

Local Search Particle Filter for a Video Surveillance System

A. Sánchez, R. Cabido, J.J. Pantrigo, A.S. Montemayor, J. Gutiérrez, F. Fernández

Dpto. CC. Computación

ESCET

U. Rey Juan Carlos

28933 Móstoles (Madrid)

Dpto. Tecnología Fotónica

Fac. Informática

U. Politécnica de Madrid

28660 Boadilla del Monte (Madrid)

{angel.sanchez.raul.cabido.juanjose.pantrigo.antonio.sanz}@urjc.es

jgr@dtf.fi.upm.es felipe.fernandez@es.bosch.com

Abstract

This paper presents a work in progress for indoor and outdoor target detection and feature extraction in video sequences. The framework can be applied to AmI systems related to surveillance activities. The system is based on a Local Search Particle Filter (LSPF) algorithm, which tracks a moving target and calculates its bounding box. Possible applications of this prototype include assisted monitoring to supervise video sequences from different cameras, and scene analysis for domotic environments.

1 Introduction

The term Ambient Intelligence (AmI) was coined by Philips Research [1] and it is beyond Ubiquitous Computing. This recent research field aims to build digital environments that are aware of the human presence, their behaviours and needs [2]. Thai [3] characterizes some of the key properties in AmI systems as: context awareness, system personalization, system anticipation, embedded devices and adaptability. Based on this paradigm, different AmI frameworks have been developed [3], mainly focused on the context-awareness issue that centres on the fact that the system can recognize people, their situational context, their actions and interactions. Different scenarios for the application of AmI principles are related to public transport environments [4], intelligent buildings [5] and other public places [4]. For example, in the case of intelligent buildings some technical innovations are embedded to the building to adapt it to changing conditions, to increase the comfort of the people in the building

or to make a more efficient usage of resources (air conditioning, lighting, humidity subsystems, etc). Another important issue in this context is security. The safety of people in these buildings can be increased by embedding an appropriate video surveillance subsystem to prevent uncontrolled access to the building area (both indoor and outdoor regions). This system should ideally be able to track the movements of a particular suspicious subject or a sequence of people (also for a car in the parking area), and to detect the actions performed by suspicious target. Video analysis in the context of AmI would also be useful to accurately count the number of people in specific building halls in order to smartly adapt the temperature or air conditioning conditions to the changing presence of persons. This dynamic system adaptability requires from near real-time video analysis requirements to make the surveillance tasks practical. We present in this paper a work in progress focused on analysis of image sequences. The resulting subsystem could be embedded into a video surveillance system for AmI applied to the indoor and outdoor security of a controlled environment (i.e. a public building and their parking regions).

The considered video analysis task consists in extracting quantitative measures from a moving target (human or vehicle) in a video sequence. This work is related to several important areas in computer vision which can be classified into three groups: low-level image preprocessing, object analysis and representation, and feature extraction from the tracked target shape or from its movement. A subsequent analysis of the target actions in surveillance tasks based on the extracted information of each

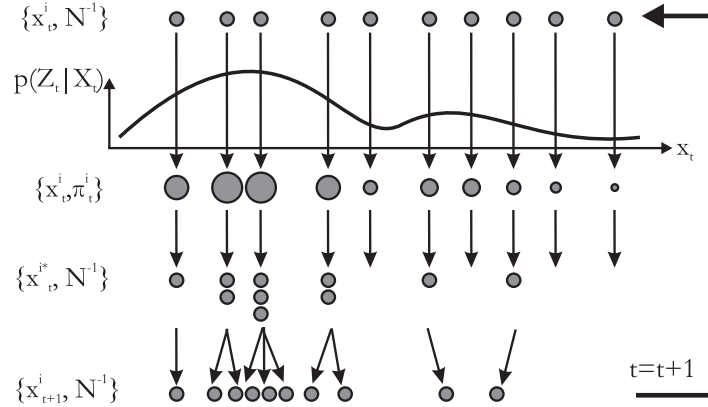


Figure 1: Particle Filter scheme.

frame is also further required but is not the goal of this work.

The scope of automatic visual surveillance technologies has shifted to the preventive tasks [6]. However, new technical challenges appear for the actual and potential applications of surveillance systems. These challenges include video processing capabilities, acceptable trade-off between system performance and involving costs, robust multiple object detection and tracking, and adaptability to uncontrolled changing environments [7][8].

Visual tracking provides a useful tool in surveillance systems such as the extraction of regions of moving targets [9]. Two of the most popular tracking methods include the Kalman filter (KF) and the particle filter (PF) algorithms. The KF is a recursive solution to the discrete-data linear filtering problem. KF models stochastic processes with Gaussian probability density functions (pdf) parameterized by their respective mean and covariance. The PF algorithm, enables the modeling of sequential stochastic processes with an arbitrary pdf, by approximating it numerically with a set of points (called particles) in a state-space process [10]. In Computer Vision, PF is known as Condensation algorithm and it has been successfully applied to many video surveillance systems [9][11].

In this work, we have designed and implemented a video-based feature extraction module as a component of visual tracking system. For the proposed

implementation, we make use of the Local Search Particle Filter (LSPF) framework to perform an accurate and fast tracking. The rest of the paper is described as follows: Section 2 presents the particle filter framework and extends it including an optimization stage, Section 3 describes the proposed video analysis system and Section 4 its evaluation. Finally Section 5 summarizes the conclusions and future work.

2 Local Search Particle Filter

To make inference about a dynamic system, two different models are necessary: (i) a measurement model requiring an observation vector (\mathbf{z}) and a system state vector (\mathbf{x}), and (ii) a system model describing the evolution of the state of the system [12]. The objective in the Bayesian approach to dynamic state estimation is to construct the posterior *pdf* of the state.

Particle filters (PF) approximates the theoretical distributions on the state-space are by simulated random measures [13]. This *pdf* is represented by a set of weighted samples, called particles $\{(\mathbf{x}_t^0, \pi_t^0), \dots, (\mathbf{x}_t^N, \pi_t^N)\}$, where the particle weights $\pi_t^i = p(\mathbf{z}_t | \mathbf{x}_t = \mathbf{x}_t^i)$ are normalized. In Figure 1 an outline of the Particle Filter scheme is shown.

PF algorithm starts by initiating a set $\mathbf{x}_0 = \{\mathbf{x}_0^i | i = 1, \dots, N\}$ of N particles using a known

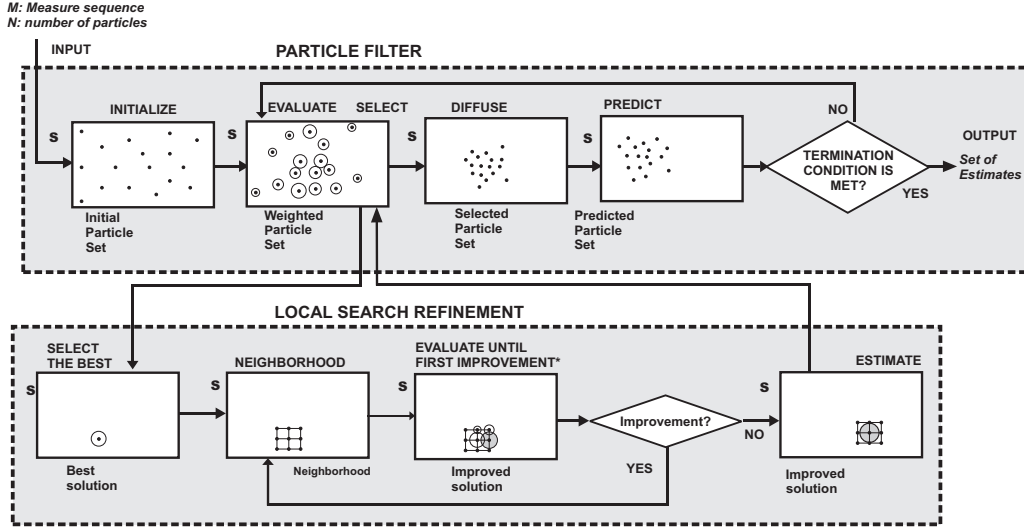


Figure 2: Local Search Particle Filter scheme. Weight computation is required during EVALUATE and EVALUATE UNTIL FIRST IMPROVEMENT stages (*)

pdf. The measurement vector \mathbf{z}_t at time step t is obtained from the system. Particle weights at time step t , $\pi_t = \{p_t^i \mid i = 1, \dots, N\}$ are computed using a fitness function. Weights are normalized and a new particle set \mathbf{x}_t^* is selected. As particles with higher weights can be chosen several times, a diffusion stage is applied to avoid the loss of diversity in \mathbf{x}_t^* . Finally, particle set at time step $t + 1$, \mathbf{x}_{t+1} , is predicted using a defined motion model.

Local Search Particle Filter (LSPF) algorithm is introduced to be applied to estimation problems in sequential processes that can be expressed using the state-space model abstraction. The aim of this algorithm is to improve the efficiency of the standard particle filters, by means of a local search procedure. This proposal is specially suitable for applications requiring high quality estimations. LSPF integrates both local search (LS) and particle filter (PF) frameworks in two different stages:

- In the *particle filter stage*, a particle (solution) set is propagated over the time and updated with measurements to obtain a new one. This stage is focused on the time evolution of the best solutions found in previous time steps.

- In the *local search stage*, the best solution from the particle set is selected and its neighborhood is evaluated searching for a better solution. This stage is devoted to improve the quality of the PF estimate.

2.1 Local Search Particle Filter Basic Algorithm

Figure 2 shows a graphical template of the LSPF method. Dashed lines separate the two main components in the LSPF scheme: PF and LS, respectively. LSPF starts with an initial population of N particles drawn from a known pdf (Figure 2: INITIALIZE stage). Each particle represents a possible solution of the problem. Particle weights are computed using a weighting function and a measurement vector (Figure 2: EVALUATE stage). LS stage is later applied for improving the best obtained solution of the particle filter stage:

First, a neighborhood of the best solution is defined (Figure 2: NEIGHBORHOOD stage). Then, solution weights are computed until a better solution is found in the neighborhood of the initial one (Figure 2: EVALUATE UNTIL FIRST IMPROVEMENT stage). This procedure is repeated un-

til there are no better solutions in the neighborhood than the initial one.

Once the LS stage is finished the process continues with the rest of the particle filter stages: In the selection one, a new population of particles is created by selecting the individuals from the whole particle set with probabilities according to their weights (Figure 2: SELECT stage). To avoid the loss of diversity, a diffusion stage is applied to the particles of the new set (Figure 2: DIFFUSE stage). Finally, particles are projected into the next time step by making use of the update rule (Figure 2: PREDICT stage).

2.2 Implementation details of the LSPF

In particular, LSPF can use different weighting functions and also different state-space topologies in PF and LS stages. In this work, PF uses a 2D search-space where the state of the individual i at time t is defined by two variables (x_t^i, y_t^i) describing the position of the target in the image. The quality or weight of a solution is proportional to the quantity of pixels detected as object in the background subtraction result, taking into account a window or bounding box of predetermined size (l_x^0, l_y^0) .

However, LS stage performs a local search in a 4D search space, instead of the 2D one shown in Figure 2. Therefore, the state of the individual i is defined by four variables (x^i, y^i, l_x^i, l_y^i) . The new variables l_x^i and l_y^i determine the size of the bounding box that fits the target. The local search procedure performs an iterative exploration of the neighborhood of the best solution (x^{best}, y^{best}) and initial bounding box dimensions (l_x^0, l_y^0) . Now, the quality or weight of a solution is not only proportional to the quantity of pixels detected as object but also inversely proportional to the background pixels in the bounding box. In this way, given two bounding boxes with the same number of object pixels, the larger one will represent a lower quality solution. The local search procedure tries to find the best solution using this fitness criterion. First, LS performs an evaluation of the weight of every solution in an 8-neighbor space from the initial position (x^i, y^i) . For each neighbor it is evaluated the weight while changing the size of the bounding box (l_x^i, l_y^i) . This process is repeated until no better solution is found in the neighborhood. As a result, we obtain a bound-

ing box centered and fitting the target.

3 Overview of the developed video analysis system for automatic surveillance

This section sketches the feature extraction subsystem corresponding to the surveillance application for both indoor and outdoor areas. For example, in the indoor case a video camera can be placed at the end of a corridor in order to capture the complete area, where we expect people to walk without stopping many times or during a long period of time.

We have developed a MATLAB prototype for the visual tracking feature extraction component that works under a standard PC equipped with a webcam. Figure 3 shows the graphic user interface (GUI) of the implemented subsystem. This GUI visualizes different measures of interest extracted from the video sequence, related to shape and kinematics of a target being tracked.

On the top left side of the GUI, we show the actual video frame. In this image, we draw the smallest enclosing rectangle and the convex hull of the moving target. We also represent the background subtraction image (left bottom) in which our LSPF algorithm is applied to detect and track the moving target along the video sequence. The foreground moving object is represented in white while the background is set to black at each photogram.

In order to compute measurements of the trackable target, a background subtraction is used as measurement model for the LSPF algorithm. Given the background image I_B and the actual video frame at time t , I_F^t , a new binary image results from applying a threshold to the difference image $|I_F^t - I_B|$. This binary image is the measurement model.

The set of extracted features for the target are grouped into four categories: position, shape, kinematics and occupation area. The position features correspond to the coordinates of the target centroid and its orientation with respect to the horizontal axis. The extracted shape features are: major and minor axis length, perimeter in pixels, solidity (computed as the ratio between the target area and its corresponding convex hull) and its Euler number (computed as a difference between the number of connected components and the number of holes in the target). We extract two kinematic features: velocity and acceleration of the considered object.

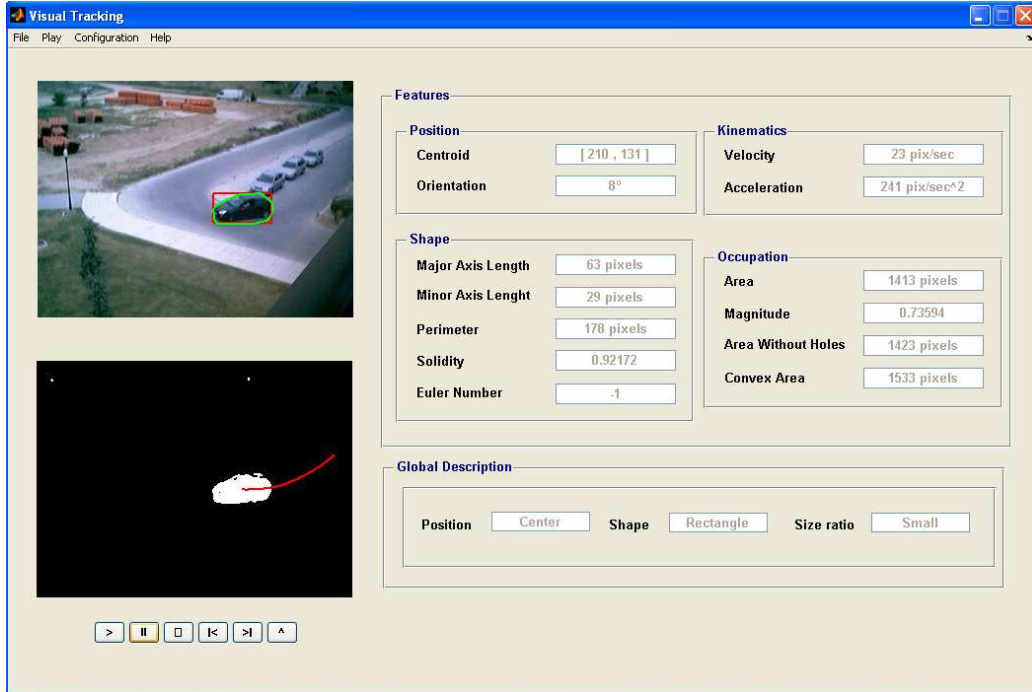


Figure 3: Application GUI.

The occupation features are related to the different area measures of the target.

Finally, we also define three global description parameters (right-bottom part of the GUI): position, shape and size ratio. The position feature means the region of the image where the target is placed. We consider nine possible positions: north-east, north, north-west, east, cent, west, south-east, south and south-west. The shape feature considers the global shape of the target that is extracted from its convex hull (it can be a rectangle, a square, a circle or another shape). The size ratio represents how big is the object area with respect to the image area (we considered three possibilities: small, medium and large). This set of qualitative features can easily be extended (i.e. to incorporate the target trajectory in the video sequence), and these features can be used to establish a set of surveillance rules to support decision-making. In particular, we can define different security levels depending on the po-

sition and/or the recognized actions performed by the target: “safe”, “warning” and “alert” levels. For example, in an indoor video sequence (as presented in figure 4.b) a person walks along a corridor. The “safe” level is activated when the person advances in normal conditions, that is, when the person is approaching to or moving away the video camera at reasonable speed. The “warning” level is activated when the person stays in the place without noticeable motion. Finally, the system turns to “alert” when the person adopts a suspicious attitude (i.e. he/she throws away an object).

4 System evaluation

We tested our visual tracking feature extraction subsystem on multiple indoor and outdoor video sequences. Figures 4.a and 4.b respectively show two different surveillance situations where a car is moving through a parking area and where a man ap-

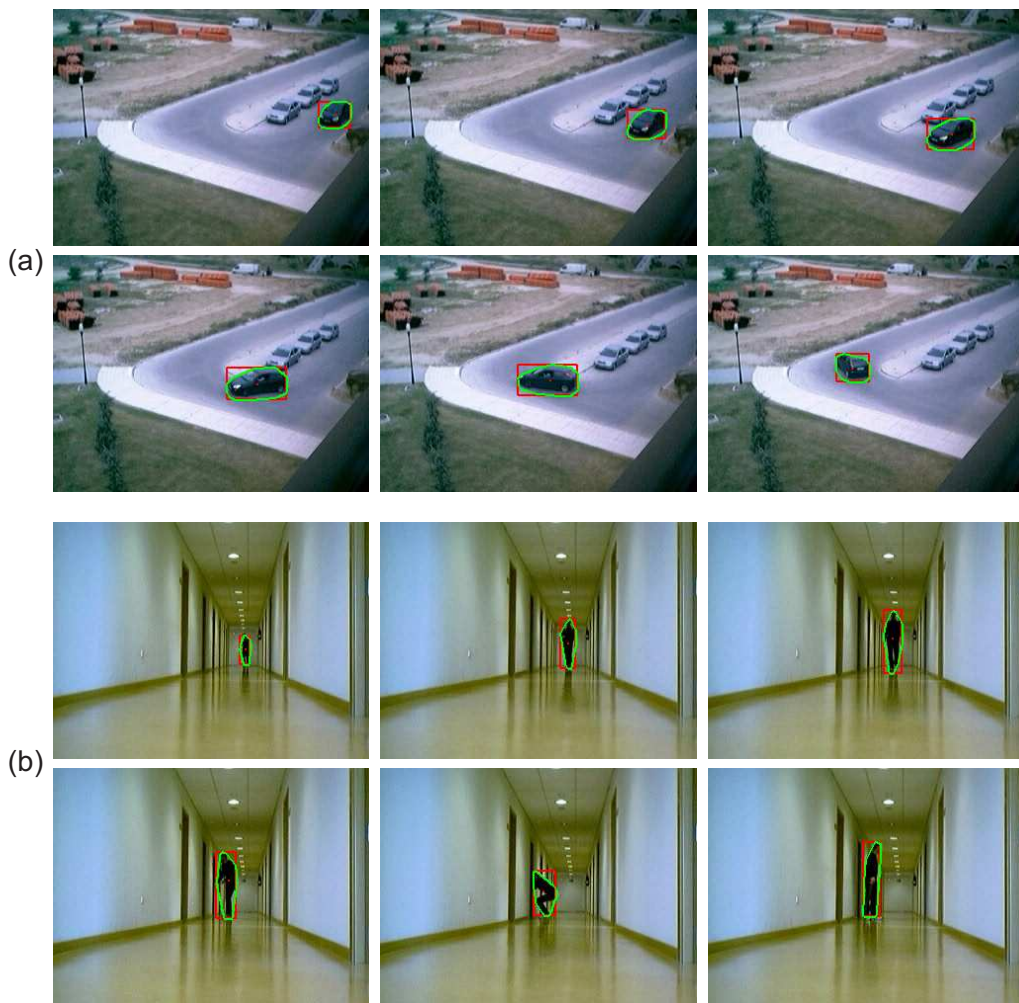


Figure 4: (a) Outdoor sequence, (b) Indoor Sequence.

pears at the end of a corridor and he is moving ahead to the camera. For simplicity, we only show the top-left image in the GUI (i.e. the actual photogram where the tracked target is perfectly delimited by its convex hull and its smallest enclosed rectangle). In most analyzed image sequences the target position and size are perfectly described by the LSPF algorithm. We have successfully tested our prototype using different video sequences with varying illumination conditions and tracking only one moving target.

5 Conclusions

This paper shows a work in progress of indoor and outdoor target (person or vehicle) detection and feature extraction in video sequences. It can be applied to AmI frameworks related to the security of intelligent buildings. The core of the system is a LSPF algorithm, which tracks the moving target, calculates its bounding box and computes different shape and motion parameters. This system can be integrated as monitoring tool in order to help and assist to human operators to supervise video capture from many different cameras. It can also be used as an analysis component of a domotic environment.

As future works we propose to establish a complete set of rules to enrich the identification of dangerous situations. Also, as the system increase its functionality it would be desirable to work under a well established rule combination framework such as a fuzzy rule-based system. In the proposed prototype we handle only one trackable object, so a multiple object tracking system would improve the system capabilities.

Acknowledgments.

This research has been supported by the Spanish projects TIN2005-08943-C02-02 (2005-2008) and TIN2005-08943-C02-01 (2005-2008).

References

- [1] Philips Research, Philips Research Technologies, Ambient Intelligence, 2007. http://www.research.philips.com/technologies/syst_softw/ami/vision.html
- [2] M. Lindwer et al, "Ambient Intelligence Visions and Achievements: Linking Abstract Ideas to Real-World Concepts", Proc. Intl. Conf. on Design Automation & Test in Europe (DATE'03), Vol. 1, 2003.
- [3] V.T. Thai, "A Survey on Ambient Intelligence in Manufacturing Environment", Technical Report, National University of Ireland, 2006.
- [4] S.A. Velastin, B.A. Boghossian, B.P.L. Lo, J. Sun and M.A. Vicencio-Silva, "PRISMATICA: Towards Ambient Intelligence in Public Transport Environments", IEEE Trans. on Systems, Man, and Cybernetics - Part A, 35(1), pp. 164-182, 2005.
- [5] L. Snidaro, C. Micheloni and C. Chiadevale, "Video Security for Ambient Intelligence", IEEE Trans. on Systems, Man, and Cybernetics - Part A, 35(1), pp. 133-144, 2005.
- [6] Haritaoglu, I., Harwood, D., Davis, L.S. (2000) W4: Real-Time Surveillance of People and Their Activities, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **22**: 809-830.
- [7] Iannizzotto, G., Vita, L. (2002) On-line Object Tracking for Colour Video Analysis, *Real-Time Imaging*, **8**: 145-155.
- [8] Sebe, I.O., Hu, J., You, S., Neumann, U. (2003), 3D Video Surveillance with Augmented Virtual Environments, *Proc. 1st ACM SIGMM Int. Workshop on Video Surveillance, Berkeley, CA, USA*, pp. 107-112.
- [9] Dockstader, S.L., Tekalp, M. (2001) On the Tracking of Articulated and Occluded Video Object Motion, *Real-Time Imaging*, **7**: 415-432.
- [10] Zotkin, D., Duraiswami, R., Davis, L. (2002) Joint Audio-Visual Tracking Using Particle Filters. *EURASIP Journal on Applied Signal Processing*, **11**: 1154-1164
- [11] KaewTrakulPong, P.; Bowden, R. (2003) A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes. *Image and Vision Computing*, **21**: 913-929

- [12] Arulampalam, M., Maskell, S., Gordon, N., Clapp, T. (2002) A Tutorial on Particle Filter for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Trans. On Signal Processing*, **50** (2): 174-188 (2002)
- [13] Carpenter, J., Clifford, P., Fearnhead, P. Building robust simulation-based filters for evolving data sets. Tech. report, Univ. of Oxford, Dept. of Statistics (1999).
- [14] Torma, P. and Szepesvári, C. LS-N-IPS: An Improvement of Particle Filters by means of Local Search. Proc. of the Non-linear Control Systems (2001).