2017

Keywords:
Video surveillance
Video
Intrusion detection
Global features
Machine learning
Event
Recognition

ABSTRACT

The increasing number of video surveillance cameras is challenging video control systems. Monitoring centers require tools to guide the process of supervision. Different video analysis methods have effectively met the main requirements from the industry of perimeter protection. High accuracy detection systems are able to process real time video on affordable hardware. However some problematic environments cause a massive number of false alerts. Many approaches in the literature do not consider this kind of environments while others use metrics that dilute their impact on results. An intelligent video solution for perimeter protection must select and show the cameras which are more likely witnessing a relevant event but systems based only on background modeling tend to give importance to problematic situations no matter if an intrusion is taking place or not. We propose to add a module based on machine learning and global features, bringing adaptability to the video surveillance solution, so that problematic situations can be recognized and given the right priority. Tests with thousands of hours of video show how good an intruder detector can perform but also how a simple fault in a camera can flood a monitoring center with alerts. The new proposal is able to learn and recognize events such that alerts from problematic environments can be properly handled.

1. Introduction

Surveillance means close observation or supervision maintained over a person or group. From this definition we could easily think that video is the best technology to empower surveillance. Video surveillance gives people the opportunity to see what is going on in remote places, moreover, it allows to watch several remote places at the same time. Since the very first Closed Circuit Television (CCTV) systems, cameras formed networks of sensors. Not surprisingly, cameras were after computers and printers, one of the first kind of devices to embrace IP technology. The review of Intelligent Surveillance (Valera & Velastin, 2005) groups the evolution of surveillance systems into three generations. The first one based on analog CCTV systems faced all the limitations of analog techniques for information distribution. The second generation adopted digital video uncovering a new world of possibilities for communications and processing. These possibilities multiplied the demand of cameras for different environments: airports, railways, banks, supermarkets, even homes. The third generation of surveillance systems faced the challenges created by new networks with thousands of cameras, that could be monitored from different places.

Multiplying the number of sensors multiplies as well the amount of information generated, thus increasing dramatically the requirements of bandwidth. But even with the upcoming of megapixel cameras, the biggest bottleneck is not related to communications but to video processing capabilities. In the beginning the only video processing unit was the operator's brain analyzing a matrix with images in monitors, however staring at a monitor is not a task humans can execute efficiently during a long period of time (Rankin, Cohen, Maclennan-Brown, & Sage, 2012). Efficient supervision of video requires visual attention, a process by which the human brain selects the elements that will be analyzed. Monitoring centers usually have to control dozens, hundreds or even thousands of cameras. The challenge for intelligent video surveillance systems is to select those most likely to be witnessing relevant events.

There are several possibilities for the selection criteria, but most of them can be grouped into one of these classes of methods: motion detection and pattern recognition. Systems based on motion detection should select cameras where some element or elements are moving. Systems based on pattern recognition should select cameras where a particular pattern has been recognized. In this work we will focus on intruder detection systems for perimeter

* Corresponding author.
E-mail addresses: eduardo.cm@vaelsys.com (E. Cermeño), ana.pf@vaelsys.com (A. Pérez), j.alberto.siguenza@uam.es (J.A. Sigüenza).

http://dx.doi.org/10.1016/j.eswa.2017.08.052
0957-4174/

protection solutions. Motion and people detection are two natural approaches for this kind of systems.

In the literature we can find different works showing good results for both, motion detection (Xu, Dong, Zhang, & Xu, 2016) and people detection (García-Martín, Hauptmann, & Martínez, 2011). However these promising results might be misleading. In real world sites "intelligent video surveillance solutions" (IVSS) must face complex environments where many of these promising approaches fail. In the article "Video Surveillance: Past, Present, and Now the Future" (Porikli et al., 2013) Hoogs states that "I cannot remember the last time I saw a video surveillance paper that showed results on a scene with rain or snow, or blowing dust, or water on the lens, or horrible video quality from transmission dropouts or image plane artifacts. Occasionally a paper will appear that tries to deal with one or more of these conditions independently, but not in the context of an end-to-end system". Moreover, many of the most successful approaches are computationally too expensive to be implemented in real installations with tens or hundreds of cameras.

In this paper we present the work done with thousands of hours of video from 76 cameras used to protect perimeters of real sites. We describe how an efficient intruder detection system can be built using state-of-the-art methods and test its performance and accuracy. Instead of trying to improve the robustness of the method like Tian, Senior, and Lu (2012) or Javed, Oh, Bouwmans, and Jung (2015), we propose to evaluate how useful computationally simple algorithms can be used to build intrusion detection systems. We also introduce a new approach to deal with challenges that make some intrusion detection systems useless.

1.1. Contributions

The contributions of this paper could be summarized as follows. To the best of our knowledge this work is the first research paper to analyze and test methods to build an intrusion detection system (IDS) based on intelligent video surveillance techniques from an operational point of view. We discuss what makes an IDS good, how to compare it with others and how to evaluate its performance in different environments. The experiments are conducted in real sites with thousands of hours of video. The results would have a real impact in a monitoring center.

We suggest that for a monitoring center specialized in intruder detection the main value of an intelligent video surveillance system is the reliability of its camera selection process. The accuracy of the image processing algorithms is relevant because the system relies on them to select which camera is more likely to be witnessing a relevant event. Most of the methods in the literature focus their efforts on increasing the robustness of the background model. However sometimes modeling is not possible. We have created a new dataset with extreme sudden illumination changes to show that background modeling might be impossible in real environments.

We propose that an intelligent video surveillance system has to be able to learn. Instead of increasing the robustness of image processing algorithms, we have added a module based on supervised learning of global features to learn the problematic scenes. The results suggest that global features are a good choice to capture knowledge by acquaintance, the knowledge that is difficult to express with propositions (propositional knowledge). With our approach the operator does not have to figure out which rules would be more suited to describe a problematic scene, he only needs to tag it as such. Without a problematic scene detector, cameras capturing extreme illumination changes would get an unfair relevance over other cameras and would therefore damage the reliability of the whole system.

The following sections are organized as follows. In Section 2 we review relevant literature related to intrusion detection systems based on video surveillance. In Section 3 we present the objectives and requirements for such a system, how to implement and evaluate it. In Section 4 we describe a new method to manage problematic situations found in outdoor sites. Section 5 illustrates our contributions with experiments undertaken in real sites. Results are discussed in Section 6. Finally we present our conclusions in Section 7.

2. Related works

2.1. Motion detection

An intruder is someone moving in an area where he is not supposed to be. Motion detection algorithms are therefore a natural approach to intruder detection. Kim and Street (2004) list conventional approaches for motion detection: background subtraction (Piccardi, 2004), temporal filtering (Lipton, Fujiyoshi, & Patil, 1998), and optical flow (Barron, Fleet, & Beauchemin, 1994).

Optical flow is an approximation to image motion defined as the projection of velocities of 3D surfaces points onto the imaging plane of a visual sensor (Beauchemin & Barron, 1995). Different optical flow techniques are detailed by Barron and Beauchemin in Barron et al. (1994), most of them are computationally complex. Another important weakness is that optical flow algorithms are very sensitive to noise, which is very common in video from CCTV cameras (Hu, Tan, Wang, & Maybank, 2004).

Temporal filtering, is based on temporal differencing (Lipton et al., 1998). This method uses a thresholded difference of pixel between consecutive images (two or three) to extract the moving object, so it shows high computing performance. However its detection accuracy may be weak, failing in extracting all the relevant pixels of a target object or leaving holes inside moving objects (Kim & Street, 2004).

Background subtraction techniques are probably the most popular choice from vendors of motion detection systems, but are also a recurrent topic in scientific conferences. The idea is to extract foreground objects from an image by subtracting a "background model" image from the original one. The main challenge is to generate "background model" fast and with robust results. Brutzer, Höferlin, and Heidemann (2011) and Bouwmans (2014) describe the main challenges for background subtraction (BS) methods. We reproduce (Brutzer et al., 2011) detailed description:

- **Gradual illumination changes.** It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day.
- **Sudden illumination changes.** Sudden once-off changes are not covered by the background model. They occur for example with sudden switch of light, strongly affect the appearance of background, and cause false positive detections.
- **Dynamic background.** Some parts of the scenery may contain movement, but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees).
- **Camouflage.** Intentionally or not, some objects may poorly differ from the appearance of background, making correct classification difficult. This is especially important in surveillance applications.
- **Shadows.** Shadows cast by foreground objects often complicate further processing steps subsequent to BS. Overlapping shadows of foreground regions for example hinder their separation and classification. Hence, it is preferable to ignore these irrelevant regions.