A COLOR-BASED PARTICLE FILTER FOR JOINT DETECTION AND TRACKING OF MULTIPLE OBJECTS

Jacek Czyz[a], *Branko Ristic*[b] *and Benoît Macq*[a]

[(a)]UCL-TELE, Batiment Stévin, Place du Levant 2, 1348 Louvain-la-Neuve, Belgium
[(b)]DSTO, ISRD 200 labs, PO Box 1500, Edinburgh SA 5111, Australia

ABSTRACT

Recent works have shown that the particle filter using color as observation feature is a powerful technique for tracking deformable objects in image sequences with complex backgrounds. This paper presents a hybrid valued sequential state estimation algorithm, and its particle filter-based solution, that extends the standard color particle filter in two ways. Firstly, track initialization is embedded in the particle filter without relying on an external target detection algorithm. Secondly, the algorithm is able to track multiple objects sharing the same color description. We evaluate the performance of the proposed filter on various real-world video sequences with appearing and disappearing targets.

1. INTRODUCTION

Tracking moving objects in video sequences has many applications ranging from video surveillance to human-computer interfaces. Particle filters (PF) have proven to be very effective for solving tracking problems [1, 2]. In these approaches, tracking is modelled by a state-space time series estimation problem, which is solved using a sequential Monte Carlo estimation method [2].

Our work evolves from the adaptive color-based particle filter [3]. This color based tracker can efficiently and successfully handle non-rigid deformation of the target and rapidly changing dynamics in complex unknown background. However, it was designed for tracking a single object and uses an external mechanism to initialize the track. When several objects sharing the same color description are present in the scene, the color particle filter approach fails because particles are attracted by the different objects and the computed state estimates are meaningless. Thus using color histograms as object features, in this paper we develop a particle filter which *integrates* detection and tracking of multiple objects. The key feature in our approach is the augmentation of the state vector by discrete-valued variable which represents the number of existing objects in the video sequence. This random variable is incorporated into the state vector and modelled as an M-state Markov chain.

Mixed or hybrid valued (continuous-discrete) sequential state estimation, and its PF-based solution, has been successful in many video sequence analysis problems. In [4], the proposed tracking algorithm switches between different motion models depending on a discrete label, included in the state vector, which encodes which one of a discrete set of motion models is active. Black and Jepson proposed a mixed state-space approach to gesture/expression recognition [5]. First several models of temporal trajectories are trained. Next the models are matched against new unknown trajectories using a PF-based algorithm in which the state vector contains a label of the model that matches the observed trajectory. The Bayesian Multiple-Blob tracker [6], BraMBLe, manages multiple blob tracking by incorporating the number of objects present in the state vector. The multi-blob observation likelihood is based on filter bank responses which may come from the background image or one of the object models.

In contrast to BraMBLe, which requires background and foreground models, the method that we propose does not need a background estimation and can be used directly in camera moving sequences. Moreover, the use of color description leads to a small state vector size which allows the algorithm to track many similar objects (i.e. that have the same color description) given some computational resources. The tracker can detect objects entering or leaving the scene; it keeps an internal list of observable objects (than can vary from 0 to a predefined number) without the need of an external detection mechanism.

The paper is organized as follows. In the next section we formulate joint detection and tracking as a sequential estimation problem. The conceptual solution and its particle filter implementation are given is Section 3. Section 4 is devoted to experiments. Conclusions are given is the last section.

2. PROBLEM FORMULATION

The aim is to sequentially perform simultaneous detection and tracking of objects described by the same specified color histogram q^*, in a video sequence $\mathbf{Z}_k = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}$, where $\mathbf{z}_j$, is the image at discrete-time (sequence) index $j = 1, \ldots, k$. The time series state space approach requires to specify a motion model, i.e. the evolution of the state $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, and a measurement model, i.e. the link between the state and the current measurement $p(\mathbf{z}_k|\mathbf{x}_k)$. The next three subsections describe the modelling of object motion, of object appearance and disappearance, and the measurement likelihood function.

State vector and dynamic model

The state vector at frame k of a single object typically consists of kinematic and region (or shape) parameters, as in [3]. For simplicity we use the random walk model

$$\mathbf{x}_k = [x_k \quad y_k \quad H_x \quad H_y]^T, \tag{1}$$

where (x, y) denotes the center of the image region (in our case a rectangle) used for the color histogram computation, H_x and H_y denote the region width and height respectively. Note that other variables can be added, such as velocities and scale change rate, depending on the application. The state dynamics is typically described by a linear model: $\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}$, where $\mathbf{F}$ is the transitional matrix and $\mathbf{w}_{k-1}$ is process noise, assumed to be white, zero-mean Gaussian, with the covariance matrix $\mathbf{Q}$.

Number of existing objects

A discrete-valued random variable $E \in \mathbb{E} = \{0, 1, \ldots, M\}$ denotes the number of existing objects, with M being the maximum expected number. This random variable is modelled by an M-state Markov chain, whose transitions are specified an $(M+1) \times (M+1)$ transitional probability matrix (TPM) $\mathbf{\Pi} = [\pi_{ij}]$, where

$$\pi_{ij} = Pr\{E_k = j | E_{k-1} = i\}, \qquad (i, j \in \mathbb{E}) \tag{2}$$

is the probability of a transition from i objects existing at time $k-1$ to j objects at time k. The elements of the TPM satisfy $\sum_{j=0}^{M} \pi_{ij} = 1$ for each $i \in \mathbb{E}$. Random variable E is fully specified by the TPM and initial probabilities at time $k = 0$, i.e. $\mu_i = Pr\{E_0 = i\}$, for $i = 0, 1, \ldots, M$.

For the case of joint tracking and detection of a single object, i.e. $M = 1$, let P_b and P_d represent the probability of object "birth" (i.e. entering the scene) and "death" (leaving the scene). The corresponding TPM is given by:

$$\mathbf{\Pi} = \begin{bmatrix} (1 - P_b) & P_b \\ P_d & (1 - P_d) \end{bmatrix}.$$

Color measurement model

Following [7, 8, 3], we do not use the entire image $\mathbf{z}_k$ as the measurement, but rather extract from the image the color histogram q_k, computed inside the image region that is specified by the state vector $\mathbf{x}_k$. The center and the size of the rectangle, are defined by (x_k, y_k) and (H_x, H_y), respectively. Furthermore, we adopt the Gaussian density for the likelihood function of the measured color histogram as follows:

$$p(q_k|\mathbf{x}_k) \propto \mathcal{N}(D_k; 0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{D_k^2}{2\sigma^2}\} \tag{3}$$

where D_k is the distance between the reference histogram q^* of objects to be tracked and the histogram q_k computed from current frame $\mathbf{z}_k$ in the region specified by the state vector $\mathbf{x}_k$. The standard deviation σ of the Gaussian density in (3) is a design parameter.

If the two histograms are calculated over U bins, the distance D_k between two histograms is derived in [7] from the Bhattacharyya similarity coefficient and defined as: $D_k^2 = 1 - \sum_{u=1}^{U} \sqrt{q^*(u)\, q_k(u)}$.

3. BAYESIAN RECURSIVE SOLUTION

In this section we briefly outline the conceptual solution to integrated detection and tracking of multiple objects in the Bayesian framework, for the models described in the previous section. We then present the particle filter that implements this solution.

Conceptual recursive solution

We define an augmented state vector

$$\mathbf{y}_k = (E_k, \mathbf{x}_{1,k}^T, ..., \mathbf{x}_{M,k}^T)^T$$

where $\mathbf{x}_{m,k}$ is the state vector for object m. Given the posterior density $p(\mathbf{y}_{k-1}|\mathbf{Z}_{k-1})$, and the latest available image $\mathbf{z}_k$ in the video sequence, the goal is to construct the posterior density at time k, $p(\mathbf{y}_k|\mathbf{Z}_k)$. Once the posterior pdf $p(\mathbf{y}_k|\mathbf{Z}_k)$ is known, the probability P_m that there are m objects in a video sequence at time k is computed as the marginal of $p(\mathbf{y}_k|\mathbf{Z}_k)$, i.e.:

$$P_m = \int p(\mathbf{x}_{1,k}, ..., \mathbf{x}_{m,k}, E_k = m|\mathbf{Z}_k)\, d\mathbf{x}_{1,k} ... d\mathbf{x}_{m,k} \tag{4}$$

for $m = 1, \ldots, M$. The MAP estimate of the number of objects at time k is then determined as:

$$\hat{m}_k = \arg \max_{m=0,1,\ldots,M} P_m. \tag{5}$$

This estimate provides the means for automatic detection of new object appearance and the existing object disappearance. The posterior pdfs of state components corresponding to individual objects in the scene are then computed as the marginals of pdf $p(\mathbf{x}_{1,k}, \ldots, \mathbf{x}_{\hat{m},k}, E_k = \hat{m}|\mathbf{Z}_k)$.

The formal Bayesian recursive solution to the sequential hybrid estimation can be presented as a two step procedure consisting of prediction and update. For the prediction step, assuming that the object states (kinematics, size parameters) are mutually independent, the prediction pdf $p(\mathbf{x}_{1,k}, \ldots, \mathbf{x}_{m,k}, E_k = m|\mathbf{Z}_{k-1})$ can be expressed as a

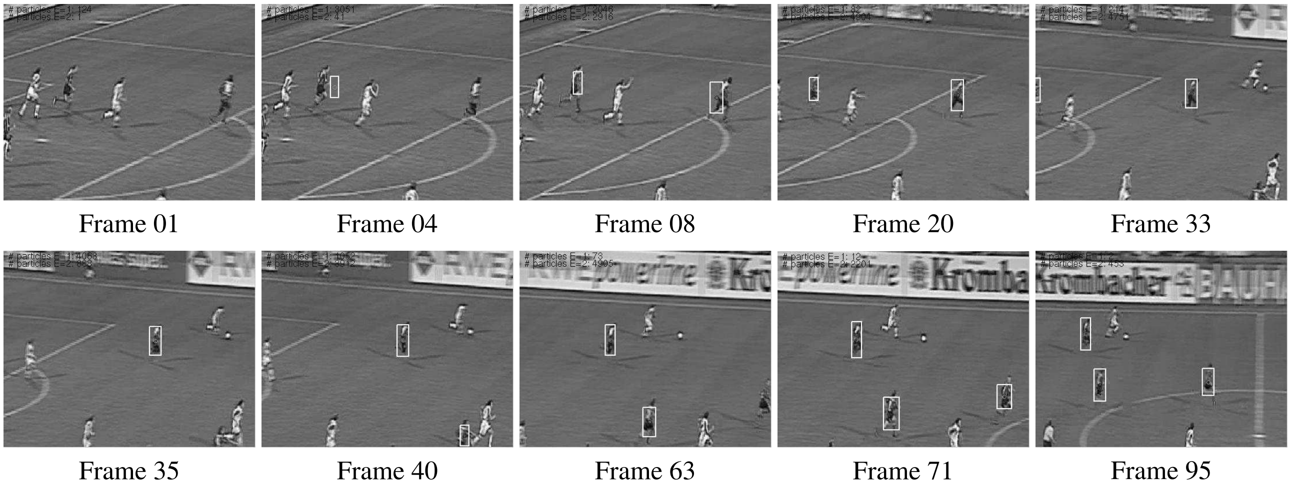

Fig. 1. Tracking results on a soccer sequence. The aim is to detect and track red players as they enter and leave the scene. Detected players are marked with a rectangle showing the region where the histogram is computed.

function of the transitional density $p(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})$ defined by the dynamic model, the elements of the transition probability matrix π_{ml} and an initial object pdf on its appearance $p_b(\mathbf{x}_{i,k})$ which is assumed to be known. For example we can assume the objects to appear in certain region (corridor or image edge) with a certain size.

The update step results from the application of the Bayes rule and allows to write the posterior pdf $p(\mathbf{y}_k|\mathbf{Z}_k)$ as a function of the prediction pdf (computed in the prediction step) and the observation density $p(\mathbf{z}_k|\mathbf{y}_k)$. As described in Section 2, we extract from image $\mathbf{z}_k$ color histograms and use them as the measurements. This means that the observation density can be expressed as

$$p(q_{1,k},\ldots,q_{m,k}|\mathbf{y}_k) = \prod_{i=1}^{m} p(q_{i,k}|\mathbf{x}_{i,k}), \tag{6}$$

where $q_{i,k}$ is a color histogram computed from $\mathbf{z}_k$ in the region specified by $\mathbf{x}_{i,k}$. Thus based on (3) we have

$$p(\mathbf{z}_k|\mathbf{y}_k) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m} D_{i,k}^2\} \tag{7}$$

where $D_{i,k}$ is the distance between i-th object color histogram and the reference color histogram.

Particle filter

We implement the conceptual solution described in the previous section in the form of a particle filter. Particle filters approximate the posterior density $p(\mathbf{y}_k|\mathbf{Z}_k)$ by a weighted set of random samples or particles. In our case, a particle of index n is characterized by a certain value of E_k^n variable and the corresponding number of state vectors $\mathbf{x}_{i,k}^n$ where $i = 1,\ldots,E_k^n$, i.e.

$$\mathbf{y}_k^n = \left[E_k^n, \mathbf{x}_{1,k}^n, \ldots, \mathbf{x}_{E_k^n,k}^n\right] \qquad (n = 1,\ldots,N)$$

where N is the number of particles. The pseudo-code of the main steps of this filter (single cycle) are presented in Table 1. The first step in the algorithm represents random transition of E_{k-1} to E_k^n based on the TPM $\mathbf{\Pi}$. Step 2.a of Table 1 follows from the update step in the previous section. If $E_{k-1}^n = E_k^n$, then we draw $\mathbf{x}_{i,k}^n$ from the transition prior $p(\mathbf{x}_k|\mathbf{x}_{i,k-1}^n)$ for $i = 1,\ldots,E_k^n$. If $E_{k-1}^n < E_k^n$, then for the objects that continue to exist we draw $\mathbf{x}_{i,k}^n$ using the transitional prior (as above), but for the newborn objects we draw particles from $p_b(\mathbf{x}_k)$. Finally if $E_{k-1}^n > E_k^n$, we select at random E_k^n objects from the possible E_{k-1}^n, with equal probability. The selected objects continue to exist and for them we draw particles using the transitional prior (as above) while the remaining objects are deleted. Step 2.b follows from the update step. Using (7) the unnormalized importance weights are computed as:

$$\tilde{w}_k^n = \begin{cases} 1, & \text{if } E_k^n = 0 \\ C_B \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{E_k^n}\left(D_{i,k}^n\right)^2\right\}, & \text{if } E_k^n > 0 \end{cases} \tag{8}$$

where C_B is a design parameter which takes into account the similarity between target and background histogram, and $D_{i,k}^n$ is the distance between the reference histogram q^* and $q_{i,k}^n(\mathbf{z}_k)$. In step 6, the number of objects is estimated based on (5), where P_m is computed in the PF as

$$P_m = \frac{1}{N}\sum_{n=1}^{N}\delta(E_k^n, m)$$

where $\delta(i,j)$ is the Kronecker delta. The estimate of the state vector of object $i = 1, \ldots, \hat{m}$ is then

$$\hat{\mathbf{x}}_{i,k|k} = \frac{1}{N_i} \sum_{n=1}^{N} \mathbf{x}_{i,k}^n \; \delta(E_k^n, i), \tag{9}$$

where $N_i = \sum_{n=1}^{N} \delta(E_k^n, i)$.

Table 1. Particle filter pseudo-code (single cycle)

$[\{\mathbf{y}_k^n\}_{n=1}^N]$ = PF$[\{\mathbf{y}_{k-1}^n\}_{n=1}^N, \mathbf{z}_k]$

1. Transitions of E_{k-1} variable (random transition of the number of existing objects):
   $[\{E_k^n\}_{n=1}^N]$ = ETrans $[\{E_{k-1}^n\}_{n=1}^N, \mathbf{\Pi}]$
2. FOR $n = 1 : N$
   - a. Based on (E_{k-1}^n, E_k^n) pair, draw at random $\mathbf{x}_{1,k}^n, \ldots, \mathbf{x}_{E_k^n,k}^n$;
   - b. Evaluate importance weight $\tilde{w}_k^n$ (up to a normalizing constant) using (8).
3. END FOR
4. Normalize importance weights
   - a. Calculate total weight: t = SUM $[\{\tilde{w}_k^n\}_{n=1}^N]$
   - b. FOR $n = 1 : N$
     - Normalize: $w_k^n = t^{-1}\tilde{w}_k^n$

     END FOR
5. Resample:
   $[\{\mathbf{y}_k^n, -, -\}_{n=1}^N]$ = RESAMPLE $[\{\mathbf{y}_k^n, w_k^n\}_{n=1}^N]$
6. Compute the output of the PF (for reporting purposes)

4. EXPERIMENTS

The color histograms were computed in the RGB color space using 8x8x8 bins as in [3]. The transition probability from m objects to $m\pm 1$ was set to 0.05. This ensures that more or less 5% of the particles are in a state with $E_k = m \pm 1$. For simplicity we did not allow transitions from m to $m \pm 2$ by setting this probability to zero (i.e. the TPM is tri-diagonal). This means that when two objects appear at the same time, the estimate of object number $\hat{m}$ will be incremented in two steps. The number of particles required by the filter depends on M, the maximum number of objects. We observed that with $M = 1$, 200 particles were enough to obtain accurate estimates. With $M = 3$, 5000 particles were necessary with 300x400 images. This number is also related to the prior knowledge on where objects are likely to appear and expressed by $p_b(\mathbf{x}_{i,k})$. In our case we have used a uniform density over the state vector variables for $p_b(\mathbf{x}_{i,k})$ which is equivalent to no prior knowledge.

Experiments where conducted on real world data sequences. Figure 1 shows a soccer sequence where the red players have to be tracked. At the beginning, there are two red players in the scene. At frame 4 the first player is detected ($\hat{m}$ switches to 1). At frame 8 the second player is detected ($\hat{m} = 2$). This shows the quick response of the filter. At frame 33, player 1 leaves the scenes. At the same time a third player is partially visible at the bottom of the image. At frame 35, $\hat{m}$ switches back to 1. Player 3 is detected at frame 40. Frame 63 shows a fourth player appearing. It is detected at frame 71 ($\hat{m} = 2$). Note that the camera moves rapidly on this sequence.

5. CONCLUSION

The paper presented an algorithm for detection and tracking of multiple similar objects. Joint detection and tracking is formulated as a hybrid sequential estimation problem with color histograms as measurement features. The problem is solved conceptually and implemented as a particle filter. Experimental results show that the proposed multiple object tracking method can successfully detect and follow objects as they enter and leave the image scene. This basic tracker can be improved in several ways. Color histograms can be computed in different regions of the target to take into account topological information [8]. For efficiency, many variants of the particle filter exist and could improve the basic scheme.

6. REFERENCES

[1] M. Isard and A. Blake, "Visual tracking by stochastic propagation of conditional density," in *Proc. European Conf. Computer Vision*, pp. 343–356. 1996.

[2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications*, Artech House, 2004.

[3] K. Nummiaro, E. Koller-Meier, and L. Van-Gool, "An adaptive color-based particle fi lter," *Image and Vision Computing*, vol. 21, pp. 99–110, 2003.

[4] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model-switching," in *Proc. Int. Conf. Computer Vision*, pp. 107–112. 1998.

[5] M. J. Black and A. D. Jepson, "Recognizing temporal trajectories using the condensation algorithm," in *Proc. of the 3rd Int. Conf. Automatic Face and Gesture Recognition*, pp. 16–21. 1998.

[6] M. Isard and J. MacCormick, "BraMBLe: a bayesian multiple blob tracker," in *Proc. Int. Conf. Computer Vision*, pp. 34–41. 2001.

[7] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. Comp. Vision Pattern Recog.*, pp. II:142–149. Hilton Head, SC, June 2000.

[8] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Proc. European Conf. Computer Vision (ECCV)*, A. Hayden et al., Ed., pp. 661–675. Springer-Verlag, 2002, LNCS 2350.