USING AI AND VIDEO ANALYTICS TO MAKE CITIES SAFER

Today's municipal authorities are placing extra emphasis on providing a safe environment for their citizens, on the understanding that this is a basic requirement to help their cities prosper. A range of efforts are required to attain this goal, including an increased budget for police and other first responders, and the investment in various technologies that can assist in detecting, preventing, and managing safety and security threats.

As part of this effort, and recognising that video surveillance is key to helping achieve this goal, a high percentage of cities around the world have deployed surveillance cameras in the streets and in government buildings. In a given city, the number of these cameras can range from a few dozen to hundreds of thousands, and while they all record the city's happenings, precious few of them are actually monitored, due to lack of human resources. The main benefits provided by these cameras is the collecting of evidence for debriefing or investigation, as well as the ability to view events remotely in real-time, presuming that operators know which cameras to look at, and when. Sadly, the truth is that the majority of cameras deployed in cities provide little value, most of the time.

By Zvika Ashani, CTO and Co-Founder of Agent Video Intelligence (Agent Vi)

The huge leaps forward that have been made in the past few years in Artificial Intelligence (AI) and Intelligent Video Analysis (IVA) are beginning to impact the market. The idea is actually very simple: a computer never gets tired, bored or distracted, and can continuously monitor a camera without pause.

Yet, imagine a hypothetical situation where each camera is watched continuously by someone, who never gets tired or distracted, and who also becomes extremely familiar with the camera that they are monitoring. For example, they immediately understand that a line forming in front of an ATM at lunchtime is nothing special, but that three people standing near that same ATM at one o'clock in the morning looks highly suspicious, and should be reported. Over time, this hypothetical person would develop a comprehensive understanding of what constitutes accepted behavior for people and vehicles, at any point throughout the day, on every day of the week. As a result, this person would be able to intuitively identify events that require attention, and potentially also a real-time response.

Naturally, this scenario cannot feasibly be implemented, even in countries with very low wages, since multiple personnel would be needed for constant 24x7 monitoring of each camera. Unfortunately, therefore, the best-case scenario is that certain cameras in key locations are monitored continuously, while the rest simply consume electricity.

All this is about to change. The huge leaps forward that have been made in the past few years in Artificial Intelligence (AI) and Intelligent Video Analysis (IVA) are beginning to impact the market. The idea is actually very simple: a computer never gets tired, bored or distracted, and can continuously monitor a camera without pause. While this is arguably nothing new from a technological standpoint, multiple barriers have prevented this kind of solution from being deployed thus far. These can be summarised as follows:

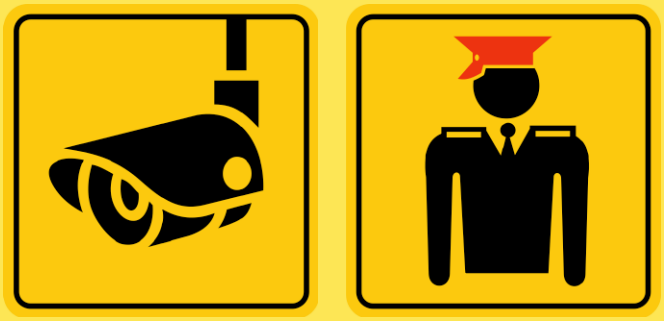
Lack of understanding

How to cause a computer to accurately recognise the different objects in a video stream is a very well researched topic, yet until now the quality of the results (particularly in a busy urban environment) has been somewhat underwhelming. For this solution to work well, the software needs to be able to differentiate between a person, a car, a motorcycle, a static object, a dog and so on, and under various circumstances: during the day, at night, and under year-round weather conditions. The technology required for this to be executed at an acceptable level of accuracy has only recently matured, and is known as Deep Learning. Deep Learning has several uses, but one of its most exciting applications is the ability to accurately detect and classify objects both in still images and in video. This technology is fast becoming the basic building block for IVA.

Inability to learn

The ability to detect and classify objects is an important step, but it is certainly not enough. Within an urban setting, there could be multiple objects moving in the field of view of a camera at any given time, yet it is highly unlikely that we would need to be alerted to each of them. Traditionally, IVA

The rapid increase in GPU computational capacity, coupled with mass market adoption, has reduced prices to the point where a single server, with the correct implementation, can handle hundreds of cameras, which makes large-scale deployment a feasible option.



applications have relied on a rule-based approach, meaning that for each camera monitored, someone needs to configure the software and specify the types of events that generate an alert. For example, a crowd of people standing on a certain street corner after nine o'clock at night, or a vehicle parked in front of a fire hydrant any time during the day, should both trigger alerts. While a rule-based approach can be effective in some scenarios, it cannot be practically applied in a setting with hundreds or hundreds of thousands of cameras, because of the prohibitive amount of manual labor required to configure and maintain these rules, as well as the fact that in most cases, one would not know which rules to set for each camera.

There needs to be a different solution for this problem, one which requires minimal (if any) input from a human operator. This is where AI (or as it is better known in technological circles, Machine Learning) comes into play. An AI solution can gather data about all the people and vehicles observed by each camera over extended periods of time (including days, weeks and even months). Machine Learning techniques can then be applied to this data, to generate models for what can be regarded as "normally observed" behavior. These models can then be applied in real-time to check whether certain behaviors observed by each camera either fit the normal pattern, or deviate from it – in which case they should be flagged as suspicious events that require review by a human operator. This technique allows the solution to scale to any quantity of cameras, without placing a burden on the users to configure each new device.

Cost

There is no simple ROI model that can be applied to increasing safety and security, which can then be measured in terms of financial gain. The result of this is that budgets will always be constrained, and the cost per camera needs to be low enough so that the decision to deploy this across thousands of cameras does not become unworkable. Until recently, implementing real-time AI was extremely cost-prohibitive, and could require an entire server per camera. The rapid increase in GPU computational capacity, coupled with mass market adoption, has reduced prices to the point where a single server, with the correct implementation, can handle hundreds of cameras, which makes large-scale deployment a feasible option.

The convergence of Deep Learning for video analysis, advances in AI for fully automated event detection, plus the significant reduction in cost to implement these techniques, means that the fully automated video surveillance solution for cities is becoming a reality. We expect to start seeing this type of solution being deployed over the coming months, and to become the de facto standard during the next couple of years, for any smart city deployment. SST