WHITE PAPER – Providing a comprehensive 100% surveillance coverage

The lack of 100% surveillance coverage

A 1999 study (Green, 1999) found that after only 20 minutes, guards watching a video scene will miss up to 95 percent of all activity. And we are assuming that this is motivated person being paid to do the job. Part of the problem is that people monitoring security cameras are usually watching several displays at once.

A further 2008 study (Sulman, Sanocki, Goldgof, & Kasturi, 2008) on the effectiveness of human video surveillance performance showed that there are real limitations in people's ability to monitor several signals at once. From the experiments [1] conducted, they found that the observers missed 60% of the target events when they were monitoring 9 displays, and 20% when monitoring 4 displays.

Together, these studies show that human operators cannot focus effectively on a security monitor after 20 minutes, and the more scenes the person has to monitor, the more activity they are likely to miss.

In many central monitoring centres today, the ratio is easily few hundred cameras per human operator. Extrapolating those results, it is not difficult to surmise that up to 99% of all surveillance video is never really watched by anyone, much less detect any meaningful activities from these videos. Hence, surveillance monitoring largely produces little to no value whatsoever.

This is best observed in a report that only 3% of street robberies in London are solved using CCTV, despite the fact that Britain has more cameras than any other country on Europe (Bowcott 2008).

Rule-based analytics cannot provide 100% coverage

The industry has touted the use of rule-based video analytics over the past 10 years to automate this gargantuan task. The rule-based approach with its ability to define pre-set rules seemed to be an answer for alleviating the human operator for having to monitor all these cameras. Or is it? Question is do we always know what rules to define in every camera so as to automate its monitoring? Even if we knew which rule to apply, can it be applied effectively to capture all possible variations of events in a scene? As a rule of thumb, we can only define a rule if we knew what we were looking for beforehand. Furthermore, in order to effectively monitor all the surveillance cameras, we need analytics which can effectively cover all 100% of deployed cameras.

If we take a closer look at the rule sets offered by most leading rule-based video analytic companies in the world today (table below), we will find that there are about 12 rules which we are supposed to use to help us with automating the task for city surveillance.

Category	Rule Type
Person	crossing a line (virtual trip wire), perimeter intrusion
	occupancy, dwell time, loitering
	people count
	crowd density, estimation
Vehicle	occupancy, dwell time
	crossing a line
	illegal parking, stopped vehicles
	wrong direction
	tailgating
	vehicle count
Object	abandoned objects
	removed objects

We will find that out of these, only 2 rules (i.e. line crossing and occupancy) are used for tracking people and possibly together with a third rule, the "abandoned objects" rule are meant to address all the security needs related to possible events that could occur in scenes with suspicious persons. However, a crowded scene usually renders the abandoned object rule ineffective.

Next, we consider all the possible locations for CCTV placement in a city surveillance:

Type	Location	Rule
Outdoors	Public transport interchanges, platforms (e.g. bus stops)	
	Public areas, open spaces (parks, squares, beaches etc)	
	Recreational (swimming pool, playground etc)	
	Protected perimeters (fences and barricades), Entrances and exits, checkpoints	Perimeter Intrusion
	Out-of-bound and restricted areas	Loitering
	Pedestrian walkways in city centres	
	Residential neighbourhood streets	
	Outdoor public parking areas	
	Areas outside public facilities such as sports arenas and subway stations	

	Roads, open streets, highway, thoroughfare	
	Traffic intersections	
Indoors	Reception and lobby areas	
	Service counters	
	Office area	
	Meeting rooms, halls and auditoriums	
	Loading and Unloading areas	
	Corridors, walkways	
	Lift	
	Courtyards, atriums	
	Escalators	

Out of all the different locations, we can possibly only apply rules with certainty to 2 out of the 20 scenarios to effectively address security needs. For the rest of the locations, it is not so obvious if the application of a rule can effectively filter out suspicious activity or only lead to false alarms.

Therefore, we find that rules are usually applicable in a very small fraction of the deployed cameras in a city since the scenes are usually not suitable for setting a rule or the user does not know beforehand what rules will be useful. Besides this, if the place is not restricted access, we could perhaps apply rules to measure occupancy rate, people counting or crowd density level. However, these are more for business intelligence or retail use and not for security purposes as it does not distinguish anything about behaviours of these persons. In other cases, because these are public areas with people movement, setting any rule in place will simply lead to high false alarm rates. The customer may be tempted to just apply 'blanket rules' or 'all-rules' but this is usually too costly for practical deployment. The customer is then often left thinking what rule to apply to his cameras or may have a scenario or specific event he wants detected but not possible to be done with any of the rules or even a combination of them.

We set forth the argument that out of all the thousands of cameras deployed in a city, only a small fraction (possibly less than 10%) are actually looking at perimeters, entrances or exits where an video analytic rule can effectively be applied. If we look at this 10% of cameras, we can see that out of the all the possible behaviours or events that can be detected, perhaps only another 10% of these events can be detected by applying simple analytic rules. Putting it together means that we are only detecting 1% of all possible events, essentially missing out on the other 99% of possible events which could be of interest to us. This can be illustrated in the 3 figures below.

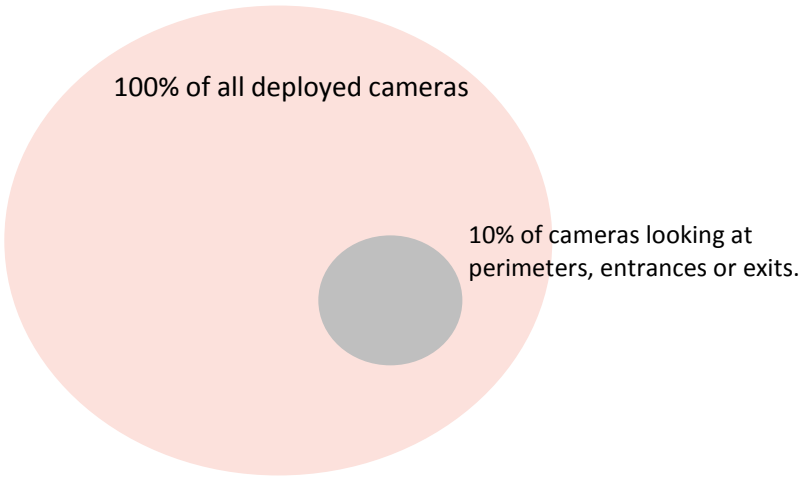

Figure 1

Figure 1 shows that out of all the possible thousands of deployed cameras in a city surveillance, only about 10% are cameras looking at perimeters, entrances/exits or restricted zones where an analytic rule can be effectively applied.

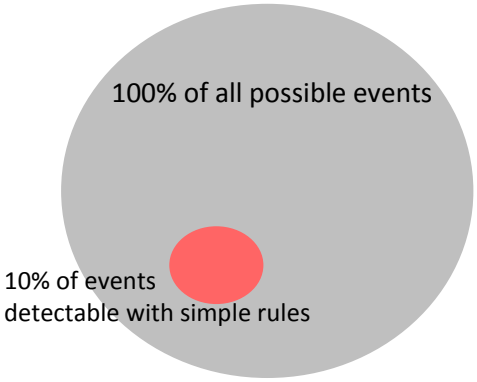

Figure 2

Figure 2 shows out of all the possible events and behaviours which could occur in this 10% of cameras, only another 10% are simple behaviours which can be detected using a rule-based system.

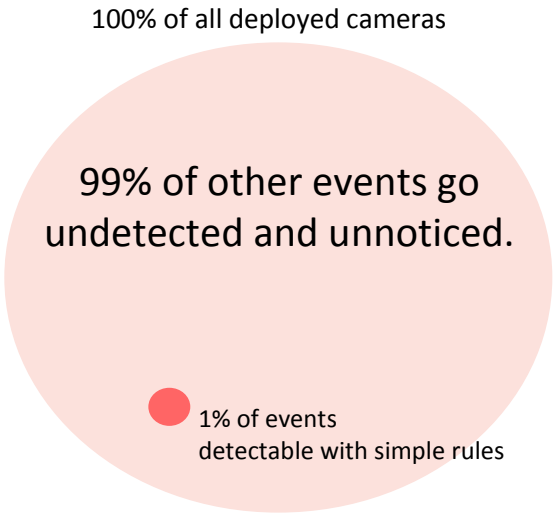

Figure 3

Figure 3 shows that combining Figures 1 and 2, we can show that 99% of all possible events and behaviours in all deployed cameras (whether with analytics or not) largely go undetected or unnoticed. This argument is supported from the statistics gathered from 2 reports:

"The World Market For CCTV & Video Surveillance Equipment – 2013 Edition" by IMS Research estimated that the total number of security cameras shipped worldwide will grow from 42.9 million in 2012 to 81.5 million in 2017.

In another report by IMS Research titled "The World Market for Video Content Analysis in Security & Business Intelligence Applications – 2013" the world market for VCA software is estimated to be $110.3 million in 2012 and will grow to $170.2 in 2017.

In terms of physical units this represents approximately 250,000 units in 2012 and 510,000 units in 2017. This represents a growth of about 55% in market value and a growth of 100% in total units. This roughly translates to a penetration of video analytics to about 0.6% (i.e. 510,000/81.5 million) of the total cameras shipped.

Therefore, using 1% to describe rule-based video analytics coverage of all deployed cameras is conservative.

This inability to monitor all the cameras has also led customers to either deploy less cameras or stream back less cameras for live monitoring, keeping only those critical few for post-event forensic and investigative purposes. This results in a reactive rather than pro-active approach to security.

In a report by "Urbaneye: CCTV Systems in London. Their Structures and Practices", it is found that 66% of targeted surveillances were initiated from outside the system. This reflects the extremely low level of proactive use of the system by CCTV operators.

ARVAS for 100% coverage

If these camera sensors do not have built-in intelligence to convey meaningful alarms, then they are essentially ineffective in alerting users to incidents of interest. Hence, rule-based video analytics only provide sporadic and limited coverage. A truly effective analytics system must be able to be applied to all 100% of cameras with ability to surface abnormalities in unlimited possibilities of behaviours and events for scrutiny.

Unlike a rule-based system, ARVAS utilizes a completely automatic, unsupervised algorithm to learn frequently occurring activity patterns in the scene. The functionality of ARVAS is the automatic detection of abnormal activities by looking for data out of the ordinary. With the set of frequently occurring activity patterns recovered by the unsupervised algorithm, the detection of abnormal activity will correspondingly be automatic. This detection is performed by matching the observed activity against the activity patterns recovered. In contrast to event or object centric methods used in rule-based systems described in the previous section, this method works automatically without requiring any human input.

In addition, ARVAS is ideal because:

1) It is easy to setup and can be applied to thousands of cameras with little to no effort. There is absolutely no need to specify or pre-define any rules beforehand which significantly reduces the time needed for a full-scale deployment.
2) It is fully autonomous once set up and does not require users to set conditions for positive detections.
3) It does not require prior knowledge, special skills or experience from human operators to set up the system.
4) Unlike rules which can be applied only to certain camera views, angles or scenarios thereby limiting their appropriateness to all cameras, ARVAS can find abnormalities in almost any camera view.

5) It is not specifically targeted at any particular behaviour which means it can discover a limitless range of possible behaviours even those not known beforehand.

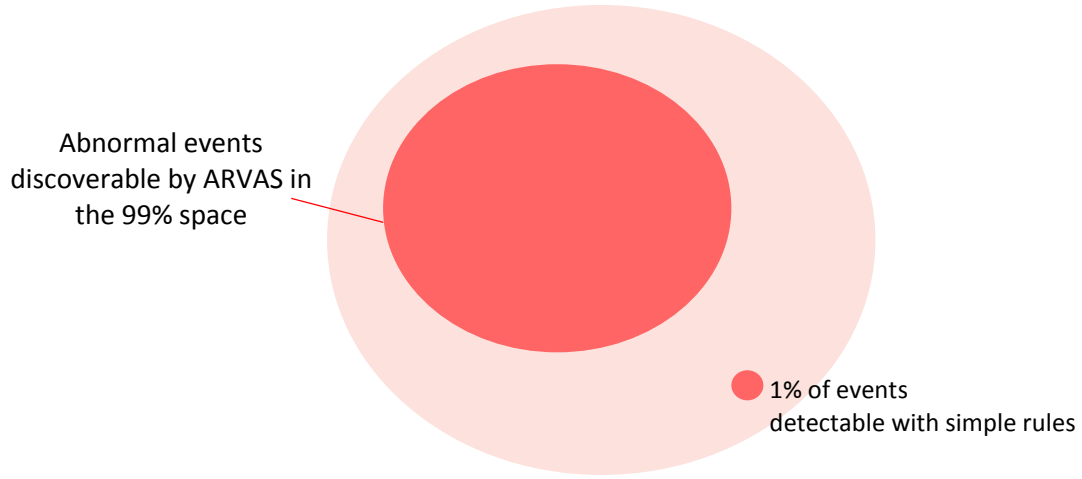

Figure 4

ARVAS enables every camera sensor to intelligently transmit information about its surroundings whether it is related to security, safety or operations. This means that ARVAS will be able to discover more abnormal behaviours and events in the 99% space as shown in Figure 4 far more effectively and efficiently than any CCTV human operator or rule-based video analytic systems can.

This holistic approach provides valuable and actionable insights not achievable in other systems. This enables any city to effectively monitor 100% of its surveillance cameras, converting better value from its constantly streaming video.

---

[1] - In the experiment, a single computer monitor was split into either nine display cells or four display cells. Each cell contained a stream of between 2 and 4 moving objects. Participants in the study were told to signal when a target event occurred, i.e., one of the objects entered a small square area in the center of the display. Target events could occur alone or in groups of 2 or 3 events happening close together.

About Vi Dimensions

Vi Dimensions was founded in 2015 with the simple idea that video analytics can be done in a much better and efficient way with the ultimate goal to revolutionize safe city surveillance harnessing thousands of cameras.

The company uses its patented algorithms and proprietary unsupervised Machine Learning techniques to derive meaningful information and actionable insights from live streaming video data. This translates to immediate value to the customer not only in terms of security and surveillance but also improves the organisation's safety, operational and maintenance aspects.

Our advanced and innovative system analyses vast amounts of real-time streaming (or archived) data autonomously for abnormal behavior and events. It does not require human intervention to automatically discover dominant motion patterns which means that unlike conventional systems, it does not require a human to specify rules necessary for detection.

Contact

Vi Dimensions Pte Ltd
16 Ayer Rajah Crescent
Tempco Technominium
#05-04
Singapore 139965
Tel: +65 65702331
www.vidimensions.com
enquiries@vidimensions.com