Multi-sensor Data Fusion for Accurate Modelling of Mobile Objects

Nuria Sánchez, Alejandro Cuerdo, Daniel Sastre, Jorge Alfonso, José Manuel Menéndez
Grupo de Aplicación de Telecomunicaciones Visuales (G@TV)
E.T.S.I. de Telecomunicación. Lab. C-300
Av. Complutense, 30. Madrid - 28040 (Spain)

{nsa, acb, dsg, jak, jmm}@gatv.ssr.upm.es

ABSTRACT

In the last decade, multi-sensor data fusion has become a broadly demanded discipline to achieve advanced solutions that can be applied in many real world situations, either civil or military. In Defence, accurate detection of all target objects is fundamental to maintaining situational awareness, to locating threats in the battlefield and to identifying and protecting strategically own forces. Civil applications, such as traffic monitoring, have similar requirements in terms of object detection and reliable identification of incidents in order to ensure safety of road users. Thanks to the appropriate data fusion technique, we can give these systems the power to exploit automatically all relevant information from multiple sources to face for instance mission needs or assess daily supervision operations. This paper focuses on its application to active vehicle monitoring in a particular area of high density traffic, and how it is redirecting the research activities being carried out in the computer vision, signal processing and machine learning fields for improving the effectiveness of detection and tracking in ground surveillance scenarios in general. Specifically, our system proposes fusion of data at a feature level which is extracted from a video camera and a laser scanner. In addition, a stochastic-based tracking which introduces some particle filters into the model to deal with uncertainty due to occlusions and improve the previous detection output is presented in this paper. It has been shown that this computer vision tracker contributes to detect objects even under poor visual information. Finally, in the same way that humans are able to analyze both temporal and spatial relations among items in the scene to associate them a meaning, once the targets objects have been correctly detected and tracked, it is desired that machines can provide a trustworthy description of what is happening in the scene under surveillance. Accomplishing so ambitious task requires a machine learning-based hierarchic architecture able to extract and analyse behaviours at different abstraction levels. A real experimental testbed has been implemented for the evaluation of the proposed modular system. Such scenario is a closed circuit where real traffic situations can be simulated. First results have shown the strength of the proposed system.

1.0 INTRODUCTION

In the last decade, multi-sensor data fusion has become a broadly demanded discipline to achieve advanced solutions that can be applied in many real world situations, either civil or military. This technique consists of the integration and analysis of data from multiple sensors to develop a more accurate understanding of a situation and determine how to respond to it.

Particularly in Defence, accurate detection of all target objects is fundamental to maintaining situational awareness, to locating threats in the battlefield and to identifying and protecting strategically own forces. Civil applications, such as traffic monitoring, have similar requirements in terms of object detection and reliable identification of incidents in order to ensure safety of road users. We consider recent advances in data fusion techniques in this area an emerging disruptive technology that can be extrapolated to different domains.

Most of the sensor-based systems, whichever the infrastructure under surveillance where they are deployed, use to be passive in nature and do not have the capability by their own to support any of the above commented requirements. Thanks to the appropriate data fusion technique, we can give these

systems the power to exploit automatically all relevant information from multiple sources to face for instance mission needs or assess daily supervision operations.

This paper focuses on its application to active vehicle monitoring in a particular area of high density traffic, where you have multiple targets, and how it is redirecting the research activities being carried out in the computer vision, signal processing and machine learning fields for improving the effectiveness of detection and tracking in ground surveillance scenarios in general.

Specifically, our system proposes fusion of data coming from a video camera and a laser scanner. On one hand, video cameras provide a lot of visual information but are quite sensitive to illumination and weather changes. On the other hand, laser scanners offer robust and accurate distance information even in poor lighting conditions although they do not provide visual information that allow verifying the result of the detection. It can be seen that these two families of sensors offer complementary features; consequently a technical solution based on fusion of laser and camera streams will be more robust than traditional approaches using a single sensor. Thus, we propose a feature level fusion approach which combines both technologies benefits and overcomes their disadvantages.

In addition, although the detection results with the previous fusion of data coming from different sensors are good, in case of some occlusions happen, the detection and tracking accuracy is inevitably lost without an additional processing module, like the particle filtering we propose to use to deal with such uncertainty.

Finally, once the targets objects have been correctly detected and tracked, it is desired that machines can provide a trustworthy description of what is happening in the scene under surveillance. Accomplishing so ambitious task requires a machine learning-based hierarchic architecture able to extract and analyse behaviours at different abstraction levels. Figure 1, shows the architecture of the proposed system at the highest level of abstraction.

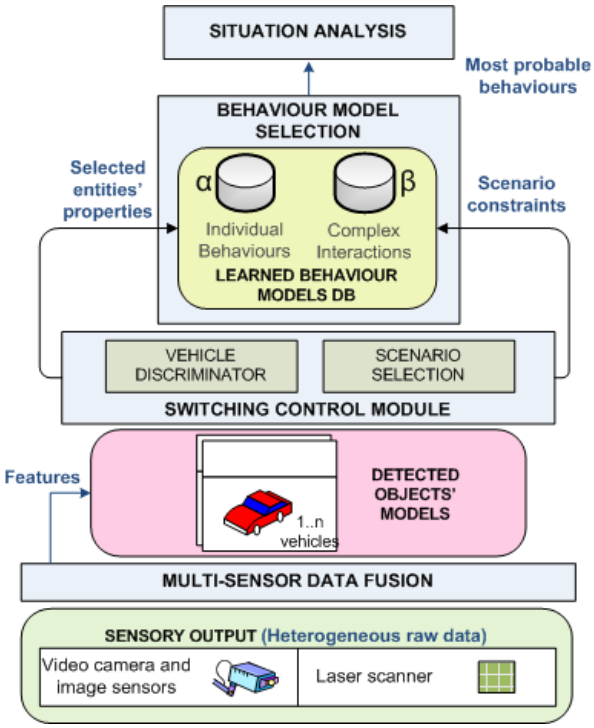

Figure 1: Illustration of the proposed generic system for analysing target's behaviours at different levels of abstraction

2.0 MULTI-SENSOR DATA FUSION FOR OBJECT SEGMENTATION AND TRACKING

In this Section, the necessary steps for processing at low level data provided by measurement sensors, in our case cameras and lasers, available in the infrastructure under surveillance are first shown. One of the functionalities of these independent modules is the localisation and segmentation of moving regions in the scene, serving this information as input for the accurate tracking of the objects involved.

In addition, a revision of the technical challenges allows us to immediate notice complementarities presented by LIDAR and video cameras, which motivated us for proposing a multi-sensor data fusion approach, shown at the end of this Section.

2.1 Motion Detection and Tracking

2.1.1 Using data coming from cameras

Several approaches in the Computer Vision field have been proposed along the last thirty years to solve the problem of robust motion detection and tracking outdoors using a single camera or a network of cameras [1]. However, there are still many challenges for which no solution has been found yet.

In particular, video analytics has been widely used for different applications related to Intelligent Transportation Systems (ITS) such as traffic monitoring. In such an uncontrolled environment, the related algorithms have shown to be very sensitive to changes in lighting and weather conditions, while having the advantage of providing a large amount of visual information understandable by a human operator.

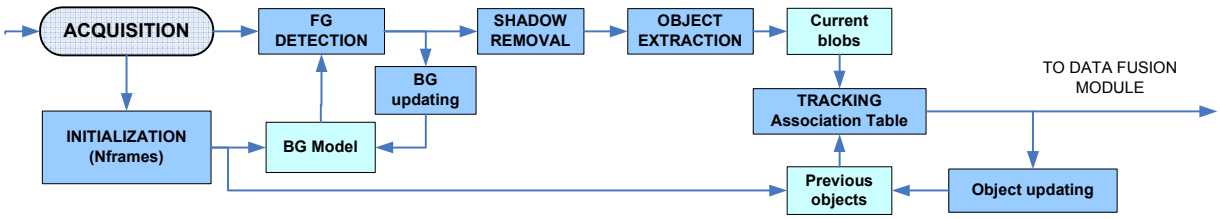

Figure 2: Approach for target detection using cameras on the basis of Mixture of Gaussians technique

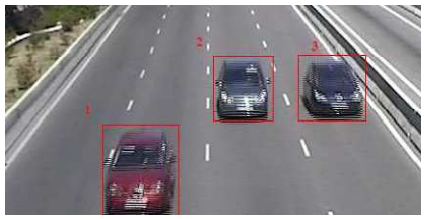

Figure 3: Visual results provided by processing camera information in a traffic monitoring scenario

2.1.2 Using data coming from a laser scanner

LIDAR (*Laser Imaging Detection And Ranging*) is another widespread technology in the field of object detection and tracking, especially in Robotics but also used in the ITS field [2]. These sensors are powered by the technology of flight time; the sensor emits a laser beam which, upon encountering an obstacle, is reflected back to sensor. The distance between the sensor and the object is calculated taking into account the time between the laser emission and reception processes.

LIDAR sensors, also known as laser scanners, are much more robust than video cameras to changes in

lighting and weather conditions, providing more accurate measures. However, they have the disadvantage of providing little visual information compared to the cameras.

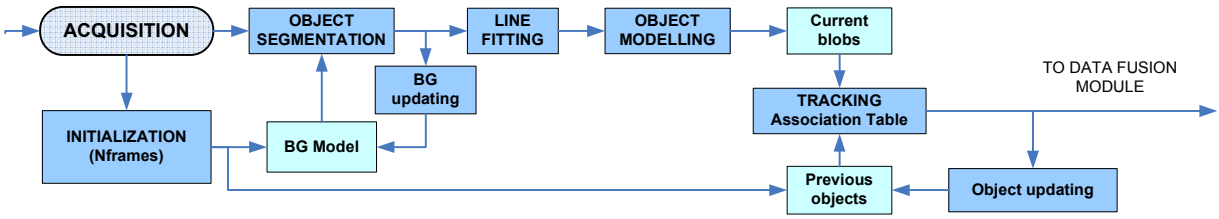

Figure 4: Approach for target detection using laser scanners

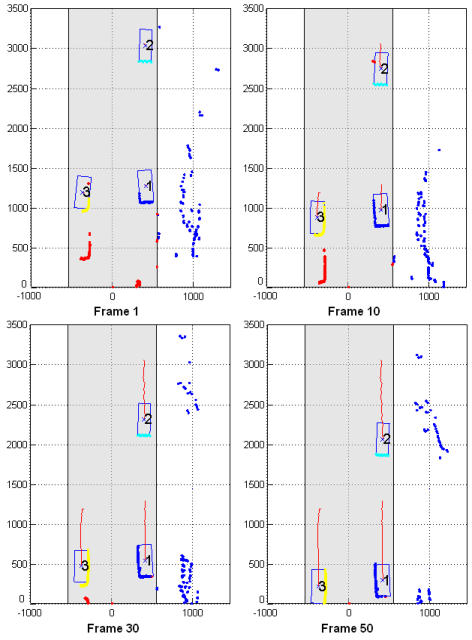

Figure 5: Visual results provided by processing laser scanners information in a traffic monitoring scenario, in which three vehicles are correctly detected and tracked along the sequence

2.2 Technical Challenges

The conclusions extracted from the analysis carried out in the previous Section, where Vision and LIDAR-based technology are presented separately, are:

- Cameras are one of the sensors most commonly used on all the highways around the World for monitoring purposes, due to the fact they offer rich visual information from the environment.
- The application of machine vision algorithms to a video sequence can provide individualized information for each vehicle, like the speed or the trajectory it is following.
- Environmental conditions such as changes in lighting or other atmospheric factors such as rain and fog can decrease the reliability of machine vision – based systems
- The laser sensors, meanwhile, provide less visual information than cameras but the information in terms of location of objects uses to be very accurate and reliable.

- LIDAR are less sensitive (more robust) to the environmental conditions than the cameras used for video surveillance.

- Both video cameras as the LIDAR suffer from the problem of occlusion in which an object concealed the existence of another covering the "field of view" of the sensor.

From the conclusions drawn is immediate notice complementarities presented by LIDAR and video cameras. The first get precise spatial information, enabling reliable and robust set of objects in the correct coordinates, while the latter get a lot of visual information that will uniquely identify each object and determine whether they are vehicles or not, the type of vehicle colour, registration, etc.

2.3 Multi-sensor data fusion

Sensors described in the previous Section have complementary characteristics and consequently a technical solution based on their fusion can provide more accurate and robust results to the problem of object detection and tracking than traditional approaches using a single sensor due to redundancy of information and intelligent fusion itself. On one hand, video cameras provide a lot of visual information but are quite sensitive to illumination and weather changes. On the other hand, laser scanners offer robust and accurate distance information even in poor lighting conditions although they do not provide visual information that allow verifying the result of the detection.

Although many works have been carried out in the field of data fusion, using LIDAR data and video for segmentation and object tracking [3][4], the fusion of information coming from fixed sensors has been less explored. Thus, we propose a feature level fusion approach which combines both technologies benefits and overcomes their disadvantages.

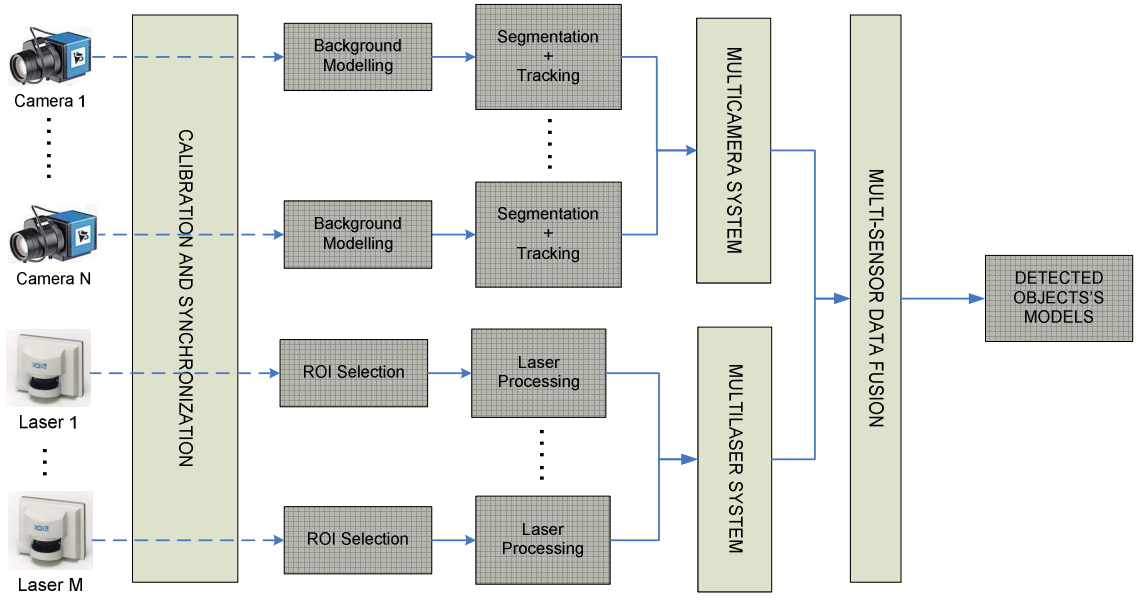

Figure 6: Multi-sensor Data Fusion Approach

Before processing the available data, an off-line calibration step is carried out in order to define a global coordinate system which allows the system to establish spatial correspondence between camera coordinate system points and laser points. Once the scenario is calibrated the on-line process is performed. Data from both sensors is acquired in a synchronized way and is processed in parallel. First of all, the computer vision module detects, models and tracks moving objects. After the previous step, laser processing module

segments, models and tracks laser data. Finally, a fusion step merges features from previous modules. Each object is matched against a predefined model mapped in the above commented global coordinate system. This model is built from laser and visual features applying a coordinate transformation based on the corresponding calibration matrices. The redundant information provides robustness and high confidence level to the system in order to detect and track objects in the global coordinate system.

4.0 IMPROVED TRACKING OF OBJECTS USING PARTICLE FILTERING

Although the detection results with the previous fusion of data coming from different sensors are good, there are situations in which the detection and tracking accuracy is inevitably lost without an additional processing module. In fact, one of the most common problems for this kind of systems when using a multi-target model is that they are hardly able to deal with occlusions. Normally, these situations happen when deterministic trackers based on motion cues are used and consequently, the system will be dragged by an inaccurate global object model.

To solve this issue, we make use of a stochastic-based tracking [5] which introduces some random elements, i.e. particles filters, into the model to deal with such uncertainty due to occlusions. The aim of these trackers is to estimate recursively the position of a target or multiple targets in a first prediction step, and to measure the new objects locations to check the global location estimation of the objects in a second update step. The proposed architecture is presented in a Figure 7.

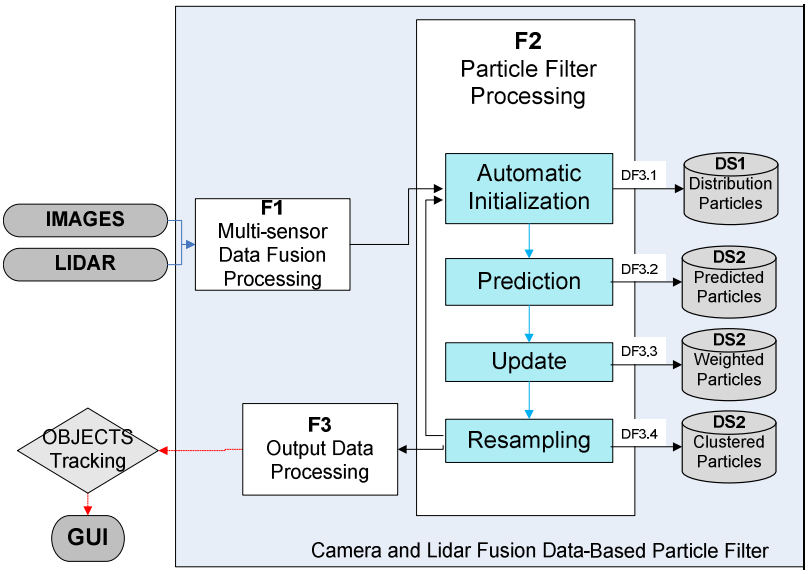

Figure 7: Camera and LIDAR Fusion Data-Based Particle Filter Architecture

The particle filter (PF) is a sampling weighted representation of the Bayesian filter, where each one of the samples $\left\{\vec{x}_t^{(i)}\right\}_{i=1}^{n}$ taken from the continuous probabilistic distribution is called particle. In our multi-target tracking scheme each basic particle is characterized by a set of features such as pixel location, speed and colour (1).

$$\vec{x}_t^{(i)} = \left\{x, y, \dot{x}, \dot{y}, \vec{I}_c(x, y)\right\} \qquad (1)$$

Basically, a particle filter is a hypothesis tracker that approximates the filtered posterior distribution $p\left(\vec{x}_t \middle| \vec{y}_{1t}\right)$ by a set of weighted particles $S_t = \left\{\vec{x}_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{n}$ or a sum of n dirac functions centered in $\left\{\vec{x}_t^{(i)}\right\}_{i=1}^{n}$

as shown in (2). The term $\{w_t^{(i)}\}_{i=1}^n$ represents the weights associated to the particles which are calculated as shown in (3) where $p(\vec{y}_t|\vec{x}_t)$ is the likelihood of the current measurement and $q(\vec{x}_t|\vec{x}_{0:t-1},\vec{y}_{t-1})$ is the "importance function" which usually is computed from state transition probabilities $p(\vec{x}_t|\vec{x}_{t-1})$ and leads $w(\vec{x}_{0t}) = w(\vec{x}_{0t-1}) \cdot p(\vec{y}_t|\vec{x}_t)$.

$$p(\vec{x}_t^{(1:n)}|\vec{y}_{1:t}) \approx \sum_{i=1}^{n} w_t^{(i)} \cdot \delta(\vec{x}_t^{(i)}) \quad (2), \qquad w(\vec{x}_{0t}) = w(\vec{x}_{0t-1}) \cdot \frac{p(\vec{y}_t|\vec{x}_t) \cdot p(\vec{x}_t|\vec{x}_{t-1})}{q(\vec{x}_t|\vec{x}_{0t-1},\vec{y}_{t-1})} \quad (3)$$

Once data from camera and LIDAR is processed, a Particle Filter Processing step is applied:

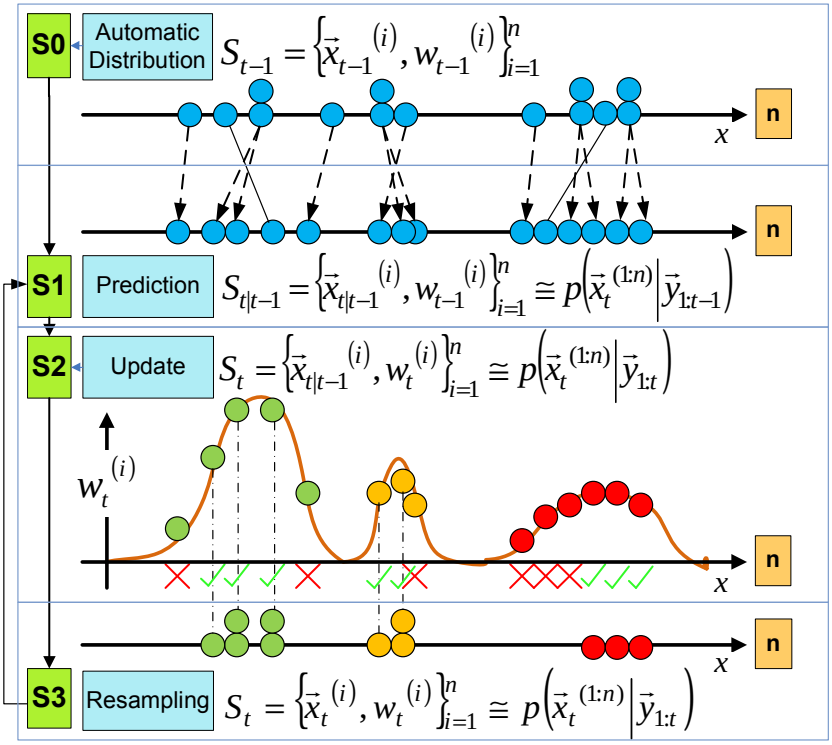

Figure 8: Camera and LIDAR Fusion Data-Based Particle Filter Steps

S0 Automatic initialization: Using a laser and camera fusion to extract features from the pixels of the image where the objects are certainly located, and colour cues from the object model, filter's initial particles are processed each frame to find the means new objects forming the set S_{t-1} as initial distribution

S1 Prediction: Next, the locations of the particles are predicted. The particle filter propagates (the particle locations are moved) the particles according to a motion model $p(\vec{x}_t|\vec{x}_{t-1})$ and the previous states distribution $p(\vec{x}_{t-1}|\vec{y}_{1:t-1})$. First, the current state $\vec{x}_t$ is predicted as output of the evolution function which depends on the previous state $\vec{x}_{t-1}$ and the noise state $\vec{u}_{t-1}$ (4) (blue particles in S1). Second, the prior of the state $\vec{x}_t$, $p(\vec{x}_t|\vec{y}_{1:t-1})$, at time t without knowledge of the measurement y_t is calculated as shown in equation (5).

$$\vec{x}_t = f_t(\vec{x}_{t-1}, \vec{u}_{t-1}) \quad (4) \qquad p(\vec{x}_t|\vec{y}_{1:t-1}) = \int p(\vec{x}_t|\vec{x}_{t-1}) \cdot p(\vec{x}_{t-1}|\vec{y}_{1:t-1}) \cdot \partial\vec{x} \quad (5)$$

S2 Update: This step consists of calculating the weights of each particle according to their features' distance (colour and position) to last global object model. In particle filtering this step combines likelihood of this current measurement $p(\vec{y}_t|\vec{x}_t)$ with predicted state $p(\vec{x}_t|\vec{y}_{1:t-1})$, which usually ensures convergence of measured states towards true states. First, the measurement y_t (from current image features) is obtained depending on the predicted current state $\vec{x}_t$ (the predicted locations into the image) and the noise measurement $\vec{u}_t$ (6). Second, the particle filter weights particles based on a likelihood score $p(\vec{y}_t|\vec{x}_t)$ just processed (weight function in S2) (7).

$$\vec{y}_t = h_t(\vec{x}_t, \vec{u}_t) \quad (6)\,, \qquad p(\vec{x}_t|\vec{y}_{1:t}) = \frac{p(\vec{y}_t|\vec{x}_t, \vec{y}_{1:t-1}) \cdot p(\vec{x}_t|\vec{y}_{1:t-1})}{p(\vec{y}_t|\vec{y}_{1:t-1})} \quad (7)$$

S3 Resampling: Finally, using all collected data from the steps above a resampling step is executed, deleting and cloning the particles according to their weights and their nearest modes. Furthermore, each particle is clustered to the nearest mode forming the tracked objects.

An Output Data Processing module then administers all processed data to obtain the detected object tracking over time. Moreover, this functionality takes care of preparing the data to the Graphical User Interface to verify the results.

Using the proposed tracker based on particle filters, the previous multi-sensor data fusion is improved following the objects through their colour and position features. The proposed particle filter contributes to the global algorithm with parallel tracked data which helps to detect the objects in scenarios under poor visual information or under occlusion circumstances where LIDAR data is not correctly extracted.

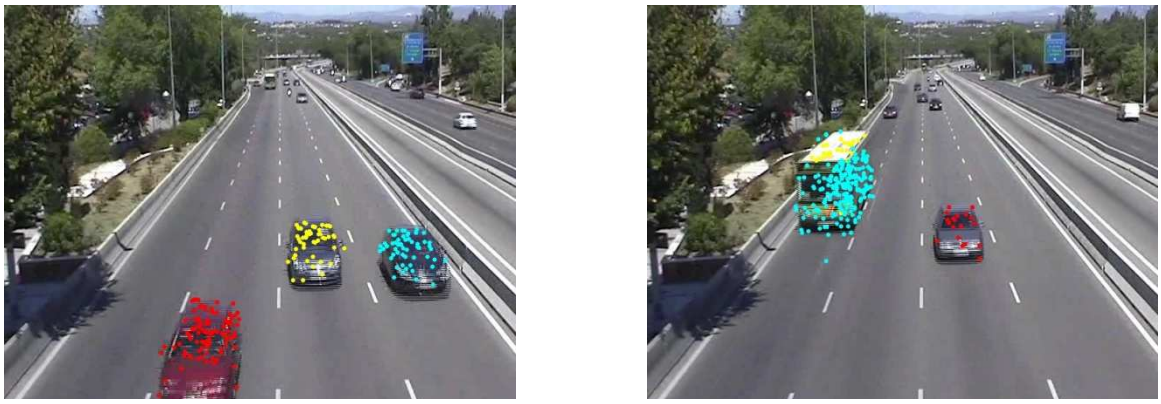

Figure 9: Particle filter steps applied to improve performance of basic multi-sensor data fusion

5.0 SCENE UNDERSTANDING

Finally, in the same way that humans are able to analyze both temporal and spatial relations among items in the scene to associate them a meaning, once the targets objects have been correctly detected and tracked, it is desired that machines can provide a trustworthy description of what is happening in the scene under surveillance [6]. Accomplishing so ambitious task requires a machine learning-based hierarchic architecture able to extract and analyse behaviours at different abstraction levels.

Depending on the application, it will be more suitable to model activities statistically or more simply to discover them by matching the information provided by the multi-sensor data fusion step against a set of predefined rules.

Thus, according to Figure 1, a Switching Control Module discriminates the type of entities present in the scenario under analysis, i.e. vehicle or group of vehicles, and selects for each of them characteristic properties from the object model according to the specific application which the proposed architecture will be used for. In addition, a set of logical and spatio-temporal constraints is also defined in this module. Another module for Behaviour Selection takes those selected properties and predefined constraints to analyse the correspondences between them and a set of previously learned models in a supervised way. Finally, a decision on the situation (incident or activity being carried out) is made.

6.0 EXPERIMENTAL RESULTS

Following similar initiatives in Europe, real experimental testbeds in the framework of different national initiatives in Spain are currently being implemented. Particularly, our testbed is initially located in Madrid, Spain, where a firewire camera and a SICK LMS221 LIDAR have been deployed for the evaluation of the proposed modular system. It is a closed circuit where real traffic situations are planned to be simulated (e.g. stopped vehicle, road leaving, jams, etc.). First results have shown the strength of the proposed system.

7.0 CONCLUSION

In this paper, a new architecture for the accurate detection and tracking of target objects and the analysis of critical situations both in Civil and Defence applications to ensure safety of road users is proposed. Several Computer Vision, signal processing and machine learning approaches are applied for improving the effectiveness of the system. Thanks to the appropriate multi-sensor data fusion technique, all relevant information is first extracted from multiple sources (cameras and laser scanners) and then combined to assess daily supervision operations in an active vehicle monitoring scenario. In addition, a stochastic-based tracking which introduces some particle filters into the model to deal with uncertainty due to occlusions and improve the previous detection output is presented. It has been shown that this computer vision tracker contributes to detect objects even under poor visual information. Finally, once the targets objects have been correctly modelled, the system establish the mechanisms to provide a trustworthy description of what is happening in the scene under surveillance analysing behaviours at different abstraction levels and thus emulating human cognitive processing.

8.0 REFERENCES

[1] Lai, Jin-Cyuan; Huang, Shih-Shinh; Tseng, Chien-Cheng; , "Image-based vehicle tracking and classification on the highway," *Green Circuits and Systems (ICGCS), 2010 International Conference on* , vol., no., pp.666-670, 21-23 June 2010

[2] Zhao, H.; Shao, X.W.; Katabira, K.; Shibasaki, R.; , "Joint Tracking and Classification of Moving Objects at Intersection Using a Single-Row Laser Range Scanner," *Intelligent Transportation Systems Conference, 2006. ITSC '06.* pp.287-294, 17-20 Sept. 2006

[3] Kaempchen, N.; Buehler, M.; Dietmayer, K.; , "Feature-level fusion for free-form object tracking using laserscanner and video," *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* , vol., no., pp. 453- 458, 6-8 June 2005

[4] Jae Pil Hwang; Seung Eun Cho; Kyung Jin Ryu; Seungkeun Park; Euntai Kim; , "Multi-Classifier Based LIDAR and Camera Fusion," *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE* , vol., no., pp.467-472, Sept. 30 2007-Oct. 3 2007

[5] M. Isard and A. Blake, "Condensation—Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision*, vol. 29, nº 1, 1998.

[6] Weiming Hu; Tieniu Tan; Liang Wang; S. Maybank; , "A survey on visual surveillance of object motion and behaviors," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on , vol.34, no.3, pp.334-352, Aug. 2004