Information Sciences 388–389 (2017) 172–190

Contents lists available at ScienceDirect

Information Sciences

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

Clustering for filtering: Multi-object detection and estimation using multiple/massive sensors

Tiancheng Li[a,b,*], Juan M. Corchado[a,c], Shudong Sun[b], Javier Bajo[d]

[a] *School of Science, University of Salamanca, Salamanca 37008, Spain*
[b] *School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China*
[c] *Osaka Institute of Technology, Asahi-ku Ohmiya, Osaka 535-8585, Japan*
[d] *Department of Artificial Intelligence, Technical University of Madrid, Madrid 28660, Spain.*

ARTICLE INFO

Article history:
Received 8 April 2016
Revised 7 January 2017
Accepted 9 January 2017
Available online 10 January 2017

Keywords:
Target detection
Object recognition
Filtering
Tracking
Sensor fusion

ABSTRACT

Advanced multi-sensor systems are expected to combat the challenges that arise in object recognition and state estimation in harsh environments with poor or even no prior information, while bringing new challenges mainly related to data fusion and computational burden. Unlike the prevailing Markov-Bayes framework that is the basis of a large variety of stochastic filters and the approximate, we propose a clustering-based methodology for multi-sensor multi-object detection and estimation (MODE), named clustering for filtering (C4F), which abandons unrealistic assumptions with respect to the objects, background and sensors. Rather, based on cluster analysis of the input multi-sensor data, the C4F approach needs no prior knowledge about the latent objects (whether quantity or dynamics), can handle time-varying uncertainties regarding the background and sensors such as noises, clutter and misdetection, and does so computationally fast. This offers an inherently robust and computationally efficient alternative to conventional Markov–Bayes filters for dealing with the scenario with little prior knowledge but rich observation data. Simulations based on representative scenarios of both complete and little prior information have demonstrated the superiority of our C4F approach.

1. Introduction

Multiple object/target detection and estimation (MODE), which lies at the core of multi-object recognition and tracking, involves the joint estimation of the number of targets and their states from noisy observations that are received at discrete time instants in the presence of clutter/false-alarms and misdetection. The standard and prevailing solution, whether for a single object or multiple objects, is based on the Markov–Bayes framework, which employs a hidden Markov model to describe the object's dynamics, and an observation function to link observed variables to the hidden states, namely the *state space model* (SSM). This SSM formulation, which functions by finding a Markov–Bayes filter, has been well demonstrated in the field for decades since the pioneering work of the Kalman filter. It has been forty years since the first survey of multi-object tracking methods [2] and the related topics are attracting more attention than ever with the introduction of the finite set statistics [30].

[*] Corresponding author at: School of Science, University of Salamanca, Salamanca 37008, Spain.
E-mail addresses: t.c.li@usal.es, tiancheng.li1985@gmail.com (T. Li), corchado@usal.es (J.M. Corchado), sdsun@nwpu.eud.cn (S. Sun), jbajo@fi.upm.es (J. Bajo).

http://dx.doi.org/10.1016/j.ins.2017.01.028

The concept of "simultaneous signal detection and estimation" appeared since as early as [34] where detection was treated as a special case of estimation, or estimation as a generalized detection process. The primary challenge for MODE involves a wide range of latent statistical models/uncertainties regarding object quantity and dynamics (including appearance, movement, deformation and disappearance), sensor profiles (such as irregular revisit frequency, misdetection and out-of-sequence data) and background (such as clutter); simply, real world noises can be distorted, correlated, colored and/or multiplicative but hardly be additive Gaussian which is required by Kalman filters. Most of the time, these statistical models are unknown and can only be approximated on the basis of online or offline data [35], namely *system modeling* (which is also referred to as *system identification* or the *design of the prior* [43]). Even if they can be accurately established to a sufficient extent, the full Bayesian, or even only the necessary likelihood [16,33,37,44], computation could be prohibitively intractable and approximation must be resorted to.

More specifically, the instability and inconsistency of the Bayesian inference for either linear or nonlinear models has been recognized [15,17]. It has been demonstrated that the use of a stochastic filter could be inefficient or counteractive in certain cases [23,36]. Recently, we proposed the concept of *probability of filter benefit* (PoFB) [23] based on the data-driven observation-only (O_2) inference to clarify this issue: it is only when the PoFB is larger than 0.5 that the filter is more preferable compared to the O_2 inference.

Despite immense ongoing efforts spent on developing new Markov–Bayes filters, MODE in a harsh environment with little prior information remains an open challenge, regardless of the real time computing burden. Nowadays, the advent of multiple/massive sensor systems provides very rich observation at high frequency yet low financial cost. This facilitates a novel perspective based on sensor data clustering to deal with false alarms and misdetection, named *clustering for filtering* (C4F) in this work, without relying on unrealistic statistical assumptions about the objects, sensors and the background. The proposed multi-sensor C4F methodology has comparably better 'frequentist' interpretation, and is expected to prove its mettle in the challenging scenario with little prior information.

The paper is organized as follows. Section 2 summarizes the motivation and outline of our methodology. Section 3 presents the idea of the multi-sensor O_2 inference in which typical sensor models used for object recognition and tracking are also discussed. Section 4 presents the C4F methodology in detail, which constitutes the core contribution of this paper. Section 5 presents simulation demonstrations including a comparison with state-of-the-art algorithms. Section 6 concludes the paper and draws attention to our future work.

2. Motivation and outline

The SSM design for Markov–Bayes inference involves choosing a dynamical function to describe the state evolving process over time, and an observation function to link observed variables to the hidden states. Hereafter, the uncertainties about the object quantity and dynamics, the background and sensor profiles (e.g. detection uncertainty and noise) all of which are inherently vague in reality, are jointly referred to as the *Statistical Model* information, which excludes the deterministic part of the observation function that is precisely given a priori. When designing optimal estimators it is unrealistic to assume that the statistical model is known perfectly. The issue is then to design a robust estimator that is optimal relative to an uncertainty class of processes [6]. Efforts have been devoted from various aspects, including the following, to name a few:

- Non-cooperative/intractable object motion: constrained [28,35,41]/sparse [1]/circular [19] state sequence and maneuvers [14,27,43], etc.,
- Unknown background and sensor uncertainties: noise characteristics [8], detection probability [13], clutter rate and detection profiles [31,46,49], network-induced delay [7], etc., and
- Imperfect data: asynchronous sensor data [14,42], intractable likelihood [16,33,37,44], measurement random latency [48], outlier [32] and continuous-time observation [12], etc.

Given that we are particularly interested in the mean-square error (MSE) of the estimation, the optimal estimator is the least mean squares (LMS, also referred to as minimum MSE, MMSE) estimator, which calculates the conditional expectation of the state $\boldsymbol{x}_t$ given all the collected observations $\boldsymbol{z}_{1:\,t}$ to the newest time t as follows:

$$\hat{\boldsymbol{x}}_t = E(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{z}_{1:t}) d\boldsymbol{x} \tag{1}$$

where the conditional/posterior distribution $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$ consists of the prior information (related to the state transition function) and the likelihood (related to the observation uncertainty), both of which depend on the statistical models over the time series $\mathfrak{m}_{1:t}$.

When the statistical model is not given a priori, a general Bayesian solution shall estimate the statistical model $\mathfrak{m}_{1:t}$ jointly with the state $\boldsymbol{x}_t$ (also referred to as hierarchical Bayesian modeling), i.e.,

$$p(\boldsymbol{x}_t, \mathfrak{m}_{1:t}|\boldsymbol{z}_{1:t}) = p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}, \mathfrak{m}_{1:t}) p(\mathfrak{m}_{1:t}|\boldsymbol{z}_{1:t}) \tag{2}$$

Given a model set $\mathcal{M}$ consisting of a sufficient number of candidate models, the full probability formula and the Bayes' rule show us

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) &= \sum_{\forall i\in[1,t]:\,\mathfrak{m}_i\in\mathcal{M}} p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t},\mathfrak{m}_{1:t})p(\mathfrak{m}_{1:t}|\boldsymbol{z}_{1:t}) \\ &= \frac{1}{p(\boldsymbol{z}_{1:t})}\sum_{\forall i\in[1,t]:\,\mathfrak{m}_i\in\mathcal{M}} p(\boldsymbol{x}_t|z_{1:t},\mathfrak{m}_{1:t})p(\boldsymbol{z}_{1:t}|\mathfrak{m}_{1:t})p(\mathfrak{m}_{1:t}) \\ \text{where } p(\boldsymbol{z}_{1:t}) &= \sum_{\forall i\in[1,t]:\,\mathfrak{m}_i\in\mathcal{M}} p(\boldsymbol{z}_{1:t}|\mathfrak{m}_{1:t})p(\mathfrak{m}_{1:t}). \end{aligned} \tag{3}$$

Clearly, the model set $\mathcal{M}$ shall be complete to contain all significant potential candidates (if not impossible), which however will cause remarkable computing requirements. A useful strategy in this aspect is to adaptively interact between the candidate models assumed in the set $\mathcal{M}$ which can save the required number of parallel filters, such as the so-called interacting multiple model (IMM) [27]. Still, the real system may change so frequently that the algorithm can hardly catch up, or even the best models may be inadequate for representing key features of the real system, suffering from erroneous fractions. The situation will become much worse in the case of tracking an unknown and time-varying number of maneuvering objects by using heterogeneous sensors. In any case, the performance of the Markov–Bayes estimator depends greatly on the coinciding degree between the real system and the model assumed [27,45].

Compared with single object detection and estimation (SODE), clutter and misdetection will significantly affect the object detection (which is related to determining the number of objects). Even in an ideal situation where the dynamic models are established and objects are detected perfectly, the correspondence between the observations and the objects in the cluttered environment is often ambiguous and poses another fundamental challenge. The way in which this correspondence is identified distinguishes two main groups of existing solutions. First, conventional solutions decompose the MODE problem into multiple SODE problems based on observation-to-object association, which is inherently able to provide complete 'track' information. One of the main obstacles in this aspect is that the number of possible associations will exponentially increase over time and must be merged or pruned at the expense of reducing precision for real time filtering. Secondly, solutions based on finite set statistics such as the probability hypothesis density (PHD) filters and multi-Bernoulli filters [30,31,46] provide an approximation of the multi-object Bayesian recursion without explicit observation-to-object association. In both cases, approximation is indispensable for both modeling and computing in nonlinear systems.

On the other hand, the rapid development and joint deployment of sensors (with a gradually improved revisit frequency and accuracy but a lower cost) facilitate combating either the poor prior knowledge or individual sensor failures, providing richer and more reliable observation [5,18]. For instance, an ordinary smartphone encompasses a variety of sensors that permit multi-source WIFI, Bluetooth, GPS signals as well as altitude, acceleration, direction and even context information based on embedded software, which can all be explored for the same task of "mobile positioning". While the hardware strength will increase the capacity of the system, leading to a favorable advantage for more accurate estimation, it will also pose new challenges for optimal data fusion and real time computation. It is worth noting that, the more accurate the data are, the more sensitive the estimator is to model errors [23].

Given all the problems, challenges and opportunities for multi/massive-sensor MODE in harsh environments with little prior information, an interesting question is whether we have other efficient solutions that can avoid the need to design a sophisticated, computationally intensive and model-sensitive filter in order to maximally meet the requirement for reliable and real time MODE. This forms the starting point of our work, by which we advocate the use of clustering technologies for filtering, named C4F. The present C4F solution comprises three major steps:

1) Project all (synchronous) sensor observations from the observation space to the state space, to generate multiple detections that correspond to real objects and false alarms; the detail is given in Sections 3.1–3.2.
2) Cluster the detections according to their spatial distribution density and certain constraints: those congregating/ clustering correspond to the same or close object(s) while those that are scattered will be identified as false alarms (and discarded); the detail is given in Sections 4.1–4.2.
3) The detections that are identified as belonging to the same object will be fused in the manner of least squares estimation, which forms a final state-estimate of that object; the detail is given in Section 4.3.

The present multiple sensor C4F approach differs from existing filter-based target detection algorithms by releasing the need to make statistical assumptions about the object quantity and dynamics, sensor detection profiles and even observation noises. We noticed a few works that employ clustering within filters, e.g., [20] and various adaptive/robust estimators that have been proposed to accommodate little/poor prior information; however, their core is a model-specific filter. While a limited part of this work appeared earlier in [21] for identically distributed sensors only, new and detailed content will be presented in this paper.

3. Multi-sensor O$_2$ inference

The concerned multi-sensor MODE scenario is general, which can be described as follows: an unknown and time-varying number of objects evolve in terms of birth, disappearance, deformation (including merging and splitting) and maneuvering both in state space and in time. A number of conditionally independent sensors are used for observing the scenario, whose observation function and positions are known. The objects may be detected or missed by the sensors with unknown probabilities, perhaps at irregular observation rates, in the presence of latent and time-varying significance levels of clutter and noises. To avoid deeper technical issues that detract from the core idea of our work, some common assumptions are made to the objects, clutter and sensors, respectively:

(A.1) Each **object** generates observations independently of the others, and one object generates no more than one observation at each sensor at each scan. In other words, we contend with only the point-object (regardless of extended object or multiple-observation due to reflection) and neither convoys nor engagements are involved.

(A.2) The **clutter** will not congregate or cluster in the view field of each sensor. This assumption is restrictive albeit common in the literature since in special cases clutter can congregate in closely placed sensors. For instance, spikes due to the sea surface returns are likely to be detected by multiple maritime radars; however, we will not account for tricky cases like this at the present time.

(A.3) All **sensors** work independently, synchronously and the average object detection probability they provide is not too low, otherwise tracking before detection (TBD) is advised.

3.1. Projecting observations to the state space

In the general formulation, the sensor performs periodic scans based on a known observation function $\boldsymbol{h}_t(\cdot)$

$$\boldsymbol{z}_t = \boldsymbol{h}_t(\boldsymbol{x}_t, \boldsymbol{v}_t) \tag{4}$$

where t indicates the observation time – instant, $\boldsymbol{x}_t$ denotes the state, $\boldsymbol{z}_t$ the observation, and $\boldsymbol{v}_t$ the observation noise.

As the concern of this paper, no statistical information is feasible about the target motion and the clutter. We resort to the pure sensor data for detection and estimation, namely O$_2$ inference. It can be conceptually written as follows (as long as the observation function is generally invertible):

$$\hat{\boldsymbol{\chi}}_t = \boldsymbol{h}_t^{-1}(\boldsymbol{z}_t, \bar{\boldsymbol{v}}_t) \tag{5}$$

where $\boldsymbol{h}_t^{-1}$ is the "generalized" inversing function of $\boldsymbol{h}_t$ in real variable space, $\hat{\boldsymbol{\chi}}_t$ is the inference of "the observed component of" the state $\boldsymbol{x}_t$ and $\bar{\boldsymbol{v}}_t$ is an average used to compensate the observation error and can be specified as the mean of the noise distribution $E(\boldsymbol{v}_t)$. In a fully observed system, $\hat{\boldsymbol{\chi}}_t$ is a full-dimensional state estimate $\hat{\boldsymbol{x}}_t$ which should be the best fit of the observation data, i.e.,

$$\hat{\boldsymbol{x}}_t = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\| \boldsymbol{z}_t - \boldsymbol{h}_t\left(\boldsymbol{x}, \bar{\boldsymbol{v}}_t\right) \right\| \tag{6}$$

or in a time series manner,

$$\hat{\boldsymbol{x}}_{t_1:t_2} = \underset{\boldsymbol{x}_{t_1:t_2}}{\operatorname{argmin}} \sum_{t=t_1}^{t_2} \left\| \boldsymbol{z}_t - \boldsymbol{h}_t\left(\boldsymbol{x}_t, \bar{\boldsymbol{v}}_t\right) \right\| \tag{7}$$

where $\|\boldsymbol{x}-\boldsymbol{y}\|$ is a measure of the distance to be defined between $\boldsymbol{x}$ and $\boldsymbol{y}$, $[t_1, t_2]$ is the time window under consideration and the summation is over all the sampling instants in the time window.

However, if any target motion model information is given a priori or can be learned from the online sensor data (e.g., the state evolving is a stationary process), such as the deterministic part of the state transition [17], state constraints [47], model error [3] or even only a context-information like "trajectory is smooth" [25], it should be taken into account in the formulation of (7), to be addressed in Section 6.

When the statistics of the sensor noise is available, the Monte Carlo (MC) inversing approach can be applied to accommodate nonlinear observation function $\boldsymbol{h}_t$ for unbiased estimation, for which a set of samples are randomly sampled from the noise distribution $\boldsymbol{v}_t^{(i)} \sim \boldsymbol{v}_t, i = 1, 2, \ldots I$ in order to carry out the inversing calculation of (5), i.e., for i=1, 2, …I

$$\hat{\boldsymbol{\chi}}_t^{(i)} = \boldsymbol{h}_t^{-1}\left(\boldsymbol{z}_t, \boldsymbol{v}_t^{(i)}\right) \tag{8}$$

Then, it is straightforward to calculate the unbiased estimate and its covariance as follows

$$\hat{\boldsymbol{\chi}}_t = \frac{1}{I}\sum_{i=1}^{I} \hat{\boldsymbol{\chi}}_t^{(i)} \tag{9}$$

$$\operatorname{Cov}\left(\hat{\boldsymbol{\chi}}_t\right) = \frac{1}{I-1}\sum_{i=1}^{I}\left(\hat{\boldsymbol{\chi}}_t^{(i)} - \hat{\boldsymbol{\chi}}_t\right)\left(\hat{\boldsymbol{\chi}}_t^{(i)} - \hat{\boldsymbol{\chi}}_t\right)^{\mathrm{T}} \tag{10}$$

To speed up the computation, the samples can be created in a deterministic manner such as sigma-points. We must note that the sensor uncertainty cannot be exactly known in the real world, not to say that only suffering from a single white noise. If the statistics of $\boldsymbol{v}_t$ is unknown, an approximate error such as that caused by omitting the noise will not affect the O$_2$ inference as much as does to the sequential Bayesian filters.

More challenging, the observation function may be mathematically irreversible and therefore prevents the direct inversing calculation. Practical treatments to this issue, referred to as dynamic inverse problem [38], can be found. However, in the context of object tracking, sensors are limited to a few types and the joint use of multiple or even massive sensors can improve the observability to circumvent the irreversibility. In the case of massive sensors, it would be interesting to find an optimal observation configuration by properly grouping the sensors based on their spatial distribution (such as minimizing the squared position error bound [40]) and optimally fusing their results. In this paper, we concentrate on the case that each sensor is a sufficient sensory unit for simplicity.

3.2. O_2 inference on typical tracking sensors

We consider three representative sensors for observation in the two dimensional Cartesian coordinate.

Position observation (e.g., visual camera): the observation is directly made on the planar position $[p_{x,\,t},\, p_{y,\,t}]^{\mathrm{T}}$ of the object, given by

$$\boldsymbol{z}_t = \begin{bmatrix} z_{x,t} \\ z_{y,t} \end{bmatrix} = \begin{bmatrix} p_{x,t} \\ p_{y,t} \end{bmatrix} + \boldsymbol{v}_t \tag{11}$$

where $[z_{x,\,t},\, z_{y,\,t}]^{\mathrm{T}}$ is the observed position of the object in x and y dimensions, respectively.

As addressed, the unbiased O_2 inference is given by

$$\begin{bmatrix} \hat{p}_{x,t} \\ \hat{p}_{y,t} \end{bmatrix} = \begin{bmatrix} z_{x,t} \\ z_{y,t} \end{bmatrix} - E(\boldsymbol{v}_t) \tag{12}$$

The variance of the estimate is equivalent to that of the observation noise $\boldsymbol{v}_t$.

Range-bearing observation (active radar): the observation is a noisy range and bearing vector, given by

$$\boldsymbol{z}_t = \begin{bmatrix} r_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \sqrt{(p_{x,t} - S_{x,t})^2 + (p_{y,t} - S_{y,t})^2} \\ \arctan\left(\frac{p_{y,t} - S_{y,t}}{p_{x,t} - S_{x,t}}\right) \end{bmatrix} + \boldsymbol{v}_t \tag{13}$$

where $[p_{x,t},\, p_{y,t}]^{\mathrm{T}}$ and $[S_{x,t},\, S_{y,t}]^{\mathrm{T}}$ are the position of the object and of the sensor, respectively.

To implement the O_2 inference, one estimate can be inferred by inversing (13) after taking off $\boldsymbol{v}_t$, yielding

$$\begin{bmatrix} \hat{p}_{x,t} \\ \hat{p}_{y,t} \end{bmatrix} = +/- \begin{bmatrix} \sqrt{\frac{r_t^2}{1+\theta_t^2}} \\ \tan(\theta_t)\sqrt{\frac{r_t^2}{1+\theta_t^2}} \end{bmatrix} + \begin{bmatrix} S_{x,t} \\ S_{y,t} \end{bmatrix} \tag{14}$$

As shown, inversing the arctan function involves a sign problem, denoted by ¨+/−¨ in (14). Often, the state is constrainted in either positive or negative state space and the sign problem is avoided; see our simulation in Section 5.1. Otherwise at least two active sensors placed at different positions are needed to infer the sign by triangulation.

As stated, the estimate given by (14) by simply abandoning the noise is actually biased even if $E(\boldsymbol{v}_t)=0$. If the statistics of the noise $\boldsymbol{v}_t=[v_{1,\,t},\, v_{2,\,t}]^{\mathrm{T}}$ is known, the MC debiasing scheme can be applied, i.e., sampling a set of samples according to the noise distribution $[v_{1,t}^{(i)}, v_{2,t}^{(i)}]^{\mathrm{T}} \sim p(\boldsymbol{v}_t), i = 1, 2, \ldots I$, yielding simulating estimates as follows:

$$\begin{bmatrix} \hat{p}_{x,t}^{(i)} \\ \hat{p}_{y,t}^{(i)} \end{bmatrix} = +/- \begin{bmatrix} \sqrt{\frac{\left(r_t - v_{1,t}^{(i)}\right)^2}{1+\left(\theta_t - v_{2,t}^{(i)}\right)^2}} \\ \tan\left(\theta_t - v_{2,t}^{(i)}\right)\sqrt{\frac{\left(r_t - v_{1,t}^{(i)}\right)^2}{1+\left(\theta_t - v_{2,t}^{(i)}\right)^2}} \end{bmatrix} + \begin{bmatrix} S_{x,t} \\ S_{y,t} \end{bmatrix} \tag{15}$$

The mean and variance of the estimate (for each object) can be obtained respectively by

$$\begin{bmatrix} \hat{p}_{x,t} \\ \hat{p}_{y,t} \end{bmatrix} = \begin{bmatrix} \frac{1}{I}\sum_{i=1}^{I} \hat{p}_{x,t}^{(i)} \\ \frac{1}{I}\sum_{i=1}^{I} \hat{p}_{y,t}^{(i)} \end{bmatrix} \tag{16}$$

$$\begin{bmatrix} Var\left(\hat{p}_{x,t}\right) \\ Var\left(\hat{p}_{y,t}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{I-1}\sum_{i=1}^{I} \left(\hat{p}_{x,t}^{(i)} - \hat{p}_{x,t}\right)^2 \\ \frac{1}{I-1}\sum_{i=1}^{I} \left(\hat{p}_{y,t}^{(i)} - \hat{p}_{y,t}\right)^2 \end{bmatrix} \tag{17}$$

Bearing-only observation (passive radar): the observation is bearing only, given by

$$\theta_t = \text{actan}\left(\frac{p_{y,t} - S_{y,t}}{p_{x,t} - S_{x,t}}\right) + v_t \tag{18}$$

To improve the observability, two bearing sensors located at different positions can be coordinated to infer the position of objects by triangulation. Assuming that they are located at positions $[S_{x1,t}, S_{y1,t}]^{\mathrm{T}}$ and $[S_{x2,t}, S_{y2,t}]^{\mathrm{T}}$ and report bearing $\theta_{1,t}$ and $\theta_{2,t}$, respectively, the estimate $[\hat{p}_{x,t}, \hat{p}_{y,t}]^{\mathrm{T}}$can be calculated by solving the following equation set

$$\begin{cases} \theta_{1,t} = \text{actan}\left(\frac{\hat{p}_{y,t} - S_{y1,t}}{\hat{p}_{x,t} - S_{x1,t}}\right) \\ \theta_{2,t} = \text{actan}\left(\frac{\hat{p}_{y,t} - S_{y2,t}}{\hat{p}_{x,t} - S_{x2,t}}\right) \end{cases} \tag{19}$$

which gives (ignoring the sign problem here by treating it as positive)

$$\begin{cases} \hat{p}_{x,t} = \frac{\tan\theta_1 S_{x1,t} - \tan\theta_2 S_{x2,t} + S_{y2,t} - S_{y1,t}}{\tan\theta_1 - \tan\theta_2} \\ \hat{p}_{y,t} = \frac{\tan\theta_1 \left(\tan\theta_2 S_{x1,t} - \tan\theta_2 S_{x2,t} + S_{y2,t} - S_{y1,t}\right)}{\tan\theta_1 - \tan\theta_2} + S_{y1,t} \end{cases} \tag{20}$$

To guarantee that (20) uniquely exists, neither $[S_{x1,t}, S_{y1,t}]^{\mathrm{T}}$ and $[S_{x2,t}, S_{y2,t}]^{\mathrm{T}}$ nor $\theta_{1,t}$ and $\theta_{2,t}$ can be the same, namely the equation set needs to be consistent. As stated, if the noise v_t is exactly given, debiasing shall be applied.

Likewise, if the observation is only a range (or equivalent, such as range-rate [50]), at least two sensors placed at different positions are required to infer the planar position/velocity, which is omitted here. In fact, the idea of projecting the sensor data to the state space has been involved in wireless triangulation, trilateration and multilateration positioning technologies based on angle of arrival, signal strength observations and time difference of arrival information, respectively; see e.g., [29].

3.3. Observability

One active range-bearing sensor or one position-observed sensor suffices to conduct O_2 inference and is referred to as a sufficient sensory sensor; to do so, at least two range-only or bearing-only sensors located at different positions are required to work jointly for position estimation. We will not emphasize this issue here as a sufficient number of sensors are used in our case and the observability is always strongly guaranteed. The challenge here is not under-observability but over-observability.

Clearly, the basic O_2 approach only estimates the dimensions of the state that have been observed. Note that even in filters the unobserved dimensions are inferred from the observed dimensions based on the physical law contained in the Markov transition matrix, e.g., the differentiation of the position over time is the velocity and the differentiation of the velocity is the acceleration. Therefore, based on successive estimates of position or velocity, we can infer one from the other. When further target motion information is available, such as “nearly constant velocity”, the estimates can be coordinated by fitting them over time-series and their accuracy be improved [25]. In this paper, we are only interested in object detection and position estimation.

4. Multi-sensor data clustering for MODE

As assumed in (A.1–3), all sensors work synchronously, reporting massive conditionally independent detections in the observation space. Realizing the O_2 inference on these detections will generate corresponding potential estimates (also called detections in the following) in the state space, consisting of the real estimates of objects and false alarms because of the clutter. In the sequence, we will present a constrained clustering approach to distinguish the real estimates of individual objects from each other and from the false alarms, which is the core task required for MODE as compared with SODE, and forms a novel perspective for “decluttering”.

4.1. Expected cluster size

The cluster size is given by the number of detections in the cluster, which depends on both the number of sensors whose field of view (FoV) covers that cluster, and their respective detection probabilities. In most cases, the sensor’s detection capacity is state-sensitive and therefore, we denote the object detection probability of sensor s at area a as $p_{D,s}(a)$ which is smaller than 1. Denoting the set of all sensors as S and the set of sensors whose FoV covers area a as S_a, the expected number of sensors that detect an object existing at area a is given by

$$E(n_a) = \textstyle\sum_{s \in S_a} p_{D,s}(a) \lesssim |S_a| \tag{21}$$

where n_a is the number of detections reported (for a single object) at area a, $|S_a|$ gives the number of elements in set S_a and “ $\lesssim$ ” means “smaller than or approximately equal to” in which “approximately equal to” is because clutter may fall in the cluster.

Algorithm 1 Multi-sensor data clustering.

```
Step 1. Project all sensor observations into the same Cartesian coordinate based on the $O_2$ inference, obtaining undistinguished detections
    corresponding to either real objects or false alarms; see Section 3.
Step 2. Detections from different sensors will be identified as connected if they satisfy (22) and (24).
Step 3. Calculate the number of detections in each cluster and identify whether the cluster is over-sized via (25). If $k_a \geq 2$, cluster $C_a$ will be further
    partitioned into $k_a$ sub-clusters conditioned on (21–22).
```

4.2. Decluttering via clustering

Based on (A1–3), the detections of a single object given by different sensors will cluster locally, while false alarms will not. Consequently, we have the following criterion to distinguish the real objects from false alarms:

Criterion 1 *The detections are of high density in the area containing objects and are of low density anywhere else.*

While a general density-based clustering method (such as DBSCAN [9]) that distinguishes the areas of high data density from low-density regions will be helpful to distinguish distant objects, we also need to distinguish between close objects (namely overlapping clusters) that share the same cluster. To do so, two constraints should be satisfied:

Constraint 1 *The detections reported by the same sensor cannot be clustered into the same cluster.*

Constraint 2 *The size of each cluster shall be determined such that (21) is respected.*

Constraint 1 is due to the point-object assumption (A.1) and is referred to as the 'cannot link' (CL) constraint. Denote all the detections generated in sensor s as a set $\mathbf{X}_s = \{\boldsymbol{x}_1^s, \boldsymbol{x}_2^s, \ldots \boldsymbol{x}_{m_s}^s\}$ where $m_s = |\mathbf{X}_s|$ is the total number of detections. The CL constraint states that

$$c_{\neq}\left(\boldsymbol{x}_i^s, \boldsymbol{x}_j^s\right), \ \forall i \neq j \tag{22}$$

where $c_{\neq}(\boldsymbol{x}_i^s, \boldsymbol{x}_j^s)$ means that $\boldsymbol{x}_i^s, \boldsymbol{x}_j^s$ cannot belong to the same cluster/object. To this end, their distance $\|\boldsymbol{x}_i^s - \boldsymbol{x}_j^s\|$ can be treated as infinite in the distance-based clustering algorithm design. In the Cartesian coordinate for $\boldsymbol{x}_i^s = [p_{x,i}^s, p_{y,i}^s]^{\mathrm{T}}, \boldsymbol{x}_j^m = [p_{x,j}^m, p_{y,j}^m]^{\mathrm{T}}$, the ℓ_2 norm Euclidean distance $\| \boldsymbol{x}_i^s - \boldsymbol{x}_j^m \|_2$ is given by

$$\| \boldsymbol{x}_i^s - \boldsymbol{x}_j^m \|_2 := \left(\boldsymbol{x}_i^s - \boldsymbol{x}_j^m\right)^{\mathrm{T}}\left(\boldsymbol{x}_i^s - \boldsymbol{x}_j^m\right) = \sqrt{\left(p_{x,i}^s - p_{x,j}^m\right)^2 + \left(p_{y,i}^s - p_{y,j}^m\right)^2} \tag{23}$$

The proposed constrained density-based clustering method can be described as shown in Algorithm 1. In Step 2 of the algorithm, a 'neighbor distance' threshold (which resembles the parameter "neighbor radius ε" used in DBSCAN and gives the maximum distance between two detections from different sensors for their direct connection to be included in one cluster) is required to associate detections from different sensors that very likely correspond to the same object. Our solution is to determine it according to the covariance of the distribution of all detections for that object, reflecting the average observation accuracy of all sensors. Considering the most common case of ellipsoidal cluster shape, we propose to use the rule of ellipsoidal validation, i.e., two detections $\boldsymbol{x}_i^s, \boldsymbol{x}_j^m$ are considered to be included in the same cluster if they satisfy

$$\left(\boldsymbol{x}_i^s - \boldsymbol{x}_j^m\right)^{\mathrm{T}}\left(\mathrm{Cov}\left(\hat{\boldsymbol{x}}_t\right)\right)^{-1}\left(\boldsymbol{x}_i^s - \boldsymbol{x}_j^m\right) \leq l_1 \tag{24}$$

where $\mathrm{Cov}(\hat{\boldsymbol{x}}_t)$ is the covariance of the estimation given by the proposed MC sampling method as shown in (10), on average over all sensors and l_1 is the first parameter needed in our approach. We suggest a scaling scope $l_1 \in [1, 4]$.

When multiple objects appear closely, their detections can be easily clustered into one cluster. Therefore, to partition the over-sized cluster, another threshold ρ is needed to give the average number of detections in a single-object cluster. It will be designed with respect to the expected number $E(n_a)$ of detections for a single object at local, e.g., $\rho_a = l_2 \times E(n_a)$, where $0 < l_2 < 1$ is scalable and involves a trade-off between missing detections and causing false alarms. We recommend $l_2 \in [0.6, \ 0.9]$. According to Constraint 2, the number k_a of sub-clusters contained in cluster C_a satisfies

$$k_a \rho_a \leq |C_a| < (k_a + 1)\rho_a \tag{25}$$

Given that the number of sub-clusters is determined by (25), the over-sized cluster can be then easily partitioned since false alarm is no longer involved and the standard k-means clustering algorithm is applicable. To note, each sub-cluster shall have approximately ρ_a but no more than $|S_a|$ detections while satisfying (22).

Remark 1. Scaling parameters l_1 and l_2 imply confidence levels assigned to coordinating the *neighbor distance* ε and the *cluster size* ρ, with reference to the sensors' observation noise variance and detection probabilities. They are essentially correlated, as a larger l_1 indicates more detections to be associated in each cluster and correspondingly a larger l_2 to be used.

Remark 2. In the case of extended/group objects, the cluster can be of non-ellipsoidal shapes, for which extra constraints have to be made on the cluster shape and taken into account for finer partitioning of the overlapped cluster in Step 3 of Algorithm 1.

In this paper, we assume the sensors' profiles are either known or easy to approximate in order to facilitate the calculation of (21) and (24), which is indeed a fair assumption commonly made. As long as not too many objects simultaneously

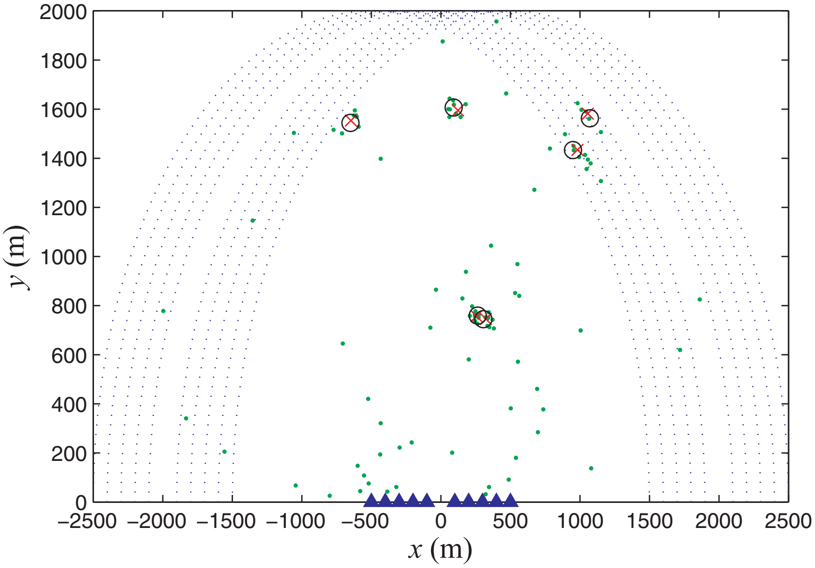

Fig. 1. Real object-positions (red "x"), detections (green '.'), clustering estimates (black "o") and FoV bounds (blue circle) of 10 sensors (located at '▲'). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

appear at the same place (which are hard to differentiate), a slight change of l_1 or l_2 will not affect the result much. This adds to the flexibility and robustness of our approach, which accommodates inaccurate knowledge of the sensors and is little sensitive to the accuracy of the parameters.

Moreover, adaptive online learning procedures were proposed to deal with unknown sensor profiles including the noise statistics and detection probability in [22], in which the neighbor distance ε and the cluster size ρ can not be calculated directly as addressed, but must be learned from the sensor data. To this end, the average number of detections that originate from the same sensor in the over-sized cluster was used to approximate the single-target cluster size. Then, instead of (25), for over-sized cluster C_a, the number of sub-clusters can be given as

$$k_a \approx \left[\frac{1}{|S_a|} \sum_{s \in S_a} |\boldsymbol{X}_s \cap C_a| \right] \tag{26}$$

where $[\cdot]$ gives the nearest integer to the content. Clearly, a larger $|S_a|$ indicates a better estimation. Given that only one object exists in cluster C_a, we have $E(|\boldsymbol{X}_s \cap C_a) = 1$.

For illustration purposes, an example is given in Fig. 1 for one scenario of the simulation conducted in Section 5.1. In this example, 10 homogeneous range-bearing sensors (marked as '▲') of different FoVs (their bounds are given in individual blue circles) are placed on a horizontal line. The detections from different sensors are all projected into the same planar space as shown by the green dots in the figure. Using the proposed constrained clustering method (with parameters $l_1 = 2$, $l_2 = 0.6$), detections are clustered and fused (see the next subsection), obtaining the final estimates (marked by black 'o') which are apparently quite close to the true position of the six objects (marked by red 'x'). In particular, close objects are correctly distinguished from each other and the number of objects is correctly estimated. Quantitative analysis will be given in Section 5.1.

4.3. MMSE cluster fusion

Given that the real object-detections are distinguished from false alarms and from each other by the given clustering algorithm, and that their variances are calculated by the proposed MC sampling approach, we can next fuse these relevant detections contained in each cluster to extract the final state-estimate for each object.

Denoting the mean and (co)variance of the detections (which are not necessary Gaussian variables) in the same cluster as $\boldsymbol{m}_i$ and $\mathbf{P}_i$, $i = 1, \ldots N$ respectively, the best linear unbiased estimator (BLUE, also referred to as linear MMSE estimator) fuses these conditionally independent detections according to their (co)variance as

$$\boldsymbol{m}_{\{1,2,\ldots N\}} = \frac{\sum_{i=1}^{N} \mathbf{P}_{\boldsymbol{i}}^{-1} \boldsymbol{m}_i}{\sum_{i=1}^{N} \mathbf{P}_{\boldsymbol{i}}^{-1}} \tag{27}$$

If all these detections are subject to Gaussian distribution, the final estimate is also Gaussian and its variance is

$$\mathbf{P}_{\{1,2,\ldots N\}} = \left(\sum_{i=1}^{N} \mathbf{P}_{\boldsymbol{i}}^{-1}\right)^{-1} \tag{28}$$

Furthermore, if these sensors are of the same accuracy over the surveillance area (namely the same observation noise of variance), then (27–28) reduce to

$$\boldsymbol{m}_{\{1,2,\ldots N\}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{m}_i \tag{29}$$

$$\mathbf{P}_{\{1,2,\ldots N\}} = \frac{\mathbf{P}}{N} \tag{30}$$

Remark 3. Eq. (29) is free of **P** indicating that the fusion can be optimally carried out without exactly knowing the statistics of the sensors as long as they are known to be of the same accuracy. This facilitates MMSE data fusion of homogeneous sensors of unknown statistics.

When the detection is defined in constrained, bounded or circular state spaces, constrained fusion will be involved. The reader is referred to [19,35] and the references therein. Compared to a Markov–Bayes filter, the proposed C4F approach may not be precise when few sensors of low quality are used, but its accuracy will be surely improved with the increase in the number of unbiased sensors as shown in (28)/(30). This however may not hold in practice for a Markov-model-specific filter, to be shown in our simulation study next.

While (parametric) Cramér–Rao and Weiss–Weinstein lower bounds [10] imply the best possible performance (in the sense of lowest mean square error) of any estimator, the O_2 inference gives the worst performance of any *effective* estimators with respect to a particular observation sequence. In this sense, the C4F approach that extends the O_2 approach to the multi-object case in cluttered environments therefore indicates the worst performance of any *effective* MODE solutions with respect to particular sequences of observations of multiple sensors. We will show next that based on this definition, the suboptimal filters are not always *effective*.

5. Simulations

In this section, two representative scenarios of respective complete and little prior information are considered. The well-defined optimal sub-pattern assignment (OSPA) metric [39] is used to evaluate the multi-object estimation accuracy. For finite subsets $\boldsymbol{X}=\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ and $\boldsymbol{Y}=\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n\}$ where $m, n \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$, the OSPA metric of order p between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as (if $m \le n$)

$$\bar{d}_p^{(c)}(\boldsymbol{X},\boldsymbol{Y}) = \left(\frac{1}{n}\left(\min_{q\in\Pi_n}\sum_{i=1}^{m} d_c\big(\boldsymbol{x}_i,\boldsymbol{y}_{q(i)}\big)^p + c^p(n-m)\right)\right)^{\frac{1}{p}} \tag{31}$$

where $d_c(\boldsymbol{x}, \boldsymbol{y})=\min(c, d(\boldsymbol{x}, \boldsymbol{y}))$, the cut off parameter $c > 0$ and $d(\boldsymbol{x}, \boldsymbol{y})$ is the Euclidean distance as defined by (23). $\bar{d}_p^{(c)}(\boldsymbol{X},\boldsymbol{Y}) = \bar{d}_p^{(c)}(\boldsymbol{Y},\boldsymbol{X})$ if $m > n$ and $\bar{d}_p^{(c)}(\boldsymbol{X},\boldsymbol{Y}) = 0$ if $m=n=0$.

The order parameter p determines the sensitivity to outliers, and the cut-off parameter c determines the relative weighting of the penalties assigned to cardinality and localization errors. Clearly, a better target detection capacity that renders more accurate estimation of the target number will significantly reduce the OSPA metric.

5.1. Completely known background

To set up the simulation, the real object trajectories are given in Fig. 2 where each starts at 'Δ' and ends at '□' (with both the origin and ending times noted in the figure). The object birth process is subject to Poisson with intensity $\gamma_t = \sum_{i=1}^{4} r_{t,i}\mathcal{N}(\cdot; \boldsymbol{B}_i, \boldsymbol{Q})$, where $\boldsymbol{B}_1=[-1500, 0250, 0, 0]^{\mathrm{T}}$, $\boldsymbol{B}_2=[-250, 0, 1000, 0, 0]^{\mathrm{T}}$, $\boldsymbol{B}_3=[250, 0750, 0, 0]^{\mathrm{T}}$, $\boldsymbol{B}_4=[1000, 0, 1500, 0, 0]^{\mathrm{T}}$, $\boldsymbol{Q}$=diag$([10, 10, 10, 10, \pi/180]^{\mathrm{T}})^2$ and $r_{t,\,1}=r_{t,\,2}=0.02$, $r_{t,\,3}=r_{t,\,4}=0.03$.

The object state $\boldsymbol{x}_t = [\tilde{\boldsymbol{x}}_t, w_t]^{\mathrm{T}}$consists of the planar position and velocity $\tilde{\boldsymbol{x}}_t = [p_{x,t}, \dot{p}_{x,t}, p_{y,t}, \dot{p}_{y,t}]^{\mathrm{T}}$ and the turn rate w_t. Each object either continues to exist with survival probability $p_S=0.98$ and move to a new state based on a nearly constant turn-rate (NCT) transition model, or disappears with probability 0.02. The NCT object state transition model can be written as

$$\tilde{\boldsymbol{x}}_t = F(w_{t-1})\tilde{\boldsymbol{x}}_{t-1} + U_{t-1},\; w_t = w_{t-1} + \Delta u_w \tag{32}$$

where

$$F(w) = \begin{bmatrix} 1 & \sin(w\Delta)w^{-1} & 0 & w^{-1}(\cos(w\Delta)-1) \\ 0 & \cos(w\Delta) & 0 & -\sin(w\Delta) \\ 0 & w^{-1}(1-\cos(w\Delta)) & 1 & \sin(w\Delta)w^{-1} \\ 0 & \sin(w\Delta)w^{-1} & 0 & \cos(w\Delta)w^{-1} \end{bmatrix}, U_{t-1} = \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ \Delta & 0 \\ 0 & \frac{\Delta^2}{2} \\ 0 & \Delta \end{bmatrix}\begin{bmatrix} u_x \\ u_y \end{bmatrix}$$

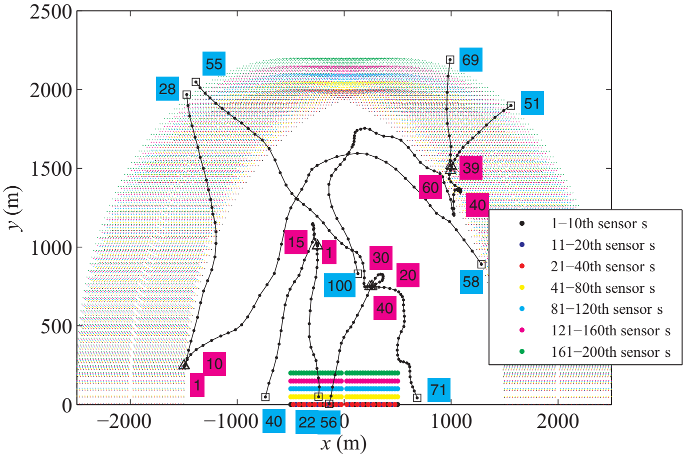

Fig. 2. An unknown number of objects observed by a sensor array: different colors indicate different sets of sensors that are used gradually in order; origin time (highlighted in magenta) and ending time (highlighted in cyan) of each trajectory are noted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

$\{u_x, u_y\} \sim \mathcal{N}(\cdot; 0, \sigma_u^2)$, $u_w \sim \mathcal{N}(\cdot; 0, \sigma_w^2)$, $\Delta = 1\text{s}$, $\sigma_u = 15\text{m/s}^2$ and $\sigma_w = \pi/180$ rad/s.

An array of homogeneous active radars (totally $40 \times 5 = 200$ as shown in Fig. 2) are used and all have a FoV of the half disc of radius 2000 m, centralized with their respect positions $[s_{x,i}, s_{y,i}]^{\mathrm{T}}$, $i = 1, 2, \ldots 200$. Without loss of generality, these sensors are used gradually in order as noted in Fig. 2. The object detection probability function of sensor i is $p_{D,i}(\boldsymbol{x}_t) = 0.95\mathcal{N}([p_{x,t} - s_{x,i}, p_{y,t} - s_{y,i}]^{\mathrm{T}}; 0, 6000^2\boldsymbol{I}_2)/\mathcal{N}(0; 0, 6000^2\boldsymbol{I}_2)$. The observation function of each sensor is given as shown in (13) with $\boldsymbol{v}_t \sim \mathcal{N}(\cdot; 0, \boldsymbol{R}_t)$, where $\boldsymbol{R}_t = \mathrm{diag}([\sigma_r^2, \sigma_\theta^2]^{\mathrm{T}})$, $\sigma_r = 20$ m, $\sigma_\theta = \pi/90$ rad/s. Clutter is uniformly distributed over the region with an average rate of r_t points per scan and we will set different r_t (but constant over time for simplicity).

In our approach, the O_2 detections and their variances are given in (16) and (17) where we use $I = 100$ for debiasing and $l_1 = 2$, $l_2 = 0.6$ are used in Algorithm 1 for clustering. The C4F approach will be compared with the SMC-PHD (sequential Monte Carlo-PHD) filter [24,30]. In the latter, 500 particles per expected object are used and the total number of particles is hard-limited to be not fewer than 300. The OSPA parameters are set as $c = 500$, $p = 2$.

The filter is provided with completely correct models and suitable parameters for the most favorable performance. There are several ways to incorporate different sensors to work for the SMC-PHD filter, differing from one another in terms of how to fuse the sensor data [4]. First, we apply a naive track-to-track (T2T) fusion that will run the SMC-PHD filter separately on each sensor, and then fuse their estimates based on the nearest neighbour association at the end of each filtering step. The overestimation of the number of objects by one filter as compared to the average of others will be simply eliminated, although this may not be the best choice. Secondly, a filter is assumed with the use of a single "super" sensor positioned at $[0, 0]^{\mathrm{T}}$which is affected by a smaller observation noise $\boldsymbol{v}_t \sim \mathcal{N}(\cdot; 0, \boldsymbol{R'}_t)$, where $\boldsymbol{R}'_t = \mathrm{diag}([\sigma_r'^2, \sigma_\theta'^2]^{\mathrm{T}}) = \boldsymbol{R}_t/N$, i.e., $\sigma_r' = \frac{20}{\sqrt{N}}$m, $\sigma_\theta' = \frac{\pi}{90\sqrt{N}}$rad/s. This corresponds to an observation accuracy that is equivalent to the MMSE fusion of N sensors and is an ideal assumption (in practice, such good a sensor may not exist). This is referred to as observation fusion and tracking (OFT).

First, we set the clutter rate $r_t = 10$. Different parameter N from two to 200 will be used for simulating different numbers of sensors in the C4F approach and the T2T SMC-PHD filter or different magnitudes of the observation noise for the OFT SMC-PHD filter. Fig. 3 gives the average OSPA (over 100 steps × 10 MC runs) and average processing time for each run (over 10 MC runs), obtained by different estimators for different N. The results show that if only two sensors are used, the C4F approach performs really badly, but with an increase in the number of sensors N, better results can be gained up to a stable level. This does not hold, however, with the SMC-PHD filters. Specifically, the OFT SMC-PHD filter remarkably degrades with the increase of N. We conjecture that this is because a very small observation noise corresponds to a sharp likelihood distribution that can easily cause severe sample degeneracy (this is a notorious problem of the SMC [26]). The naive version of the T2T SMC-PHD filter is slightly better but still significantly underperforms the C4F approach when more than six sensors are used. The results indicate that the filter may not benefit from the use of more precise or a greater number of sensors for one reason or another. Advanced implementations of the SMC, e.g., using the newest observation to design the proposal that best matches with the posterior distribution to accommodate sharp likelihood may help to correct this, yielding better filtering results; but the opposite can also occur as shown in the simulation study presented in Section 4.2 of [23]. In contrast, the proposed C4F approach avoids this problem and guarantees a reliable performance that is consistent with the observation condition.

The following theoretical and experimental evidences are highly relevant:

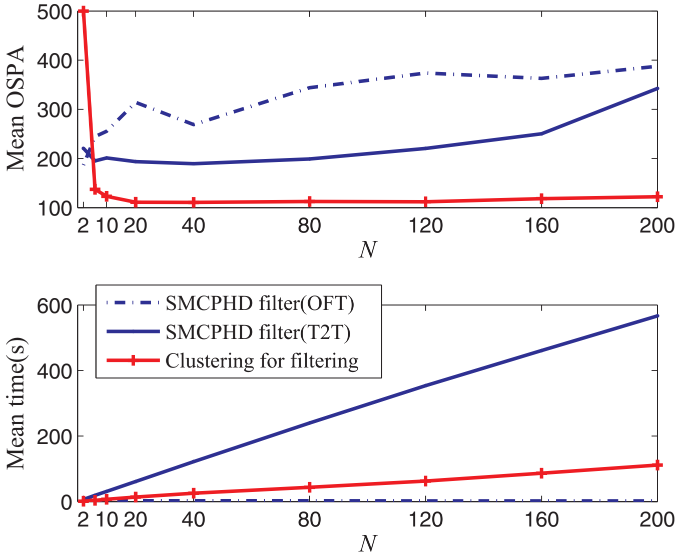

Fig. 3. Mean OSPA (over 100 steps × 10 MC runs) and average processing time of 100 steps (over 10 MC runs), for different N.

- An analytical tracker performance model has been proposed in [5], which suggests that "*scan-based tracking performance improves with increasing numbers of sensors, but only to a certain point beyond which degradation is observed.*" The optimal choice for the number of sensors depends on the false track requirement.
- Seeking increasingly precise approximations of the Kalman filter (e.g. extended, unscented, cubature, etc.) can be of limited benefit, particularly when the measurement noise is low enough and nonlinearities of the measurement function are non-negligible [11].
- When the objects do not display any significant random motions at the length and time scales of interest, variational methods based on deterministic-model trajectory-optimization (however non-sequential) may be more appropriate than stochastic filters (such as Kalman filters and particle filters) [17].

On the computing speed, the C4F approach shows more obvious advantage than the T2T SMC-PHD filter, but unsurprisingly, both have a speed proportional to the number of sensors which is slower (when $N \geq 6$ in this case) than that of the single-sensor based OFT SMC-PHD filter.

For further insight, Fig. 4 shows the detections given by all sensors and the C4F estimates when 200 sensors are used for time $t = 42$, where only four sensors in the corners of the array and their FoV bounds are noted. The results show that the C4F estimates are apparently very close to the truth. Secondly, we set $N = 10$, i.e., 10 sensors (located in the bottom horizontal line) for the C4F approach and the T2T filter. The C4F estimates have been given in Fig. 1 in Section 4.2 for time $t = 42$ in one simulation run (with clutter rate $r_t = 10$). The average estimated number of objects and the mean OSPA over 10 MC runs by using different methods are given in Fig. 5. The results show clearly that the C4F approach outperforms the SMC-PHD filters.

In particular, the performance of all estimators turns out to degrade in the periods $t \in \{20 \sim 28\}$ and $t \in \{40 \sim 70\}$ because some objects move out of the FoV of sensors, for which no compensating strategies were adopted. To exclude this we next control the scenario so that all the objects to be considered are always in the common area that can be detected by all sensors, as shown in Fig. 6. We also reduce the clutter rate to $r_t = 5$. Then, for different N, the average OSPA (over 100 steps × 10 MC runs) and average processing time of 100 steps (over 10 MC runs) are given in Fig. 7 which show again, unsurprisingly, that the C4F approach outperforms the filters when $N \geq 6$. Different to the last scenario, both SMC-PHD filters seemingly achieve a relatively stable level of estimation accuracy with an increase in the number of sensors used. To get more insights, all sensors' detections and the C4F estimates in one run are given in Fig. 8. The average OSPA and average number of objects estimated by different methods against time are given for $N = 10$ in Fig. 9, which shows that the clustering approach exhibits better capacity to detect the appearance and disappearance of objects instantly (with no time-delay) and so the number of objects is better estimated (especially in the period $t \in \{20 \sim 40\}$). In contrast, the object appearance and disappearance are confirmed by successive scans in the filter due to the infinite impulse response nature of recursive filtering.

We iterate that such perfect prior statistical knowledge about the objects, the clutter and even the sensors, which ideally complies with the ground truth in the simulation, does not hold in the real world. Consequently, the same filters can hardly yield as good a result as shown here. For example, mismatches on clutter rate r_t and detection probability $p_{D,\,t}$ can cause remarkably biased estimation in the multi-object filters [31,46]. However, the C4F approach will perform the same since it

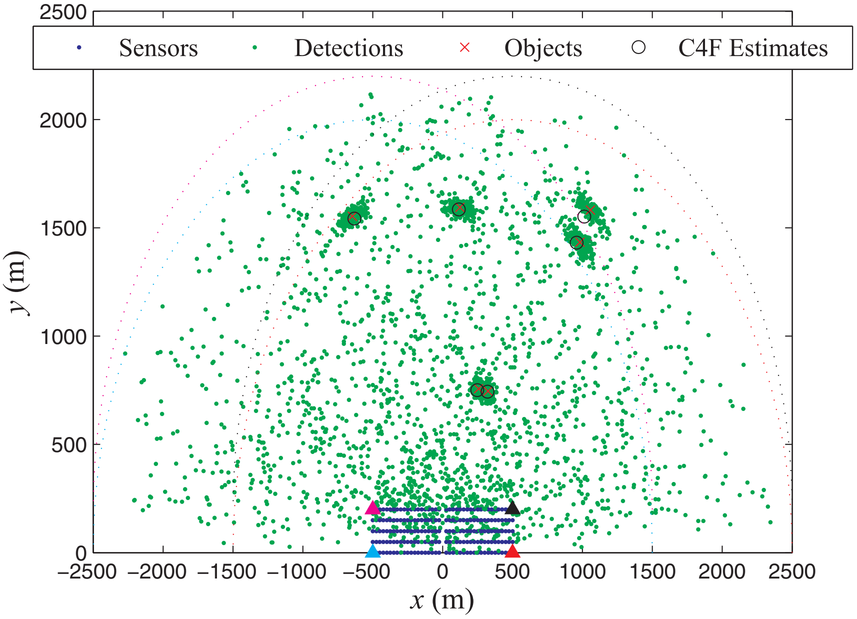

Fig. 4. Detections (green ".") of 200 sensors, real object-positions (red "x") and C4F estimates (black "o") at time $t = 42$ for $r = 10$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

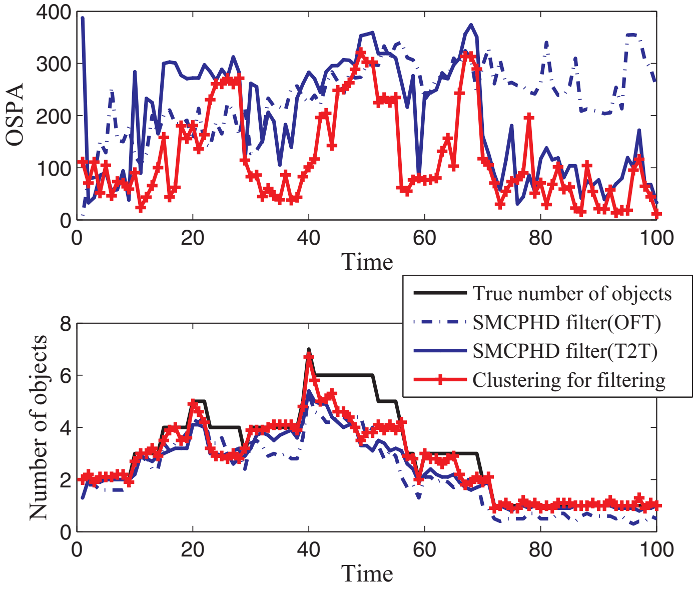

Fig. 5. Average OSPA and estimated number of objects over 10 MC runs against the filtering time, for $r = 10$, $N = 10$.

does not require rigorous and precise statistical assumptions on the objects, clutter and even the sensor uncertainty. The next simulation will consider an intractable scenario of which very little prior statistical model knowledge is available.

5.2. Poor prior information

In this simulation, we consider an extremely challenging scenario where almost no prior information is given about the objects, the clutter rate and even the sensor profiles. The latent objects may appear and disappear anywhere, anytime, whether jointly, adjacently or solitarily, and they may split, merge or cross each other, to name a few possibilities. The real trajectories of objects over the view region $[-100,\ 100]\text{m} \times [-100,\ 100]\text{m}$ are given in Fig. 10. Only the deterministic observation function is given a priori by (11) but nothing is known about $\boldsymbol{v}_k$.

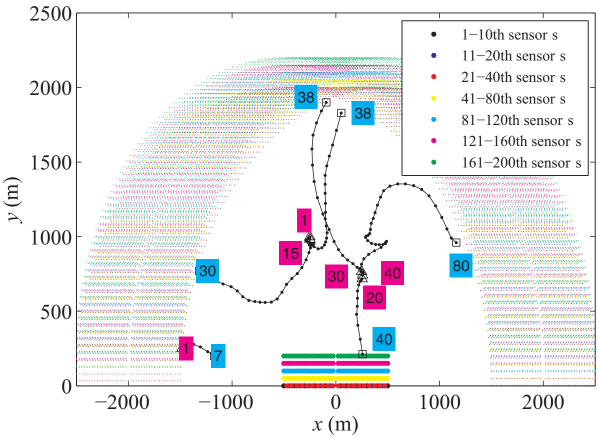

Fig. 6. An unknown number of objects whose trajectories are fully covered by the FoV of all sensors; origin time (highlighted in magenta) and ending time (highlighted in cyan) of each trajectory are noted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

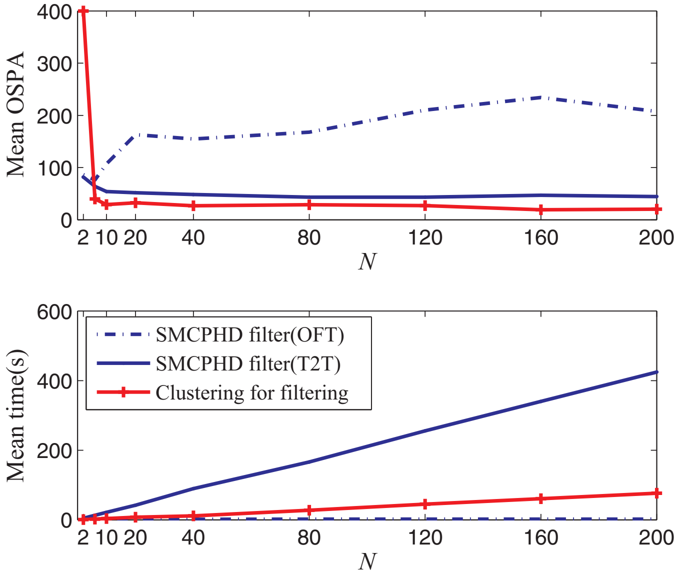

Fig. 7. Mean OSPA and average processing time of 100 steps (over 10 MC runs), for different N when $r = 5$.

As shown in Table 1, the dynamic models of the objects under concern consist of NCT, nearly constant velocity (NCV), nearly circular constant turn-rate (NCCT), constant velocity (CV) and a stationary object. The latter two deterministic models represent the case being affected with negligible noises. To note, the object motion does not have to be modeled by a Markov-transition model, e.g., neither (33) nor (34) are given in the form of a Markov jump. However, these latent models are sealed in a black box, unseen by all estimators. According to our knowledge, there is no report of a single filter that is able to handle all of these unknown statistics about the objects, the clutter and the sensor uncertainty jointly, as requested herein.

For simulation only, we assume the $x-y$ position-observation noise as $\boldsymbol{v}_t = [v_{x,\,t},\ v_{y,\,t}]^{\mathrm{T}}$, where $v_{x,\,t}$ and $v_{y,\,t}$ as mutually independent zero-mean Gaussian noise with variance 5. Clutter is uniformly distributed with an average rate of $r = 10$ points per scan of each sensor. To accommodate unknown sensor uncertainty for which (24) is prevented, two detections $\boldsymbol{x}_i^s, \boldsymbol{x}_j^m$ are assigned to the same cluster if they satisfy$\| \boldsymbol{x}_i^s - \boldsymbol{x}_j^m \|_2 \leq 10$, which is equivalent to assuming a neighbor radius $\varepsilon = 10$ (then l_1 is no more needed). We set $l_2 = 0.7$. To note, our experiments have exposed that choosing $\varepsilon = 8 \sim 15$ does not make an obvious difference, indicating that the C4F approach is less sensitive to the preciseness of the assumption on noises. In the meanwhile, it also indicates that we must have some sense of the uncertainty of the sensors, even if we cannot exactly know it – which is arguably the reality at most of the time. In the C4F approach, the estimate is simply given by the center of each cluster (assuming these sensors are of the same accuracy), regardless of the prior information given in Table 1.

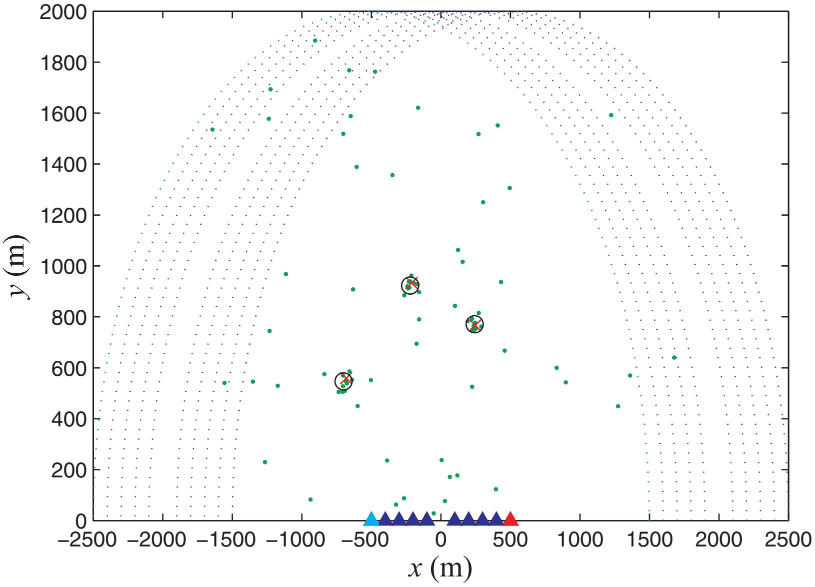

Fig. 8. Detections (green ".") of 10 sensors, real object-positions (red "x") and C4F estimates (black o) at time $t = 20$ for $r = 5$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

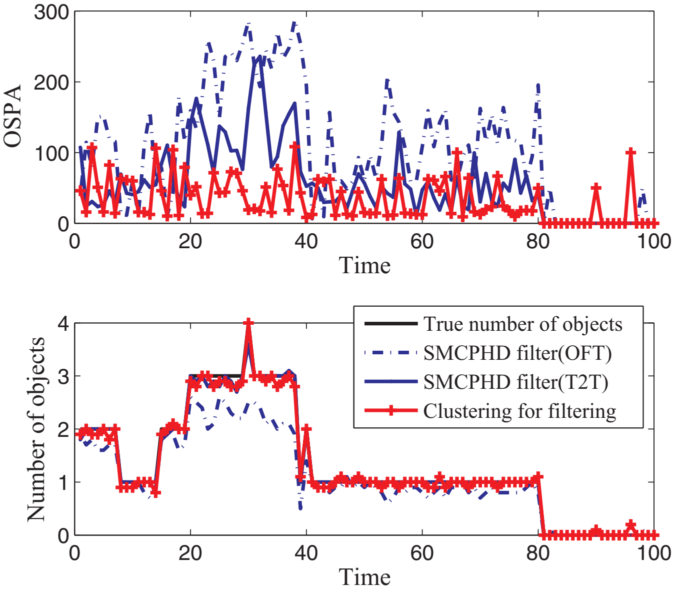

Fig. 9. Average OSPA and estimated number of objects over 10 MC runs against time for $r = 5$, $N = 10$.

First, we use 10 sensors. The detections from 10 sensors, real object positions and the clustering estimates are given in Fig. 11 for time instant $t = 16$, where different clusters are marked differently for better illustration. The result given by the clustering approach is apparently quite reasonable, overcoming the problems of clutter and misdetections. The average estimate of the number of objects and average OSPA over 10 MC runs are given in Fig. 12. The mean OSPA over 100 steps × 10 MC runs is 9.756 m, which is arguably very good with regard to OSPA parameters $c = 100$, $p = 2$, considering that almost no prior information is given for this complicated scenario.

Fig. 13 gives the mean OSPA and mean processing time of 100 steps over 10 MC runs, both against the number of sensors used. The results show again that with an increase in the number of sensors, the present clustering approach will get a gradually better estimation on average, though it will also cost proportionally more processing time.

Overall, both simulations with complete and poor prior information have demonstrated that the proposed C4F approach enjoys increasing MODE accuracy with the increase in the number of sensors. It can perform even better than the perfectly

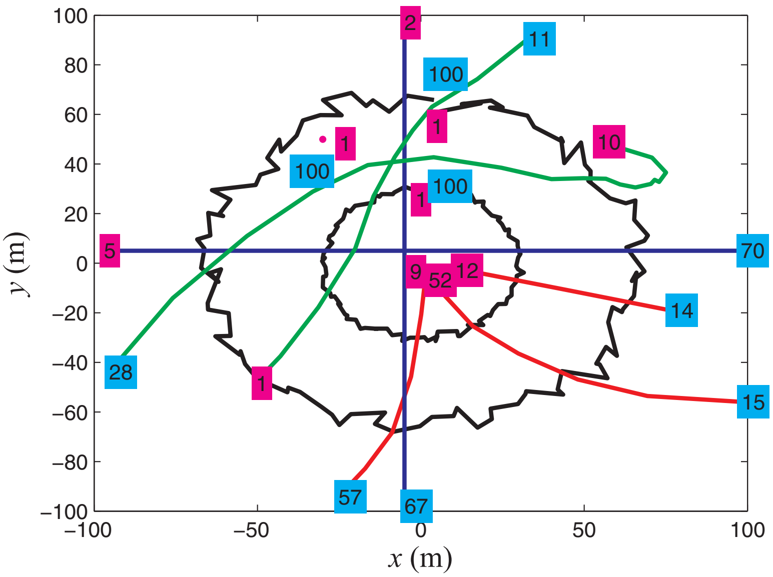

Fig. 10. Origin time (marked in magenta) and ending time (marked in cyan) of each trajectory in the scenario of no prior information. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Table 1
Black box: latent dynamic models of the objects.

Object model 1. Stationary – one object.
The position of the object $[p_{x,\,t},\ p_{y,\,t}]^{\mathrm{T}}$ is constant over time.

$$\begin{cases} p_{x,t} = p_x \\ p_{y,t} = p_y \end{cases} \tag{33}$$

where we set constant $p_x = -30$ m, $p_y = 50$ m.
Object model 2. NCCT – two objects.

$$\begin{cases} p_{x,t} = r \times \cos(\theta_t) + u_{x,t} \\ p_{y,t} = r \times \sin(\theta_t) + u_{y,t} \end{cases} \tag{34}$$

where $r = 65\mathrm{m}, u_{x,t} \sim \mathcal{N}(0,\ 9), u_{y,t} \sim \mathcal{N}(0,\ 9)$ for one object and $r = 30\mathrm{m}, u_{x,t} \sim \mathcal{N}(0, 1), u_{y,t} \sim \mathcal{N}(0, 1)$ for the other. In both cases, $\theta_t = t \times \pi/50$ rad.
Object model 3. CV – two objects.

$$\boldsymbol{x}_t = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}_{t-1} \tag{35}$$

where $\boldsymbol{x}_t = [p_{x,t}, \dot{p}_{x,t}, p_{y,t}, \dot{p}_{y,t}]^{\mathrm{T}}, \Delta = 1$ s. One object is initialized with $\boldsymbol{x}_0 = [-100$ m, 3 m/s, 5m, 0 m/s$]^{\mathrm{T}}$ and the other $\boldsymbol{x}_0 = [-5$ m, 0 m/s, 100 m, -3 m/s$]^{\mathrm{T}}$.
Object model 4. NCV – three objects.

$$\boldsymbol{x}_t = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}_{t-1} + \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ \Delta & 0 \\ 0 & \frac{\Delta^2}{2} \\ 0 & \Delta \end{bmatrix} \begin{bmatrix} u_{x,t} \\ u_{y,t} \end{bmatrix} \tag{36}$$

where $u_{x,t} \sim \mathcal{N}(0,\ 1), u_{y,t} \sim \mathcal{N}(0,\ 0.01)$. Three objects are initialized as $\boldsymbol{x}_0 = [0$ m, 3 m/s, 0 m, -3 m/s$]^{\mathrm{T}}$, $[0$ m, 1.6 m/s, 1 m, -2.1 m/s$]^{\mathrm{T}}$, and $[-20$ m, 30 m/s, 10 m, -15 m/s$]^{\mathrm{T}}$ respectively.
Object model 5. NCT – two objects.
The model is the same as given in (32) except a different turn-rate statistics $\sigma_w = 5$ m/s^2. Two objects are initialized as $\boldsymbol{x}_0 = [-50$ m, 0 m/s, -50 m, 0 m/s$]^{\mathrm{T}}$ and $[50$ m, 0 m/s, 50 m, 0 m/s$]^{\mathrm{T}}$ respectively, both with an initial turn rate $w_0 \sim \mathcal{N}(0, \pi/180)$.

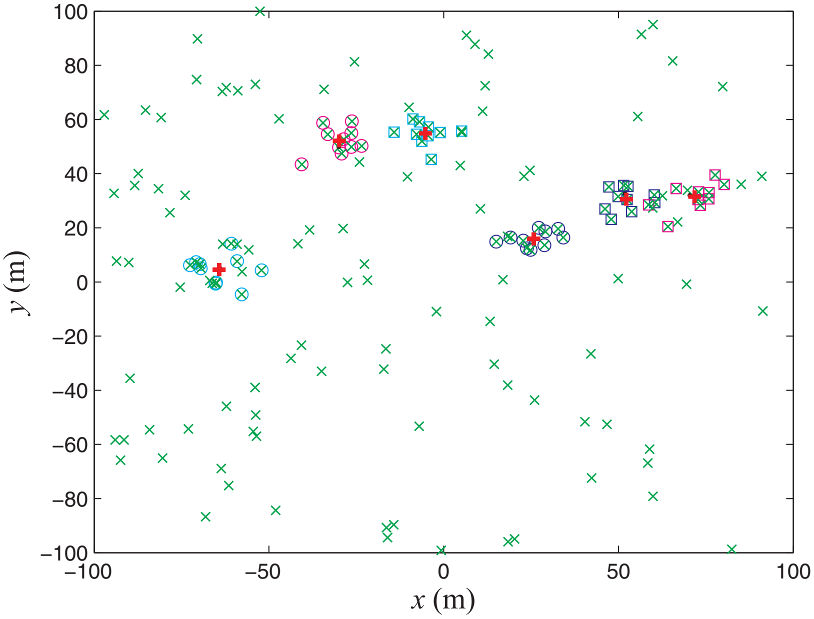

Fig. 11. Detections (green "x") of 10 sensors, true object-positions (black "•"), clusters formed ("o/□") and the corresponding C4F estimates (red "+") at time $t=16$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

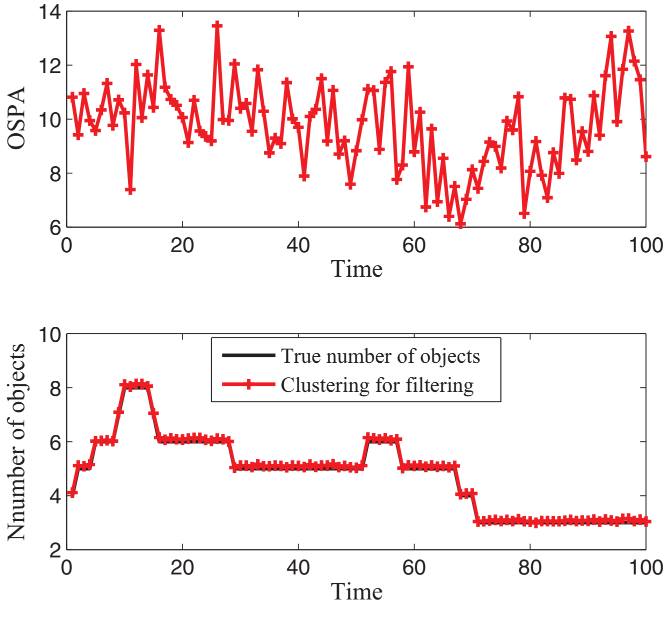

Fig. 12. Mean OSPA and estimated number of objects over 10 MC runs against time.

modelled filter when a sufficient number of sensors are used. This is not surprising, considering that the single sensor based O_2 inference can even outperform the filters in certain cases [23]. The joint use of massive sensors in the C4F approach has shown to be favorable for circumventing the poor prior knowledge and complicated scenario, which however does not readily hold in filters.

6. Conclusions and future work

We have proposed a clustering for filtering (C4F) paradigm for the purpose of multiple object detection and estimation (MODE) in a challenging cluttered environment with little prior knowledge, yet multiple or massive sensors are available for observation. Significantly different to the conventional sequential Bayesian inference, the present C4F approach solves

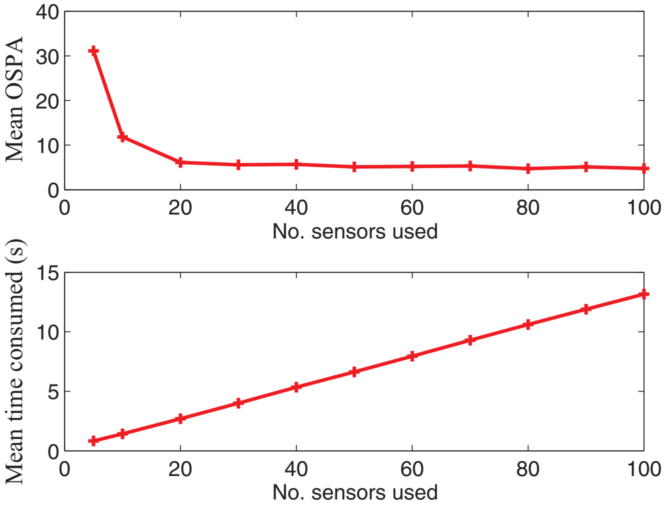

Fig. 13. Mean OSPA and processing time of 100 steps × 10 MC runs against different numbers of sensors.

MODE from the sensor data mining perspective (more precisely, clustering) which amplifies the strength of rich sensor data to circumvent the background unawareness. It can handle significant object maneuvering, unknown and time varying noises and clutter density, etc., and can do so computationally fast and reliably by releasing impractical albeit sophisticated models. Simulations have demonstrated that the C4F approach can even outperform Markov-Bayes filters that use correct models in average multi-sensor cases, and is able to handle the unknown scenario that makes traditional filter modelling awkward. The proposed C4F approach enjoys increasing accuracy with the increase in whether the number or the accuracy of sensors, which however does not readily hold in filters. We point out that advanced object tracking solutions have to accommodate intractable sensor conditions such as heterogeneity, asynchrony, correlation, inconsistency etc., as well as account for more tracking issues such as track formation and management, for which more sensor data mining and learning technologies are worth considering.

We do not argue that the C4F approach should replace any Markov–Bayes filters when the system and background can be well modelled ensuring a high PoFB but that it should be viewed as an alternative for the case little/no background information but high-performance sensors are available. In fact, if reliable object motion information is available, the C4F estimates can be then refined and even the continuous-time target trajectory can be formed. In this regard, a part of our follow-up work appeared in [25], which assumed the object trajectory is a smooth function of time (rather than a Markov jump model) and that C4F/O$_2$ estimates over the time series can then be fitted by a function and correspondingly optimized online. The core idea is to model the state evolving by a continuous function of time $\boldsymbol{F}(t)$, which provides a unified framework for smoothing, tracking and forecasting (STF) based on optimization

$$\underset{\boldsymbol{F}(t)}{\operatorname{argmin}} \sum_{t=t_1}^{t_2} \|\boldsymbol{z}_t - \boldsymbol{h}_t\left(\boldsymbol{F}(t), \overline{\boldsymbol{v}}_t\right)\| \tag{37}$$

where $\boldsymbol{F}(t)$ is the desired trajectory function of continuous-time t which gives $\hat{\boldsymbol{x}}_t = \boldsymbol{F}(t)$ for any time instant $t \geq t_1$ that does not need to be an integer, even if the sampling instants from t_1 to t_2 are intermittent.

By placing the series of estimates in (7) in the form of a deterministic trajectory function, (37) unifies the goals of STF as simply seeking a trajectory function that best explains the sensor data series in the time domain. To accommodate any a priori model information, $F(t)$ may be specified as being a specific form of a few parameters or subject to certain constraints, denoted as $F(t; C)$ where C denotes the set of the parameters or constraints. This will integrate both useful model knowledge and online data information for estimation and forms the key of our work on a unified Markov-free framework for STF. It is worth noting that, this formula, most being non-convex, does not admit a simple exact solution in general, but it can be simplified if the sensor data are projected into the state space, as is done with the O$_2$/C4F approach. The problem then reduces to performing fitting on the intermediate O$_2$/C4F estimates $\hat{\boldsymbol{x}}_{t_1:t_2}$, i.e.,

$$\underset{\boldsymbol{F}(t;C)}{\operatorname{argmin}} \sum_{t=t_1}^{t_2} \|\widehat{\boldsymbol{x}}_t - \boldsymbol{F}(t;C)\| \tag{38}$$

This embodies one important use of the presented C4F approach.

Acknowledgments

The first author acknowledges Prof./Dr. Yu- Chi (Larry) Ho for his insightful discussions and encouragement on this series of work in a variety of ways. Draft of the work has been discussed with Dr. S. Coraluppi, Dr. R. L. Streit and Dr. H. Fan. This work is in part supported and funded by Marie Skłodowska-Curie Individual Fellowship (Grant No. 709267) under the European Union's Framework Programme for Research and Innovation Horizon 2020, Spanish Ministry of Economy, Industry and Competitiveness (I+D+i Project Ref. TIN2015-65515-C4-3-R), and National Natural Science Foundation of China (Grant No.51475383).

References

[1] A.Y. Aravkin, J.V. Burke, G. Pillonetto, Optimization viewpoint on Kalman smoothing with applications to robust and sparse estimation, in: Compressed Sensing & Sparse Filtering, Springer Berlin Heidelberg, 2013, pp. 237–280.
[2] Y. Bar-Shalom, Tracking methods in a multitarget environment, IEEE Trans. Autom. Control AC 23 (4) (1978) 618–626.
[3] P.A. Browne, Model error moment estimation via data assimilation, 2016, arXiv:1610.01226v1.
[4] H. Chen, T. Kirubarajan, Y. Bar-Shalom, Performance limits of track-to-track fusion versus centralized estimation: theory and application, IEEE Trans. Aerosp. Electron. Syst. 39 (2) (2003) 386–400.
[5] S. Coraluppi, M. Guerriero, P. Willett, C. Carthel, Fuse-before-track in large sensor networks, J. Adv. Inform. Fusion 5 (1) (2010) 18–31.
[6] L.A. Dalton, E.R. Dougherty, Intrinsically optimal Bayesian robust filtering, IEEE Trans. Signal Process 62 (3) (2014) 657–670.
[7] D. Du, B. Qi, M. Fei, C. Peng, Multiple event-triggered H2/H∞ filtering for hybrid wired–wireless networked systems with random network-induced delays, Inform. Sci. 325 (2015) 393–408.
[8] J. Duník, O. Straka, M. Šimandl, O. Kost, J. Ajgl, M. Soták, R. Baránek, Z. Kaňa, Estimation of state and measurement noise characteristics, in: Proceedings of the 18th International Conference on Information Fusion, Washington, DC, 2015, pp. 1817–1824.
[9] M. Ester, H.P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, AAAI Press, 1996, pp. 226–231.
[10] C. Fritsche, U. Orguner, F. Gustafsson, On parametric lower bounds for discrete-time filtering, in: The 41th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Shanghai, China, March 2016, pp. 4338–4342.
[11] ÁF. García-Fernández, L. Svensson, M.R. Morelande, S. Särkkä, Posterior linearization filter: principles and implementation using sigma points, IEEE Trans. Signal Process 63 (20) (2015) 5561–5573.
[12] A. Ghoreyshi, T.D. Sanger, A nonlinear stochastic filter for continuous-time state estimation, IEEE Trans. Autom. Control 60 (8) (2015) 2161–2165.
[13] A. Gning, W.T.L. Teacy, R.V. Pelapur, H. AliAkbarpour, K. Palaniappan, G. Seetharaman, S.J. Julier, The effect of state dependent probability of detection in multitarget tracking applications, in: Proceedings of SPIE 9089, Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II, 908905, 2014 June 19.
[14] S. Godsill, J. Vermaak, W. Ng, J.F. Li, Models and algorithms for tracking of maneuvering objects using variable rate particle filters, Proc. IEEE 95 (5) (2007) 925–952.
[15] P.D. Grünwald, T. van Ommen. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing It, 2014, arXiv: 1412.3730.
[16] F. Hartig, J.M. Calabrese, B. Reineking, T. Wiegand, A. Huth, Statistical inference for stochastic simulation models–theory and application, Ecol. Lett 14 (8) (2011) 816–827.
[17] K. Judd, Forecasting with imperfect models, dynamically constrained inverse problems, and gradient descent algorithms, Phys. D: Nonlinear Phenom. 237 (2) (2008) 216–232.
[18] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, multisensory data fusion: a review of the state-of-the-art, Inf. Fusion 14 (1) (2013) 28–44.
[19] G. Kurz, I. Gilitschenski, U.D. Hanebeck, Recursive Bayesian filtering in circular state spaces, arXiv:1501.05151
[20] T. De Laet, H. Bruyninckx, J. De Schutter, Shape-based online multitarget tracking and detection for targets causing multiple measurements: variational Bayesian clustering and lossless data association, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2477–2491.
[21] T. Li, J.M. Corchado, J. Bajo, G. Chen, Multi-target detection and estimation with the use of massive independent, identical sensors, in: Proceedings of SPIE Vol. 9469-15, Sensors and Systems for Space Applications VIII, Baltimore, Maryland, US, 2015 April 20-24.
[22] T. Li, J.M. Corchado, J. Bajo, S. Sun, Multi-source data clustering, in: Proceedings of the 18th International Conference on Information Fusion, Washington DC, 2015, pp. 830–837.
[23] T. Li, J.M. Corchado, J. Bajo, S. Sun, J.F. Paz, Effectiveness of Bayes filters: an information fusion perspective, Inform. Sci. 329 (2016) 670–689.
[24] T. Li, J.M. Corchado, S. Sun, H. Fan, Multi-EAP: Extended EAP for multi-estimate extraction for SMC-PHD filter, Chin. J. Aeronaut. (2016). http://dx.doi.org/10.1016/j.cja.2016.12.025.
[25] T. Li, J. Prieto, J.M. Corchado, Fitting for smoothing: A methodology for continuous-time target track estimation, 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), At Alcalá de Henares, Spain, 2016 Oct. 4-7.
[26] T. Li, M. Bolić, P. Djurić, Resampling methods for particle filtering: classification, implementation, and strategies, IEEE Signal Process. Mag. 32 (3) (2015) 70–86.
[27] X.R. Li, V.P. Jilkov, Survey of maneuvering target tracking Part I: dynamic models, IEEE Trans. Aerosp. Electron. Syst. 39 (4) (2003) 1333–1364.
[28] X.R. Li, Compatibility and modeling of constrained dynamic systems, in: 19th IEEE International Conference on Information Fusion, Heidelberg Germany, 2016, pp. 240–247.
[29] Y. Liu, Z. Yang, X. Wang, L. Jian, Location, localization, and localizability, J. Comput. Sci. Technol. 25 (2) (2010) 274–297.
[30] R. Mahler, Advances in statistical multisource-multitarget information fusion. Artech House, 2014.
[31] R. Mahler, B.T. Vo, B.N. Vo, CPHD filtering with unknown clutter rate and detection profile, IEEE Trans. Signal Process 59 (8) (2011) 3497–3513.
[32] C.S. Maíz, J. Míguez, E.M. Molanes-López, P.M. Djurić, A robustified particle filtering scheme for processing time series corrupted by outliers, IEEE Trans. Signal Process 60 (9) (2012) 4611–4627.
[33] J.M. Marin, P. Pudlo, C.P. Robert, R. Ryder, Approximate Bayesian computational methods, Stat. Comput 22 (6) (2012) 1167–1180.
[34] D. Middleton, R. Esposito, Simultaneous optimum detection and estimation of signals in noise, IEEE Trans. Inform. Theory 14 (3) (1968) 434–444.
[35] S.C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, S.L. Shah, Nonlinear Bayesian state estimation: a review of recent developments, Control Eng. Pract. 20 (10) (2012) 933–953.
[36] E. Punskaya, A. Doucet, W.J. Fitzgerald, On the use and misues of particle filtering in digital commnunications, in: Proceedings of the 11th European Signal Processing Conference (EUSIPCO), Toulouse, France, 2002 03-06 Sept..
[37] U. Picchini, Likelihood-free stochastic approximation EM for inference in complex models, 2016, arXiv:1609.03508v1.
[38] U. Schmitt, A.K. Louis, Efficient algorithms for the regularization of dynamic inverse problems: I. theory, Inverse Probl 18 (3) (2002) 645–658.
[39] D. Schuhmacher, B.T. Vo, B.N. Vo, A consistent metric for performance evaluation in multi-object filtering, IEEE Trans. Signal process 56 (8) (2008) 3447–3457.
[40] Y. Shen, M.Z. Win, Fundamental limits of wideband localization part I: a general framework, IEEE Trans. Inform. Theory 56 (10) (2010) 4956–4980.
[41] O. Straka, J. Duník, M. Šimandl, Truncation nonlinear filters for state estimation with nonlinear inequality constraints, Automatica 48 (2) (2012) 273–286.

[42] D.K. Tasoulis, N.M. Adams, D.J. Hand, Should delayed measurements always be incorporated in filtering? in: 15th International Conference on Digital Signal Processing, Cardiff, UK, 2007, pp. 264–267.

[43] F. Tobar, P.M. Djuric, D.P. Mandic, Unsupervised state-space modeling using reproducing kernels, IEEE Trans. Signal Process 63 (19) (2015) 5210–5221.

[44] M.N. Tran, D.J. Nott, R. Kohn. Variational Bayes with intractable likelihood, 2015, arXiv:1503.08621v1.

[45] A. Tulsyan, B. Huang, R.B. Gopaluni, J.F. Forbes, Performance assessment, diagnosis, and optimal selection of non-linear state filters, J. Process Control 24 (2014) 460–478.

[46] B.T. Vo, B.N. Vo, R. Hoseinnezhad, R. Mahler, Robust multi-Bernoulli filtering, IEEE J. Sel. Top. Signal Process 7 (3) (2013) 399–409.

[47] H. Wang, J. Sun, X. Zhang, X.Y. Huang, T. Auligné, Radar data assimilation with WRF 4D-Var. Part I: system development and preliminary testing, Mon. Weather Rev. 141 (2013) 2224–2244.

[48] X. Wang, Y. Liang, Q. Pan, Y. Wang, Measurement random latency probability identification, IEEE Trans. Autom. Control 61 (12) (2016) 4210–4216.

[49] S. Yildirim, L. Jiang, S.S. Singh, T.A. Dean, Calibrating the Gaussian multi-target tracking model, Stat. Comput 25 (2014) 1–14.

[50] G. Zhou, M. Pelletier, T. Kirubarajan, T. Quan, Statically fused converted position and Doppler measurement Kalman filters, IEEE Trans. Aerosp. Electron. Syst. 50 (1) (2014) 300–318.