Automatica 69 (2016) 380–394

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Autonomous crowds tracking with box particle filtering and convolution particle filtering☆

Allan De Freitas [a], Lyudmila Mihaylova [a], Amadou Gning [b], Donka Angelova [c], Visakan Kadirkamanathan [a]

[a] Department of Automatic Control and Systems Engineering, University of Sheffield, United Kingdom

[b] Department of Computer Science, University of Hull, United Kingdom

[c] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Bulgaria

ARTICLE INFO

Article history:
Received 23 January 2015
Received in revised form
11 January 2016
Accepted 26 February 2016
Available online 25 March 2016

Keywords:
Box particle filter
Convolution particle filter
Crowd tracking

ABSTRACT

Autonomous systems such as Unmanned Aerial Vehicles (UAVs) need to be able to recognise and track crowds of people, e.g. for rescuing and surveillance purposes. Large groups generate multiple measurements with uncertain origin. Additionally, often the sensor noise characteristics are unknown but measurements are bounded within certain intervals. In this work we propose two solutions to the crowds tracking problem— with a box particle filtering approach and with a convolution particle filtering approach. The developed filters can cope with the measurement origin uncertainty in an elegant way, i.e. resolve the data association problem. For the box particle filter (PF) we derive a theoretical expression of the generalised likelihood function in the presence of clutter. An adaptive convolution particle filter (CPF) is also developed and the performance of the two filters is compared with the standard sequential importance resampling (SIR) PF. The pros and cons of the two filters are illustrated over a realistic scenario (representing a crowd motion in a stadium) for a large crowd of pedestrians. Accurate estimation results are achieved.

1. Introduction

Tracking a large number of objects requires scalable algorithms that are able to deal with large volumes of data characterised by the presence of clutter. Although groups are made up of many individual entities, they typically maintain certain patterns of motion, such as in the case of crowds of pedestrians (Ali & Dailey, 2009). When the number of objects in the group is huge, e.g. hundreds and thousands, it is impractical (and impossible) to track them all individually. Instead of tracking each separate component, the group can be considered as one whole entity. Large group techniques identify and track concentrations, typically the kinematic states of the group and its extent parameters (Koch, 2008).

Recent results for the modelling, simulating and visual analysis of crowds are presented in Ali, Nishino, Manocha, and Shah (2014) from the point of view of computer vision, transportation systems and surveillance. The social force model (Ali et al., 2014; Helbing & Molnár, 1995; Mazzon & Cavallaro, 2013) has been used to model behaviour of pedestrians, including evacuation of people through bottlenecks. The social force model has also been combined with some filtering techniques for multiple-target tracking in Pellegrini, Ess, Schindler, and Van Gool (2009).

There is a wealth of approaches that are developed to track kinematic states of large crowds (e.g. the centre of the crowds) and their size (extent parameters). A recent survey (Mihaylova et al., 2014) presents key trends in the area. Although, the problem of tracking large groups has received attention in the literature, it is far from being resolved due to the various challenges that are present. Some of these challenges involve difficulties in modelling the interactions between the entities of the crowd, data association and dynamic shape changes of the crowd. Some of the approaches that have been proposed include mixtures of Gaussian components (Carmi, Septier, & Godsill, 2012) and a wealth of Random finite sets (RFS) methods, e.g., Grandström (2012), Granström, Lundquist, and Orguner (2011), Mahler (2007, 2009, 2013) and Mahler and Zajic (2002).

☆ The material in this paper was presented at the 15th International conference in Information Fusion, July 9–12, 2012, Singapore and the 17th International Conference in Information Fusion, July 7–10, 2014, Salamanca, Spain. This paper was recommended for publication in revised form by Associate Editor Huijun Gao under the direction of Editor Ian R. Petersen.

E-mail addresses: a.defreitas@sheffield.ac.uk (A. De Freitas), l.s.mihaylova@sheffield.ac.uk (L. Mihaylova), E.Gning@hull.ac.uk (A. Gning), donka@bas.bg (D. Angelova), visakan@sheffield.ac.uk (V. Kadirkamanathan).

http://dx.doi.org/10.1016/j.automatica.2016.03.009

This work proposes two novel solutions to the crowd tracking problem based on the recently developed box particle filter (PF) (Gning, Ristic, Mihaylova, & Abdallah, 2013) and convolution particle filter (CPF) (Angelova, Mihaylova, Petrov, & Gning, 2013; Campillo & Rossi, 2009; Rossi & Vila, 2006) frameworks.

The box PF (Abdallah, Gning, & Bonnifait, 2008; Gning, Mihaylova, Abdallah, & Ristic, 2012; Gning et al., 2013) relies on the concept of a box particle, which occupies a small and controllable rectangular region having a non-zero volume in the state space. The box PF affords to resolve the data association problems arising from the multiple measurements originating from the crowd. This is a common case when a UAV is flying over a region and collects data, seeing the area from above.

This paper has several novel contributions when compared with our previous works such as Petrov, Gning, Mihaylova, and Angelova (2012), Petrov, Ulmke et al. (2012) and Petrov, Mihaylova, de Freitas, and Gning (2014). These novelties include: (i) a generalised likelihood function for the box PF is derived when the state vector consists of kinematic states and extent parameters; (ii) the likelihood of the Box PF is calculated based on optimisation, by solving a constraint satisfaction problem (CSP) with multiple measurements; (iii) the online estimation of the crowd and clutter measurement rates; (iv) an adaptive CPF is proposed. The developed CPF is able to deal with multiple measurements, including a high level of clutter. It is able to resolve the data association problem without the crowd and clutter measurement rates. The adaptive CPF can estimate both dynamic kinematic states and dynamic parameters which is a different solution from the CPF based approach for static parameters presented in Angelova et al. (2013), Campillo and Rossi (2009) and Rossi and Vila (2006).

Both filters have very appealing properties in solving nonlinear estimation problems. Both filters operate in the condition of uncertain and imperfect observations: fluctuating number of sensor reports.

The performance of the box PF and CPF is evaluated for two different cases. Firstly, in a fully matched case where the models used by the filter directly match that used by the simulator, and secondly, in an unmatched scenario of a realistic crowd moving through a bottleneck. Both filters are compared with the standard sequential importance resampling PF (SIR PF) (Doucet, De Freitas, & Gordon, 2001) in terms of filter accuracy and computational complexity.

A main advantage of the box PF consists in its robustness to measurement characteristics and its ability to be implemented efficiently in a distributed way. The CPF is based on the principles of kernel based learning and can deal with problems where the likelihood is not available in an analytical form or it is difficult to calculate.

The rest of this paper is organised in the following way. Section 2 describes the state space modelling of a crowd. Section 3 is a brief overview of inference in a Bayesian framework. Section 4 presents the adaptation of the box PF for group object tracking. Section 5 introduces the CPF for crowd tracking, which is followed by a performance evaluation of the presented approaches in Section 6. Finally, conclusions are presented in Section 7.

2. State space modelling of a crowd

The characteristics of the crowd and scene that are required to be inferred at each time step k, $k = 1, 2, \ldots, K$, are represented by an augmented state vector:

$$\boldsymbol{\zeta}_k = \left(\boldsymbol{\lambda}_k^T, \boldsymbol{X}_k^T, \boldsymbol{\Theta}_k^T\right)^T, \tag{1}$$

where $\boldsymbol{X}_k$ is the kinematic vector of the centre of the crowd, and $\boldsymbol{\Theta}_k$ is the parameter vector which characterises the crowd extent. Multiple measurements are received from the crowd and from clutter at each time step, thus the state vector includes $\boldsymbol{\lambda}_k$ which is the measurement rate vector. The notation $(\cdot)^T$ is the transpose operator. In this paper we consider the two-dimensional case, where the kinematic vector consists of the position coordinates and the velocity of the centre of the crowd and the extent of the crowd is represented by a rectangle. The resulting kinematic vector has the following form:

$$\boldsymbol{X}_k = (x_k, \dot{x}_k, y_k, \dot{y}_k)^T \tag{2}$$

and the parameter vector is given by:

$$\boldsymbol{\Theta}_k = (a_k, b_k)^T \tag{3}$$

where a_k and b_k represent the lengths of the sides of the rectangle in the x and y dimensions, respectively. The measurement rate vector is represented by:

$$\boldsymbol{\lambda}_k = (\lambda_{T,k}, \lambda_{C,k})^T, \tag{4}$$

where $\lambda_{T,k}$ and $\lambda_{C,k}$ represent the crowd and clutter measurement rates, respectively.

2.1. Crowd dynamics model

The motion of the centre of the crowd is modelled by a correlated velocity model. The correlated velocity model is related to the Singer model (Singer, 1970) and jerk model (Mehrotra & Mahapatra, 1997) with the difference being that the velocity component is correlated in time and that the second and other higher order derivatives of position are negligible. The evolution model for the kinematic state of the target is represented mathematically by

$$\boldsymbol{X}_k = \boldsymbol{A}\boldsymbol{X}_{k-1} + \boldsymbol{\eta}_k, \tag{5}$$

where $\boldsymbol{\eta}_k$ represents the system dynamics noise. The state transition matrix is given by

$$\boldsymbol{A} = \begin{bmatrix} 1 & \frac{1}{\alpha}\left(1 - e^{-\alpha T_s}\right) \\ 0 & e^{-\alpha T_s} \end{bmatrix} \otimes \mathbf{I}_2 \tag{6}$$

where T_s is the sampling interval, $\otimes$ denotes the Kronecker product, $\mathbf{I}_2$ denotes the 2×2 identity matrix, and α is the reciprocal of the velocity correlation time constant. The covariance of the system dynamics noise $\boldsymbol{\eta}_k$ can be modelled as

$$\boldsymbol{Q} = 2\alpha\sigma_v^2 \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \otimes \mathbf{I}_2, \tag{7}$$

where σ_v^2 is the variance of the velocity of the crowd centroid for a single dimension and

$$\begin{aligned} q_{11} &= \frac{1}{2\alpha^3}\left(4e^{-\alpha T_s} - 3 - e^{-2\alpha T_s} + 2\alpha T_s\right), \\ q_{12} &= \frac{1}{2\alpha^2}\left(e^{-2\alpha T_s} + 1 - 2e^{-\alpha T_s}\right), \\ q_{22} &= \frac{1}{2\alpha}\left(1 - e^{-2\alpha T_s}\right). \end{aligned} \tag{8}$$

The evolution for the crowd extent is assumed to be a random walk model, described by

$$\boldsymbol{\Theta}_k = \boldsymbol{\Theta}_{k-1} + \boldsymbol{\eta}_{p,k}, \tag{9}$$

where the parameter noise $\boldsymbol{\eta}_{p,k}$ is characterised by the standard deviation $\boldsymbol{\sigma}_\theta \in \mathbb{R}^{n_\Theta}$.

2.2. Observation model

In this paper we consider the scenario where the measurements originate from within a confined area. However, other scenarios, such as the case where the measurements only come from the border of the crowd, have a similar solution.

The total number of measurements M_k, obtained at each time step from the sensor consists of the $M_{T,k}$ number of measurements, originating from the crowd and $M_{C,k}$ clutter measurements, i.e. $M_k = M_{T,k} + M_{C,k}$. The number of measurements $M_{T,k}$ originating from the crowd is considered as a Poisson-distributed random variable with mean value of the crowd rate, $\lambda_{T,k}$, i.e., $M_{T,k} \sim Poisson(\lambda_{T,k})$. Similarly, the number of clutter measurements is $M_{C,k} \sim Poisson(\lambda_{C,k})$. The $M_{T,k}$ measurements originating from the crowd are uniformly located in the area represented by the crowd. The $M_{C,k}$ clutter measurements are uniformly located in the region about the crowd.

Typically in point target tracking, an observation model which directly relates the states to the measurements is available, in the form given by:

$$\boldsymbol{z}_k = h(\boldsymbol{\zeta}_k) + \boldsymbol{\xi}_k, \tag{10}$$

where $\boldsymbol{\xi}_k$ represents the observation noise. However, since the crowd is an extended target,[1] there is no direct observation model. The observations can be indirectly related to the states through the sensor characteristics and the target model.

The sensor characteristics describe the relationship between the measurement point m, $m = 1, \ldots, M_k$ and the measurement source in a Cartesian coordinate system and is of the form:

$$\boldsymbol{z}_k^m = \tilde{h}(\boldsymbol{x}_k^m) + \boldsymbol{\xi}_k, \tag{11}$$

where $\tilde{h}(\cdot)$ is the measurement function and $\boldsymbol{x}_k^m = (x_k^m, y_k^m)^T$ denotes the Cartesian coordinates of the measurement source in a two dimensional space. In this paper we consider the following model:

$$\boldsymbol{z}_k^m = \boldsymbol{H}\boldsymbol{x}_k^m + \boldsymbol{\xi}_k, \tag{12}$$

where $\boldsymbol{H} = \mathbf{I}_2$, and the measurement noise $\boldsymbol{\xi}_k = (\xi_{1,k}, \xi_{2,k})^T$, is assumed (but not restricted) to be Gaussian, with a known covariance matrix $\boldsymbol{R} = diag(\sigma_1^2, \sigma_2^2)$. The vector of interval measurements is $[\boldsymbol{z}_k^m] = ([z_{1,k}^m], [z_{2,k}^m])^T$, where $[z_{1,k}^m]$ and $[z_{2,k}^m]$ are the intervals of the mth measurement point. One way to describe these components is by representing the noise terms in Eq. (11) as intervals:

$$[\boldsymbol{\xi}_k] = [-3\sigma, +3\sigma]. \tag{13}$$

At each time step k, the M_k interval measurements are combined into an interval matrix $[\boldsymbol{Z}_k] = \{[\boldsymbol{z}_k^1], \ldots, [\boldsymbol{z}_k^m]\} \in \mathbb{R}^{n_z \times M_k}$.

Each measurement originates from either random clutter or the crowd but its origin is unknown. The target model describes the relationship between the states and the measurement sources for the $M_{T,k}$ measurements that originate from the crowd. As previously described, the measurement sources are uniformly distributed across the region which exhibits measurements, and this region is represented by the states through the following probability density:

$$p(\boldsymbol{x}_k^m|\boldsymbol{x}_k) = U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m), \tag{14}$$

where $U_{[\mathbf{x}]}(\cdot)$ denotes the multivariate uniform probability density function (pdf) with the interval $[\mathbf{x}]$ as support. The support of the uniform distribution describes two independent regions which cover the area of the rectangle used to approximate the extent of the crowd:

$$q(\boldsymbol{x}_k) = \begin{cases} x_k - \dfrac{a_k}{2} \le x_k^m \le x_k + \dfrac{a_k}{2}, \\ y_k - \dfrac{b_k}{2} \le y_k^m \le y_k + \dfrac{b_k}{2}. \end{cases} \tag{15}$$

3. Inference in a Bayesian framework

Classic Bayesian inference relies on computing the posterior distribution from a prior distribution and measurements. The posterior distribution can be updated sequentially based on a prediction step, followed by an update step. The following equation describes the prediction:

$$p(\boldsymbol{\zeta}_k|\boldsymbol{Z}_{1:k-1}) = \int_{\mathbb{R}^{n_\zeta}} p(\boldsymbol{\zeta}_k|\boldsymbol{\zeta}_{k-1}) p(\boldsymbol{\zeta}_{k-1}|\boldsymbol{Z}_{1:k-1}) d\boldsymbol{\zeta}_{k-1}. \tag{16}$$

The measurement update is described by the following equation:

$$p(\boldsymbol{\zeta}_k|\boldsymbol{Z}_{1:k}) = \frac{p(\boldsymbol{Z}_k|\boldsymbol{\zeta}_k) p(\boldsymbol{\zeta}_k|\boldsymbol{Z}_{1:k-1})}{p(\boldsymbol{Z}_k|\boldsymbol{Z}_{1:k-1})}. \tag{17}$$

The recursive relationship of Eqs. (16) and (17) forms the optimal Bayesian solution. Utilising these equations for Bayesian filtering is generally not possible since an analytical solution rarely exists. A solution for when the state space model is linear and perturbed by Gaussian noise is referred to as the Kalman filter (Bar-Shalom, Li, & Kirubarajan, 2001). Several techniques have been used in the more general case consisting of non-linearities and non-Gaussianity in the state space model, such as the extended Kalman filter (Bar-Shalom et al., 2001), unscented Kalman filter (Wan & Van Der Merwe, 2000) and particle filter based techniques (Cappe, Godsill, & Moulines, 2007) to name a few.

For further notational convenience, the marginal state is defined as follows:

$$\boldsymbol{x}_k = \left(\boldsymbol{X}_k^T, \boldsymbol{\Theta}_k^T\right)^T. \tag{18}$$

In this application the posterior distribution can be further factored into the following form:

$$p(\boldsymbol{\zeta}_k|\boldsymbol{Z}_{1:k}) = p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}, \boldsymbol{\lambda}_k) p(\lambda_{T,k}|\boldsymbol{Z}_{1:k}) p(\lambda_{C,k}|\boldsymbol{Z}_{1:k}). \tag{19}$$

This factorisation implicitly states that the crowd and clutter measurement rates are independent of the kinematics and extent of the crowd. This is true for the clutter measurement rate but not necessarily valid for the crowd measurement rate. However, the variance of the prior distribution for the crowd rate is sufficient to represent the variation of the number of measurements over time.

It has been shown that a closed form recursive Bayesian solution exists for the estimation of the mean of a Poisson distribution, based on using the conjugate prior Gamma distribution (Granström & Orguner, 2012). The crowd and clutter measurement rates are estimated based on this concept,[2] and the focus of this paper thus lies on the calculation of the marginal posterior distribution for the states representing the kinematics and extent of the crowd, $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}, \boldsymbol{\lambda}_k)$, using the novel box particle filter and convolution particle filter algorithms.

4. The box particle filter for crowd tracking

This section begins with a review of the box PF in point target tracking without clutter, thus, the following subsection does not consider the extent of the target, i.e. $\boldsymbol{x}_k = \boldsymbol{X}_k$.

[1] An *extended target* cannot be considered as a point, but instead it has a physical extent characterising its size and volume.

[2] Refer to Appendix B for more information on crowd and clutter measurement rate estimation.

4.1. The classic box particle filter

The concept of a *box particle* is introduced where a box particle represents a small region with controllable size (or volume). The box PF approximates the posterior state pdf with a mixture of uniform pdfs (Gning, Mihaylova, & Abdallah, 2010; Gning, Mihaylova et al., 2012), i.e.

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1}) \approx \sum_{p=1}^{N} w_{k-1}^{(p)} U_{[\boldsymbol{x}_{k-1}^{(p)}]}(\boldsymbol{x}_{k-1}). \tag{20}$$

For the box PF, the time update can be written as:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) &\approx \int_{\mathbb{R}^{n_x}} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \sum_{p=1}^{N} w_{k-1}^{(p)} U_{[\boldsymbol{x}_{k-1}^{(p)}]}(\boldsymbol{x}_{k-1}) d\boldsymbol{x}_{k-1} \\ &= \sum_{p=1}^{N} w_{k-1}^{(p)} \int_{[\boldsymbol{x}_{k-1}^{(p)}]} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) U_{[\boldsymbol{x}_{k-1}^{(p)}]}(\boldsymbol{x}_{k-1}) d\boldsymbol{x}_{k-1}. \end{aligned} \tag{21}$$

For any transition function f, we can obtain an inclusion function $[f]$ where $f([\boldsymbol{x}]) \subseteq [f]([\boldsymbol{x}])$. For the inclusion function, with $\forall\, p = 1, \ldots, N$, if $\boldsymbol{x}_{k-1} \in [\boldsymbol{x}_{k-1}^{(p)}]$ we have $\boldsymbol{x}_k \in [f]([\boldsymbol{x}_{k-1}^{(p)}]) + [\boldsymbol{\eta}_k]$. Thus, for all $p = 1, \ldots, N$ we can write

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) U_{[\boldsymbol{x}_{k-1}^{(p)}]}(\boldsymbol{x}_{k-1}) = 0, \quad \forall\, \boldsymbol{x}_k \notin [f]([\boldsymbol{x}_{k-1}^{(p)}]) + [\boldsymbol{\eta}_k]. \tag{22}$$

Using interval analysis techniques, the support function[3] for the pdf terms in (21) can be approximated by $[f]([\boldsymbol{x}_{k-1}^{(p)}]) + [\boldsymbol{\eta}_k]$. In the Box PF algorithm each pdf term in (21) is approximated by one uniform pdf component having as support the interval $[f]([\boldsymbol{x}_{k-1}^{(p)}], [\boldsymbol{\eta}_k])$, i.e.,

$$\int_{[\boldsymbol{x}_{k-1}^{(p)}]} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) U_{[\boldsymbol{x}_{k-1}^{(p)}]}(\boldsymbol{x}_{k-1}) d\boldsymbol{x}_{k-1} \approx U_{[f]([\boldsymbol{x}_{k-1}^{(p)}])+[\boldsymbol{\eta}_k]}(\boldsymbol{x}_k). \tag{23}$$

Combining (21) and (23) gives

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) &\approx \sum_{p=1}^{N} w_{k-1}^{(p)} U_{[f]([\boldsymbol{x}_{k-1}^{(p)}])+[\boldsymbol{\eta}_k]}(\boldsymbol{x}_k) \\ &= \sum_{p=1}^{N} w_{k-1}^{(p)} U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k). \end{aligned} \tag{24}$$

Approximating each pdf term using one uniform pdf component may not be accurate enough. However, as for the PF, it is sufficient to approximate the first moments of the pdf. If a more accurate representation is required then each term can be approximated as a mixture of uniform pdfs as shown in Gning, Mihaylova et al. (2012).

Under the assumption that at time instant k, the time update pdf $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})$ can be represented by a mixture of N uniform pdfs with interval supports $[\boldsymbol{x}_{k|k-1}^{(p)}]$ and weights w_{k-1}^{p}, the measurement update step can be performed. A probabilistic model p_{ξ} for the measurement noise $\boldsymbol{\xi}_k$ is also available. It is assumed in general that p_{ξ} can be expressed by using a mixture of uniform pdfs. For simplicity and without loss of generality, p_{ξ} is considered here to be a single uniform pdf, such that the box measurement $[\boldsymbol{z}_k]$ contains all realisations of (10). Then we have: $p(\boldsymbol{z}_k|\boldsymbol{x}_k) = U_{[\boldsymbol{z}_k]}(h(\boldsymbol{x}_k))$ and according to Eq. (17), the measurement update can be expressed with the equation:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) &= \frac{1}{\alpha_k} p(\boldsymbol{z}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) \\ &= \frac{1}{\alpha_k} U_{[\boldsymbol{z}_k]}(h(\boldsymbol{x}_k)) \sum_{p=1}^{N} w_{k-1}^{(p)} U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k) \\ &= \frac{1}{\alpha_k} \sum_{p=1}^{N} w_{k-1}^{(p)} U_{[\boldsymbol{z}_k]}(h(\boldsymbol{x}_k)) U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k), \end{aligned} \tag{25}$$

where α_k denotes the normalising constant. Each of the terms $U_{[\boldsymbol{z}_k]}(h(\boldsymbol{x}_k)) U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)$ is also a constant function with a support being the following region $S_p \subset \mathbb{R}^{n_x}$, where

$$S_p = \left\{ \boldsymbol{x}_k \in [\boldsymbol{x}_{k|k-1}^{(p)}] \mid h(\boldsymbol{x}_k) \in [\boldsymbol{z}_k] \right\}. \tag{26}$$

Eq. (26) represents a constraint and from its expression we can deduce that predicted supports $[\boldsymbol{x}_{k|k-1}^{(p)}]$, from the time update pdf $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})$ approximation, have to be contracted with respect to the measurement $[\boldsymbol{z}_k]$. These contraction steps result in the new box particles denoted $[\boldsymbol{x}_k^{(p)}]$, which approximate the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ at time k. Following the definition of the sets S_p in (26), we can write

$$U_{[\boldsymbol{z}_k]}(h(\boldsymbol{x}_k)) U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k) \simeq U_{[\boldsymbol{z}_k]}(h(\boldsymbol{x}_k)) \frac{1}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|} \|S_p\| U_{S_p}(\boldsymbol{x}_k), \tag{27}$$

where $|\cdot|$ denotes the interval length (respectively the box volume in the multidimensional case). By combining Eqs. (25) and (27), and keeping in mind that $[\boldsymbol{x}_k^{(p)}] = [S_p]$ (i.e. by definition $[\boldsymbol{x}_k^{(p)}]$ is the smallest box containing S_p),

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) &= \frac{1}{\alpha_k} \sum_{p=1}^{N} w_{k-1}^{(p)} \frac{1}{|[\boldsymbol{z}_k]|} \frac{1}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|} \|S_p\| U_{S_p}(\boldsymbol{x}_k) \\ &\approx \frac{1}{\alpha_k} \sum_{p=1}^{N} w_{k-1}^{(p)} \frac{1}{|[\boldsymbol{z}_k]|} \frac{1}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|} |[\boldsymbol{x}_k^{(p)}]| U_{|[\boldsymbol{x}_k^{(p)}]|}(\boldsymbol{x}_k) \\ &\propto \sum_{p=1}^{N} w_{k-1}^{(p)} \frac{|[\boldsymbol{x}_k^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|} U_{|[\boldsymbol{x}_k^{(p)}]|}(\boldsymbol{x}_k). \end{aligned} \tag{28}$$

In the Sequential Importance Resampling (SIR) PF, each particle weight is updated by a factor equal to the likelihood $p(\boldsymbol{z}_k|\boldsymbol{x}_{k|k-1}^{(p)})$, followed by normalisation of weights. In the Box PF this step is very similar, i.e., after contracting each box particle $[\boldsymbol{x}_{k|k-1}^{(p)}]$ into $[\boldsymbol{x}_k^{(p)}]$, according to (28) the weights are updated by the ratio

$$L_k^{(p)} = \frac{|[\boldsymbol{x}_k^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}. \tag{29}$$

In summary, the posterior distribution is approximated by $\{(\tilde{w}_k^{(p)}, [\boldsymbol{x}_k^{(p)}])\}_{p=1}^{N}$, where $\tilde{w}_k^{(p)} \propto w_{k-1}^{(p)} \cdot L_k^{(p)}$.

4.2. Derivation of the box particle filter posterior distribution in crowd tracking

The prediction step for the crowds tracking box PF follows the same spirit as described by Eqs. (21) to (24). However, when dealing with multiple target originated measurements and clutter measurements, the update step is required to be re-derived. When dealing with an extended target in a SIR PF, the generalised

[3] The support of a function is the set of points where the function is not zero-valued or, in the case of functions defined on a topological space, the closure of that set.

likelihood is given by (Gilholm & Salmond, 2005)

$$p(\boldsymbol{Z}_k|\boldsymbol{\zeta}_k) = \prod_{m=1}^{M_k}\left(1+\frac{\lambda_{T,k}}{\rho_k}p(\boldsymbol{z}_k^m|\boldsymbol{x}_k)\right) = \prod_{m=1}^{M_k}\left(1+\frac{\lambda_{T,k}}{\rho_k}\int p(\boldsymbol{z}_k^m|\boldsymbol{x}_k^m)p(\boldsymbol{x}_k^m|\boldsymbol{x}_k)d\boldsymbol{x}_k^m\right), \quad (30)$$

where $\rho = \frac{\lambda_{C,k}}{A_C}$ represents the clutter density and A_C represents the area of the region where clutter may be emitted from. We extend the generalised likelihood for the crowd tracking box PF.

A probabilistic model p_{ξ_k} for the measurement noise $\boldsymbol{\xi}_k$ is available. It is assumed in general that p_{ξ_k} can be expressed by using a mixture of uniform pdfs. For simplicity and without loss of generality, p_{ξ_k} is considered here to be a single uniform pdf, such that the box measurement $[\boldsymbol{z}_k^m]$ contains all realisations of (11). Then we have: $p(\boldsymbol{z}_k^m|\boldsymbol{x}_k^m) = U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)$. Substituting this equation and (14) into (30), we obtain

$$p(\boldsymbol{Z}_k|\boldsymbol{\zeta}_k) = \prod_{m=1}^{M_k}\left(1+\frac{\lambda_{T,k}}{\rho_k}\int U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)d\boldsymbol{x}_k^m\right). \quad (31)$$

The updated marginal posterior distribution for crowd tracking can then be expressed with the equation:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k},\boldsymbol{\lambda}_k) &= \frac{1}{\alpha_k}p(\boldsymbol{Z}_k|\boldsymbol{\zeta}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1}) \\ &= \frac{1}{\alpha_k}\sum_{p=1}^{N}w_{k-1}^{(p)}\prod_{m=1}^{M_k}\left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)\right. \\ &\quad \left.+\frac{\lambda_{T,k}}{\rho_k}\int U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)d\boldsymbol{x}_k^m\right). \end{aligned} \quad (32)$$

Each of the M_k product terms, $U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)$, is also a constant function with a support being the following region $S_{p,m} \subset \mathbb{R}^{n_x}$, where

$$S_{p,m} = \left\{\boldsymbol{x}_k \in [\boldsymbol{x}_{k|k-1}^{(p)}] \mid \boldsymbol{x}_k^m \in q(\boldsymbol{x}_k), \tilde{h}\left(\boldsymbol{x}_k^m\right) \in [\boldsymbol{z}_k^m]\right\}. \quad (33)$$

Eq. (33) represents a constraint and from its expression we can deduce that the predicted supports $[\boldsymbol{x}_{k|k-1}^{(p)}]$, from the time update pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1})$ approximation, have to be contracted with respect to the interval measurements $[\boldsymbol{Z}_k]$. These contraction steps result in M_k new box particles denoted $[\boldsymbol{x}_{k,m}^{(p)}]$. Following the definition of the sets $S_{p,m}$ in (33), we can write

$$\begin{aligned} &U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m) \\ &= U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)\frac{1}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}\|S_{p,m}\|U_{S_{p,m}}(\boldsymbol{x}_k), \\ &\simeq U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)\frac{|[\boldsymbol{x}_{k,m}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}U_{[\boldsymbol{x}_{k,m}^{(p)}]}(\boldsymbol{x}_k) \end{aligned} \quad (34)$$

since by definition $[\boldsymbol{x}_{k,m}^{(p)}]$ is the smallest box containing $S_{p,m}$. Substituting (34) in (32) we have the following updated expression for the posterior distribution:

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k},\boldsymbol{\lambda}_k) = \frac{1}{\alpha_k}\sum_{p=1}^{N}w_{k-1}^{(p)}\prod_{m=1}^{M_k}\left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)+\frac{\lambda_{T,k}}{\rho_k}\right. \times\left.\frac{|[\boldsymbol{x}_{k,m}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}U_{[\boldsymbol{x}_{k,m}^{(p)}]}(\boldsymbol{x}_k)\int U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)d\boldsymbol{x}_k^m\right). \quad (35)$$

The integration is approximated by a uniform distribution, $\int U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right)U_{q(\boldsymbol{x}_k)}(\boldsymbol{x}_k^m)d\boldsymbol{x}_k^m = U_{r(\boldsymbol{x}_k)}\left(\boldsymbol{z}_k^m\right)$, where $r(\boldsymbol{x}_k)$ represents an interval dependent on the states and measurement function. The validity of this assumption is explored in Appendix A. The posterior distribution can thus be expanded accordingly:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k},\boldsymbol{\lambda}_k) &= \frac{1}{\alpha_k}\sum_{p=1}^{N}w_{k-1}^{(p)}\prod_{m=1}^{M_k}\left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)\right. \\ &\quad \left.+\frac{\lambda_{T,k}}{\rho_k}\frac{1}{|r(\boldsymbol{x}_k)|}\frac{|[\boldsymbol{x}_{k,m}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}U_{[\boldsymbol{x}_{k,m}^{(p)}]}(\boldsymbol{x}_k)\right) \\ &= \frac{1}{\alpha_k}\sum_{p=1}^{N}w_{k-1}^{(p)}\left(\left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)\right)^{M_k}\right. \\ &\quad +\sum_{m=1}^{M_k}\sum_{j=1}^{\binom{M_k}{m}}\left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k)\right)^{M_k-m} \\ &\quad \left.\times\prod_{i\in\mathcal{A}_j^m}\frac{\lambda_{T,k}}{\rho_k}\frac{1}{|r(\boldsymbol{x}_k)|}\frac{|[\boldsymbol{x}_{k,i}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}U_{[\boldsymbol{x}_{k,i}^{(p)}]}(\boldsymbol{x}_k)\right) \end{aligned} \quad (36)$$

where $\mathcal{A}^m = \left\{\mathcal{A}_j^m, j \in \mathcal{J}\right\}$, with $\mathcal{J} = \left\{1,2,\ldots,\binom{M_k}{m}\right\}$ and $\mathcal{A}_j^m \subseteq \mathcal{S} : |\mathcal{A}_j^m| = m$, where $\mathcal{S} = \{1,2,\ldots,M_k\}$. For example, if $M_k = 3$ and $m = 2$ then $\mathcal{A}^m = \{\{1,2\},\{1,3\},\{2,3\}\}$. The posterior pdf is a weighted sum of uniform pdfs. The number of weighted uniform pdf's increases exponentially with the number of measurements, which can render the algorithm too computationally expensive for a large number of measurements. Typically, there is a large disparity between the weights of the summed uniform pdfs, since $\frac{\lambda_{T,k}}{\rho_k}\frac{1}{|r(\boldsymbol{x}_k)|}\frac{|[\boldsymbol{x}_{k,i}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|} \gg \frac{1}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}$. This allows for the approximation of the posterior pdf by a single uniform pdf for each box particle. The dominating term in the uniform pdf weights is $\frac{\lambda_{T,k}}{\rho_k|r(\boldsymbol{x}_k)|\,|[\boldsymbol{x}_{k|k-1}^{(p)}]|}$. This term is maximised when all the measurements are assumed to originate from the crowd. If the posterior pdf was approximated by this uniform pdf, the expression would be given by:

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k},\boldsymbol{\lambda}_k) \approx \frac{1}{\alpha_k}\sum_{p=1}^{N}w_{k-1}^{(p)}\left(\prod_{i\in\mathcal{S}}\frac{\lambda_{T,k}}{\rho_k}\frac{1}{|r(\boldsymbol{x}_k)|}\frac{|[\boldsymbol{x}_{k,i}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}U_{[\boldsymbol{x}_{k,i}^{(p)}]}(\boldsymbol{x}_k)\right). \quad (37)$$

The multiplication of uniform pdfs can be further simplified to obtain a single uniform pdf with a corresponding weight. This includes the intersection of the intervals of all the uniform pdfs:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k},\boldsymbol{\lambda}_k) &\propto \sum_{p=1}^{N}w_{k-1}^{(p)}\left(\prod_{i\in\mathcal{S}}\frac{\lambda_{T,k}}{\rho_k}\frac{1}{|r(\boldsymbol{x}_k)|}\frac{|[\boldsymbol{x}_{k,i}^{(p)}]|}{|[\boldsymbol{x}_{k|k-1}^{(p)}]|}\right) \\ &\quad \times\frac{|\cap_{i\in\mathcal{S}}[\boldsymbol{x}_{k,i}^{(p)}]|}{\prod_{i\in\mathcal{S}}|[\boldsymbol{x}_{k,i}^{(p)}]|}U_{\cap_{i\in\mathcal{S}}[\boldsymbol{x}_{k,i}^{(p)}]}(\boldsymbol{x}_k) \\ &\propto \sum_{p=1}^{N}w_{k-1}^{(p)}\left(\prod_{i\in\mathcal{S}}\frac{\lambda_{T,k}}{\rho_k|r(\boldsymbol{x}_k)|\,|[\boldsymbol{x}_{k|k-1}^{(p)}]|}\right) \\ &\quad \times|\cap_{i\in\mathcal{S}}[\boldsymbol{x}_{k,i}^{(p)}]|U_{\cap_{i\in\mathcal{S}}[\boldsymbol{x}_{k,i}^{(p)}]}(\boldsymbol{x}_k). \end{aligned} \quad (38)$$

However, this intersection result typically does not exist or leads to a poor contraction due to the implicit assumption that the measurements originate from the crowd. A more robust

Table 1
The proposed box particle filter for crowd tracking.

Initialisation
Use the available prior information about the target's kinematics and extent parameters states to initialise the box particles.
Repeat for K time steps, $k = 1, \ldots, K$, the following steps:
(1) ***Prediction***
Propagate the box particles through the state evolution model to obtain the predicted box particles. Apply interval inclusion functions as described in Jaulin (2002) and Jaulin et al. (2001).
(2) ***Measurement Update***
Upon the receipt of new measurements:
(a) Form intervals around the measurements, taking into account the uncertainty of the sensor, thus obtaining the measurement boxes $[\boldsymbol{Z}_k]$.
(b) Solve the CSP, as described in Section 4.3, to obtain the contracted box particles $[\boldsymbol{x}_{k,m}^{(p)}]$.
(c) Determine $[\boldsymbol{x}_k^{(p)}]$ according to (40).
(d) Update and normalise the weights $w_k^{(p)}$, $p = 1, \ldots, N$ according to (41).
(3) ***Output***
Obtain an estimate for the state of the group object as a weighted sum of all of the particles:
$$[\hat{\boldsymbol{x}}_k] = \sum_{p=1}^{N} w_k^{(p)} [\boldsymbol{x}_k^{(p)}]. \tag{42}$$
Further, a point estimate for the state can be obtained as the midpoint of the box estimate of the state.
(4) ***Resampling***
(a) *Compute the effective sample size*:
$N_{eff} = \frac{1}{\sum_{p=1}^{N} (\tilde{w}_k^{(p)})^2}$
(b) If $N_{eff} \leq N_{thresh}$ (with e.g. $N_{thresh} = 2N/3$) resample by division of particles with high weights. Finally, reset the weights: $w_k^{(p)} = 1/N$.

approximation for the posterior pdf, which does not require explicit knowledge of the origin of a measurement is given by:

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}, \boldsymbol{\lambda}_k) \approx \sum_{p=1}^{N} w_{k-1}^{(p)} \left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k) \right)^{M_k - (|\mathcal{I}_E^{(p)}| - q)} \times \left(\prod_{i \in \mathcal{I}_E^{(p)}} \frac{\lambda_{T,k}}{\rho_k |r(\boldsymbol{x}_k)| \, |[\boldsymbol{x}_{k|k-1}^{(p)}]|} \right) \times \left| \bigcap_{i \in \mathcal{I}_E^{(p)}}^{\{q\}} [\boldsymbol{x}_{k,i}^{(p)}] \right| U_{\bigcap_{i \in \mathcal{I}_E^{(p)}}^{\{q\}} [\boldsymbol{x}_{k,i}^{(p)}]}(\boldsymbol{x}_k) \tag{39}$$

where $\mathcal{I}_E^{(p)}$ represents the set of indices for the contracted boxes, $[\boldsymbol{x}_{k,m}^{(p)}]$, that exist,[4] and q represents the maximum number of clutter measurements indexed by $\mathcal{I}_E^{(p)}$. The symbol $\bigcap^{\{q\}}$ represents the q-relaxed intersection first introduced in Jaulin (2009) to aid in the processing of clutter measurements in a purely interval framework.

The difference between the posterior pdf represented by Eqs. (36) and (39) is highlighted graphically through an example in Fig. 1.

In summary, $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}, \boldsymbol{\lambda}_k)$ is approximated by $\{(\tilde{w}_k^{(p)}, [\boldsymbol{x}_k^{(p)}])\}_{p=1}^{N}$, where

$$[\boldsymbol{x}_k^{(p)}] = \bigcap_{i \in \mathcal{I}_E^{(p)}}^{\{q\}} [\boldsymbol{x}_{k,i}^{(p)}] \tag{40}$$

and

$$\tilde{w}_k^{(p)} \propto w_{k-1}^{(p)} \left(U_{[\boldsymbol{x}_{k|k-1}^{(p)}]}(\boldsymbol{x}_k) \right)^{M_k - (|\mathcal{I}_E^{(p)}| - q)} \times \left(\prod_{i \in \mathcal{I}_E^{(p)}} \frac{\lambda_{T,k}}{\rho_k |r(\boldsymbol{x}_k)| \, |[\boldsymbol{x}_{k|k-1}^{(p)}]|} \right) |[\boldsymbol{x}_k^{(p)}]|. \tag{41}$$

The algorithm for crowd tracking is summarised in Table 1.

[4] Measurements which result in a contraction of the state that does not exist are located at a significant distance from the state and are considered to be clutter measurements.

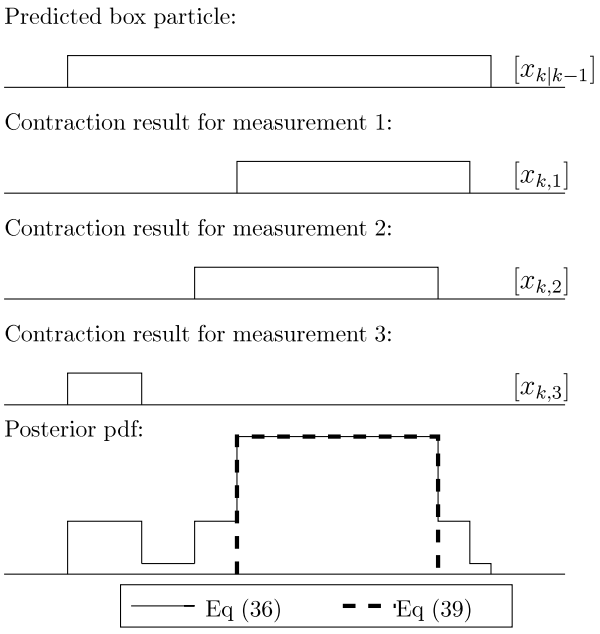


Fig. 1. Illustration of the difference between the posterior pdf represented by Eqs. (36) and (39). This example consists of 3 measurements (measurement 3 represents a clutter measurement), a single state dimension, and a single box particle.

4.3. Box particle filter implementation considerations

In general, an important step in interval based techniques used for state estimation is in interval contraction (Jaulin, 2009). In the box PF it is required to obtain the contracted box particles by solving the CSP described by Eq. (33). For the crowds tracking box PF, contraction is achieved by implementing the Constraints Propagation (CP) technique. The main advantage of the CP method is its efficiency in the presence of high redundancy of data and equations. The CP algorithm, which in this application is the calculation of the intersection of the box states for each particle with all the interval measurements, is illustrated in Table 2. For notational convenience, Table 2 refers directly to the supports of the uniform distributions found in the posterior distribution, for example in Eq. (35).

Generally, in particle filtering, there are a variety of different resampling schemes available (Li, Bolic, & Djuric, 2015). Based on

the weights, a particle is replicated a specific number of times. The box PF differs by dividing a selected box particle into smaller box-particles as many times as it was to be replicated. Several subdivision strategies exist. In this paper we subdivided based on the dimension with the largest box face.

The parameter q is introduced in Eq. (39). This specifies the maximum number of clutter measurements that still result in a contraction of the states that exists. These are the clutter measurements which are located close in vicinity to the crowd. The area in the measurement space where a measurement can result in a contraction of the state that exists is dependent on the size of the box particle. An estimate for q can then be determined through:

$$q = \frac{\rho_k A_{CT}}{4}. \tag{43}$$

The estimated clutter measurement rate is used to obtain an approximate ρ_k:

$$\rho_k = \frac{\lambda_{C,k}}{A_{CR}}, \tag{44}$$

where the area of the clutter region is given by $A_{CR} = A_S - A_T$, A_S is the total area observed by the sensor, and A_T is the area of the crowd, approximated from the estimate of the crowd at the previous time instant, $k-1$. For the given crowd tracking problem, the area A_{CT} is given by:

$$\begin{aligned} A_{CT} = &\left(\left(\left(\overline{x_k^{(p)}} + \frac{\overline{a_k^{(p)}}}{2}\right) - \left(\underline{x_k^{(p)}} - \frac{\overline{a_k^{(p)}}}{2}\right)\right)\right. \\ &\times \left.\left(\left(\overline{y_k^{(p)}} + \frac{\overline{b_k^{(p)}}}{2}\right) - \left(\underline{y_k^{(p)}} - \frac{\overline{b_k^{(p)}}}{2}\right)\right)\right) \\ &- \left(\left(\left(\underline{x_k^{(p)}} + \frac{\underline{a_k^{(p)}}}{2}\right) - \left(\overline{x_k^{(p)}} - \frac{\underline{a_k^{(p)}}}{2}\right)\right)\right. \\ &\times \left.\left(\left(\underline{y_k^{(p)}} + \frac{\underline{b_k^{(p)}}}{2}\right) - \left(\overline{y_k^{(p)}} - \frac{\underline{b_k^{(p)}}}{2}\right)\right)\right), \end{aligned} \tag{45}$$

where the notation $\underline{x}$ and $\overline{x}$ refers to the infimum and supremum of box x, respectively. The factor of 4 in Eq. (43) was introduced to take into account that the area A_{CT} also includes the region inside of the crowd, where no clutter measurements are found. It is important to note that the algorithm is fairly robust to the value of q as this represents a maximum number of clutter points, and not the actual number of clutter points.

5. The convolution particle filter for crowd tracking

This paper develops an adaptive CPF algorithm for crowds tracking. The CPF approach relies on convolution kernel density estimation and regularisation of the distributions, respectively, of the state and observation variables (Campillo & Rossi, 2009; Rossi & Vila, 2006; Vila, 2012). The CPF belongs to a class of particle filters with valuable advantages: simultaneous estimation of state variables and unknown parameters and continuous approximation of the corresponding pdf. Being likelihood free filters makes them attractive for solving complex problems where the likelihood is not available in an analytical form.

The key novelty of the proposed adaptive CPF algorithm stems from: (1) its ability to deal with multiple measurements, including high level of clutter, (2) ability to resolve data association problems, without the need to estimate clutter parameters, (3) estimation of dynamically changing parameters of crowds jointly with the dynamic kinematic states.

For the purposes of crowds tracking the marginal posterior state distribution has to be calculated and can be expressed to be

Table 2
CSP for contraction of rectangularly shaped crowds.

Solve the CSP to contract each box particle with all of the measurements.
$[x_m^{(p)}] = [x^{(p)}] \cap \left([z_1^m] \mp \frac{[a^{(p)}]}{2} \cdot [0,1]\right)$,
$[\tilde{\dot{x}}_m^{(p)}] = [\dot{x}^{(p)}] \cap \left(\frac{[x_m^{(p)}(k)] - [x_m^{(p)}(k-1)]}{\frac{1}{\alpha_x}(1-e^{-\alpha_x T_s})}\right)$,
$[\tilde{y}_m^{(p)}] = [y^{(p)}] \cap \left([z_2^m] \mp \frac{[b^{(p)}]}{2} \cdot [0,1]\right)$,
$[\tilde{\dot{y}}_m^{(p)}] = [\dot{y}^{(p)}] \cap \left(\frac{[\tilde{y}_m^{(p)}(k)] - [\tilde{y}_m^{(p)}(k-1)]}{\frac{1}{\alpha_y}(1-e^{-\alpha_y T_s})}\right)$,
$[\tilde{a}_m^{(p)}] = [a^{(p)}] \cap \pm 2\left(\frac{[z_1^m] - [x_{m,s}^{(p)}]}{[0,1]}\right)$,
$[\tilde{b}_m^{(p)}] = [b^{(p)}] \cap \pm 2\left(\frac{[z_2^m] - [\tilde{y}_{m,s}^{(p)}]}{[0,1]}\right)$,
$[\tilde{z}_1^{m,(p)}] = [z_1^m] \cap \left([x_m^{(p)}] \pm \frac{[\tilde{a}_m^{(p)}]}{2} \cdot [0,1]\right)$,
$[\tilde{z}_2^{m,(p)}] = [z_2^m] \cap \left([\tilde{y}_m^{(p)}] \pm \frac{[\tilde{b}_m^{(p)}]}{2} \cdot [0,1]\right)$.

(46)

independent of the clutter and measurement rates, reducing the expression from Eq. (19) to:

$$p(\boldsymbol{\zeta}_k|\boldsymbol{Z}_{1:k}) = p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) p(\lambda_{T,k}|\boldsymbol{Z}_{1:k}) p(\lambda_{C,k}|\boldsymbol{Z}_{1:k}). \tag{47}$$

The CPF relies on the following representation of the conditional state density:

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) = \frac{p(\boldsymbol{x}_k, \boldsymbol{Z}_{1:k})}{\int p(\boldsymbol{x}_k, \boldsymbol{Z}_{1:k}) d\boldsymbol{x}_k}. \tag{48}$$

Suppose, that we can sample from the state and measurement pdfs, $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ and $p(\boldsymbol{z}_k^m|\boldsymbol{x}_k)$, respectively. Then we can obtain a sample from the joint distribution $\{\boldsymbol{x}_k^{(i)}, \boldsymbol{Z}_k^{(i)},\ i = 1, \ldots, N\}$ at time step k by k successive simulations, starting from the sample of the initial distribution $p_0(\boldsymbol{x})$. We can obtain the following empirical estimate of the joint density

$$p(\boldsymbol{x}_k, \boldsymbol{Z}_{1:k}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}, \boldsymbol{Z}_{1:k} - \boldsymbol{Z}_{1:k}^{(i)}). \tag{49}$$

The kernel estimate $p_k^N(\boldsymbol{x}_k, \boldsymbol{Z}_{1:k})$ of the true density $p(\boldsymbol{x}_k, \boldsymbol{Z}_{1:k})$ is obtained by convolution of the empirical estimate (49) with an appropriate kernel

$$p_k^N(\boldsymbol{x}_k, \boldsymbol{z}_{1:k}) = \frac{1}{N} \sum_{i=1}^{N} K_h^x(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}) K_h^{\bar{Z}}(\boldsymbol{Z}_{1:k} - \boldsymbol{Z}_{1:k}^{(i)}), \tag{50}$$

where

$$K_h^{\bar{Z}}(\boldsymbol{Z}_{1:k} - \boldsymbol{Z}_{1:k}^{(i)}) = \prod_{j=1}^{k} K_h^Z(\boldsymbol{Z}_j - \boldsymbol{Z}_j^{(i)}) \tag{51}$$

and K_h^x and K_h^Z are the Parzen–Rosenblatt kernels of appropriate dimensions and bandwidth h. According to Eq. (48), the estimate of the posterior conditional state density has the following form:

$$p_k^N(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) = \frac{\sum_{i=1}^{N} K_h^x(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}) K_h^{\bar{Z}}(\boldsymbol{Z}_{1:k} - \boldsymbol{Z}_{1:k}^{(i)})}{\sum_{i=1}^{N} K_h^{\bar{Z}}(\boldsymbol{Z}_{1:k} - \boldsymbol{Z}_{1:k}^{(i)})}. \tag{52}$$

When dealing with point targets, the measurements are modelled as points in the measurement space. However, in the application of crowd tracking, a single point in the state space translates into a region in the measurement space. The role of the kernel in the point target case can be interpreted as a conversion of the

Table 3
The Convolution Particle Filter for Crowd Tracking.

I. Initialisation:
$k = 0$, *for* $i = 1, \ldots, N$ generate particles
$\boldsymbol{x}_0^{(i)} \sim p_0(\boldsymbol{x})$, $w_0^{(i)} = 1/N$, $k = k + 1$
II. Iterate: over steps (1) to (5) for $k \geq 1$
if $k = 1$: Prediction: *for* $i = 1, \ldots, N$
$\boldsymbol{x}_k^{(i)} \sim f(\boldsymbol{x}_k | \boldsymbol{x}_0^{(i)})$—state sampling
$\boldsymbol{Z}_k^{(i)} \sim p(\boldsymbol{z}_k^m | \boldsymbol{x}_0^{(i)})$—measurement region sampling
go to step (3)
if $k > 1$:
(1) Resampling: *for* $i = 1, \ldots, N$
$\tilde{\boldsymbol{x}}_{k-1}^{(i)} \sim p_{k-1}^N(\boldsymbol{x}_{k-1}, |\boldsymbol{Z}_{1:k-1})$, $w_{k-1}^{(i)} = 1/N$
(2) Prediction: *for* $i = 1, \ldots, N$
$\boldsymbol{x}_k^{(i)} \sim p(\boldsymbol{x}_k | \tilde{\boldsymbol{x}}_{k-1}^{(i)})$—state sampling
$\boldsymbol{Z}_k^{(i)} \sim p(\boldsymbol{z}_k^m | \tilde{\boldsymbol{x}}_k^{(i)})$—measurement region sampling
(3) Weights updating: *for* $i = 1, \ldots, N$
Update the weights according to (53),
(5) Estimating the output state:
$\hat{\boldsymbol{x}}_k = \sum_{i=1}^N \bar{w}_k^{(i)} \boldsymbol{x}_k^{(i)}$,
where $\bar{w}_k^{(i)}$ are the normalised weights.

measurement point to a measurement region. Thus there is no need for the specification of a kernel in the crowd tracking CPF framework, as the densities that describe the sensor characteristics and target model can be used to obtain an approximate region in the measurement space for each predicted particle, and are thus equivalent to the kernel. The bandwidth h of the kernel varies according to the state, resulting in a variable bandwidth which adds additional flexibility to the CPF while also removing the need to specify a bandwidth parameter. In this application the kernel is approximated as a variable uniform distribution.

An advantage of the proposed CPF framework is that it implicitly resolves the data association problem. Since there are multiple measurements assumed to be independent, the weights of individual measurements are multiplied to obtain a single weight for the particle. However, clutter measurements may occur outside of the support of the adaptive uniform kernel. This would result in particles having a weight of 0 when evaluated by the kernel. To overcome this, the adaptive uniform kernel based on the crowd is added with a uniform distribution which covers the entire observation area of the sensor. The advantage to such an approach is that it removes the need for the estimation of the clutter and measurement rates when only the kinematic states and extent parameters are of interest.

The weights are updated sequentially according to

$$w_k^{(i)} = w_{k-1}^{(i)} \prod_{m=1}^{M_k} K_h^Z \left(\boldsymbol{z}_k^m - \boldsymbol{Z}_k^{(i)} \right). \tag{53}$$

For the crowd tracking problem presented, the kernel $K_h^Z(\boldsymbol{z}_k^m - \boldsymbol{Z}_k^{(i)})$ in Eq. (53) is a compositional kernel comprised of a sum of two uniform pdfs:

$$K_h^Z(\boldsymbol{z}_k^m - \boldsymbol{Z}_k^{(i)}) = U_{CS}(\boldsymbol{z}_k) + U_{SS}(\boldsymbol{z}_k), \tag{54}$$

where the support SS is the entire region observed by the sensor, and the support CS is related to the location of crowd measurements given the particle state. In this paper we utilised the region, $r(\boldsymbol{x}_k)$, as described in Appendix A.

A detailed description of the CPF algorithm is given in Table 3.

6. Performance evaluation

In this work the performance evaluation is done using simulated measurements data. All simulations were performed on a mobile computer with Intel(R) Core(TM) i7-4702HQ CPU @ 2.20 GHz with 16 GB of RAM.

6.1. Test environment

Two different crowd simulations were used to demonstrate the performance of the crowd tracking box PF and CPF.

Rectangular group object simulator: A crowd with a rectangular extent located in a two dimensional plane. The centre of the crowd undergoes motion according to a correlated velocity model. The lengths of the sides of the crowd vary at each time step according to a random walk. Crowd measurements comprise of a number of points uniformly located within the confines of the crowd at each time step. In addition to the crowd measurements, clutter measurements are also present, uniformly located in a region about the crowd.

Realistic crowd simulator: Individuals within the crowd are represented as points moving in a two dimensional space. The dynamics of the group is determined by forces acting on those individuals: Forces of attraction towards one or more static goal points; constrained forces of repulsion between the elements of the group; constrained forces of repulsion from a set of linear contextual constraints. The net effect is that a crowd of individuals will move in a reasonably realistic manner between constraints. The simulator outputs a set of points corresponding to the positions of each individual in the crowd at each sampling step. The positions of the individuals represent the measurement sources. Additionally, clutter measurements are also present, uniformly located in a region about the crowd.

6.2. Rectangular group object simulator results

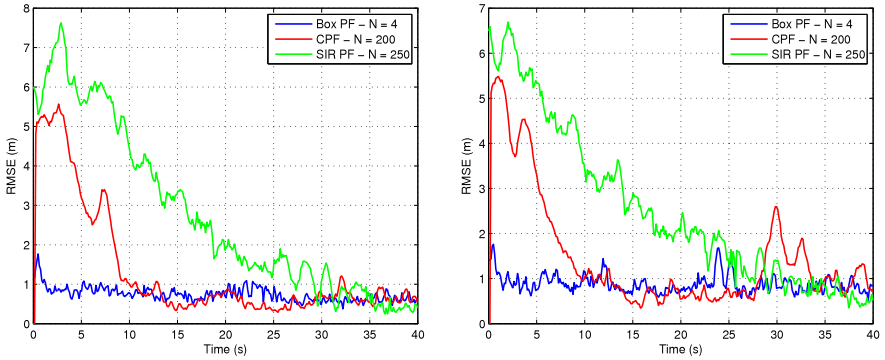
This section presents results based on the Rectangular group object simulator. The parameters are as follows:

- *Simulation*: The mean number of measurement sources: $\lambda_T = 100$, Simulation time duration: $T_{tot} = 40$ s, Sampling time, $T_s = 0.125$ s, Initial rectangular object kinematic state: $\boldsymbol{X}_0 = [100\,\text{m}, 0\,\text{m/s}, 100\,\text{m}, 0\,\text{m/s}]^T$, Initial rectangular object extent parameters: $\boldsymbol{\Theta}_0 = [40\,\text{m}, 40\,\text{m}]^T$, Crowd centre dynamics parameters: Velocity correlation time constant, $T_{cv} = 15$ s, Velocity standard deviation parameters, $\sigma_{v,x} = \sigma_{v,y} = 10$ m/s, Group extent dynamics parameters $\sigma_a = \sigma_b = 1$ m per time step.
- *Sensor*: Measurement uncertainty: $\sigma_{z_1} = \sigma_{z_2} = 0.1$ m. Clutter parameters: Clutter density, $\rho = 1 \times 10^{-2}$. Clutter area = Circular region with radius of 100 m about the centre of the crowd subtracted by the area of the crowd.
- *Filter parameters*: The CPF and SIR PF utilise a uniform distribution for each state to initialise the particles. In the case of the Box PF, the same uniform region where the CPF and SIR PF randomly generate particles from is subdivided so that the entire region is encompassed by all the box particles. This region for each state is: $x_0^{(p)} = [x_0 - 50; x_0 + 50]$ m, $\dot{x}_0^{(p)} = [\dot{x}_0 - 10; \dot{x}_0 + 10]$ m/s, $y_0^{(p)} = [y_0 - 50; y_0 + 50]$ m, $\dot{y}_0^{(p)} = [\dot{y} - 10; \dot{y} + 10]$ m/s, $a_0^{(p)} = [a_0 - 30; a_0 + 30]$ m, and $b_0^{(p)} = [b_0 - 30; b_0 + 30]$ m.

The root mean square error (RMSE) of the box PF and CPF estimates are illustrated in this section. The RMSE values for each time step are calculated over a number of Monte Carlo simulation runs according to

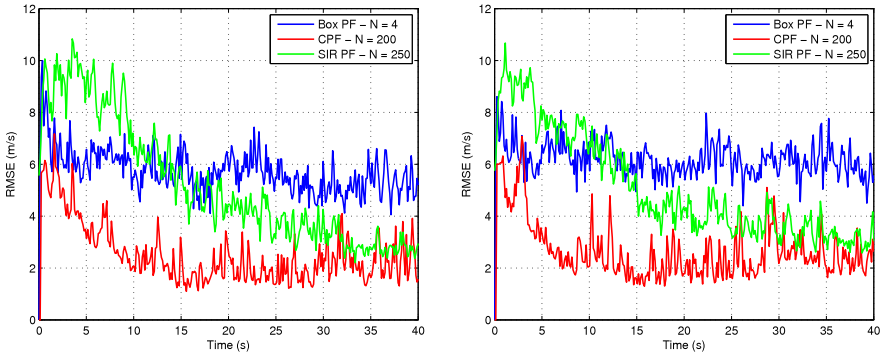
$$\text{RMSE} = \sqrt{\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left\| \hat{\boldsymbol{x}}_i - \boldsymbol{x}_i \right\|^2}, \tag{55}$$

where $\boldsymbol{x}_i$ represents the ground truth, $\hat{\boldsymbol{x}}_i$ represents the filter estimate, and N_{MC} represents the number of Monte Carlo runs.



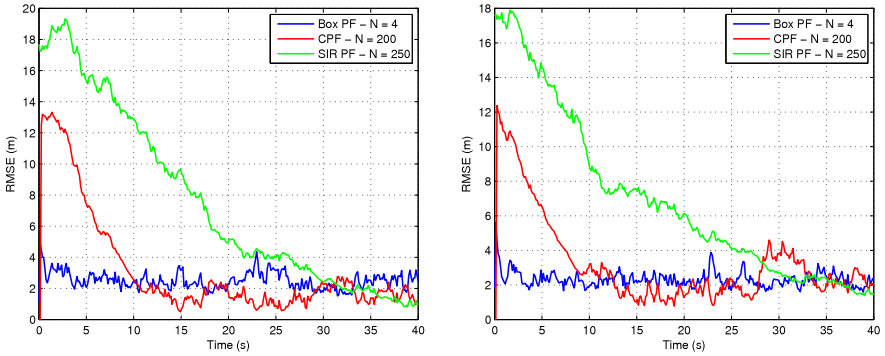
(a) RMSE corresponding to the location of the centre in the x-coordinate.

(b) RMSE corresponding to the location of the centre in the y-coordinate.



(c) RMSE corresponding to the velocity of the centre in the x-coordinate.

(d) RMSE corresponding to the velocity of the centre in the y-coordinate.



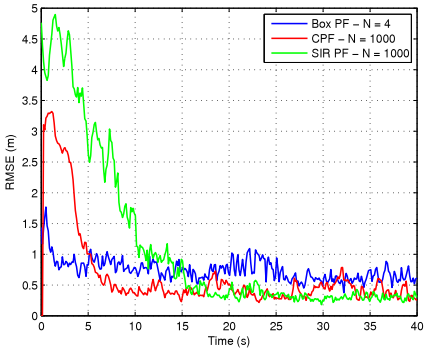
(e) RMSE corresponding to the length of side A.

(f) RMSE corresponding to the length of side B.

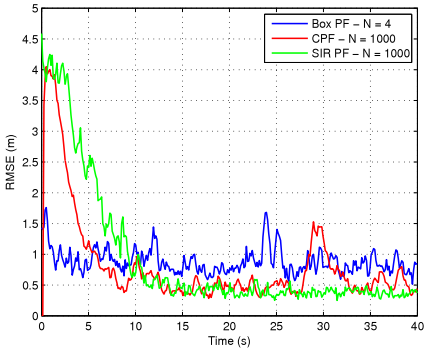
Fig. 2. Comparison of the RMSE for the states of the box PF, CPF and SIR PF with equal computational complexity.

The first set of results illustrate how the box PF, CPF and SIR PF perform when estimating the marginal posterior distribution, $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}, \boldsymbol{\lambda}_k)$, with measurement and clutter rates assumed known. Only 4 box particles are required to track the crowd. For comparison, the CPF and SIR PF were also run with 4 particles, however, this resulted in consistent filter divergence due to particle degeneracy. Instead the number of particles were selected based on achieving a similar computational expense for all algorithms. The number of Monte Carlo runs is 100. The resultant RMSE values are illustrated in Fig. 2. The comparison of the computational complexity for these results are presented in Table 4. It is worth noting that the implementation of the box PF utilises the INTLAB toolbox for performing interval operations. INTLAB was initially designed and optimised for estimating rounding errors. We believe that utilising alternative methods for the interval operations could significantly reduce the computational complexity of the box PF. The box PF and CPF are able to lock on to the crowd significantly faster than the SIR PF. It is noted that the RMSE is generally higher for the box PF once all filters have locked onto the crowd. This can be attributed to the approximations made in the derivation of the marginal posterior pdf. The SIR PF is also matched in terms of the model noise and likelihood expression.

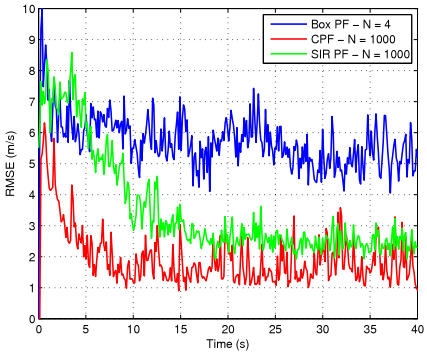
The interested reader is referred to our previous works for Box PFs for point targets where a detailed comparison is presented, with Bernoulli filters in Gning, Ristic, and Mihaylova (2012) and with a Probability Hypothesis density (PHD) filter (Schikora et al., 2014). In these works it is shown that the Box PF for point target tracking requires a significantly smaller number of box particles compared with the particles needed in the Bernoulli and PHD filters, including the computational cost. The Box PF can be used in sensor network systems as it has been shown in Haj Chadé, Gning, Abdallah, Mougharbel, and Julier (2014) and its key advantage is that it provides accurate estimation results with a small number



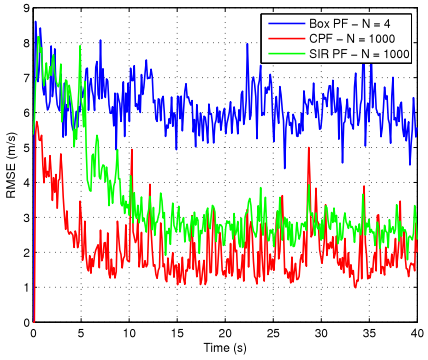
(a) RMSE corresponding to the location of the centre in the *x*-coordinate.



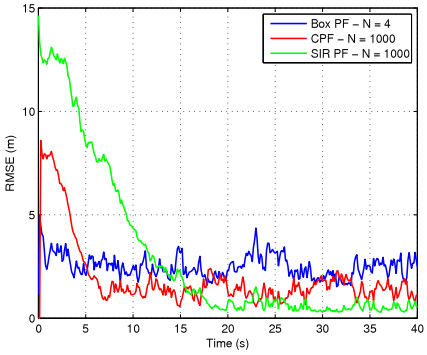
(b) RMSE corresponding to the location of the centre in the *y*-coordinate.



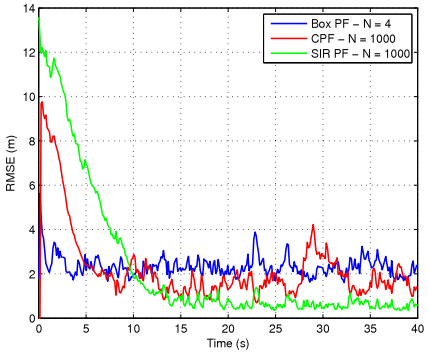
(c) RMSE corresponding to the velocity of the centre in the *x*-coordinate.



(d) RMSE corresponding to the velocity of the centre in the *y*-coordinate.



(e) RMSE corresponding to the length of side A.



(f) RMSE corresponding to the length of side B.

Fig. 3. Comparison of the RMSE for the states of the box PF, CPF and SIR PF for maximised performance.

Table 4
Matlab computational time corresponding to the results in Fig. 2.

Algorithm	Computation time (s)
Box PF	13.47
CPF	14.43
SIR PF	13.01

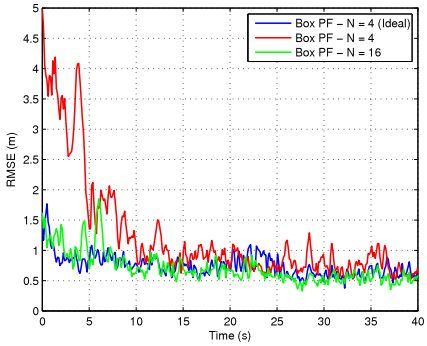
Table 5
Matlab computational time corresponding to the results in Fig. 3.

Algorithm	Computation time (s)
Box PF	13.47
CPF	42.16
SIR PF	45.58

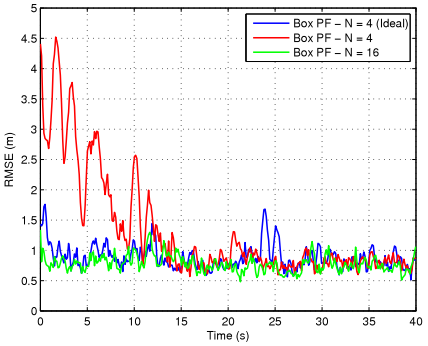
of particles. The Box PF can also be used in industrial applications, e.g. such as those in Liu et al. (2014) and Wang, Gao, and Qiu (2016) and other network control systems.

The second set of results re-iterate the experiment with a significant increase in the number of particles for the CPF and SIR PF in order to improve tracking performance with an increase in computational expense. The resultant RMSE values are illustrated in Fig. 3, and the computational cost comparison for these results are presented in Table 5. Increasing the number of particles in the CPF and SIR PF decrease the amount of time required to lock on to the crowd, however, the faster lock comes at a significantly larger computational burden.

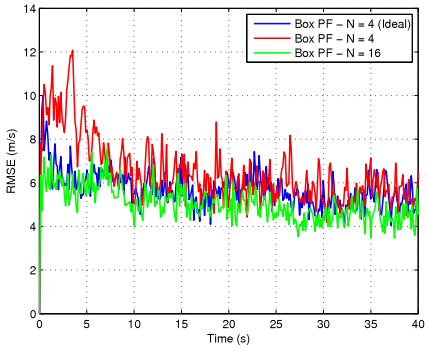
The third set of results focus on the effect of jointly estimating the crowd and clutter measurement rates on the box PF performance. This is compared with the performance of the box PF for the ideal case where crowd and clutter measurement rates are known. The resultant RMSE values are illustrated in Fig. 4. The computational cost comparison for these results are presented in Table 6. The joint estimation results in an increase in the time required to lock onto the crowd, however, this is overcome by



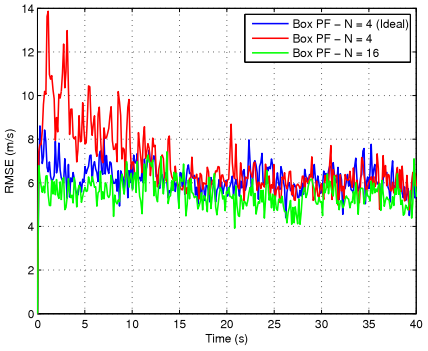
(a) RMSE corresponding to the location of the centre in the *x*-coordinate.



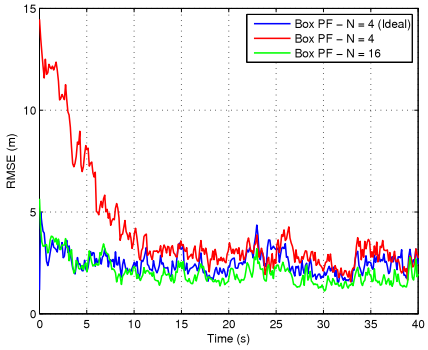
(b) RMSE corresponding to the location of the centre in the *y*-coordinate.



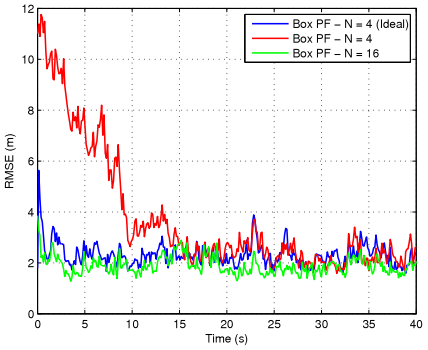
(c) RMSE corresponding to the velocity of the centre in the *x*-coordinate.



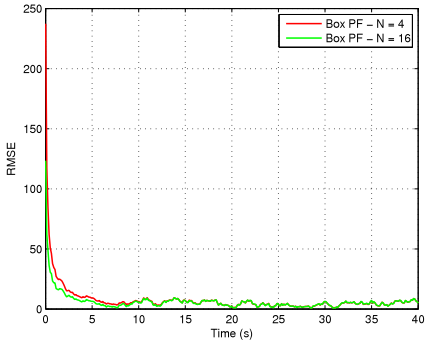
(d) RMSE corresponding to the velocity of the centre in the *y*-coordinate.



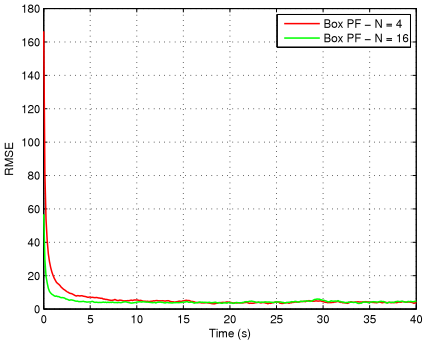
(e) RMSE corresponding to the length of side A.



(f) RMSE corresponding to the length of side B.



(g) RMSE corresponding to the crowd measurement rate.



(h) RMSE corresponding to the clutter measurement rate.

Fig. 4. Comparison of the RMSE for the states of the box PF with crowd and clutter rate estimation.

Table 6
Matlab computational time corresponding to the results in Fig. 4.

Number of Box particles	Computation time (s)
4	13.47
16	25.22

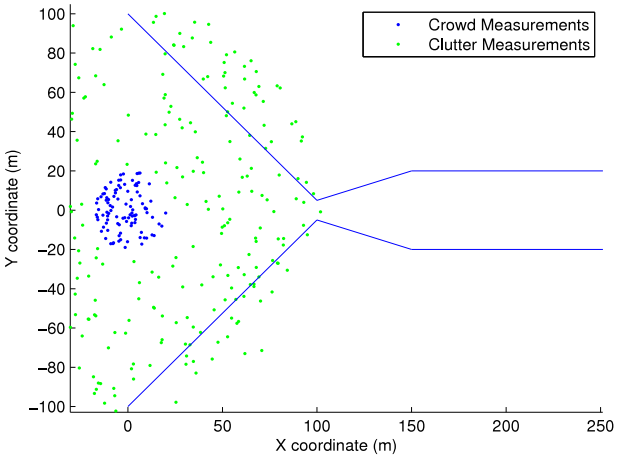


Fig. 5. Initialisation of the realistic crowd simulator.

increasing the number of box particles at the cost of an increased computational burden.

6.3. The realistic crowd simulator results

In the realistic crowd simulator the crowd moves through a corridor which consists of a bottleneck. The crowd is initialised at the entrance of the bottleneck. This is illustrated in Fig. 5. In this section a comparison between the box PF and CPF is presented to illustrate the filters operation on the realistic crowd simulator. The SIR PF is not included since it is incapable of operating without knowledge of the crowd and clutter measurements which are not available in a realistic situation. The parameters for the simulations are as follows:

- *Simulation*: The number of entities in the crowd: $N_T = 100$, Simulation time duration: $T_{tot} = 150$ s, Sampling time, $T_s = 0.125$ s,
- *Sensor*: Measurement uncertainty: $\sigma_{z_1} = \sigma_{z_2} = 0.1$ m, Clutter parameters: Clutter density, $\rho = 1 \times 10^{-3}$, Clutter area = Circular region with radius of 100 m about the centre of the crowd,
- *Filter parameters*: Number of box particles: $N = 16$, Number of CPF particles: $N = 1000$, Crowd centre dynamics parameters: Velocity correlation time constant, $T_{cv} = 30$ s, Velocity standard deviation parameters, $\sigma_{v,x} = \sigma_{v,y} = 1$ m/s, Group extent dynamics parameters $\sigma_a = \sigma_b = 0.1$ m per time step. Measurement uncertainties: matched to the sensor parameters. Initialisation: Initialised in the same manner as for the rectangular group object simulator.

The RMSE for each state, based on the ground truth extracted from the crowd measurements, are illustrated in Fig. 6 for both the box PF and CPF. The number of Monte Carlo runs is 50. The crowd moves through the bottleneck in the vicinity of 60 s. Initially, the CPF struggles to lock on to the target. Once locked, and after the crowd has passed through the bottleneck, the RMSE for the length corresponding to side a is increased. This is due to several crowd entities spreading out further away from the majority of the crowd and thus being mistaken as a clutter measurements.

7. Conclusions

This paper proposes a box PF and CPF framework for tracking a large crowd of entities. A theoretical derivation for the generalised likelihood function for the box PF is presented when the state vector consists of kinematic states and extent parameters. The likelihood is calculated based on optimisation, by solving a constraint satisfaction problem (CSP) with multiple measurements. An adaptive CPF is proposed able to deal with multiple measurements, including a high level of clutter. It is able to resolve the data association problem without the need to estimate the clutter parameters.

The filters adaptively track the envelop of a crowd. Both filters resolve the data association problem in an efficient way. These are two different types of filters—the Box PF works with box particles, whereas the CPF represents the probabilistic distributions with point samples. The Box PF and the CPF are compared with the generic SIR PF. The filters are both robust to sensor error characteristics. The experiments show that the Box PF is also robust to initialisation errors. The Box PF requires a significantly smaller number of (box) particles than the SIR PF and the CPF.

Acknowledgements

We acknowledge the support from the UK Engineering and Physical Sciences Research Council (EPSRC) for the support via the Bayesian Tracking and Reasoning over Time (BTaRoT) grant EP/K021516/1, the EC Seventh Framework Programme [FP7 2013–2017] TRAcking in compleX sensor systems (TRAX) Grant agreement no.: 607400 and Selex ES under grant: "Information Fusion: Framework architectures for dynamic heterogeneous information fusion".

Appendix A. Likelihood derivation

In Section 4.2 the following approximation is presented:

$$\int U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right) U_{q(\boldsymbol{x}_k)}\left(\boldsymbol{x}_k^m\right) d\boldsymbol{x}_k^m \approx U_{r(\boldsymbol{x}_k)}\left(\boldsymbol{z}_k^m\right). \tag{A.1}$$

In this Appendix a detailed description supporting this approximation is presented.

In order to evaluate the integral, it is required to transform the domain of the uniform distribution relating a measurement to a measurement source. The explicit expression for the pdf of this distribution is given by:

$$U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right) = \begin{cases} \dfrac{1}{6\sigma} & : \tilde{h}\left(\boldsymbol{x}_k^m\right) \in [\boldsymbol{z}_k^m - 3\sigma, \boldsymbol{z}_k^m + 3\sigma] \\ 0 & : \text{elsewhere.} \end{cases} \tag{A.2}$$

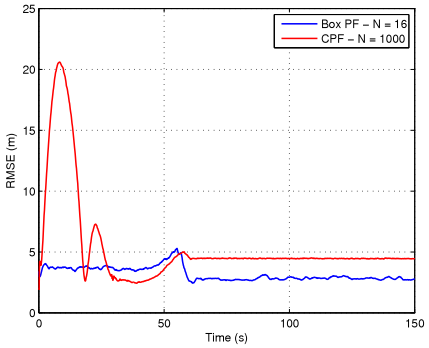
We define $\tilde{h}^{-1}(\cdot)$ as the inverse function of $\tilde{h}(\cdot)$. When the inverse function exists, a change of variable can be straightforwardly made that results in:

$$g\left(\boldsymbol{x}_k^m\right) = \begin{cases} \dfrac{1}{6\sigma}\left|\dfrac{d\left(\tilde{h}(\boldsymbol{x}_k^m)\right)}{d\boldsymbol{x}_k^m}\right| & : \boldsymbol{x}_k^m \in \mathcal{X} \\ 0 & : \text{elsewhere} \end{cases} \tag{A.3}$$

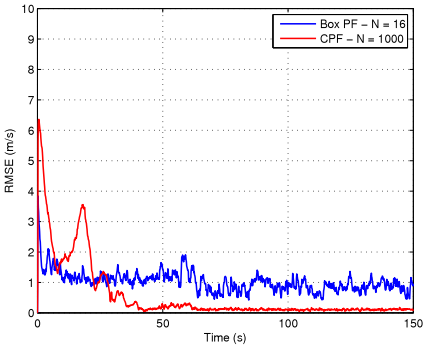
where $\mathcal{X} = [\tilde{h}^{-1}(\boldsymbol{z}_k^m - 3\sigma), \tilde{h}^{-1}(\boldsymbol{z}_k^m + 3\sigma)]$. Thus the integral in Eq. (A.1) is directly solvable in the following form:

$$\int U_{[\boldsymbol{z}_k^m]}\left(\tilde{h}\left(\boldsymbol{x}_k^m\right)\right) U_{q(\boldsymbol{x}_k)}\left(\boldsymbol{x}_k^m\right) d\boldsymbol{x}_k^m = \int g\left(\boldsymbol{x}_k^m\right) U_{q(\boldsymbol{x}_k)}\left(\boldsymbol{x}_k^m\right) d\boldsymbol{x}_k^m. \tag{A.4}$$

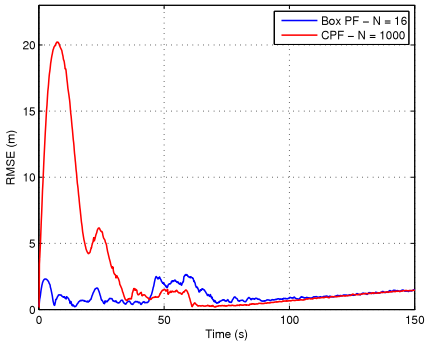
It is worth noting that after the transformation, the expression in (A.3) is not necessarily uniform.



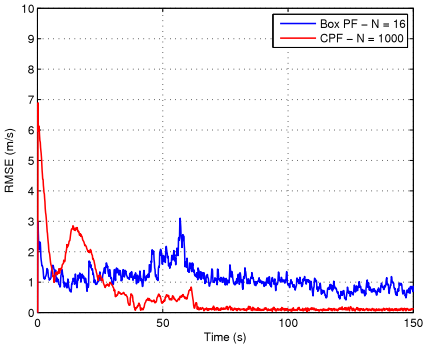
(a) RMSE corresponding to the location of the centre in the *x*-coordinate.



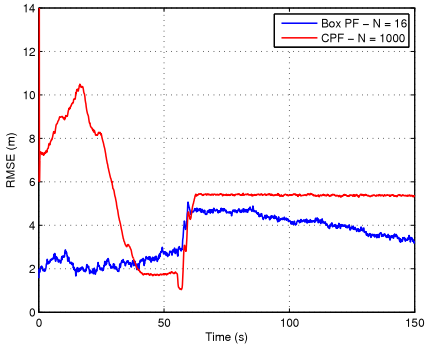
(b) RMSE corresponding to the velocity of the centre in the *x*-coordinate.



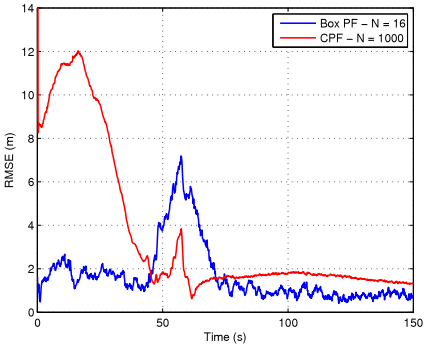
(c) RMSE corresponding to the location of the centre in the *y*-coordinate.



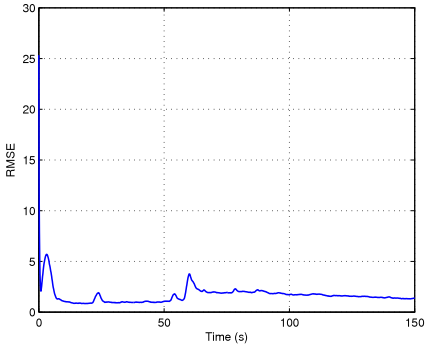
(d) RMSE corresponding to the velocity of the centre in the *y*-coordinate.



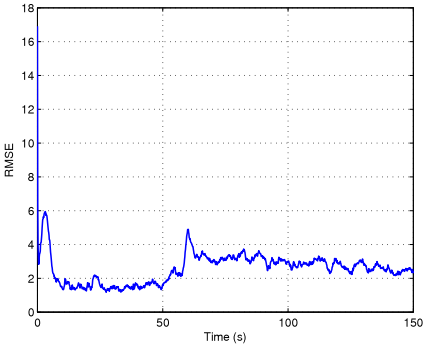
(e) RMSE corresponding to the length of side A.



(f) RMSE corresponding to the length of side B.



(g) RMSE corresponding to the crowd measurement rate estimated by the Box PF.



(h) RMSE corresponding to the clutter measurement rate estimated by the Box PF.

Fig. 6. RMSE of the box PF and CPF estimates for the realistic crowd simulator.

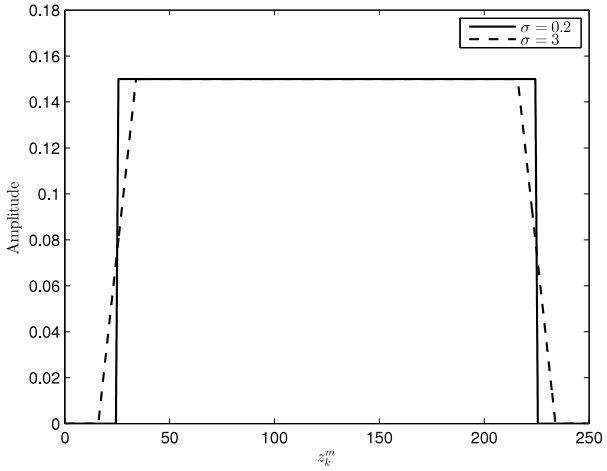


Fig. A.1. Example solution of Eq. (A.1) with $x_k = 10$ and $a_k = 10$ with varying σ.

A.1. The linear case

For the linear observation model, given in (12), the expression in (A.3) remains uniform:

$$g\left(\boldsymbol{x}_k^m\right) = \begin{cases} \dfrac{1}{6\boldsymbol{\sigma}} & : \boldsymbol{x}_k^m \in [\boldsymbol{z}_k^m - 3\boldsymbol{\sigma}, \boldsymbol{z}_k^m + 3\boldsymbol{\sigma}] \\ 0 & : \text{elsewhere.} \end{cases} \tag{A.5}$$

The range of the uniform distribution is dependent on the noise characteristics of the sensor. The range of the second uniform distribution, $U_{q(\boldsymbol{x}_k)}\left(\boldsymbol{x}_k^m\right)$, in (A.1) is dependent on the extent parameters of the target. The integral in (A.4) only exists when the two uniform distributions overlap. The overlapping region is defined by:

$$r(\boldsymbol{x}_k) = \begin{cases} x_k - \dfrac{a_k}{2} - 3\sigma_1 \leq z_{1,k}^m \leq x_k + \dfrac{a_k}{2} + 3\sigma_1 \\ y_k - \dfrac{b_k}{2} - 3\sigma_2 \leq z_{2,k}^m \leq y_k + \dfrac{b_k}{2} + 3\sigma_2. \end{cases} \tag{A.6}$$

The approximation in (A.1) is based upon the assumption that the length of the extent is significantly larger than the sensor noise characteristics. For instance, the case when the extent tends towards an infinite length with fixed sensor noise is also equivalent to an extent with a fixed length size, and with a sensor noise tending towards zero. In this case the uniform distribution in (A.5) tends towards the Dirac delta function, i.e.

$$g\left(\boldsymbol{x}_k^m\right) = \begin{cases} +\infty & : \boldsymbol{x}_k^m = \boldsymbol{z}_k^m \\ 0 & : \text{elsewhere.} \end{cases} \tag{A.7}$$

Consequently resulting in equivalence in (A.1):

$$\int g\left(\boldsymbol{x}_k^m\right) U_{q(\boldsymbol{x}_k)}\left(\boldsymbol{x}_k^m\right) d\boldsymbol{x}_k^m = U_{r(\boldsymbol{x}_k)}\left(\boldsymbol{z}_k^m\right). \tag{A.8}$$

In reality, the extent is not infinite, however in general, it is considered significantly larger than the range of the sensor noise. This is the motivating factor for the result in (A.1).

A.2. The non-linear case

A toy example is presented to illustrate the effect of a non-linear relationship between the sensor and a measurement source. Considering a single dimension with the following relationship,

$$z_k^m = \tilde{h}(x_k^m) + \xi_{1,k} = (x_k^m)^2 + \xi_{1,k}, \tag{A.9}$$

results in the following transformation:

$$g\left(x_k^m\right) = \begin{cases} \dfrac{1}{2\boldsymbol{\sigma}} x_k^m & : x_k^m \in [\sqrt{z_k^m - 3\sigma}, \sqrt{z_k^m + 3\sigma}] \\ 0 & : \text{elsewhere.} \end{cases} \tag{A.10}$$

In this case, the function $g\left(x_k^m\right)$ is clearly no longer uniform. An example of the solution of the integration in (A.4) is illustrated in Fig. A.1.

This example illustrates that although non-linearities may result in the non-uniformity of $g(x)$, when the extent parameters are significantly larger than the measurement error noise, a uniform approximation for Eq. (A.1) may still be valid. The effect of greater non-linearities is a topic for future research.

Appendix B. Crowd and clutter measurement rate estimation

The clutter rate λ_C and the crowd rate λ_T can be updated based on the assumption that they can be drawn from the Gamma distribution, similarly to Granström and Orguner (2012). For λ_C we have

$$p(\lambda_C|\boldsymbol{Z}_k) = \mathcal{GAM}(\lambda_C; \alpha_{k|k}^C, \beta_{k|k}^C)\mathcal{L}(\alpha_{k|k-1}^C, \beta_{k|k-1}^C, M_{C,k}) \tag{B.1}$$

and the updated parameters of the Gamma distribution for the clutter measurement rate are:

$$\begin{aligned} \alpha_{k|k}^C &= \alpha_{k|k-1}^C + M_{C,k}, \\ \beta_{k|k}^C &= \beta_{k|k-1}^C + 1. \end{aligned} \tag{B.2}$$

For λ_T the same relations as (B.2) are valid for the Gamma distribution parameters. In the box PF implementation, $M_{T,k} = \min_p \left|\mathcal{S}_E^{(p)}\right|$ and for the clutter measurement rate, $M_{C,k} = M_k - M_{T,k}$.

References

Abdallah, F., Gning, A., & Bonnifait, P. (2008). Box particle filtering for nonlinear state estimation using interval analysis. *Automatica*, *44*(3), 807–815.

Ali, I., & Dailey, M. N. (2009). Multiple human tracking in high-density crowds. In *LNCS: Vol. 5807. LNCS from advanced concepts for intelligent vision systems* (pp. 540–549).

Ali, S., Nishino, K., Manocha, D., & Shah, M. (Eds.) (2014). *Modeling, simulation and visual analysis of crowds*. Springer.

Angelova, D., Mihaylova, L., Petrov, N., & Gning, A. (2013). A convolution particle filtering approach for tracking elliptical extended objects. In *Proc. of the 16th international conference on information fusion, July* (pp. 1542–1549).

Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation* (1st ed.). Wiley-Interscience.

Campillo, F., & Rossi, V. (2009). Convolution particle filter for parameter estimation in general state-space models. *IEEE Transactions on Aerospace and Electronic Systems*, *45*(3), 1063–1072.

Cappe, O., Godsill, S. J., & Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, *95*(5), 899–924.

Carmi, A., Septier, F., & Godsill, S. J. (2012). The Gaussian mixture MCMC particle algorithm for dynamic cluster tracking. *Automatica*, *48*(10), 2454–2467.

Doucet, A., De Freitas, N., Gordon, N., et al. (2001). *Sequential Monte Carlo methods in practice. Vol. 1*. New York: Springer.

Gilholm, K., & Salmond, D. (2005). Spatial distribution model for tracking extended objects. *IEE Proceedings—Radar, Sonar and Navigation*, *152*(5), 364–371.

Gning, A., Mihaylova, L., & Abdallah, F. (2010). Mixture of uniform probability density functions for nonlinear state estimation using interval analysis. In *Proc. of 13th international conference on information fusion*. Edinburgh, UK.

Gning, A., Mihaylova, L., Abdallah, F., & Ristic, B. (2012). Particle filtering combined with interval methods for tracking applications. In M. Mallick, V. Krishnamurthy, & B.-N. Vo (Eds.), *Integrated tracking, classification, and sensor management: theory and applications* (pp. 43–74). New Jersey, USA: John Wiley & Sons.

Gning, A., Ristic, B., & Mihaylova, L. (2012). Bernoulli particle/box-particle filters for detection and tracking in the presence of triple measurement uncertainty. *IEEE Transactions on Signal Processing*, *60*(5), 2138–2151.

Gning, A., Ristic, B., Mihaylova, L., & Abdallah, F. (2013). An introduction to box particle filtering. *IEEE Signal Processing Magazine*, *30*(4), 1–7.

Grandström, K. (2012). *Extended object tracking using PHD filters* (Ph.D. dissertation), Sweden: Linköping University, No. 1476.

Granström, K., Lundquist, C., & Orguner, U. (2011). Tracking rectangular and elliptical extended targets using laser measurements. In *Proc. of the 14th international conference on information fusion* (pp. 1–8).

Granstrom, K., & Orguner, U. (2012). Estimation and maintenance of measurement rates for multiple extended target tracking. In *Proc. of 15th international conference on information fusion, July* (pp. 2170–2176).

Haj Chadé, H., Gning, A., Abdallah, F., Mougharbel, I., & Julier, S. (2014). Non parametric distributed inference in sensor networks using box particles messages. *Mathematics in Computer Science*, *8*(3–4), 455–478.
Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*, 4282–4286.
Jaulin, L. (2002). Nonlinear bounded-error state estimation of continuous-time systems. *Automatica*, *38*(6), 1079–1082.
Jaulin, L. (2009). Robust set-membership state estimation; application to underwater robotics. *Automatica*, *45*(1), 202–206.
Jaulin, L., Kieffer, M., Didrit, O., & Walter, E. (2001). *Applied interval analysis*. Springer-Verlag.
Koch, J. W. (2008). Bayesian approach to extended object and cluster tracking using random matrices. *IEEE Transactions on Aerospace and Electronic Systems*, *44*(3), 1042–1059.
Li, T., Bolic, M., & Djuric, P. M. (2015). Resampling methods for particle filtering: Classification, implementation, and strategies. *IEEE Signal Processing Magazine*, *32*(3), 70–86.
Liu, F., Gao, H., Qiu, J., Yin, S., Fan, J., & Chai, T. (2014). Networked multirate output feedback control for setpoints compensation and its application to rougher flotation process. *IEEE Transactions on Industrial Electronics*, *61*(1), 460–468.
Mahler, R. P. S. (2007). *Statistical multisource-multitarget information fusion*. Boston: Artech House.
Mahler, R. (2009). PHD filters for nonstandard targets, I: Extended targets. In *Proc. of the 12th international conference on information fusion* (pp. 915–921). Seattle, WA, USA, July.
Mahler, R. (2013). Statistics 102 for multisource-multitarget detection and tracking. *IEEE Journal of Selected Topics in Signal Processing*, *7*(3), 376–389.
Mahler, R.P.S., & Zajic, T. (2002). Bulk multitarget tracking using a first-order multitarget moment filter. In *Proceedings of SPIE 4729* (pp. 175–186).
Mazzon, R., & Cavallaro, A. (2013). Multi-camera tracking using a multi-goal social force model. *Neurocomputing*, *100*(1), 41–50.
Mehrotra, K., & Mahapatra, Pravas R. (1997). A jerk model for tracking highly maneuvering targets. *IEEE Transactions on Aerospace and Electronic Systems*, *33*(4), 1094–1105.
Mihaylova, L., Carmi, A. Y., Septier, F., Gning, A., Pang, S. K., & Godsill, S. (2014). Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking. *Digital Signal Processing*, *25*(1), 1–16.
Pellegrini, S., Ess, A., Schindler, K., & Van Gool, L. (2009). You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proc. of the IEEE 12th international conference on computer vision* (pp. 261–268).
Petrov, N., Gning, A., Mihaylova, L., & Angelova, D. (2012). Box particle filtering for extended object tracking. In *Proc. of the 15th international conference on information fusion, July* (pp. 82–89).
Petrov, N., Mihaylova, L., de Freitas, A., & Gning, A. (2014). Crowd tracking with box particle filtering. In *Proc. of 17th international conference on information fusion, July* (pp. 1–7).
Petrov, N., Ulmke, M., Mihaylova, L., Gning, A., Schikora, M., Wieneke, M., & Koch, W. (2012). On the performance of the box particle filter for extended object tracking using laser data. In *Workshop on sensor data fusion: trends, solutions, applications, September* (pp. 19–24).
Rossi, V., & Vila, J.-P. (2006). *Annales de l'Institut de Statistique de l'Université*, *50*(3), 71–102.
Schikora, M., Gning, A., Mihaylova, L., Cremers, D., Koch, W., & Streit, R. (2014). Box-particle intensity filter for multi-target tracking. *IEEE Transactions on Signal Processing*, *50*(3), 1660–1672. ieee transactions on signal processing.
Singer, R. A. (1970). Estimating optimal tracking filter performance for manned maneuvering targets. *IEEE Transactions on Aerospace and Electronic Systems*, *AES-6*(4), 473–483.
Vila, J.-P. (2012). Enhanced consistency of the resampled convolution particle filter. *Statistics & Probability Letters*, *82*(4), 786–797.
Wan, E. A., & Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *IEEE com. & control symp. adaptive systems signal proc.* (pp. 153–158). IEEE.
Wang, T., Gao, H., & Qiu, J. (2016). A combined adaptive neural network and nonlinear model predictive control for multirate networked industrial process control. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(2), 416–425.

Allan De Freitas received the B.Eng., B.Eng. (Hons) and M.Eng. degrees in Electronic Engineering from the University of Pretoria, South Africa in 2009, 2010 and 2013 respectively. He is currently working towards the Ph.D. degree in the Automatic Control and Systems Engineering department at the University of Sheffield, United Kingdom. His principal scientific interests are in the areas of signal processing in object tracking, communications, image processing, and complex systems.

Lyudmila Mihaylova (M'98, SM'2008) is Associate Professor (Reader in Advanced Signal Processing and Control) at the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom. Her research is in the areas of machine learning and autonomous systems with various applications such as navigation, surveillance and sensor network systems. Dr Mihaylova is an Associate Editor of the IEEE Transactions on Aerospace and Electronic Systems and of the Elsevier Signal Processing Journal.

Amadou Gning received the M.S. degree in "Technologies of Information and System" in 2002 and defended his Ph.D. in 2007 at the University of Technology of Compiegne, France. He is currently a Lecturer in the Department of Computer Science at the University of Hull, United Kingdom. He previously worked several years as a Research Engineer at Dyson Ltd, Research Associate at University College London and Lancaster University. His main interests are on Robotics applications and exploring mathematical tools ranging from Statistical methods to Set theoretic methods.

Donka Angelova received the M.S. degree in Mechanical Engineering from the Technical University of Sofia, Bulgaria, and the Ph.D. degree in Electrical Engineering from the Bulgarian Academy of Sciences, Institute for Parallel Processing, where she is an academic rank Associate Research Professor. She works as a key researcher and is involved in academic and industry projects in the areas of multitarget tracking and multisensor data fusion. Her current research interests include the application of sequential Monte Carlo methods to the problems of applied estimation and filtering.

Visakan Kadirkamanathan obtained his B.A in Electrical and Information Sciences at Cambridge University Engineering Department and went on to complete his Ph.D. in Information Engineering at the same institution. He is currently Professor in Signal and Information Processing and is also the Director of the Rolls-Royce University Technology Centre for Control and Monitoring Systems Engineering at Sheffield. He has published more than 170 papers in peer reviewed journals and conferences, and was the recipient of the PNAS Cozzarelli Prize in 2013.