A Context-Based Surveillance Framework for Large Infrastructures

Oscar Ripolles and Julia Silla and Josep Pegueroles and Juan Saenz and José Simó and Cristina Sandoval and Mario Viktorov and Ana Gomez

Abstract In this paper we present the control and surveillance platform that is currently being developed within the ViCoMo project. This project is aimed at developing a context modeling system which can reconstruct the events that happen in a large infrastructure. The data is presented through a 3D visualization where all the information collected from the different cameras can be displayed at the same time. The 3D environment has been modeled with high accuracy to assure a correct simulation of the scenario. A special emphasis has been put on the development of a fast and secure network to manage the data that is generated. We present some initial results obtained from seven cameras located in a Port Terminal in Spain.

Oscar Ripolles
Inst. of Control Systems and Industrial Computing (ai2), Universitat Politécnica de Valencia, Spain, e-mail: oripolles@ai2.upv.es

Julia Silla
Visual-Tools, Spain e-mail: mjsilla@visual-tools.com

Josep Pegueroles
Dep. d'Enginyeria Telemàtica, Universitat Politècnica de Catalunya, Spain e-mail: josep.pegueroles@upc.edu

Juan Saenz
Acciona Trasmediterránea, Spain e-mail: jsaenz@acciona.es

José Simó
Inst. of Control Systems and Industrial Computing (ai2), Universitat Politécnica de Valencia, Spain, e-mail: jsimo@disca.upv.es

Cristina Sandoval
Visual-Tools, Spain e-mail: csandoval@visual-tools.com

Mario Viktorov
Dep. d'Enginyeria Telemàtica, Universitat Politècnica de Catalunya, Spain e-mail: mario.viktorov.mechoulam@gmail.com

Ana Gomez
Acciona Infraestructuras, Spain e-mail: agomez40@acciona.es

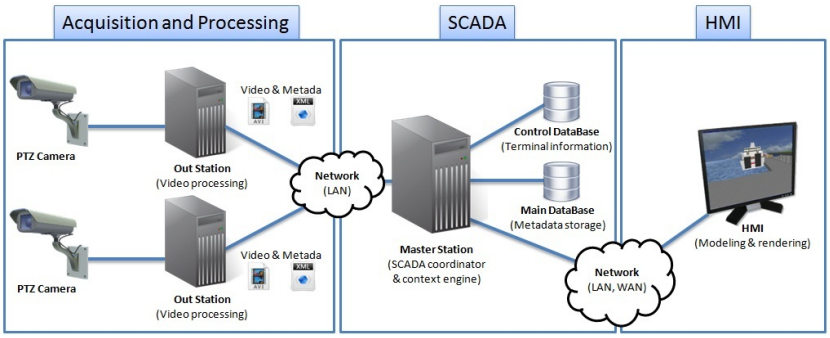


Fig. 1 Main workflow of the ViCoMo proposed platform.

1 Introduction

Computer vision is a research area which has been investigated for a long time. As a result, it is possible to find in the literature and also in the market numerous computer-vision-based systems that relieve humans from a cumbersome work and also improve efficiency, security and safety. Consequently, it would be desirable to develop intelligent surveillance systems that supported this kind of tasks. A visual context modeling system could be used to construct realistic context models to improve the decision making.This context would be a perfect candidate to be displayed within a 3D reconstruction of the scenario to be analyzed. The proposed approach can be useful, for example, to follow a moving object or to select an appropriate camera position to visualize the 3D scenario from the most adequate point of view.

There have been some previous attempts to develop such a system, where 3D information is used to improve surveillance. On the one hand, some authors propose combining 3D environments with real videos. These solutions aim at offering situational awareness so that the user has a clearer understanding of where the cameras are located [16, 14] or to offer an easy navigation between adjacent cameras [7]. On the other hand, some authors propose using computer graphics techniques to estimate the position of the objects. Authors commonly use very rough 3D scenarios and create a simple textured polygon to represent the object using images from the video source [8, 14, 15].

In this paper we present the ongoing results of a research project which has the challenging aim of offering visual interpretation and reasoning using context information. The research project[1] is developing a global context modeling system with the final purpose of finding the context of events that were captured by cameras or image sensors. Then, the extracted context is modeled so that reliable reasoning about an event can be established. Finally, modeling these events and their surroundings in a very detailed 3D scenario supposes an integral part of the system, offering

[1] ViCoMo stands for Visual Context Modeling (MICINN Project TSI-020400-2011-57, ITEA2 Project IP08009)

new means for visualization. Moreover, having a 3D feedback can improve the context modeling as it would be possible to find interpretation errors and impossible situations that, without 3D information, would not be possible to detect. An important aim of this project is the development of a trial platform for logistic control and surveillance of a large infrastructure.

We offer the results obtained in the Port Terminal of Acciona in Valencia (Spain), which has 50,000 m^2 perimeter and 2 docks whose mooring lines are over 300 meters long. Although the proposed framework is mainly aimed at large infrastructures, it could be scaled and used in a more reduced space to give way to smart homes.

This paper is structured as follows. In Section 2 we offer an overview of the whole ViCoMo platform. Section 3 presents a detailed description of the main components of the platform. Finally, in Section 4 we conclude on the solution proposed and outline the main lines for future work.

2 Details of the ViCoMo platform

To achieve these purposes, the ViCoMo system is comprised of multiple cameras, a communication network, a context engine that creates the context model, a database, and a client for retrieval and navigation through the content. The visualization client must show present events in real-time as well as past events on user's demand. This application offers an augmented virtuality environment where the 3D simulation is combined with real-life information and images.

Figure 1 offers a diagram that covers the main elements of the proposed platform. In the scenario we propose, we call *Out Stations* to the computers that receive and process video streams. The *Out Stations*, which can be embedded in the cameras themselves, send periodically the output of the processing algorithms (usually as metadata) in response to the requests performed by a computer (the *Master Station*). Once the information arrives at the *Master Station*, it is stored in a historical database in order to be able to consult it afterwards. Additionally, and depending on the system configuration, the video streams and metadata can be relayed to one or several machines (labeled as *HMI* in Figure 1) for video surveillance purposes. In this Section we describe these three main stages of this platform.

2.1 Acquisition and Processing

The first stage of the system consists in analyzing the video images to extract metadata. The analysis involves three main tasks:

- **Tracking**. We have developed a multi-camera tracking algorithm that is based on a combination of motion detection and background subtraction techniques [18]. Nevertheless, the high density of vehicles during the load/unload operation of a

ship makes the standard tracking algorithms fail due to occlusions. This is the reason why we propose a three level processing using tracking information along time:

1. Intra-camera: for each track of the objects we remove impossible behaviors, such as small splits coming from a unique object and ending in a unique object. We also remove objects of small duration or with impossible shapes [3].
2. Multi-camera: we use the method described in [6] to pair objects between views in a multi-camera system with overlapped fields of view. After that, we apply the same criteria used for improving intra-camera information. In this case, if an object is paired with an object in another view but, at one instant, it is paired with two objects, we can fix the error by removing the wrong pairing.
3. 3D model: there is a final processing in the 3D representation that is capable of correcting the tracking results using the information of the 3D environment (e.g. a car cannot be over the water, a ship is always over the water, etc.).

- **License Plate Recognition (LPR)**. LPR is an interesting feature that enables the system to link a detected object with information provided by the Port Terminal, like driver's id, destination, type of cargo, responsible company, etc. The basis of the LPR for the vehicles in the Port Terminal is the open source Optical Character Recognition (OCR) engine Tesseract [9, 17], currently developed by Google. It is considered as one of the most accurate free OCR engines available and it is able to process many different languages. In the case of plate reading, there are specific syntaxes that can be present. For doing that, Tesseract can be adapted to be able to recognize the characters that follow the desired formats.
- **Classification**. In some cases the information regarding the type of object is necessary. Together with the LPR data, this information is useful for obtaining some statistics about the presence and number of operations of a vehicle. Using basic parameters that can be extracted from the objects, such as the shape, area and position, we are able to classify between truck, vehicle, person and group of people. For the classification we have tried different classifiers from the Machine Learning libraries from OpenCV [1] and the best classification results for our sequences are achieved by the random tree classifier.

2.2 SCADA and Data Management

To achieve the data transmission and storage in a scalable fashion a Supervisory Control and Data Acquisition (SCADA) system is used. SCADAs have been extensively used for data mining, control and management of resources or monitoring and surveillance. However, the SCADA system proposed in ViCoMo innovates in the way information security is supported.

As SCADA protocol we chose the DNP3 [11], which is free and evolving into an increasingly complex protocol with more features. Another important decision to make is deciding how cameras are handled from the SCADA. Usually each man-

ufacturer provides the camera with proprietary software and protocols to manage them. Nevertheless, IP cameras manufacturers are divided into two large consortia to create standard protocols: ONVIF [12] and PSIA [4]. Currently, there is a competition to see who specifies the best standard and win the battle. Since the end of the battle is not clear we decided to integrate both (ONVIF and PSIA) in the ViCoMo project. Finally, we complete the solution with the possibility of sending video streams using protocols designed for this purpose (RTP).

One of the cornerstones of the work developed involves integrating the networking protocols with the output of the video processing protocols. The interfaces between the different components of the proposed platform (see Figure 1) have been implemented as follows:

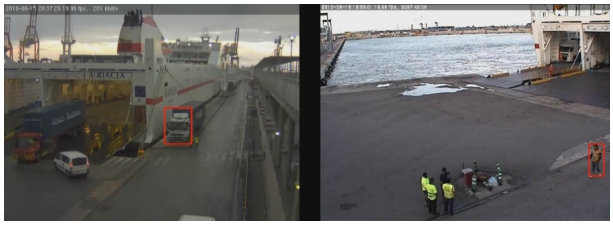
- between the cameras and the *Out Stations*: it is done through Ethernet cabling, using protocols such as MPEG4 or JPG over HTTP or RTP.
- between the *Out Stations* and the *Master Station*: it is done using DNP3 including the result of video processing in specific containers deposited within the *Out Station* in XML format. These files are parsed and embedded in DNP3 frames.
- between *Master Station* and the database: the data must be extracted from the DNP3 frame and entered to the database using a single channel whose session remains open to facilitate operations.
- between the *Master Station* and the *HMIs*: it is done via DNP3, analogously, on the other end data is extracted and passed to the TCP layer - an entry point to the libraries of the 3D rendering software.

Given the relevance of the data transmitted in ViCoMo, it is necessary to ensure encrypted communication. We originally thought about using security mechanisms such as high level TLS, which would encrypt the streams individually, in combination with a chain of trust or certificates for each component. However, several drawbacks raised: the need to establish encrypted connections for each connection, the computational burden for the *Master Station* to manage hundreds of encrypted flows or the delay introduced by public key cryptography.These were the reasons why we decided to secure the communication channel by encrypting all the communications with IPSec. Although having the security at a lower layer results in a small overhead in traffic, the benefits of IPSec are propagated to all other applications and protocols that are used.

2.2.1 Context engine

One of the main contributions of the ViCoMo platform is the development of a context engine, which derives simple conclusions with the information stored in the databases and acquired from the cameras. For limiting the scope of the context extraction, different use cases have been defined so that the proposed platform is applied from a functional point of view. Use cases defined in the Port Terminal scenario come up from the analysis of opportunity of technology involved in the research project, in order to improve the exploitation of port activities through the

(a) Cargo and vehicles tracking.



(b) Risk situation detection.

Fig. 2 Example of use cases covered by the proposed platform.

reduction of operative costs and improved security conditions. As an example of these use cases, Figure 2 presents two possible applications. On the one hand, Figure 2(a) shows an example of tracking vehicles; on the other hand, in Figure 2(b) a risk detection example is shown where a person without safety vest is traversing the terminal. In the Port Terminal we are going to detect different types of events:

- Access control: this is done by comparing the license plate of each vehicle at the entrance of the terminal with a database of unauthorized vehicles.
- Cargo loading and unloading operations: using the tracking and classification information it is possible to monitor the number of load and unload operations of a ship. We do this by using the tracking information and defining a virtual barrier to count the number of objects in each direction.
- Passenger boarding and landing monitoring: we use the method described in [2] for counting the number of people boarding and landing. The results are compared with the database of passengers that boarded in the previous port to check that everybody has left the ship.
- Risk situation detection: the tracking information allows us to detect some risky situations, such as the detection of people or vehicles in forbidden areas.

2.3 HMI and 3D modeling

The visualization system that ViCoMo proposes must display 3D scenes enriched with the context information mentioned in the previous sections. This way, the ren-

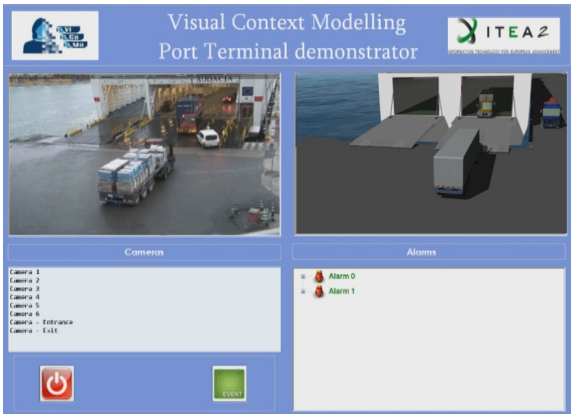


Fig. 3 HMI proposal for ViCoMo.

dering engine must be capable of displaying non 3D information like text of live video streams acquired from the cameras. The HMI we propose is based on the OpenSceneGraph (OSG) rendering engine [13], which includes some interesting features for visualizing 3D information. Taking into consideration that the ViCoMo project requires very precise 3D models, we decided to use Building Information Modeling (BIM) software. All objects/models have been based on CAD plans which must be precise in their measurements. In our case we used ArchiCAD [10], and additional texturing was made with 3DStudio [5] using real pictures taken from the Terminal.

The HMI application must be capable of retrieving information from the SCADA. The data repository presents an interface which offers different queries. Firstly, we must be able to know, for a given timestamp, which elements are present at the scenario. This query is necessary, for example, when initializing the simulation or when accessing historical data. Then, the application must also be able to update the elements on the scenario. This second query uses a time interval (defined by two timestamps) to offer the required information. Figure 3 presents a proposal for HMI where the 3D scene reconstruction from a real image can be observed, displaying the information from a snapshot where three trucks are detected.

3 Conclusions

In this paper we have presented the ongoing work of the ViCoMo research project. This project aims at developing a new control and surveillance platform which models a context of the events to allow for simple conclusions. Moreover, an improved HMI application has been described, as it will still be necessary to have human interaction for managing logistics control and surveillance. The user application triggers

alarms to get the attention from the surveillance staff to act when a set of events are detected. When managing these alarms, the system can offer a 3D simulation of the environment as well as real time camera access.

There have been some previous attempts to develop similar systems, but our proposal offers a complete solution for large infrastructures using an accurate 3D visualization system. As future work we would like to continue developing this platform, making a special effort to exploit the 3D context to improve the results from the image-processing algorithms. Moreover, the reasoning unit is continuously updated with new rules and actors to improve the context analysis and extend the functionalities of the platform.

Acknowledgements This work has been funded by the Spanish Government (TSI-020400-2011-57) and by the European Union (ITEA2 IP08009).

References

1. Open Source Computer Vision. `http://opencv.willowgarage.com/wiki`.
2. A. Albiol, I. Mora, and V. Naranjo. Real-time high density people counter using morphological tools. *IEEE Transactions on Intelligent Transportation Systems*, 2(4):204–217, 2001.
3. A. Albiol, J. Silla, A. Albiol, J. Mossi, and L. Sanchis. Automatic video annotation and event detection for video surveillance. *IET Seminar Digests*, 2009(2):P42–P42, 2009.
4. P. S. I. Alliance. PSIA standard. `http://www.psialliance.org`.
5. Autodesk Inc. Autodesk 3ds max. `http://usa.autodesk.com/adsk`.
6. J. Black, T. Ellis, and P. Rosin. Multi view image surveillance and tracking. In *Proceedings of the Workshop on Motion and Video Computing*, MOTION '02, pages 169–, 2002.
7. G. de Haan, J. Scheuer, R. de Vries, and F. Post. Egocentric navigation for video surveillance in 3D virtual environments. In *IEEE workshop on 3D User Interfaces*, pages 103–110, 2009.
8. S. Fleck, F. Busch, P. Biber, and W. Strasser. 3D surveillance a distributed network of smart cameras for real-time tracking and its visualization in 3D. In *CVPRW06*, page 118, 2006.
9. Google Inc. Tesseract OCR. `http://code.google.com/p/tesseract-ocr`.
10. Graphisoft. ArchiCAD 15. `http://www.graphisoft.com/products/archicad`.
11. D. U. Group. Distributed Network Protocol (DNP3). `http://www.dnp.org`.
12. ONVIF. ONVIF core specification ver 2.1. `http://www.onvif.org`, 2011.
13. R. Osfield and D. Burns. OpenSceneGraph. `http://www.openscenegraph.org`.
14. E. G. Rieffel, A. Girgensohn, D. Kimber, T. Chen, and Q. Liu. Geometric tools for multicamera surveillance systems. In *IEEE Int. Conf. on Distributed Smart Cameras*, 2007.
15. I. Sebe, J. Hu, S. You, and U. Neumann. 3D video surveillance with augmented virtual environments. In *ACM SIGMM Workshop on Video Surveillance*, pages 107–112, 2003.
16. Sentinel AVE LLC. AVE video fusion. `http://www.sentinelAVE.com`, 2010.
17. R. Smith. An Overview of the Tesseract OCR Engine. In *Proceedings of the 9th International Conference on Document Analysis and Recognition - Volume 02*, pages 629–633, 2007.
18. A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38:1–45, 2006.