Visual tracking of multiple objects using a local particle filter

Mehdi Oulad Ameziane[1,2], Christelle Garnier[1], François Septier[1], Emmanuel Duflos[2]

[1]IMT Lille Douai, Univ. Lille, CNRS, UMR 9189 - CRIStAL, 59000 Lille, France

[2]Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, 59000 Lille, France

mehdi.ouladameziane@imt-lille-douai.fr, christelle.garnier@imt-lille-douai.fr, francois.septier@imt-lille-douai.fr, emmanuel.duflos@centralelille.fr

Résumé – Les filtres particulaires sont largement utilisés pour le suivi visuel d'objets multiples. Pour améliorer leur performance dans les espaces d'état de grande dimension, nous proposons d'utiliser un filtre particulaire local. L'idée est de partitionner l'espace d'état en plusieurs sous-espaces de dimension plus faible associés à un objet ou un groupe d'objets. L'algorithme de suivi proposé permet de se ramener à un problème mono-dimensionnel tout en modélisant les interactions entre objets.

Abstract – Particle filters (PFs) are currently widely used for visual tracking of multiple objects. In order to improve the performance of the PFs in high dimensional state spaces, we propose to use a local particle filter. The idea is to partition the large state space into smaller subspaces associated with an object or a group of objects. The proposed tracking algorithm allows to reduce the tracking to a one dimensional problem while considering the inter-object interactions.

1 Introduction

Multiple object tracking (MOT) is still a challenging task in computer vision. Handling multiple objects in visual tracking introduces additional complexity mainly due to the varying number of objects and the interactions between the objects. In past years, sequential Monte Carlo (SMC) methods have gained a great interest and have shown capabilities to address the MOT problem. These methods can be divided into two main categories according to the state space representation.

The first category defines a single object state space and runs several independent particle filters (PFs) in parallel, one for each object [1,2]. The dimension of the state space remains small, but no inter-object interaction can be modelled within the PFs. To account for interactions, a separate processing should be added at the expense of an extra cost. Using independent trackers also requires solving a data association problem to assign the observations to the objects [3].

The second family offers a more rigorous formulation of the problem based on a joint state space made from the concatenation of the multiple objects [4,5]. In this framework, the different objects can be labelled and the interactions can be explicitly taken into account. But it induces a considerable growth of the state space dimension. PFs require a number of particles that increases exponentially with the number of objects, so the computational cost quickly becomes too expensive. This problem is known as the curse of dimensionality.

Therefore alternative methods have been developed. Sequential Markov chain Monte Carlo (MCMC) methods are known to be more effective than PFs in high dimensional state spaces [6]. They have been much used for visual tracking of multiple objects [7, 8] since the initial work of Khan et al. [9], but the number of tracked objects is still limited. Another solution consists in using PFs with partitioned sampling (PS) [10] based on a decomposition of the state space into a partition. The algorithm successively performs sampling and resampling on each subspace. However this process leads to an impoverishment of the particles due to the numerous resampling procedures and the order in which the subspaces are explored has a strong impact on the performance. To limit this impact, dynamic and ranked PS have been proposed [11, 12].

In this paper, to overcome the dimension problem in the joint state space configuration, we propose to use a local PF which combines the interaction modelling and the partition of the large state space into separate subspaces of smaller dimension. The idea recently developed by Rebeschini et al. [13] is to exploit the fact that interactions are local. Unlike PS, each subspace is sampled independently and the treatment order does not matter. Experimental results demonstrate the benefits of the proposed method.

2 MOT problem formulation

This paper deals with MOT along a sequence of images. The aim is to estimate the joint state of N_o individual objects $X_t = \{x_t^j\}_{j=1}^{N_o}$ from a sequence of observations $y_{1:t} = (y_1, ..., y_t)$. In the Bayesian framework, the distribution of interest is the filtering density:

$$p(X_t|y_{1:t}) \propto p(y_t|X_t)\cdot\int p(X_t|X_{t-1})\cdot p(X_{t-1}|y_{1:t-1})\cdot dX_{t-1}$$

where the prior density $p(X_t|X_{t-1})$ represents the dynamic evolution of the state X_t and the observation likelihood $p(y_t|X_t)$ measures the matching of the observation y_t given the state X_t.

Each object j is represented by a rectangular bounding window with a fixed size. Then $x_t^j = \{c_t^j, v_t^j\}$ with $c_t = \{c_t^x, c_t^y\}$ the position of the top left corner and $v_t = \{v_t^x, v_t^y\}$ the velocity between two successive images.

Here we assume that the number of objects N_o is fixed, but the approach could be easily extended to deal with a time varying number of objects by using a random finite set or by fixing a maximum number of objects and associating to each object an existence variable [7–9, 14].

2.1 Dynamic model

The usual quasi constant velocity model is combined with an interaction model. As in [15, 16], the objects which are close to one another and move in the same direction tend to form a group and to adopt similar dynamics. Accordingly the objects interact inside a group and the groups are supposed to evolve independently.

Hence we consider that the N_o objects are divided into N_G^t independent groups $G_t = \{G_t^g\}_{g=1}^{N_G^t}$ with $\cup_{g=1}^{N_G^t} G_t^g = \{1 : N_o\}$. We denote $x_t(G_t^g) = \{x_t^j : j \in G_t^g\}$.

Each object j moves at a velocity equal to the average velocity of all the objects of the group G_{t-1}^g to which the object j belongs between times $t-1$ and t. Then the evolution of each object state $x_t^j = \{c_t^j, v_t^j\}$ is given by:

$$\begin{cases} c_t^j = c_{t-1}^j + v^{G_{t-1}^g} + \epsilon_c \\ v_t^j = v_{t-1}^j + \epsilon_v \end{cases} \tag{1}$$

where the group velocity $v^{G_{t-1}^g}$ is equal to:

$$v^{G_{t-1}^g} = \frac{1}{|G_{t-1}^g|}\cdot\sum_{k\in G_{t-1}^g} v_{t-1}^k \tag{2}$$

with $|G_{t-1}^g|$ the number of objects in the group G_{t-1}^g. The state noises ϵ_c and ϵ_v are independent white Gaussian noises with $\Sigma_c = diag(\sigma_c^2, \sigma_c^2)$ and $\Sigma_v = diag(\sigma_v^2, \sigma_v^2)$ the respective covariance matrices defining the uncertainty region around the previous states.

Since the group motions are independent, the joint prior density can be written as:

$$p(X_t|X_{t-1}) = \prod_{g=1}^{N_G^t} p\left(x_t(G_{t-1}^g)|x_{t-1}(G_{t-1}^g)\right) \tag{3}$$

Because the current state of an object only depends on the previous joint state, it can also be written as:

$$p(X_t|X_{t-1}) = \prod_{j=1}^{N_o} p\left(x_t^j|X_{t-1}\right) \tag{4}$$

In this work, we assume that the groups of objects are known, but they could be jointly evaluated with the state of the objects as in [15, 16].

2.2 Observation model

The observation model is based on the usual colour information. For each object j, a set of RGB histograms $h_t^j = h(y_t, x_t^j)$ is extracted from the image region $R(x_t^j)$ defined by the object state x_t^j. As in [17], $R(x_t^j)$ is divided into multiple subregions to take into account the colour spatial distribution. A histogram is then computed for each colour and each subregion.

The likelihood associated to an object j is defined from the Bhattacharyya distance D_B between the candidate histograms h_t^j and the reference histograms H_t^j for the 3 RGB channels and the S subregions of $R(x_t^j)$:

$$p_t^j(y_t|x_t^j) \propto \exp\left(-\lambda\sum_{p=1}^{3}\sum_{r=1}^{S} D_B^2\left(h_t^j(p,r), H_t^j(p,r)\right)\right) \tag{5}$$

where λ is a tuning parameter that determines how peaked the likelihood is.

Then the joint likelihood results from the contribution of the N_o individual objects:

$$p(y_t|X_t) = \prod_{j=1}^{N_o} p_t^j(y_t|x_t^j) \tag{6}$$

3 Local particle filters

3.1 Principle of local particle filters

The initial idea is that in high dimensional filtering models, a decay of correlation is generally observed between the regions of the state space which are distant enough from one another. This leads to a locally low dimensional model. In [13], Rebeschini and al. propose to exploit this property to design local particle filters.

The principle is to partition the state space into separate subspaces of smaller dimension, also called blocks, under the following assumptions. We consider a HMM $(x_{1:t}, y_{1:t})_{t\geq 0}$ such that, at each time t, the state x_t with dimension d can be divided into N_B^t independent and non overlapping subsets or blocks $\{B_t^j\}_{j=1}^{N_B^t}$. These blocks verify $\cup_{j=1}^{N_B^t} B_t^j = \{1 : d\}$ and $B_t^j \cap B_t^{j'} = \emptyset\ \forall j, j' \in \{1 : N_B^t\}$ with $j \neq j'$. The HMM is assumed to satisfy the following factorization [6]:

$$p(y_t|x_t)\cdot p(x_t|x_{t-1}) = \prod_{j=1}^{N_B^t} f_t^j\left(y_t, x_{t-1}, x_t(B_t^j)\right) \tag{7}$$

for appropriate functions $f_t^j(.)$ and with $x_t(B_t^j)$ the set of the state components belonging to the subset B_t^j, thus $x_t(B_t^j) = \{x_t^j : j \in B_t^j\}$.

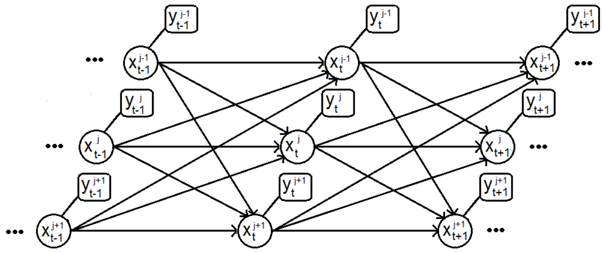

FIG. 1: Dependence graph of a HMM satisfying the factorization (7).

The dependence graph is illustrated in Figure 1 in the simplest case for $|B_t^j| = 1$, $\forall j, t$, that means $x_t(B_t^j) = x_t^j$. The dynamics of the state x_t is local in the sense that the components are mutually independent and only depend on the components of the previous state x_{t-1}: $p(x_t|x_{t-1}) = \prod_{j=1}^{N_B^t} p_t^j(x_t^j|x_{t-1})$. Similarly, the observations are also local: $p(y_t|x_t) = \prod_{j=1}^{N_B^t} p_t^j(y_t^j|x_t^j)$.

Then by running a PF on each non overlapping subset, the local PF approximates the filtering distribution as a product of marginal distributions on the N_B^t subsets:

$$p(x_t|y_{1:t}) \approx \bigotimes_{j=1}^{N_B^t} p\left(x_t(B_t^j)|y_{1:t}\right) \tag{8}$$

This strategy introduces some bias in the Monte Carlo estimation, because the approximation (8) does not converge to the exact filtering distribution as the number of particles tends to infinity. Nevertheless, the variance reduction due to the small dimension of the subsets is significant compared to the small amount of bias which is introduced in the algorithm.

3.2 Local PF for visual tracking of multiple objects

According to the MOT model described by equations (4) and (6), the product of the prior density and the likelihood can be factorized as in equation (7) with $|B_t^j| = 1$, $\forall j, t$, that is $x_t(B_t^j) = x_t^j$, and $N_B^t = N_o$.

Thus the local PF consists in running a PF for each object, while considering the dynamics dependence within each group of objects in the sampling step. Consequently the dimension of the estimation problem is divided by the number of objects.

The algorithm of the local PF obtained by using the prior density as the importance function is summarized in Table 1.

Initialisation ($t = 0$) sample $\{X_0^{(i)}\}_{i=1}^{N_p} \sim p(X_0)$ initialise $\{w_0^{j,(i)}\}_{i=1}^{N_p} = \frac{1}{N_p}$, $\forall j = 1 : N_o$ initialise the object groups $\{G_0^g\}_{g=1}^{N_G^0}$ initialise the group velocity $v^{G_0^g,(i)} = \frac{1}{\lvert G_0^g\rvert} \cdot \sum_{k \in G_0^g} v_0^{k,(i)}, \forall g = 1 : N_G^t, i = 1 : N_p$

Sequential processing ($t > 0$)
for $j = 1 : N_o$
  for $i = 1 : N_p$ do
    sample $c_t^{j,(i)} \sim p(c_t^j \lvert c_{t-1}^{j,(i)}, v^{G_{t-1}^g,(i)})$
    sample $v_t^{j,(i)} \sim p(v_t^j \lvert v_{t-1}^{j,(i)})$
    evaluate $w_t^{j,(i)} = p_t^j(y_t \lvert x_t^{j,(i)})$
  end for
  normalise the importance weights $w_t^{j,(i)}$
end for
estimate $\hat{X}_t = \bigotimes_{j=1}^{N_o} \sum_{i=1}^{N_p} w_t^{j,(i)} \cdot x_t^{j,(i)}$
resample $\{x_t^{j,(i)}, w_t^{j,(i)}\}_{i=1}^{N_p}, \forall j = 1 : N_o$
update the object groups $\{G_t^g\}_{g=1}^{N_G^t}$
update the group velocity
$v^{G_t^g,(i)} = \frac{1}{\lvert G_t^g\rvert} \cdot \sum_{k \in G_t^g} v_t^{k,(i)}, \forall g = 1 : N_G^t, i = 1 : N_p$

TAB. 1: Local particle filter algorithm for visual tracking.

4 Experimental results

To show the relevance of the local PF against the curse of dimensionality, several simulations have been conducted on four synthetic image sequences, each with 100 images and a size 500×500. In the sequences, each object is represented by a rectangular patch with a specific color. The sequences S1, S2 and S3 contain 4 objects forming three separate groups of size (2,1,1) and the sequence S4 contains 6 objects forming three separate groups of size (4,1,1). The number of groups and their composition are known and do not change in time.

Three trackers are considered for this experiment : the **joint PF** based tracker, which defines a joint state space model, the tracker based on **multiple independent PFs** per object group, which defines a state space per group and the **local PF**, which considers each object as a subset of the joint state space model.

All the trackers use the prior density as the importance function. The number of particles N_p is respectively 20, 20 and 50 for the local, independent and joint PFs.

Table 2 shows the performance results expressed as a F-measure averaged over the whole sequences and 100 simulations. The local PF achieves the best results in all the scenarios. Its superiority is due to its capability to handle interactions between objects while reducing the state space of dimension N_o to a one dimensional state space. The multiple independent PFs per object group have also good performance, but lower than the local PF. These performances are explained by switching from the state space of dimension N_o to a state space per group with smaller dimension. Finally the joint PF has the lowest performances as it applies no dimension reduction and directly deals with the high dimensional state space.

	Joint PF	Independent PFs per group	Local PF
S1	78.54	90.89	93.10
S2	76.24	89.24	92.60
S3	77.68	90.38	92.94
S4	74.90	86.18	93.87

TAB. 2: Average F-measure for the synthetic sequences.

If we look closer at the performance difference between the three trackers, we observe that for sequences S1, S2, S3, the performance obtained with the independent PFs per object group is slightly below the performance of the local PF. For the sequence S4, the F-measure difference between the independent PFs and the local PF becomes more important and the results of the independent PFs get closer to the performance of the joint PF. This is due to the increase of the number of objects within one group in sequence S4. The dimension of the subspace corresponding to the group of four objects is significant and the independent PFs per group suffer from the same dimension problem as the joint PF. The major limitation of the tracker based on the independent PFs is that the performance depends on the size of the object groups. This limitation is overcome by the local PF, which accounts for the interactions between the objects in a simple framework which is affected neither by the number of objects in the scene nor the size of the object groups.

5 Conclusion

In this paper, we propose to use the local PF for visual tracking of multiple objects. In our implementation, the local PF partitions the large state space into one dimensional subspaces associated to one object while considering the interactions between the different objects belonging to a same group. Consequently the dimension of the MOT problem is divided by the number of objects. Experimental results on synthetic videos show that the local PF is a promising solution to make PFs more effective in high dimensional applications such as MOT. These results have to be confirmed on real image sequences.

References

[1] O. Lanz, "Approximate bayesian multibody tracking," *IEEE trans. on PAMI*, vol. 28, pp. 1436–1449, 2006.

[2] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *IEEE trans. on PAMI*, vol. 33, pp. 1820–1833, 2011.

[3] K. C. Smith, *Bayesian methods for visual multi-object tracking with applications to human activity recognition.* PhD thesis, EPFL, 2007.

[4] A. Doucet, B.-N. Vo, C. Andrieu, and M. Davy, "Particle filtering for multi-target tracking and sensor management," in *FUSION*, pp. 474–481, IEEE, 2002.

[5] J. Vermaak, S. J. Godsill, and P. Perez, "Monte Carlo filtering for multi target tracking and data association," *IEEE trans. on Aerospace and Electronic systems*, vol. 41, pp. 309–332, 2005.

[6] F. Septier and G. W. Peters, "An Overview of Recent Advances in Monte-Carlo Methods for Bayesian Filtering in High-Dimensional Spaces," in *Theoretical Aspects of Spatial-Temporal Modeling*, Springer, 2015.

[7] J. Yao and J.-M. Odobez, "Multi-person Bayesian tracking with multiple cameras," *Multi-Camera Networks*, pp. 363–388, 2009.

[8] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," *IEEE Transaction on PAMI*, vol. 30, pp. 1198–1211, 2008.

[9] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," *IEEE trans. on PAMI*, vol. 27, pp. 1805–1819, 2005.

[10] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *International Journal of Computer Vision*, vol. 39, pp. 57–71, 2000.

[11] K. C. Smith, "Order matters: a distributed sampling method for multi-object tracking," tech. rep., IDIAP, 2004.

[12] N. Widynski, *Intégration d'informations spatiales floues dans un filtre particulaire pour le suivi mono-et multi-objets dans des séquences d'images 2D.* PhD thesis, Université Paris 6, 2010.

[13] P. Rebeschini, R. Van Handel, *et al.*, "Can local particle filters beat the curse of dimensionality?," *The Annals of Applied Probability*, vol. 25, pp. 2809–2866, 2015.

[14] F. Septier, S. K. Pang, A. Carmi, and S. Godsill, "On MCMC-based particle methods for Bayesian filtering: Application to multitarget tracking," in *CAMSAP*, pp. 360–363, IEEE, 2009.

[15] A. Gning, L. Mihaylova, S. Maskell, S. K. Pang, and S. Godsill, "Group object structure and state estimation with evolving networks and Monte Carlo methods," *IEEE trans. on Signal Processing*, vol. 59, pp. 1383–1396, 2011.

[16] L. Bazzani, M. Zanotto, M. Cristani, and V. Murino, "Joint individual-group modeling for tracking," *IEEE trans. on PAMI*, vol. 37, pp. 746–759, 2015.

[17] D. N. T. Cong, F. Septier, C. Garnier, L. Khoudour, and Y. Delignon, "Robust visual tracking via MCMC-based particle filtering," in *ICASSP*, pp. 1493–1496, IEEE, 2012.