VORTEX

Supervised Machine Learning Video Intelligence Platform and Knowledge Management for improved Situational Awareness

Dominique Patrick VERDEJO[1]

[1] Personal Interactor, Montpellier, France
[2] INHESJ, Paris, France
dominique.verdejo@personalinteractor.eu +33608579220

Abstract. Video Surveillance has grown and evolved from a commodity security tool up to the most efficient way to trace back perpetrators when terrorism hits our modern urban centers. As number of cameras soars, one could expect the system to leverage the huge amount of data carried through the video to provide fast access to video evidences, actionable intelligence for monitoring and enable predictive capacities to assist operators in their surveillance tasks. This paper explores the system architecture and challenges at stake to create a platform dedicated to video intelligence capture, automated extraction, processing and exploitation for urban video surveillance. We emphasize the importance of ergonomics and interoperability as pillars of usability for technology in existing operations centers. We bear in mind that information sharing is key to the efficiency of decision support systems providing assistance to people on the field as well as in operations centers. Eventually, we focus our approach on a new video based security ontology as a structuring way to initiate a standardization in video intelligence.

Keywords: video, surveillance, ontology, intelligence, annotation, labeling, predictive analytics, situational awareness, big data.

1 Introduction

VORTEX program was initiated in 2010 in France after a large number of interviews with system architects, security managers and police officers. The aim was to unveil the flaws of contemporary metropolitan video surveillance systems and a synthetic article about public-private partnerships in new video surveillance services, for the French National Institute of Security and Justice (INHESJ), was written subsequently.

As video surveillance moves through its digital transformation from analog cables, monitors and tapes to a complete computer based environment, we see a quantum leap in both numbers of video sensors and geographic scale of systems deployed. Large French urban areas like Paris, Lyon, Marseille and Nice have set up or are in the process of setting up systems with more than one thousand cameras, without mentioning the thousands of cameras already scattered along the public transportation lines and inside the vehicles [1]. While the need for police activity monitoring surges, these numbers are also increased by the new bodycams worn by police officers. Information Technology provides solutions to record and visualize all these cameras, but it does not meet the day to day exploitation needs made more complex by the multitude of video sensors. In a word, the capacity to have an eye everywhere does not spare the people watching. A global rethinking of the balance between people in front line and people in operations centers must be undertaken. We introduce the need for a rethinking and rationalization of the human role in image interpretation, based on the finding that we can deploy much more than we can actually monitor. It is made necessary to define how a human operator can collect and preserve intelligence [2] from video sources, with the aid of the machine, assuming the large number of video feeds creates a rich potential information source. This in turns requires new training procedures and new tools to be created to cope with system scale and carry out this strategic task.

Hence, in an attempt to rationalize the global security knowledge management, we propose a data model based on an ontology definition, to cope with information flowing from diverse sources cooperating in the security process. We also introduce a supervised machine learning approach based on close man machine interaction and big data fusion to create a virtuous mutually enforcing context between operators and machine in order to cope with the most pressing challenges of the next generation of control centers: monitoring exponential sensors input.

In this context, Vortex concept objectives are to keep the human operator at the heart of the system and decision process. This requires the development of computer aided monitoring automation, providing advises and recommendations as to what should be watched first in the continuous flow of contextual real-time and recorded events.

2 The key role of video surveillance in homeland security

Besides traffic and cleanliness control, the role of cameras in urban areas and public transportation systems is mostly to deter crime, theft and vandalism. This is achieved through two distinct activities, one being real-time and the second, post-event. In real time, the operators in urban control centers use techniques to follow individuals of interest or to monitor specific areas or persons to protect from attacks. Those activities are often carried out in close cooperation with police staff.

Post-event, the forensic video investigation consists in analyzing video recordings to locate the meaningful footages that can be used as evidences in a court or intelligence to track perpetrators. These post-event video investigations are often long and fastidious, but prove more and more efficient to identify the perpetrators and lead to their arrest as image quality and resolution steadily improve. In both contexts, the tremendous increase of the number of cameras represents a major challenge for the overall efficiency of the system. It has been shown that a human operator can monitor 16 cameras over a period of 20 minutes [3]. Whichever activity, real-time or post-event, requires attention of the operator on a number of video feeds. This highlights the need for computer based operator assistance. Since 2000, many researches have been made in semantic video indexing [4] and European Research has been funded to create tools to annotate and retrieve video [5]. It is commonly agreed today that we need an abstract layer of representation, a language, to describe and retrieve video. Ontologies have been proposed as an adapted tool to capture observations but also to shift domains as surveillance can be operated on media from many different natures depending on the activity [6][8][9] (satellite, urban, cyber, etc.)

2.1 Evolution and revolution of video analytics

Over the past 15 years, numerous tests and benchmarks were undertaken to assess feasibility of using algorithms to perform recognition of specific situations to ease the task of video operators. It is expected that automating video monitoring can lead to a less heavy mental workload for operators as their attention can be focused only on identified problems. In fact, false alarms tend to overcrowd the video environment and have rendered those technologies quite useless in most operational cases.

Traditional video analytics, based on bitmap analysis can be useful to identify line crossing or counter-flow. They can count individuals and detect crowds and abandoned luggage. But they fail providing insights on more complex situations like fights, tagging, thefts and carjacking where more context and common sense is required [10]. Highly focused European FP7 2010-2013 research project, VANAHEIM [11] has been developed in the context of two metro stations and revealed the difficulty to use inputs from video analytics modules to automate the displays on video walls. Nevertheless this project has been pivotal in demonstrating the huge potential of unsupervised video analysis and the detection of abnormality from long recordings.

Since 2010, a revolution has begun in video analytics. Thanks to Convolution Neural Networks, Deep Learning techniques, object recognition, image segmentation and labeling has shown impressively efficient, up to the point where the machine, using a software built on top of GoogLeNet has demonstrated in 2015 an ability to identify objects in still images that is almost identical to humans[12]. This was made possible thanks to the availability of a very large image dataset called Imagenet (over 1,4 million images from over 1000 classes) manually annotated and a challenge that took place annually between 2010 and 2015, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). In 2015, the Chinese company BAIDU also claimed actual superiority of machine image recognition compared to human on the same image dataset [13]. Learning from human generated annotations, the machine has shown a capacity to generalize, identifying objects and generating text sentences describing image and scenes. In 2017, this effort is extended to video with the DAVIS [14] challenge on video object segmentation. The downside of this approach, is that human contribution is highly necessary to generate meaningful datasets. Initiatives to deploy crowd-annotations platform have been recently undertaken to improve and speed-up the ground truth collection from users on the Web [15], fostering the need for creation of machine learning datasets.

Those breakthroughs in image and video analysis and labeling are the cornerstones of VORTEX concept. But as we witness the need for developing supervised machine learning processes that can lead to development of video intelligence expert modules, we also realize that in the ever changing very complex metropolitan environment, the patterns of normality and abnormality and their relationship to the images captured by the cameras are difficult to express. The role of the human operator in the heart of the semantic system is mandatory to reconcile volumes of data captured by computerized video sensors with contextual situation awareness.

We therefore propose an approach based on two distinct annotation processes. One being conducted through the most modern labeling algorithms running on state of the art, dedicated hardware platforms, or inference platforms, the second being performed by the operators. We introduce a third knowledge based situational awareness module or recommendation module that uses insights produced from the analysis of combined human and machine generated annotations and communicates back its recommendations to the operator. This system is able of maintaining long term memory of what is a "normal" or "abnormal" and in addition, it has the essential capability to take into account human generated alerts and comments to adapt to new situations as they happen.

Replaced in the context of video surveillance, the human operator appears even more important to machine learning as he not only recognizes objects and people but assesses the level of risk of a particular situation and correlates scenes monitored by different cameras.

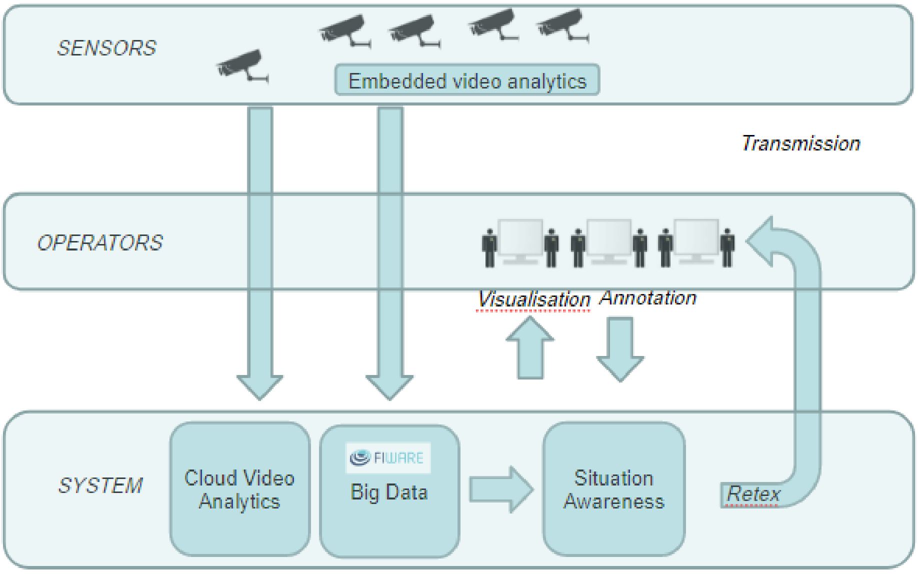

Fig. 1. Human Centric Design for man-machine interoperability

From the picture in fig.1, it is made clear that the System is fed with annotations coming primarily from human operators' interactions witnessing events happening on their surveillance screens (traffic incidents, aggressions and thefts, vandalism, tags, terrorist attacks, smuggling…). This human input is key to providing a common sense context to information that is provided automatically by labeling modules either embedded directly in the cameras or located centrally in the cloud System.

The supervised learning is then operated by human rating of situations on a severity scale, enabling the System to learn and anticipate situations contexts leading to potential risky situations.

The RETEX (Feedback) describes that anticipation data which is deduced by the system while streams of new data flows continuously from both operators and cameras. These streams can also be completed by auxiliary data streams incoming from contextual communication systems and metadata concerning the sensors. The RETEX provides a predictive capacity based on the supervised learning achieved continuously by the interaction of operators and System. An important part of the System analytics is dedicated to transform the RETEX, primarily made of textual content, into actual operational data that can be actioned by operators. In the context of video surveillance, this is achieved by highlighting those cameras that are most important to watch.

3 Situational awareness increased with video surveillance

The RETEX illustrates how the predictive capabilities of the System can be turned into prescriptive surveillance actions. Still, state of the art video surveillance management system provide poor interfaces to enable operators to capture their annotations on the fly and store them in a workable format.

It is thus the key objective of the VORTEX approach to study the conditions of an efficient real-time annotation to enable the operators to achieve the necessary supervised learning and initialize the RETEX loop. Information captured by operators are essential to a sound indexing of video and participate to the overall indexing required by both forensic investigations and day to day exploitation.

3.1 Scientific originality of the VORTEX concept

VORTEX is proposing a genuine approach to the man machine cooperation by leveraging the recent breakthroughs in Machine Learning technologies that allows processing video from traditional stationary cameras as well as mobile or handheld devices [16]. Based on the essential finding that the video media needs to be translated to be workable, we propose to organize a data model for streaming information generated by video analytics labeling algorithms. We propose to define the annotation interfaces necessary to capture operators' annotation in real-time in such a way as it can be used as input for a supervised learning input. Eventually, we propose a predictive analytics System capable of issuing recommendations to the human operators and interacting with them in a feedback loop (RETEX) of reinforcement learning.

It is important to note that annotations as well as any other indexing data may be stored and preserved much beyond the limits of video retention periods, as advised for instance by the European Data Protection Supervisor (EDPS). This means that real-time automated labeling data are key to provide large datasets of surveillance contexts without the burden of keeping the video they originated from. VORTEX is an attempt to rationalize the capture of human feedback in surveillance and crisis context. Putting face to face these data with large volume of sensors data enable the System to correlate and generalize on a rich dataset and leads to the emergence of predictive alerts and prescriptive surveillance actions that increase considerably the situational awareness in operations centers.

We are confronted with several difficulties, notably in scene description and modeling. But we are confident that we can circumvent those difficulties by using machine learning techniques rather than going through a scene data modeling exercise.

4 Conclusion

Following the trend of predictive policing, we introduce a system that will help gathering intelligence from existing and future video surveillance systems and using it to anticipate terrorism and decrease safety risk in metropolitan areas.

Based on supervised Machine Learning and RETEX interaction loop, human operators will contribute to building System cognitive computing capacities and will be augmented in return by its prescriptive analytics.

VORTEX is an independent solution that does not depend on video surveillance technology infrastructure but complements them with new video analytics labeling systems, new annotation and communication tools and new predictive capabilities.

The security ontology definition is the basis of the underlying knowledge management required to provide a consistent framework that will serve as an interoperability guide to extend the approach to different countries and open intelligence cooperation between agencies, both nationally and internationally, representing a potential benefit for global organizations like EUROPOL.

Adopting an ontology and developing automated labeling capacities provides the ground for generating a continuous stream of data flowing from the many and highly diversified sources of information available, both video sensors and human inputs. Among human inputs we can cite metropolitan security control centers operators, but also social networks OSINT (Open Source Intelligence) which play an ever increasing role in situational awareness. A mixt fusion approach based on cognitive computing, could then benefit large scale proven systems like IBM Watson [17] to extract early signals and anticipate risks from the very large textual information generated in such context.

4.1 Dual use of VORTEX technology for Defense

VORTEX framework has been conceived for aiding urban video surveillance operations, but similar initiatives have been undertaken in the field of aerial image analytics [18] and the knowledge based information fusion proposed for the System has been under scrutiny in numerous other papers [19]. The range of sensors providing field data is not limited to stationary cameras. Ground vehicle cameras, aerial drone cameras, body-worn cameras, microphones and general presence detection sensors output information streams that can be injected in the RETEX interaction loop.

The real-time annotation tool may be utilized by operators supervising media different from urban surveillance cameras, i.e. thermal cameras, radars, LIDARs as well as front-line operators located directly in the zone of interest and providing direct field intelligence to the System.

Different application field also requiring human surveillance, like cyber security, may be using VORTEX framework by adapting the vocabulary of annotations. This is made possible by using domain dependent Ontologies, as mentioned in previous projects 68.

VORTEX approach, was presented to the Aerospace competitiveness cluster PEGASE [22], now part of the larger “Pôle Risques” [23] where it was recognized for its “usefulness in the aerial vehicles data processing allowing drones and stratospheric machines to achieve their mission”.

4.2 Academic partnerships

VORTEX concept was elaborated in cooperation with two laboratories, the LIRMM from Montpellier University, expert in machine learning and the LUTIN from Paris VIII, specialized in man machine interfaces, detection and semantics of human perceived actions.

References

1. CODREANU, Dana, Thèse de doctorat IRIT, UMR 55, Université Paul Sabatier, sous la direction de Florence SEDES, “Modélisation des métadonnées spatio-temporelles associées aux contenus vidéos et interrogation de ces métadonnées à partir des trajectoires hybrides : Application dans le contexte de la vidéosurveillance », 2015.
2. [Bremond 08] Francois BREMOND, « Interprétation de scène et video-surveillance », AViRS 2008 (Analyse Video pour le Renseignement et la Sécurité, Paris, 2008
3. Le Goff, T., Malochet, V., & Jagu, T. (2011). Surveiller à distance. Une ethnographie des opérateurs municipaux de vidéosurveillance (p. 62). IAU-IDF.
4. Golbreich, C. (2000). Vers un moteur de recherche évolué de documents multimédia par le contenu. Rapport interne, Université Rennes, 2.
5. Vezzani, R., & Cucchiara, R. (2008, June). ViSOR: Video surveillance on-line repository for annotation retrieval. In Multimedia and Expo, 2008 IEEE International Conference on (pp. 1281-1284). IEEE.

6. Francois, A. R., Nevatia, R., Hobbs, J., Bolles, R. C., & Smith, J. R. (2005). VERL: an ontology framework for representing and annotating video events. IEEE multimedia, 12(4), 76-86.
7. Francois, A. R., Nevatia, R., Hobbs, J., Bolles, R. C., & Smith, J. R. (2005). VERL: an ontology framework for representing and annotating video events. IEEE multimedia, 12(4), 76-86.
8. Luther, M., Mrohs, B., Wagner, M., Steglich, S., & Kellerer, W. (2005, April). Situational reasoning-a practical OWL use case. In Autonomous Decentralized Systems, 2005. ISADS 2005. Proceedings (pp. 461-468). IEEE.
9. Hernandez-Leal, P., Escalante, H. J., & Sucar, L. E. (2017). Towards a Generic Ontology for Video Surveillance. In Applications for Future Internet (pp. 3-7). Springer International Publishing.
10. CODREANU, Dana, Thèse de doctorat IRIT, UMR 55, Université Paul Sabatier, sous la direction de Florence SEDES, "Modélisation des métadonnées spatio-temporelles associées aux contenus vidéos et interrogation de ces métadonnées à partir des trajectoires hybrides : Application dans le contexte de la vidéosurveillance », 2015.
11. Odobez, J. M., Carincotte, C., Emonet, R., Jouneau, E., Zaidenberg, S., Ravera, B., ... & Grifoni, A. (2012). Unsupervised activity analysis and monitoring algorithms for effective surveillance systems. In Computer Vision–ECCV 2012. Workshops and Demonstrations (pp. 675-678). Springer Berlin/Heidelberg.
12. Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015
13. Wu, R., Yan, S., Shan, Y., Dang, Q., & Sun, G. (2015). Deep image: Scaling up image recognition. arXiv preprint arXiv:1501.02876, 7(8).
14. http://davischallenge.org/index.html
15. Kavasidis, I., Palazzo, S., Di Salvo, R., Giordano, D., & Spampinato, C. (2014). An innovative web-based collaborative platform for video annotation. Multimedia Tools and Applications, 70(1), 413-432.
16. "Convolutional-Features Analysis and Control for Mobile Visual Scene Perception" Silvia Ferrari∗, Mark Campbell†, and Kilian Q. Weinberger‡∗Professor, Mechanical and Aerospace Engineering, Cornell University †Mellowes Professor and Sze Director, Mechanical and Aerospace Engineering, Cornell University ‡Associate Professor, Computer Science, Cornell University
17. "IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research" - Clinical Therapeutics - 2016/04/01/Chen, Ying, Elenee Argentinis,, Weber, Griff, http://www.sciencedirect.com/science/article/pii/S0149291815013168
18. Solbrig, P., Bulatov, D., Meidow, J., Wernerus, P., & Thonnessen, U. (2008, June). Online annotation of airborne surveillance and reconnaissance videos. In Information Fusion, 2008 11th International Conference on (pp. 1-8). IEEE.
19. Smart, P. R., Shadbolt, N. R., Carr, L. A., & Schraefel, M. C. (2005, July). Knowledge-based information fusion for improved situational awareness. In Information Fusion, 2005 8th International Conference on (Vol. 2, pp. 8-pp). IEEE.
20. Francois, A. R., Nevatia, R., Hobbs, J., Bolles, R. C., & Smith, J. R. (2005). VERL: an ontology framework for representing and annotating video events. IEEE multimedia, 12(4), 76-86.

21. Luther, M., Mrohs, B., Wagner, M., Steglich, S., & Kellerer, W. (2005, April). Situational reasoning-a practical OWL use case. In Autonomous Decentralized Systems, 2005. ISADS 2005. Proceedings (pp. 461-468). IEEE.
22. http://competitivite.gouv.fr/identify-a-cluster/a-cluster-s-datasheet-910/pegase-59/pegase-62/pegase-63.html?cHash=8fd7e29039de6042eb42f8768d51f8df
23. http://www.safecluster.com/