Particle filter–based visual tracking for real time videosurveillance

Fabrizio Dini

Media Integration and Communication Center, University of Florence, Italy

`http://www.micc.unifi.it - dini@dsi.unifi.it`

1. Introduction

Visual tracking for videosurveillance applications still deserves the researchers effort for the particular requirements it has to meet. In fact, those applications often requires a quick "reaction" to some observed event in the real world, and this means that the system has to be able to follow the target in real time while deciding whether a certain action (i.e. ringing an alarm) has to be taken or not, coping with occlusion, cluttered background, and highly erratic motion at the same time. It's obviously a challenging task, and often a proper tradeoff between accuracy and low computational cost must be found.

2. Color–based particle filter with uncertainty management

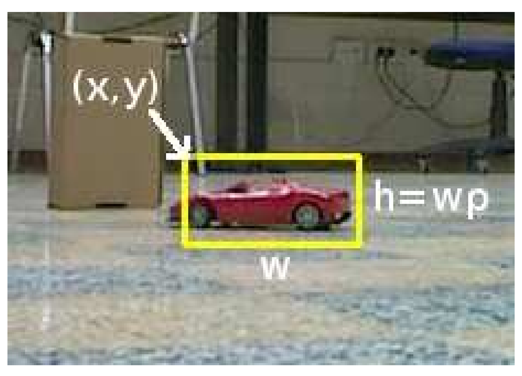
We focused our attention on particle filters, which presents a number of advantages: they make no assumption on the posterior density, work with linear and non-linear models as well, and they handle multiple hypoteses about the actual state vector value.

In order to achieve the aforementioned tradeoff between tracking accuracy and low computation cost, a computationally efficient, though weak appearance model has been chosen: color histograms. This can lead to sometimes inaccurate, and generally noisy, estimates of the state vector's value.

That is why we developed a novel technique to adapt the magnitude of the noise included in the state update equation, based on a tracking quality measure.

Given the state vector:

$$\mathbf{x}_k = \Big[\underbrace{x_k \quad y_k \quad w_k \quad \rho_k}_{\mathbf{s}_k} \quad \underbrace{\dot{x}_k \quad \dot{y}_k \quad \dot{w}_k \quad \dot{\rho}_k}_{\mathbf{d}_k}\Big]^T$$



and the state update equation:

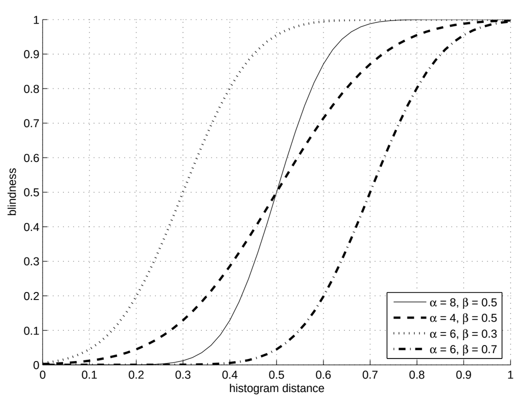
$$\mathbf{x}_k = \begin{bmatrix} \mathcal{I}_4 & \mathcal{I}_4 \Delta t \\ 0 & \mathcal{I}_4 \end{bmatrix} \cdot \mathbf{x}_{k-1} + \mathbf{v}_{k-1}$$

where $\mathbf{v}_{k-1}$ is an additive, zero mean, isotropic Gaussian noise, characterized with a vector of standard deviations:

$$\Sigma = \Big[\underbrace{\sigma^x \quad \sigma^y \quad \sigma^w \quad \sigma^\rho}_{\Sigma^{\mathbf{s}}} \quad \underbrace{\sigma^{\dot{x}} \quad \sigma^{\dot{y}} \quad \sigma^{\dot{w}} \quad \sigma^{\dot{\rho}}}_{\Sigma^{\mathbf{d}}}\Big]^T$$

we module the amplitude of the noise $\mathbf{v}_{k-1}$ according to the tracking quality, which is computed by comparing the reference color histogram with that corresponding to the last estimated target position. The Bhattacharyya distance ψ_k between the two histograms is passed trough a sigmoid function in order to smooth the histogram distance fluctuations. The result is a parameter we call *blindness*, with values in $[0, 1]$:

$$\zeta_k = \frac{\mathrm{erf}\left(\alpha\left(\psi_k - \beta\right)\right) + 1}{2}$$



A low blindness value means a high tracking quality (the histograms are similar), while a high blindness value means a low tracking quality (the histograms are different).

3. Reasons for the online parameters adaptation

1. Using 1^{st}–order dynamic model may cause the uncertainty on the velocity to be propagated to the position when applying the state update equation. Thus, the dynamic model should not include both position and velocity noises, or at least not with the same order of magnitude.
2. Target scaling causes changes in the target apparent motion magnitude. This means that the tracker may have to cope with greater motion variations (i.e. accelerations) as long as the target gets close to the camera. In 1^{st}–order dynamic models those variations can only be compensated through noise in the state update equation.
3. A weak appearance model (e.g. colour histograms) may cause the posterior pdf the filter is trying to estimate to be strongly multi-modal, making the estimate very noisy or inaccurate. In such situations, a method to limit the state space interval where the filter's particles are spread is needed.

4. Methods for the online parameter estimation

1. The maximum magnitude of position noise is chosen to be at least an order of magnitude greater than the maximum magnitude of velocity noise.
2. The noise must be proportional to the distance between the target and the camera. This could be achieved making the noise proportional to the target size: the greater, the nearest.
3. An evaluation of the tracking quality can be used to decide if the tracker can rely only on the velocity noise, or if it needs more position noise to cope with increased uncertainty on the target position. The blindness parameter is used for this purpose.

The above methods can be summarized in the following formulas:

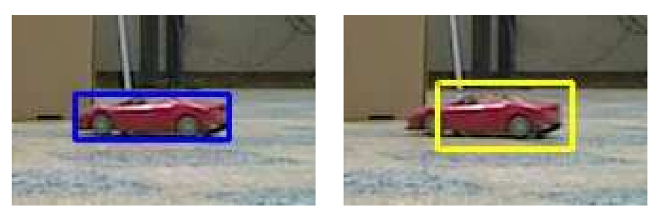
$$\begin{cases} \Sigma^{\mathbf{s}}_k &= \zeta_k \min(w_k, h_k)\, \Sigma^{\mathbf{s}} \\ \Sigma^{\mathbf{d}}_k &= (1 - \zeta_k) \min(w_k, h_k)\, \Sigma^{\mathbf{d}} \end{cases}$$

5. Experiments

The tracker performance has been tested on 10 video sequences expressly taken in our lab, comparing the segmentation provided by the particle filter with that corresponding to the background truth, frame by frame.

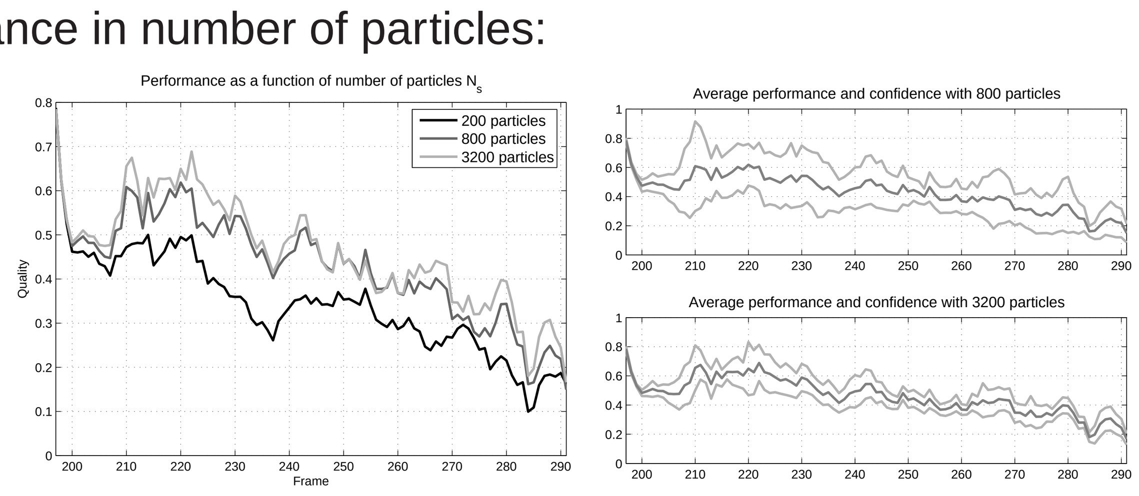
$$\mathcal{Q}_k = \frac{\left|Rect^k_g \cap Rect^k_t\right|}{\left|Rect^k_g \cup Rect^k_t\right|}$$

(*Ihsin T. Phillips and Atul K. Chhabra. Empirical performance evaluation of graphics recognitions systems.* IEEE Trans. Pattern Anal. Mach. Intell., *21(9):849-870, 1999.)*

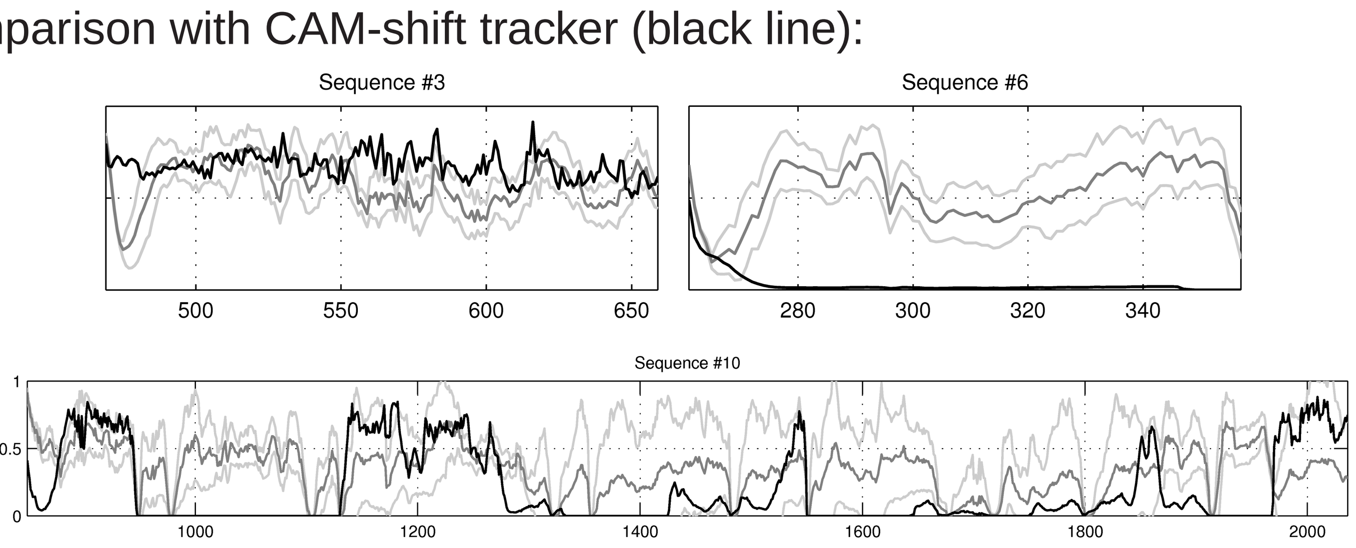


6. Results

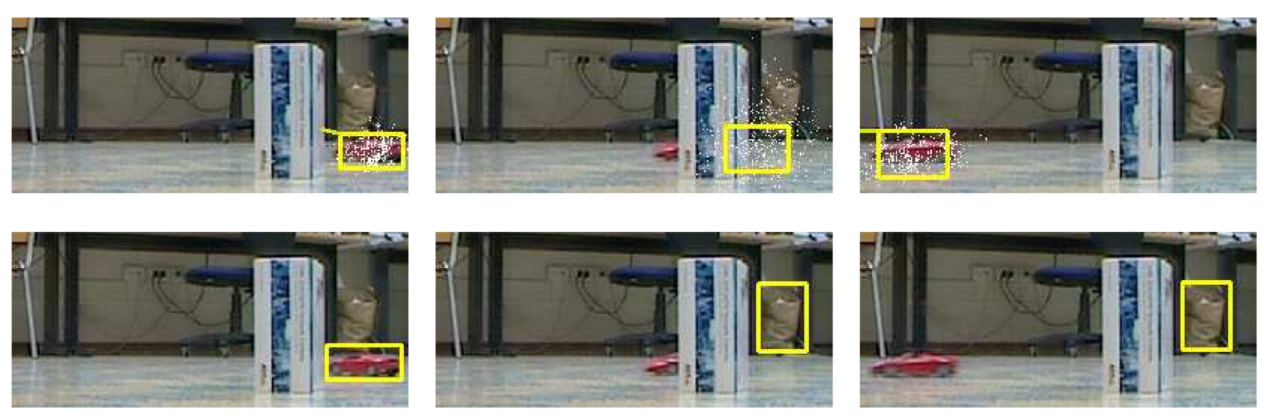
1. Performance in number of particles:



2. Comparison with CAM-shift tracker (black line):



3. Robustness to occlusions (above, the particle filter tracker; below, the CAM-shift tracker – the figures refers to sequence #10):



(a) frame 939 (b) frame 951 (c) frame 969

7. Applications

A face logger that tracks faces and stores only the best quality face images has been developed. It uses a Viola-Jones algorithm to detect faces, and decides if each face belongs to an already tracked target or to a new one. When the face detector fails to detect a face, the tracker relies on colour histogram to keep tracking the target. Face obfuscation is also possibile.

ICVSS 2008, 2nd International Computer Vision Summer School, 14-19 July 2008, Scicli (Ragusa), Sicily.