Urban Traffic Surveillance in Smart Cities Using Radar Images

J. Sánchez-Oro, David Fernández-López, R. Cabido,
Antonio S. Montemayor, and Juan José Pantrigo

Dept. Ciencias de la Computación
Universidad Rey Juan Carlos
Spain
jesus.sanchezoro@urjc.es

Abstract. The Smart City concept arises from the need to provide more intelligent and optimized applications for the development of future urban centers. Traffic monitoring including surveillance is becoming a problem as cities are getting larger and crowded with vehicles. Intelligent video applications for outdoor scenarios need for good quality, stable and robust signal in every moment or climate condition. In this paper we present a radar signal surveillance application that works in real-time, in 360 degrees, with long range up to 400 meters away from the detector, with daylight or night, or even with adverse climatology like fog presence, detecting and tracking high speed vehicles in urban areas.

Keywords: smart city, particle filter, computer vision, radar processing, visual tracking.

1 Introduction

The detection and tracking of moving objects are two of the most important tasks in the context of video surveillance systems. With the appearance of smart cities, the desirable features for video surveillance systems have been increased to satisfy the user needs. Traffic monitoring including surveillance is becoming a problem as cities are getting larger and crowded with vehicles. For this reason, new video surveillance systems should be able to analyze the scene and extract more information than the classical systems, mostly based on the scene context. Furthermore, lighting and weather conditions usually limit the functionality of the common techniques. Radar devices are robust against adverse climatological conditions, and they present long and wide range detection (up to 400 meters far away from the device and 360°).

This work describes a complete system for detecting and tracking vehicles in urban environments using marine radar images, with the objective of generating alerts when an unusual situation occurs. The visual detection and tracking problem has been tackled using different algorithm techniques, mostly based on the particle filter framework [4], Kalman filters [9] and combinations of probabilistic and evolutionary strategies [7]. Object detection and tracking are two areas of interest in the radar-assisted remote surveillance context [6].

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part II, LNCS 7931, pp. 296–305, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Most of the existing radar-based detection and tracking methods in the literature work in aerial or marine environments, in which there are very low level of noise caused by structural elements. The main aim of this work is to present a new radar-based target detection and tracking system to control unusual vehicle behaviors in urban areas, which contain obstacles that can produce interference and noise in the radar signal. Furthermore, the system must be able to tackle with adverse climatological and lighting conditions, as it should be working all the day. The main proposals of this work can be summarized as a representation of vehicles in radar images; a detection system based on the definition of relevant areas in the image, a dual background model and an adaptive sliding window algorithm to save computing time in the detection; and a radar tracking algorithm based on a particle filter scheme, with the improvements of an intelligent diffusion stage and an optimized moving average model devoted to be robust against occlusions produced by the signal.

2 System Overview

The system proposed in this work is divided in two views (hardware and software). The hardware view consists of the deployment of fourteen radars placed in strategical locations of the city of Torrevieja (Spain), and two central servers placed in the control center. The radar locations have been selected by the local police department in order to cover the most problematic places of the city. Each radar is connected through a WiMAX dedicated network to one of the central servers.

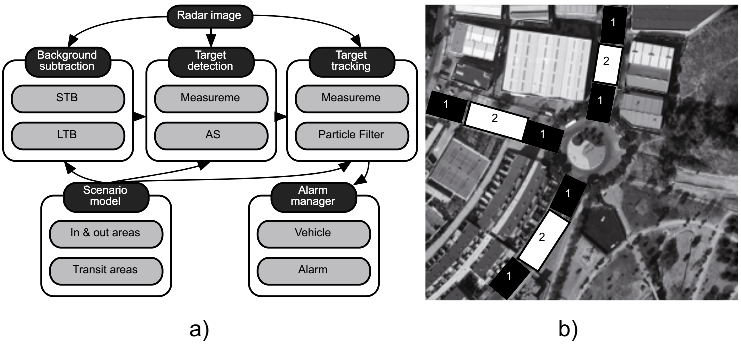


Fig. 1. a) Software architecture overview and b) expert knowledge modeling of a roundabout with in & out (1) and transit (2) areas

Software architecture of the system is depicted in Figure 1.a, describing the relations among the five main modules: scenario modeling, target detection, target tracking, background subtraction and alert manager.

3 Scenario Modeling

Radar devices are placed in strategical locations, fixed by the police department, usually in roundabouts where we can observe different ways with the same detector (see Figure 1.b). Therefore the system should be able to make the most of the location knowledge and use the structural information of the area to define regions of interest. In the context of detection and tracking, there are two important areas: the in and out areas and the transit areas. The former represents the zones where the vehicles can enter and exit of the scenario, while the latter defines the areas where vehicles can move along (generally roads). The radar device used also presents the function to stop receiving data for some angle aperture. This feature is used by the system to eliminate non-interesting parts of the image, reducing its noise and the size of the data that must be analyzed. Figure 1.b shows an example of an scenario modeled by an expert. Scenario represents a roundabout with six in-out areas (represented with a "1" label, highlighted in black) and three transit areas (represented with a "2" label, highlighted in white). The scenario models are configured using a user-friendly interface provided with the system.

4 Target Detection

Acquisition of the radar signal results in images with high level of noise, which is difficult to delete. Therefore, it is necessary to pre-process the image to deal with the noise without deleting real targets. The signal obtained from the KODEN MDC-2000 radar used in this work presents only eight different intensity levels. With this signal quality there are only slight intensity differences between real targets and noise, making the filtering more difficult. The background subtraction model is based on the vehicle kinematics for the target segmentation in order to tackle this problem. Furthermore, the artifacts that can be found in radar images (i.e., high levels of noise, clutter or jitter) make the classical background subtraction algorithms not suitable for these kind of images. We propose an adaptation of the dual background subtraction presented in [8] to tackle these problems. The dual background subtraction is based on the use of two different background images: Long Term Background (LTB or B_L) and Short Term Background (STB or B_S). The main difference between both background images is the updating time, being in LTB higher than in STB. Specifically,

$$B_X(x,y,t) = \begin{cases} B_X(x,y,t-\delta t(B_X)) + 1 \text{ if } I(x,y,t) > B_X(x,y,t-\delta t(B_X)) \\ B_X(x,y,t-\delta t(B_X)) - 1 \text{ if } I(x,y,t) < B_X(x,y,t-\delta t(B_X)) \end{cases}$$

where I_t is the radar image at time t, and $\delta t(B_X)$ is the updating time period for B_X (which represents both images, B_L and B_S).

The background subtraction results in two foreground images for each radar image: Long Term Foreground (LTF or I_L) and Short Term Foreground (STF or I_S), which are computed as the difference between their corresponding background image and the current frame (I). The combination of the active pixels in each foreground generates the whole set of possible events that can be detected, depicted in Figure 2.

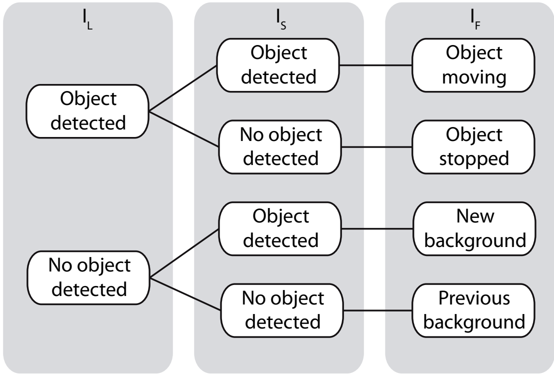


Fig. 2. Event detection using dual background subtraction

The comparison of both foreground images generates the final foreground segmentation, which is a binary image where a pixel is activated only if the value of the corresponding pixel in both outputs, STF and LTF, has a higher value than an experimental threshold. As the radar presents only eight intensity levels, the experimental threshold can only vary in a range of eight values, which makes it easy to configure.

The combination of these two foreground images can also be used for modeling traffic jams or vehicle crashes. This kind of events are detected when an object has been marked as a static object, and it remains static enough time to be added to the background image.

Vehicles are detected in the binary foreground image using the sliding window algorithm presented in [10], varying the size of the window according to the expected dimensions of the targets. The algorithm detects an object if the density of pixels of the region enclosed by the window exceeds a predefined threshold. The weight (π) of the region enclosed by a window centered in (x_p, y_p) is calculated as the area (i.e., zero-order moment) of the window as follows:

$$\pi = \sum_{x=x_p-\frac{1}{2}w}^{x_p+\frac{1}{2}w} \sum_{y=y_p-\frac{1}{2}h}^{y_p+\frac{1}{2}h} I_F(x, y) \tag{1}$$

We propose two different strategies to slide the window through the in-out areas: a standard method and an Adaptive Sliding Window (ASW) method. The former traverses the areas sequentially, resulting in an exhaustive search, which is inefficient in terms of computing time. The ASW method uses the features of the binary image to efficiently explore the in-out areas. Specifically, if the weight associated to a region is very low, then the method discards the exploration of all the consecutive windows which overlap with the former window. This strategy results in a considerable reduction of the computing time needed to explore the image, especially when the foreground image is correctly filtered and low noise is present. Furthermore, when the weight associated with a region exceeds the threshold, the method performs a local search to find the best

a) b)

Fig. 3. Adaptive Sliding Window movement (left) and local search when the window detects part of a foreground object (right)

fitting window. When the local search has finished, the method continues sliding the window in the next non-overlapping region.

Figure 3 shows the movement of the sliding window when no objects are detected (left side) and the local search procedure (right side). The dotted rectangle in the right image represents the final output of the local search (larger than the original window and centered for covering the entire object).

5 Target Tracking

This work proposes a tracking method based on the Sampling Importance Resampling (SIR) Particle Filter (PF) [3], very popular in visual tracking applications, coupled with a local search to adapt the size of the estimated region of interest. The local search method is an adaptation of the LSPF presented in [2] for radar images.

Tracked vehicles are represented by a state vector, which consists of the position (x_i^t, y_t^i) and velocity (vx_i^t, vy_i^t) of the vehicle i at time t. The state vector has an associated weight π^t, related to its likelihood [3,11], which is computed considering the foreground image I_F obtained in the detection stage. The computation is carried out by calculating the number of active pixels inside the bounding box centered in the position of the target.

The selection stage starts when all the weights have been evaluated. The particle filter replaces the particles with the lowest weights with better estimators, preserving the best particles, thus improving the quality of the particle set. The selection uses a roulette wheel selection procedure, which will probably eliminate the particles with the lowest weights, but maintaining a low probability to select them in order to escape from local minima. Once the selection stage has finished, the filter starts the diffusion stage with the aim of boosting diversity. This stage consists of altering the position of the particles to simulate the movement. This work proposes an exploitation of the scenario knowledge to improve the performance of the stage. Specifically, it uses the information stored in the scenario modeling (Section 3) to adapt the diffusion and fit it inside the transit areas. Figure 4 shows an example of this intelligent diffusion.

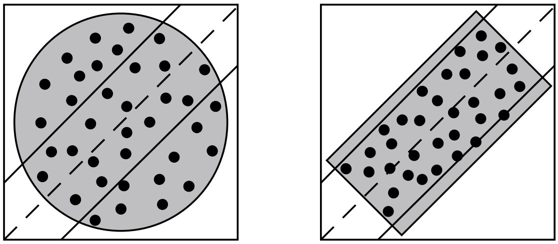


Fig. 4. Particle Filter diffusion without roads information (a) and with roads information (b)

The information stored in the scenario modeling is also used to create a heat map during the execution of the algorithm. This map contains the areas where vehicles are detected more often, based on the position of the vehicles detected and tracked. This information is used to extend the functionality of the system, like reporting statistics about the most problematic areas, or where most traffic jams are located, for instance, which can be transmitted to the authorities, and eventually resulting in improvements of the traffic, creating a kind of "smart roads".

The main problems that can be found in radar images are: jitter, occlusions and disappearances, correspondence between targets and the intensity difference. The first problem (jitter), caused by the reception of the signal, produces a drastic change on the measure taken with respect to the previous measure. This change is represented in the image as displacements of the objects in the scene, although they have not really move, which may lead to errors in the measure. The occlusions and disappearances are caused by an object that cannot be reached by the radar pulse because there is another object in the trajectory of the pulse. In radar images, two vehicles moving with similar velocity can be seen as only one object, until their velocity varies with respect to each other, which makes the identification of the targets quite difficult. Finally, the shape, materials and distance to the antenna of an object determines its intensity in the signal, which can cause that the radar signal of an object affects the signal of another object.

These problems need to be tackled in order to provide an useful system for vehicle tracking in real conditions. The jitter problem is solved with the dual background subtraction proposed, since the method is able to difference between real moving objects or jitter noise. The occlusions and changes in the intensity of the signal are tackled with an adapted computation of the particle weight. It is called Exponentially Weighted Moving Average (EWMA), where the particle weight is computed taking into account the previous weight. Specifically,

$$\pi_i^t = \alpha M_{00} + (1 - \alpha)\pi_i^{t-1}$$

where M_{00} is the area in a binary region given by the position of the bounding box associated to the particle state coordinates (zero-order moment), calculated as follows:

$$M_{00} = \sum_{x=x_i^t-\frac{1}{2}w}^{x_i^t+\frac{1}{2}w} \sum_{y=y_i^t-\frac{1}{2}h}^{y_i^t+\frac{1}{2}h} I_F(x,y)$$

The relevance of the weight of the previous time step to evaluate the new one is represented by the parameter α, giving to the totally occluded particles the possibility to appear afterwards in the image. When a target splits into two different vehicles, it is removed from the particle filter, adding the new two vehicles, thus overcoming the correspondence problem.

The generation of alerts uses the information of the vehicles tracked. The reduction of false positives given by the tracking system must be reduced, in order to avoid sending non relevant alerts. This improvement is carried out by assuming continuity in the movement of the vehicles. Therefore, abrupt changes in the movement indicate that the tracked object is not a vehicle, and it will not generate an alarm.

6 Experimental Results

This section describes the results obtained by the system in real locations. The experiments were performed on an Intel Core 2 Duo E8400 3GHz with 3 GB of RAM and Windows 7 32 bits OS. The radar images were obtained from low-cost radar devices (KODEN Electronics model MDC-2000) oriented to marine applications. The system was entirely developed in C++, using the OpenCV 2.1 (Open Source Computer Vision) library. The parameters have been experimentally obtained, and they are summarized in Table 1. See [10] for more details on the parameter selection.

Table 1. Values of the system parameters

Parameter	Value
Short term background update period - $\delta_t(B_S)$	2 frames
Long term background update period - $\delta_t(B_L)$	5 frames
Bounding box size - $w \times h$	100×100 pixels
Learning factor - α	0.75
Frames needed to generate an alarm - F	3 frames
Number of particles for each particle filter - N	100 particles

The experiments have been carried out using six different video sequences obtained directly from a radar device. The first experiment is devoted to test the quality of the target detection method, compared to a human expert. The results of the background subtraction for the two main sequences can be seen in Figure 5. Sequence 1 presents an ideal scenario with no structural elements that interfere in the radar signal. As a result, most of the foreground pixels correspond to the two targets labeled by the expert. Although some noisy pixels remain active in the foreground image, they will be discarded in the tracking stage.

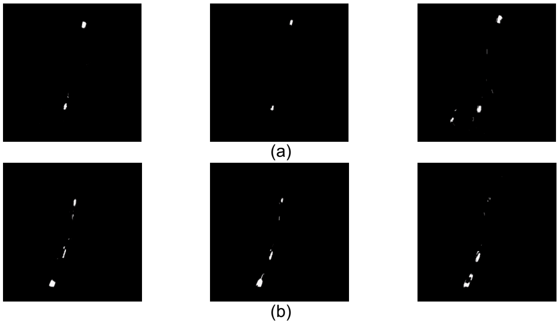


Fig. 5. Background subtraction stage over two different sequences: Sequence 1 (a) and 2 (b)

Sequence 2 shows a more complex scenario where some structural elements like buildings are present introducing higher levels of noise in the foreground. The noisy areas in this sequence are large enough to avoid being discarded in the tracking stage, resulting in some false positives. This kind of false positive is filtered by the continuity premise, as presented in Section 5. Table 2 shows the detection performance of the algorithm, including the ground truth, and the false positives rate. The system is able to detect the 92.9% of real objects (labeled by experts), while the number of false positives is about 0-2 per sequence.

Table 2. Objects detected in the analyzed sequences

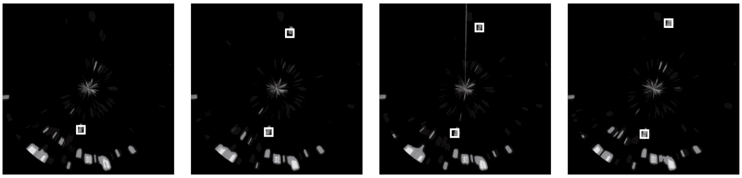
Sequence	#Frames	Real Objects	True Detections	False Detections
1	37	2	2	0
2	24	5	5	0
3	43	5	5	2
4	57	2	2	1
5	33	9	7	1
6	72	5	5	1

The aim of the next experiment is the analysis of the performance produced by the Adaptive Sliding Windows (ASW) algorithm described in Section 4. The experiment analyzes it for different radar image resolutions and window sizes. Table 3 shows the results obtained by executing the detection stage for each sequence, reporting the average time per frame obtained with the adaptive proposal versus the standard sliding window method.

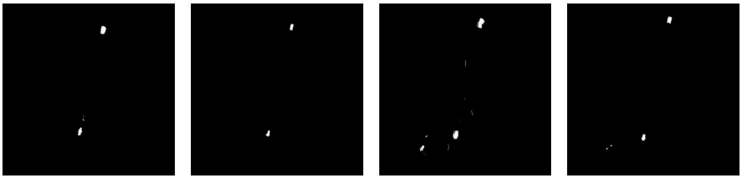
The ASW method obtains the highest speedup ($43.36\times$) in the most computing time consuming configurations (the highest image resolution and the largest window size), with an average of $32.05\times$. The increase of the window size makes the standard method slower than the ASW, as it needs to evaluate larger areas, specially if there is no noise in the image.

Table 3. Performance of the Adaptive Sliding Window (ASW) versus the standard method

Image Resolution	Window Size	Algorithm		Speedup	
		Standard (ms)	ASW (ms)	Average	Maximum
640 × 640	10 × 10	3.75	1.08	3.60×	5.32×
	20 × 20	10.04	0.98	10.22×	14.70×
	40 × 40	15.82	0.85	17.88×	24.72×
800 × 800	10 × 10	6.09	1.64	3.77×	4.89×
	20 × 20	17.04	1.45	11.63×	15.55×
	40 × 40	31.76	1.31	23.83×	31.15×
1024 × 1024	10 × 10	9.97	2.46	4.20×	5.95×
	20 × 20	30.20	2.12	14.18×	18.36×
	40 × 40	63.91	1.97	32.05×	43.36×



(a)



(b)

Fig. 6. Visual tracking results for Particle Filter (PF) in the first sequence: target tracking (a) and foreground image (b)

Figure 6.a shows the visual tracking of the particle filter over the first sequence with the targets tracked highlighted with a white rectangle. Figure 6.b shows the foreground image obtained in the background subtraction stage. This sequence shows how an occluded vehicle (upper area in the third frame) is recovered by the particle filter.

7 Conclusions

This work presents a complete intelligent traffic surveillance system based on low-cost marine radar devices, which is able to extract information of the controlled area in real time. This solution supports the concept of smart cities as it can be connected to other systems, to control and improve the traffic automatically. The main aim of this system is traffic monitoring and the generation of alarms in case of non-common situations (i.e., speed exceeds, traffic jams, etc.). The system is robust against adverse lighting and weather conditions thanks to the nature of the signal received, which makes it better than camera-based systems in those kind of conditions, extending its functionality. The system is divided into five main modules: signal filtering, detection, tracking, alarm manager and scenario modeling. The target detection module uses a dual background subtraction to segment vehicles and an adaptive sliding window method to detect them in the image. The target tracking module uses a particle filter algorithm to track the vehicles previously detected. Finally, the alarm manager filters false positives from the tracking system and generates alarms for unusual situations. The computational results show that the system is able to detect and track vehicles in real time even in difficult scenarios. Furthermore, the number of false positive alerts generated is very low in average, which is a really important feature of systems that are part of bigger ones, as would occur in smart cities. Currently the proposed system is working in a real urban environment, presenting a high performance based on the end-user analysis.

Acknowledgments. This research has been partially supported by the Spanish companies Inteomedia Mobile SLNE and Vigidar SL and the Spanish Government research funding ref. TIN 2011-28151 and TIN2012-35632, and by the Government of the Community of Madrid, grant ref S2009/TIC-1542.

References

1. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: the clear mot metrics. EURASIP Journal on Image and Video Processing 2008(1), 1–10 (2008)
2. Cabido, R., Montemayor, A.S., Pantrigo, J.J., Payne, B.R.: Multiscale and local search methods for real time region tracking with particle filters: local search driven by adaptive scale estimation on GPUs. Machine Vision & Applications 21(1), 43–58 (2009)
3. Gordon, N., Salmond, D., Smith, A.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings F. Radar and Singal Processing 140(2), 107–113 (1993)
4. Jing, L., Vadakkepat, P.: Interacting MCMC particle filter for tracking maneuvering target. Digital Signal Processing 22(1), 54–65 (2010)
5. Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Boonstra, M., Korzhova, V., Zhang, J.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics and protocol. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), 319–336 (2009)
6. Lanterman, A.D.: Passive radar imaging and target recognition using illuminators of opportunity. In: RTO SET Symposium on "Target Identification and Recognition Using RF Systems" (2004)
7. Pantrigo, J.J., Sánchez, A., Mira, J.: Representation spaces in a visual based human action recognition system. Neurocomputing 72(4-6), 901–915 (2009)
8. Porikli, F., Ivanov, Y., Haga, T.: Robust abandoned object detection using dual foregrounds. EURASIP J. Adv. Signal Process 2008 (2008)
9. Rossi, C., Abderrahim, M., Dìaz, J.C.: Tracking moving optima using Kalman-based predictions. Evolutionary Computation 16(1), 1–30 (2008)
10. Sánchez-Oro, J., Fernández-Lõpez, D., Cabido, R., Montemayor, A.S.: Radar-based road-traffic monitoring in urban environments. Digital Signal Processing 23(1), 364–374 (2013)
11. Zotkin, D.N., Duraiswami, R., Davis, L.S.: Joint audio-visual tracking using particle filters. EURASIP Journal on Applied Signal Processing 2002(1), 1154–1164 (2002)