Dynamic Partitioned Sampling for Tracking with Discriminative Features

Stefan Duffner[1]
stefan.duffner@idiap.ch
Jean-Marc Odobez[1]
jean-marc.odobez@idiap.ch
Elisa Ricci[2]
elricci@fbk.eu

[1] Idiap Research Institute
Martigny, CH

[2] Bruno Kessler Foundation - FBK-irst
Technologies of Vision Research Lab
Povo, Italy

Fusion of multiple complementary cues in the context of visual tracking has received considerable attention recently. Many works realize features integration with algorithms that dynamically lower the contribution of uncertain features and use only reliable cues [2, 4]. Cue reliability can be measured in many ways, *e.g*. by considering the spread of the particle distribution [2] or by quantifying the discrepancy between the tracking result obtained by the cue alone and by using jointly all the features [4]. Other approaches for multi-cue fusion use separate particle filters for each cue and model inter-filter dependencies explicitly with a graphical model [1]. Another series of works realize tracking with particle filters by dividing the state space into partitions, each one corresponding to a single cue, and sampling from them in a hierarchical manner according to a predefined order of cue relevance [3, 5].

This paper fits in the latter category. However, differently from previous approaches, the order of partitions is not fixed a priori but changes dynamically depending on the reliability of each cue, *i.e*. more reliable cues are sampled first. Therefore, we call our approach Dynamic Partitioned Sampling (DPS). The reliability of each cue is measured in terms of its ability to discriminate the foreground (FG) object with respect to the background (BG). Interestingly, this reliability measure is computed considering a dynamically adaptive model of BG: the BG is not described by a fixed model or by random patches but is represented by a set of informative particles which are also explicitly tracked in order to be as similar as possible to the object.

DPS. We define the state space $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_C)$ where C is the number of cues. For the i-th cue the likelihood $p(\mathbf{y}_{i,t}|\mathbf{x}_{i,t})$ is defined independently from the other cues. However, $\mathbf{x}_1..\mathbf{x}_C$ are not independent since they describe the same target. We describe their interactions introducing a MRF prior in the dynamical model, *i.e*. defining $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \prod_{c\in\mathscr{C}} p(\mathbf{x}_{c,t}|\mathbf{x}_{c,t-1}) \prod_{c_i,c_j\in\mathscr{C}} \phi(\mathbf{x}_{c_i,t}, \mathbf{x}_{c_j,t})$ where $p(\mathbf{x}_{c,t}|\mathbf{x}_{c,t-1})$ denotes the dynamics of the c-th cue, $\phi(\mathbf{x}_{c_i,t}, \mathbf{x}_{c_j,t})$ is the pairwise interaction potential between the pair of cues (c_i, c_j), and $\mathscr{C}$ denotes the set of cues.

Similarly to previous approaches [3, 5], we define a hierarchy between cues and we use the upstream models to restrict the search space of the downstream ones. However, in DPS the order of cues is not fixed but is determined at time t by their reliability $R_{c,t-1}$ (which we discuss below) estimated at time $t-1$, *i.e*. the most reliable cues are processed first. To this aim, we consider a proposal distribution that depends on this order, *i.e*. $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) = \prod_{c\in\mathscr{C}} q_c(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}, R_{c,t-1})$.

More specifically, in DPS for the most reliable cue r_1 the samples are simply drawn from its dynamical model ($q_{r_1}(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}, R_{r_1,t-1}) = p(\mathbf{x}_{r_1,t}|\mathbf{x}_{r_1,t-1})$) and weighted according to the likelihood $p(\mathbf{y}_{r_1,t}|\mathbf{x}_{r_1,t})$ as with a standard particle filter. On the other hand, for the subsequent cues r_i ($\{r_i : R_{r_i,t-1} < R_{r_{i-1},t-1}, r_i \in \mathscr{C}\}$), the samples are drawn from the proposal functions $q_{r_i}(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}, R_{r_i,t-1}) = p(\mathbf{x}_{r_i,t}|\mathbf{x}_{r_i,t-1})\phi(\mathbf{x}_{r_i,t}, \mathbf{x}_{r_{i-1},t})$, *i.e*. the product of the dynamics with the potential function w.r.t. the previous cue. As we define both $p(\mathbf{x}_{c,t}|\mathbf{x}_{c,t-1})$ and $\phi(\mathbf{x}_{c_i,t}, \mathbf{x}_{c_j,t})$ as Gaussian distributions, we can easily sample from their product. For downstream cues, the weighing is performed by their likelihoods $p(\mathbf{y}_{r_i,t}|\mathbf{x}_{r_i,t})$ and in case of the last cue r_C also by the evaluation of the potential $\phi(\mathbf{x}_{r_1,t}, \mathbf{x}_{r_C,t})$ modelling the interactions between the states of the first and the last cues. Fig. 1 illustrates DPS in the case of 3 cues. An alternative approach which adopts a *tree* representation of the hierarchy rather than a *chain* is also described in the paper.

Cue reliability. Together with the object tracking algorithm described above, we also define C auxiliary BG trackers. For cue c, a set of N_{BG} weighted BG particles $\{\mathbf{x}_{c,t}^{b,BG}, \omega_{c,t}^{b,BG}\}$ are initialised at random positions in the image region surrounding the tracked object. Then, the BG samples are updated dynamically using a simple filtering approach. First, the particles are weighted according to their likelihood $\omega_{c,t}^{b,BG} \alpha\ p(\mathbf{y}_{c,t}^{b,BG}|\mathbf{x}_{c,t}^{b,BG})$

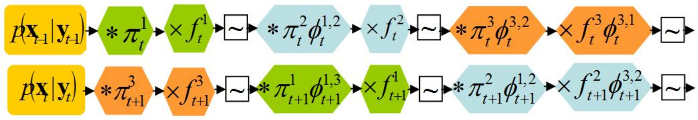

Figure 1: Dynamic Partitioned Sampling (different colours represent different partitions): the order of partitions for sampling is different at each time step. In the diagram, $\pi_t^c = p(\mathbf{x}_{c,t}|\mathbf{x}_{c,t})$ and $\phi_t^{c_i,c_j} = \phi(\mathbf{x}_{c_i,t}, \mathbf{x}_{c_j,t})$, $*$ denotes the convolution with the dynamical model, $\times$ the multiplication by the likelihood $f_t^c = p(\mathbf{y}_{c,t}|\mathbf{x}_{c,t})$, $\sim$ standard resampling.

which measures the similarity between the current BG particle and the observation which corresponds to the mean of the PDF $\bar{\mathbf{x}}_{c,t}$ estimated by the DPS tracker at current frame. In this way, the changes in appearance of the target are taken into account from the BG tracker. Then, the BG samples states are updated according to an appropriate dynamical model which excludes from the search space the target region estimated by the FG tracker. Finally, resampling eliminates uninformative BG samples.

This collaborative scheme of FG and BG trackers allows us to define the reliability:

$$R_{c,t} = \sum_{b=1}^{N_{BG}} w_b \delta \left[\log \frac{p(\mathbf{y}_{c,t}|\bar{\mathbf{x}}_{c,t})}{p(\mathbf{y}_{c,t}^{b,BG}|\mathbf{x}_{c,t}^{b,BG})} > T_c \right] \tag{1}$$

where δ is an indicator function and T_c is a user-defined threshold. If $T_c = 0$ and $w_b = 1$, $R_{c,t}$ counts how many times the observation corresponding to the current target estimate $\mathbf{y}_{c,t}(\bar{\mathbf{x}}_{c,t})$ is more similar to the FG template than to a BG sample. In practice, the weights w_b are set higher for BG samples that are spatially close to the object: the idea is that the closer a BG sample is, the more it is able to distract the FG tracker.

Head tracking. Despite the generality of the proposed approach, in this paper we demonstrate its effectiveness for the specific problem of head tracking with three different cues: colours, edge and contours. For colour, we use a multi-part representation concatenating histograms obtained by splitting the tracked region into subregions. Similarly, for texture we adopt multiple edge orientation histograms. For contour, the gradient information along the contour of a target is collected and compared with an ellipse as reference. Experimental results reported in the paper prove the robustness of our algorithm in many challenging video sequences with pose changes, partial occlusions, and moving camera. Fig. 2 shows the performance of DPS for one sequence.

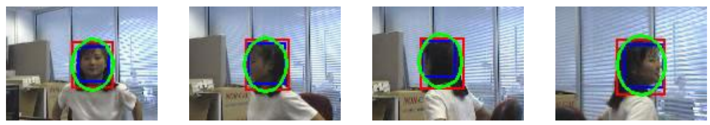

Figure 2: Results of head tracking with DPS. The circles and the rectangles are the estimates of the colour (blue), texture (red) and contour (green) cues.

[1] W. Du and J. H. Piater. A probabilistic approach to integrating multiple cues in visual tracking. In *Proc. ECCV*, pages 225–238, 2008.

[2] E. Maggio, F. Smeraldi, and A. Cavallaro. Adaptive multi-feature tracking in a particle filtering framework. *IEEE Trans. on Circuits and Systems for Video Technology*, 10(7):1348–1359, October 2007.

[3] P. Perez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. *Proceedings of the IEEE*, 92(3):495–513, 2004.

[4] C. Shen, A. van den Hengel, and A. Dick. Probabilistic multiple cue integration for particle filter based tracking. In *Proc. Digital Image Computing : Techniques and Applications*, pages 10–12, 2003.

[5] C. Yang, R. Duraiswami, and L. Davis. Fast multiple object tracking via a hierarchical particle filter. In *Proc ICCV*, pages 212–219, 2005.