MULTI-OBJECT TRACKING USING THE PARTICLE FILTER ALGORITHM ON THE TOP-VIEW PLAN

Santiago Venegas Martnez, Jean-François Knebel and Jean-Philippe Thiran

Signal Processing Institute, Swiss Federal Institute of Technology (EPFL)
CH - 1015 Lausanne, Switzerland
phone: +41 21 6934329, fax: +41 21 6937600, email:{santiago.venegas, Jean-Francois.Knebel, JP.Thiran}@epfl.ch
web: http://ltswww.epfl.ch

ABSTRACT

In this paper we address the problem of multi-object tracking in video sequences, with application to pedestrian tracking in a crowd. In this context, particle filters provide a robust tracking framework under ambiguity conditions. The particle filter technique is used in this work, but in order to reduce its computational complexity and increase its robustness, we propose to track the moving objects by generating hypotheses not in the image plan but on the top-view reconstruction of the scene. Comparative results on real video sequences show the advantage of our method for multi-object tracking.

1. INTRODUCTION

Video object tracking in dense visual clutter, although being notably challenging, has many practical applications in scene analysis for automated surveillance, such as the detection of suspicious moving objects (pedestrians or vehicles), or the monitoring of an industrial production [1][2] [3][4]. The quality of an object tracking system is very much dependent on its ability to handle ambiguous conditions, such as occlusion of an object by another one. To cope with such ambiguities, multi-hypotheses techniques have been developed [5]. In the standard techniques using multi-hypotheses for the state estimation and tracking, the Kalman filter is used under the premise that the noise distributions are Gaussian and the system dynamics are linear [6]. However, when tracking human movements, non-linear and non-stationary assumptions make it suboptimal to use. In this context particle filter algorithms are attractive because they are both simple and very general. The particle filter algorithms track objects by generating multiple hypotheses and by ranking them according to their likelihood. They suppose that the correct hypothesis is retained [7][8]. Many tracking filters have been proposed using this approach, defining the states as being each static posture or position of the objects and modeling a motion sequence by the composition of these states with some transitional probabilities [9][10][11]. Those state-of-the-art techniques perform efficiently to trace the movement of one or two moving objects but the operational efficiency decreases dramatically when tracking the movement of many moving objects because systems implementing multiple hypotheses and multiple targets suffer from a combinatorial explosion, rendering those approaches computationally very expensive for real-time object tracking. In this paper we propose an efficient approach for the track maintenance problem keeping

a low computational cost. In our algorithm, the hypotheses are generated not on the image plan but on the top-view reconstruction of the scene. A calibrated camera is necessary to get this reconstruction. On this plan, the object dynamics can be modeled more conveniently and precisely than on the image plan, allowing to considerably reduce the number of hypotheses needed to achieve a robust tracking. In our practical application of pedestrian tracking we will show a simple model where the appropriate guidance control follows a anisotropic Gaussian function oriented along the current object motion direction [12].

The article is organized as follows: In Section 2 we briefly describe the particle filter algorithm. In Section 3 the multi-object tracking system, using the reconstructed top-view plan, is introduced. In Section 4 some dynamic models are presented and tested. Results are illustrated and discussed. Conlusion and future research in Section 5.

2. PARTICLE FILTER

Particle filtering provides a robust tracking framework, as it models uncertainty. Particle filters are very flexible in that they not require any assumptions about the probability distributions of data. In order to track moving objects (e.g. pedestrians) in video sequences, a classical particle filter continuously looks throughout the 2*D*-image space to determine which image regions belong to which moving objects (target regions). For that a moving region can be encoded in a state vector.

2.1 Target regions encoded in a state vector

In the tracking problem the object identity must be maintained throughout the video sequences. The image features used therefore can involve low-level or high-level approaches (such as the colour-based image features (histograms), a subspace image decomposition or appearance models) to build a state vector.

A target region over the 2*D*-image space can be represented for instance as follows:

$$\mathbf{r} = \{\mathbf{l}, \mathbf{s}, \mathbf{m}, \gamma\} \tag{1}$$

where $\mathbf{l}$ is the location of the region, $\mathbf{s}$ is the region size, $\mathbf{m}$ is its motion and γ its direction. In the standard formulation of the particle filter algorithm, the location $\mathbf{l}$, of the hypothesis, is fixed in the prediction stage using only the previous approximation of the state density. Moreover, the importance of using an adaptive-target model to tackle the problems such as the occlusions and large-scale changes has been largely

recognized. For example, the update of the target model can be implemented by the equation

$$\bar{\mathbf{r}}_t = (1-\lambda)\bar{\mathbf{r}}_{t-1} + \lambda E[\mathbf{r}_t] \qquad (2)$$

where λ weights the contribution of the mean state to the target region. So, we update the target model model during slowly changing image observations.

2.2 The propagation algorithm

In the standard formulation of a particle filter algorithm, the aim is to estimate recursively in time the filtering density (also called posterior density) defined in a state space. Therefore the image features are modeled as an object class and they can be used in a dynamical model expressed as a temporal Markov-chain, where the hypotheses are fixed in the prediction stage using only the previous approximation to the state density. For example, a 2nd order process can be conveniently represented in discrete time t as,

$$\mathbf{r_t} - \bar{\mathbf{r}} = \mathbf{S}(\mathbf{r_{t-1}} - \bar{\mathbf{r}}) + \mathbf{N}\mathbf{w_t} \qquad (3)$$

where, $\mathbf{r_t}, \mathbf{r_{t-1}}$ are the state-vectors, $\bar{\mathbf{r}}$ is the mean value of the state vector, $\mathbf{w}$ is the noise term and $\mathbf{S}$ and $\mathbf{N}$ are the matrices representing the deterministic and stochastic components. In this way, the *learned dynamical models* are appropriate to be used in the propagation algorithms. Given a continuous-valued Markov chain with independent observations, the conditional state-density p_t at time t is defined by

$$p_t(\mathbf{r_t}) \cong p(\mathbf{r_t} \mid \mathbf{I_t}). \qquad (4)$$

This represents the whole information about the state of a region $\mathbf{r}$, and $\mathbf{I_t} = \{\mathbf{i_1}...\mathbf{i_t}\}$ the image features at time t. And the dynamical model can be re-expressed as:

$$p(\mathbf{r_t} \mid \mathbf{r_{t-1}}) \propto$$
$$\exp{-\frac{1}{2}\|\mathbf{N^{-1}}((\mathbf{r_t} - \bar{\mathbf{r}}) - \mathbf{S}(\mathbf{r_{t-1}} - \bar{\mathbf{r}}))\|^2} \qquad (5)$$

The time propagation rule is made of two steps: a prediction and a update step:

PREDICTION STEP : The prediction density is obtained by applying a dynamical model to the output of the previous time step.

$$p(\mathbf{r_t} \mid \mathbf{I_{t-1}}) = \int_{\mathbf{r_{t-1}}} p(\mathbf{r_t} \mid \mathbf{r_{t-1}})p(\mathbf{r_{t-1}} \mid \mathbf{I_{t-1}}). \qquad (6)$$

UPDATE STEP : The output measurement update stage is a set of N weighted particles.

$$p(\mathbf{r_t} \mid \mathbf{I_t}) = p(\mathbf{i_t} \mid \mathbf{r_t})p(\mathbf{r_t} \mid \mathbf{I_{t-1}}) \qquad (7)$$

where the set of image features at time t is $\mathbf{i_t}$ with history $\mathbf{I_t} = \{\mathbf{i_1}...\mathbf{i_t}\}$. In the standard particle filter, the set is re-sampled in order to discard particles with insignificant weights and multiply particles with large weights.

3. TRACKING MOVING OBJECTS ON THE TOP-VIEW PLAN

3.1 State-space over the top-view plan

In a practical particle filter implementation, the prediction density is obtained by applying a dynamic model to the output of the previous time-step. This is appropriate when the hypothesis set approximation of the state density is accurate. But the random nature of the motion model induces some non-zero probability everywhere in state-space that the object is present at that point. The tracking error can be reduced by increasing the number of hypotheses (particles) with considerable influence on the computational complexity of the algorithm. However in the case of tracking pedestrians we propose to use the top-view information to refine the predictions and reduce the state-space, which permits an efficient discrete representation. In this top-view plan the displacements become Euclidean distances. The prediction can be defined according to the physical limitations of the pedestrians and their kinematics. The calibration of such models is a work in progress in our group.[1] In this article we use a simpler dynamic model, where the actions of the pedestrians are modeled by incorporating internal (or personal) factors only. The displacements $\mathbf{M}_{\text{topview}}^{t}$ follows the expresion

$$\mathbf{M}_{\text{topview}}^{t} = \mathbf{A}(\gamma_{\text{topview}})\mathbf{M}_{\text{topview}}^{t-1} + \mathbf{N} \qquad (8)$$

where $\mathbf{A}(.)$ is the rotation matrix, $\gamma_{topview}$ is the rotation angle defined over top-view plan and follows a Gaussian function $g(\gamma_{topview}; \sigma_\gamma)$, and $\mathbf{N}$ is a stochastic component. This model proposes an anisotropic propagation of M : the highest probability is obtained by preserving the same direction. The evolution of a sample set is calculated by propagating each sample according to the dynamic model. So, that procedure generates the hypotheses.

3.2 Estimation of region size

The size of the search region represents a critical point. In our case, we use the *a-priori* information about the target object (the pedestrian) to solve this tedious problem. We assume an averaged height of people equal to 160 cm, ignoring the error introduced by this approximation. That means, we can estimate the region size $\mathbf{s}$ of the hypothetical bounding box containing the region of interest $\mathbf{r} = (l, \mathbf{s}, \mathbf{m}, \gamma)$ by projecting the hypothetical positions from top-view plan (see Fig. 1). A camera calibration step is necessary to verify the hypotheses by projecting the bounding boxes. So this automatic scale selection is an useful tool to distinguish regions. In this way for each visual tracker we can perform a realistic partitioning (bounding boxes) with consequent reduction in the computational cost. The distortion model of the camera's lenses has not been incorporated in this article. Under this approach, the processing time is dependent on the region size.

3.3 The output measurement update stage

In multi-object tracking, the hypotheses are verified at each time step by incorporating the new observations (images). A well known measure of association (strength) of the relationship between two images is the normalized correlation.

$$dc_{j,n} = corr_{nor}(target_j, hypothesis_{j,n}) \qquad (9)$$
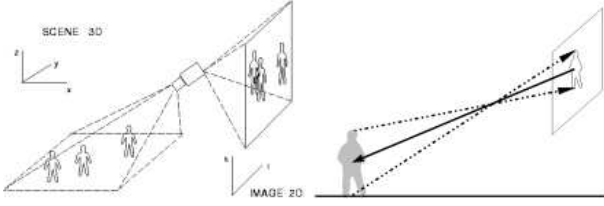

Figure 1: left: the approximation of Top-View plan by image plan with a monocular camera, right: size estimation
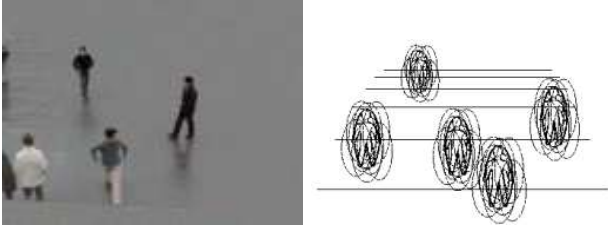

Figure 2: left: foreground image of the background subtraction, right: multi-hypotheses

where j : target region, and n : an hypothesis of the target region j. The observation of each hypothesis is weighted by a Gaussian function with variance σ.

$$h^{(j,n)} = \frac{1}{\sqrt{2\pi}\sigma_{dc}} e^{\frac{-(1-\mathbf{dc_{j,n}})^2}{2\sigma_{dc}^2}} \qquad (10)$$

where $h^{(j,n)}$ is the observation probability of the hypothesis n tracking the target j. The obvious drawback of this technique is the choice of the region size (defined in previous section) that will have a great impact on the results. Larger region sizes are less plagued by noise effects.

3.4 Background subtraction

In order to reduce background effects, the correlation is performed by using foreground image. Since the camera is fixed the background can be modelled statistically. We compute the difference between the background image and the current frame. From that we obtain a binary support layer and a foreground-object image (see Fig. 2). So, the visual correlations will be performed over the foreground-object image between current frame and the target-regions.

The proposed tracking algorithm performs the following steps.

- Recognize the moving regions based on background subtraction.
- Map the image coordinates to the real-world coordinates,(top-view plan).
- Perform a realistic partitioning on the image using the real-world coordinates.
- Compute correlation and use the mean-value locations for the consecutive video images to establish the trajectory, based on the particle filtering technique.
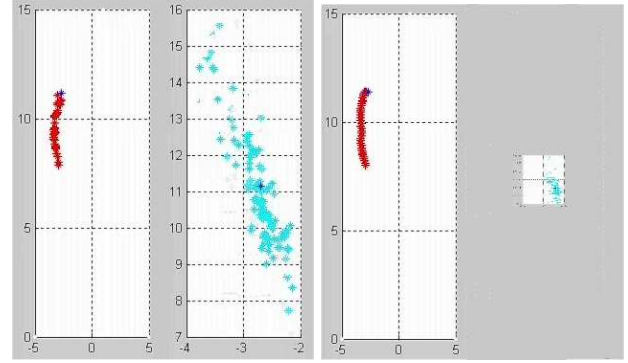
(a) **A corrupted track** and the non-compact diffusion of hypotheses at time t

(b) **An improved track** and the compact diffusion of hypotheses at time t

Figure 3: left: The top-view pedestrian trajectory of a particle propagation on image plan, right: The top-view pedestrian trajectory of a particle propagation on top-view plan

4. RESULTS

The goal of our experiments is to track moving regions (pedestrians) during the video sequences. We compare both dynamical propagation on image plan (the classical approach) and on top-view plan.

- **model₁** : The dynamic model propagates the particles over the image plan with anisotropic propagation.

- **model₂** : The dynamic model propagates the particles over the top-view plan with isotropic propagation.

- **model₃** : The dynamic model propagates the particles over the top-view plan with anisotropic propagation.

We have supposed pedestrian's height : 1.60 m and we have analysed outdoor video sequences representing the exit of a metro station, 10 *images/s* with pedestrian's displacements between 0.05 and 0.25 *m/image*. Fig. 3 shows an example of tracking a pedestrian with the same dynamical model performing on both image plan and top-view plan. We can see the projected particles on the top-view plan : the first case presents a track corrupted by particles located far from $E[r_t]$ and the zoomed area containing the particles in a range of $9mx2m$, and the second one presents an improved track performed with the compact diffusion in a range of $0.5mx1.2m$. The experiment was repeated many times varying the particles each time. Fig. 4 shows the results. N/A means that more than 100 hypotheses are needed to track moving objects. Fig. 5 and Fig. 6 show examples of the tested video sequences and illustrate the importance of an adaptive target model in cases of occlusions and large scale changes. The mean state of each object is estimated at each time step and then plotted as a box. The video sequences can be downloaded from http://ltswww.epfl.ch/ltsftp/Venegas/.

5. CONCLUSION AND FUTURE RESEARCH

In this paper we have shown how a simple behavioral model of pedestrian dynamic consisting of maximum displacement

Samples	$model_1$	$model_2$	$model_3$
$video_1$	N/A	35	30
$video_2$	N/A	N/A	60
$video_3$	N/A	30	30
$video_4$	N/A	50	30
$video_5$	N/A	N/A	100
$video_6$	N/A	100	30
$video_7$	N/A	55	35
$video_8$	N/A	30	30

Figure 4: The number of hypotheses to avoid crossing.
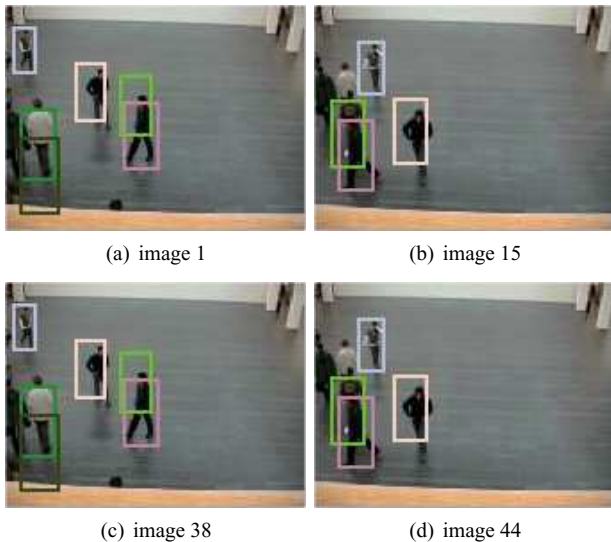

(a) image 1   (b) image 15

(c) image 38   (d) image 44

Figure 5: The mean state of the objects. $model_3$: the particles are propagated with anisotropic propagation on top-view plan

(a) image 31   (b) image 39
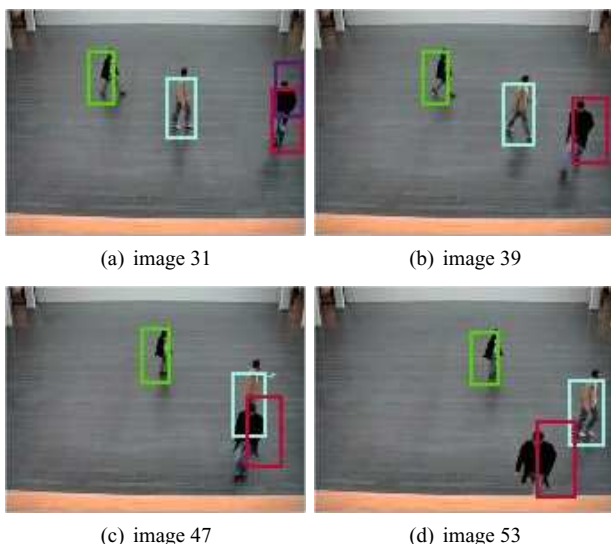
(c) image 47   (d) image 53

Figure 6: The mean state of the objects. $model_3$: the particles are propagated with anisotropic propagation on top-view plan

and change in direction, can be very usefull to solve the tracking problem, because it is directly linked to the actual pedestrian behavior. We are currently working on the specification and calibration of a more complex pedestrian behavioral model. The preceding sections have discussed a particle filter algorithm which performs propagation on top-view plan and verification on image plan. We believe that the constraints and/or models, made on the top-view plan, are more effective (realistic) that complex models made on the image plan. The pedestrian tracker can efficiently handle non-rigid objects under different appearance changes. Also, as a limitation under this approach is the real-time capability, the processing time is dependent on the region size and the number of hypotheses per pedestrian. Incorporating the coarse-to-fine hierarchy of observation is straightforward.

REFERENCES

[1] A. W. Senior, " Tracking with Probabilistic Appearance Models," in *Proc ECCV workshop on Performance Evaluation of Tracking and Surveillance Systems*, pp 48–55, June 2002

[2] M. Bierlaire, G. Antonini and M. Weber "Behavioural Dynamics for Pedestrians," *in K. Axhausen (Ed.), Moving through nets: the physical and social dimensions of travel*, 1–18, Elsevier. 2003

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," In *Proc. Conf. Comp. Vision Pattern Rec.*, vol II, pp 142–149, Hilton Head, SC, 2000

[4] D. Izquierdo,and Y. Berthoumieu, "High and low level object descriptions for video tracking process," In *Proc. EUSIPCO2002*, Toulouse, France, 2002

[5] K. Choo, and D.J. Fleet, "People tracking using hybrid Monte Carlo filtering," In *Proc. Int. Conf. Computer Vision*, vol. II, pp. 321-328, Vancouver, Canada, 2001

[6] B. Anderson, and J. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, 1979.

[7] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte-Carlo Methods in Practice*, Springer Verlag, April 2001

[8] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," Journal of Computational and Graphical Statistics, Vol. 5(1), pp. 1–25, 1996.

[9] K. Nummiaro, E. Koller-Meier, L.J. Van Gool, "Object Tracking with an Adaptive Color-Based Particle Filter," *DAGM-Symposium Pattern Recognition*, pp. 353–360, 2002

[10] M. Isard and A. Blake, "CONDENSATION - Conditional Density Propagation for Visual Tracking". *International Journal on Computer Vision*, vol. 1(29), pp. 5–28, 1998.

[11] J. Vermaak, A. Doucet and P. Perez "Maintaining Multi-Modality through Mixture Tracking," *International Conferenceon on Computer Vision*, ICCV2003, Nice, France 2003.

[12] S.Venegas, JM. Rendon and G. Stamon " Performed Anisotropic Diffusion by a Recursive Linear Convolving Method: Application to Space-Time Segmentation," *VI2002*, Calgary, Canada, 2002.