Particle flow SMC-PHD filter for audio-visual multi-speaker tracking

Yang Liu [1] *, Wenwu Wang [1], Jonathon Chambers [2], Volkan Kilic [3], and Adrian Hilton [1]

[1] Department of Electrical and Electronic Engineering, University of Surrey, Guildford, GU2 7XH, U.K.
{yangliu,w.wang,a.hilton}@surrey.ac.uk
[2] School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.
Jonathon.Chambers@newcastle.ac.uk
[3] Department of Electrical and Electronics Engineering, Izmir Katip Celebi University, 35620 Cigli-Izmir, Turkey.
volkan.kilic@ikc.edu.tr

Abstract. Sequential Monte Carlo probability hypothesis density (SMC-PHD) filtering has been recently exploited for audio-visual (AV) based tracking of multiple speakers, where audio data are used to inform the particle distribution and propagation in the visual SMC-PHD filter. However, the performance of the AV-SMC-PHD filter can be affected by the mismatch between the proposal and the posterior distribution. In this paper, we present a new method to improve the particle distribution where audio information (i.e. DOA angles derived from microphone array measurements) is used to detect new born particles and visual information (i.e. histograms) is used to modify the particles with particle flow (PF). Using particle flow has the benefit of migrating particles smoothly from the prior to the posterior distribution. We compare the proposed algorithm with the baseline AV-SMC-PHD algorithm using experiments on the AV16.3 dataset with multi-speaker sequences.

Keywords: Audio-visual tracking, PHD filter, SMC implementation, multi-speaker tracking

1 Introduction

Multi-speaker tracking for indoor environments has received much interest in the fields of computer vision and signal processing [25]. An increasing amount of attention has been paid to the use of audio-visual modalities [2, 15], which provide

* This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1), the BBC as part of the BBC Audio Research Partnership, the China Scholarship Council (CSC), and the EPSRC grant EP/K014307/1 and the MOD University Defence Research Collaboration in Signal Processing.

complementary information in addressing several challenges such as occlusion, limited view of cameras, illumination change, and room reverberations.

Several approaches using multi-modal information have been proposed. One such method is based on audio-visual diarization [19], which is only effective when the speakers continuously face the cameras. Kılıç et al. [21] addresses this problem in the framework of audio-visual speaker tracking using a particle filter (PF) and a probability hypothesis density (PHD) filter based on sequential Monte Carlo (SMC) approximation [20]. Different from the Bayesian approaches (Kalman or PF filters) [3, 4, 27], prior knowledge such as the number of targets is not required in the PHD filter. As for other SMC-PHD filters, the AV-SMC-PHD filter in [20] uses particles to represent the posterior density. However, after some updates, the prior distribution may not overlap with the target distribution [17].

Recently, the particle flow (PF) filter has been proposed for solving the non-linear and non-Gaussian problem [5, 8, 9, 12]. In this method, particle flow is created by a log-homotopy of the conditional density migrating from the prior to the posterior. Several approaches have been proposed to create the particle flow which can be categorized into five classes: incompressible flow [6], zero diffusion exact flow [7], Coulomb's law particle flow [13], zero-curvature particle flow [9] and non zero diffusion flow [12]. The zero-curvature particle flow has been used widely [18, 24, 30], as it is straightforward to implement.

Particle flow has been used to improve the accuracy of the particle filter [24], and is denoted as the particle flow particle filter (PFPF). Different from conventional particle filters, the PFPF uses a small number of particles to achieve the similar accuracy as that for particle filters with a higher effective sample size (ESS) [22]. However, for multi-target tracking, a dependent filter needs to be applied to each target, which introduces the model-data association problem [14]. In addition, prior knowledge of the number of targets is needed. In [30], a Gaussian particle flow implementation of the PHD filter (GPF-PHD) is proposed yielding good accuracy in a nonlinear tracking problem. However, in this method, the particles are generated for each target, and the computational cost could be high for a large number of targets and clutter. For non-linear and non-Gaussian problems, the auxiliary particle PHD filter proposed in [1] has better performance than the GPF-PHD filter in terms of Optimal Subpattern Assignment (OSPA) [14], since it efficiently distributes the particles by maximizing the accuracy of the cardinality estimate.

In this paper, we extend the AV-SMC-PHD filter presented in [20] by incorporating particle flow within the particle evolution in order to improve its tracking performance. The major contribution of this paper is a novel particle flow SMC-PHD filtering method for multi-speaker tracking, where the audio data are used to compute the prior distribution and the visual data are applied to compute the particle flow. The posterior distribution is calculated by the color histograms of the visual image and adjusted by the position of the direction of arrival (DOA) lines drawn from the targets. Using audio information, the computational cost for generating the particle flow can be reduced, as only the relevant particles surrounding the DOA line will be chosen; while the influence of the particles, that are likely from the clutter and distant from the DOA line, is mitigated. The proposed method is shown to outperform the baseline AV-SMC-PHD based on evaluations on the AV16.3 dataset.

The reminder of this paper is organized as follows: the next section introduces the AV-SMC-PHD filter and particle flow. Section III describes our proposed audio-visual particle flow SMC-PHD (AV-PF-SMC-PHD) filtering algorithm. In Section IV, experiments on the AV16.3 dataset are presented to show the performance of the proposed AV-PF-SMC-PHD algorithm as compared with the baseline AV-SMC-PHD algorithm.

2 AV-SMC-PHD Filter and Particle Flow

In this section, the baseline AV-SMC-PHD filter and the particle flow filter are introduced. For the discrete-time and non-linear filtering problems, we assume that the target dynamics and observations are described as a Markov state-space signal model:

$$\widetilde{\boldsymbol{m}}_k = \boldsymbol{f}_{\widetilde{\boldsymbol{m}}}\left(\widetilde{\boldsymbol{m}}_{k-1}, \boldsymbol{\tau}_k\right), \tag{1}$$

where $\widetilde{\boldsymbol{m}}_k$ is the target state vector at time-step k and $\widetilde{\ }$ is used to distinguish the target state from the particle state used later. In this paper, the state vector $\boldsymbol{m}_k = [x_k, y_k, \dot{x}_k, \dot{y}_k]^T$ consists of the target positions (x_k, y_k) and the target velocities $(\dot{x}_k, \dot{y}_k)$, and the observation is a noisy version of the position. The parameter vector $\boldsymbol{\tau}_k$ denotes the system excitation and observation noise terms and $\boldsymbol{f}_{\widetilde{\boldsymbol{m}}}$ is the transition density.

2.1 AV-SMC-PHD Filter

In the visual SMC-PHD filter [26], the surviving, spawned and born particles are used to model the existing and new speakers. For detecting the new targets, new particles need to be added randomly, leading to increase in the number of particles and hence to increase in computational load. To address this problem, the AV-SMC-PHD filter is proposed in [20].

Audio information is applied for re-locating existing particles around the DOA lines, since the DOA information shows the approximate direction of the sound emanating from the speakers. The movement distances of the particles $\hat{\boldsymbol{d}}_k$ are calculated as [20]:

$$\hat{\boldsymbol{d}}_k = \frac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_1} \odot \boldsymbol{d}_k \tag{2}$$

where $\boldsymbol{d}_k$ is the perpendicular Euclidean distances between the particles and the DOA line, $\|.\|_1$ is the l_1 norm, and $\odot$ is the element-wise product; $\hat{\boldsymbol{d}}_k$ is applied to relocate the surviving and spawned particles $\boldsymbol{m}_{s,k}$ around the DOA line [20]:

$$\boldsymbol{m}_{s,k} = \boldsymbol{m}_{s,k} \oplus \boldsymbol{h}_k \hat{\boldsymbol{d}}_k \tag{3}$$

where $\boldsymbol{h}_k = [cos(\theta_k), sin(\theta_k), 0, 0]$ and θ_k is the angle from the DOA line. $\oplus$ is the element-wise addition. As such, the particles are modified along the perpendicular movement to the DOA line, and N_Γ born particles are sampled from the new born importance function,

$$\boldsymbol{m}^i_{k|k-1} \sim p_k(\cdot|\boldsymbol{Z}_k). \tag{4}$$

where $\boldsymbol{m}^i_{k|k-1}$ is the i-th predicted particle state at time-step k.

Apart from that, audio information in the AV-SMC-PHD filter can be used to detect the new speakers effectively and the particles are born in particular directions. This reduces the number of particles and hence computational complexity. The DOA lines are determined by the relative delay between pairs of microphone signals [29]. When detecting new targets, the filter compares the number of DOA lines, N_D, with the number of estimated speakers at time $k-1$, N_{k-1}. If $N_D = N_{k-1}$, the number of the speakers remains unchanged. If $N_D < N_{k-1}$, the speakers may walk out of the camera view, or be occluded by other speakers, or the DOA line may not be detected. In this paper, if $N_D < N_{k-1}$ and $N_D \neq 0$, we assume that the number of the speakers reduces to N_D. If $N_D = 0$, we assume that the microphones do not detect the speakers successfully and the number of speakers remains the same as N_{k-1}. If $N_D > N_{k-1}$, a new speaker (or some new speakers) may appear in the scene and hence new born particles should be created. Since born particles are only generated when the detection of a new speaker occurs via audio, the computational complexity is reduced.

The pseudo code of AV-SMC-PHD filter is given in Algorithm 1 where $\{\boldsymbol{m}^i_k, \omega^i_k\}^{N_k}_{i=1}$ is the set of the particle state vectors and weights at time-step k; $\{\widetilde{\boldsymbol{m}}^j_k, \widetilde{\omega}^j_k\}^{\widetilde{N}_k}_{j=1}$ is the target set and $\widetilde{N}_k$ is the number of targets at the time-step k; N_Γ is the number of born particles, which is given as the initial value; $\boldsymbol{Z}_k$ contains observations at time-step k. The weights of the particles are predicted and updated by

$$\omega^i_{k|k-1} = \begin{cases} \dfrac{\phi_{k|k-1}(\boldsymbol{m}^i_{k|k-1}, \boldsymbol{m}^i_{k-1})\omega^i_{k-1}}{q_k\left(\boldsymbol{m}^i_{k|k-1}|\boldsymbol{m}^i_{k-1}, \boldsymbol{Z}_k\right)} & , i = 1, ..., N \\ \dfrac{\gamma_k(\boldsymbol{m}^i_{k|k-1})}{N_\Gamma p_k(\boldsymbol{m}^i_{k|k-1}|\boldsymbol{Z}_k)} & , i = N+1, ..., N+N_\Gamma \end{cases} \tag{5}$$

$$\omega^i_k = \left[1 - p_{D,k}(\boldsymbol{m}^i_k) + \sum_{\boldsymbol{z} \in \boldsymbol{Z}_k} \frac{p_{D,k}(\boldsymbol{m}^i_k) g_k(\boldsymbol{z}|\boldsymbol{m}^i_k)}{\kappa_k(\boldsymbol{z}) + C_k(\boldsymbol{z})} \right] \omega^i_{k|k-1} \tag{6}$$

where

$$C_k(\boldsymbol{z}) = \sum_{i=1}^{N+N_\Gamma} p_{D,k}(\boldsymbol{m}^i_k) g_k(\boldsymbol{z}|\boldsymbol{m}^i_k) \omega^i_{k|k-1} \tag{7}$$

in which $\omega^i_{k|k-1}$ is the i-th predicted particle weight at time-step k. $\phi_{k|k-1}(.|.)$ is the analogue of the state transition probability with the previous state. $q_k(.|.)$ is the proposal distribution. $\gamma_k(.)$ is the probability of the born particle. $\boldsymbol{Z}_k$ is the observation set at time-step k. $\kappa_k(\boldsymbol{z})$ denotes the clutter intensity of the observation $\boldsymbol{z}$ at time step k. $p_{D,k}(.|.)$ is the probability of detection at time step k. $g_k(.|.)$ is the likelihood of individual targets.

Algorithm 1 AV-SMC-PHD Filter

```
Input: ${\boldsymbol{m}_{k-1}^i, \omega_{k-1}^i}_{i=1}^{N_{k-1}}$, $N_\Gamma$, $\boldsymbol{Z}_k$ and DOA line.
Output: $\{\widetilde{\boldsymbol{m}}_k^j, \widetilde{\omega}_k^j\}_{j=1}^{\widetilde{N}_k}$, and $\{\boldsymbol{m}_k^i, \omega_k^i\}_{i=1}^{N_k}$.
  Run:
  Predict existing targets.
  if DOA exists then
    Calculate distances $\boldsymbol{d}_k$.
    Calculate movement distances $\hat{\boldsymbol{d}}_k$ by Eq. (2).
    Concentrate $\boldsymbol{m}_{s,k}$ around the DOA line by (3).
    if new speaker then
      Born $N_\Gamma$ particles uniformly around the DOA line by (4).
    end if
  end if
  Predict the weights of the particles $\omega_{k|k-1}^i$ by Eq. (5).
  (Optional) Update the states and the weights of the particles by the particle flow.
  Update the weights of the particles $\omega_{k|k}^i$ by Eq. (6) and calculate $\widetilde{N}_k = \sum_{i=1}^{N_k} \omega_{k|k}^i$.
  Get $\{\widetilde{\boldsymbol{m}}_k^j, \widetilde{\omega}_k^j\}_{j=1}^{\widetilde{N}_k}$ by the k-means method and get $\{\boldsymbol{m}_k^i, \omega_k^i\}_{i=1}^{N_k}$ by re-sampling.
```

2.2 Particle Flow

There are several particle flow algorithms. Here we use the zero diffusion exact flow [24], since it is straightforward to implement. Daum and Huang define the flow of the logarithm of the conditional probability density function with respect to step size λ [11]:

$$\log(\psi_k(\boldsymbol{m}, \lambda)) = \log(h_k(\boldsymbol{m})) + \lambda \log(g_k(\boldsymbol{m})) \tag{8}$$

where λ takes values from $[0, \triangle\lambda, 2\triangle\lambda, \cdots, N_\lambda \triangle\lambda]$, where $N_\lambda \triangle\lambda = 1$. $g_k(.)$ is the likelihood function. At the start of the flow ($\lambda = 0$), $\psi_k(\boldsymbol{m}_k, \lambda)$ represents the prior density, $h_k(.)$. At the end of the flow ($\lambda = 1$), $\psi_k(\boldsymbol{m}_k, \lambda)$ is translated into the normalized posterior density. This flow simulates the motion of the physical particles as Brownian movement [5] from the prior to the posterior density.

When the prior and the likelihood are unnormalized Gaussian probability densities, the exact solution for the particle flow is given as [10]:

$$\frac{d\boldsymbol{m}}{d\lambda} = \boldsymbol{A}(\lambda)\, \boldsymbol{m} + \boldsymbol{b}(\lambda) \tag{9}$$

where

$$\boldsymbol{A}(\lambda) = -\frac{1}{2}\boldsymbol{P}\boldsymbol{H}^T \left(\lambda \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \boldsymbol{R}\right)^{-1} \boldsymbol{H}, \tag{10}$$

$$\boldsymbol{b}(\lambda) = (\boldsymbol{I} + 2\lambda\boldsymbol{A}) \left[(\boldsymbol{I} + \lambda\boldsymbol{A})\, \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{A}\bar{\boldsymbol{m}}\right] \tag{11}$$

in which $\bar{\boldsymbol{m}}$ is the mean of the particle and $\boldsymbol{R}$ is the covariance matrix of the observation noise. For nonlinear problems, the observations need to be linearized for each particle (analogous to an extended Kalman filter). $\boldsymbol{P}$ is the covariance matrix of the particles. $\boldsymbol{H}$ is computed as the Jacobian matrix.

3 Proposed AV-PF-SMC-PHD Filter

In the AV-SMC-PHD filter as already summarised in Section 2.1, the particles need to be drawn from a proposal distribution. However, it may not be well matched to the posterior density because of the particle degeneracy issue [7]. To mitigate this problem, we add an adjustment step between the prediction step and update step, where the particle flow Eqs. (9)-(13) are applied to adjust the states and weights of the particles by smoothly migrating them from the prior to the posterior density.

In our proposed filter, audio information is used to calculate the number of particle flows. As in other multi-speaker particle filters, the particles need to be labeled [24] or cluttered [15] before updated. However, the prior information about the number of targets and the association between the targets and particles of the visual SMC-PHD filter is unknown and time-varying in multi-target tracking. In our method, such information could be provided by the DOA lines. We assume that the number of particle flows is the same as the number of DOA lines N_D. Then the particles are classified to N_D sets based on the Euclidean distance between the particles and the DOA lines. These particles are denoted as $\{\boldsymbol{m}^i_{k|k-1}, \omega^i_{k|k-1}\}_{i \in \boldsymbol{\Lambda}(\boldsymbol{z})}$, where $\boldsymbol{\Lambda}(\boldsymbol{z})$ is a subset of $\boldsymbol{E} = [1, \cdots, N + N_\Gamma]$. In practice, some particles are created due to clutter and noise. To account for the noise effect, we assume that each flow will only be influenced by the particles in the neighborhood of the DOA lines within a certain distance d.

The mean $\bar{\boldsymbol{m}}(\boldsymbol{z})$ and covariance $\boldsymbol{P}(\boldsymbol{z})$ are calculated based on different particle flows. The states of particles are adjusted by Eq. (9) and the weights of the particles also need to be adjusted as

$$\omega^i_{k|k-1} := \frac{q_k(\boldsymbol{m}^i_{k|k-1} | \boldsymbol{m}^i_{k-1}, \boldsymbol{z})}{q_k(\acute{\boldsymbol{m}}^i_{k|k-1} | \boldsymbol{m}^i_{k-1}, \boldsymbol{z})} \omega^i_{k|k-1} \tag{12}$$

where $\acute{\boldsymbol{m}}^i_{k|k-1}$ is the updated value of $\boldsymbol{m}^i_{k|k-1}$ by particle flow.

The pseudo-code of the adjustment step of the PF-SMC-PHD filter is presented in Algorithm 2. The observation of particle flow $\boldsymbol{z}$ is calculated by the color histogram matching [16]. The reference histogram $\boldsymbol{v}$ is updated with the estimate from the previous time step $k-1$. Note that $\boldsymbol{m}$ in Eq. (9) should be represented with $\boldsymbol{m}^i_{k|k-1}$, and

$$\boldsymbol{H} = \begin{bmatrix} cos\,(\theta) & -sin\,(\theta) \\ sin\,(\theta) & cos\,(\theta) \end{bmatrix} \tag{13}$$

where $\theta = arctan(\frac{\boldsymbol{m}(2)}{\boldsymbol{m}(1)})$, and $\boldsymbol{m}(1)$ and $\boldsymbol{m}(2)$ are the first and second element of $\boldsymbol{m}$, respectively.

4 Experimental Results

In this section, the proposed algorithm is compared with the visual SMC-PHD algorithm, the baseline AV-SMC-PHD algorithm in [20] using the AV16.3 dataset

Algorithm 2 Adjustment Step of the AV-PF-SMC-PHD Filter

```
Input: $\{\boldsymbol{m}^i_{k|k-1}, \omega^i_{k|k-1}\}^{N_k}_{i=1}$, $\boldsymbol{Z}_k$, $\boldsymbol{v}$ and the DOA line
Output: $\{\boldsymbol{m}^i_{k|k-1}, \omega^i_{k|k-1}\}^{N_k}_{i=1}$.
  Run:
  for each DOA line do
    Calculate $\boldsymbol{z}$ by the reference histogram $\boldsymbol{v}$ and input image $\boldsymbol{Z}_k$
    for $\lambda \in [0, \triangle\lambda, 2\triangle\lambda, \cdots, N_\lambda\triangle\lambda]$ do
      Calculate $\boldsymbol{H}$ via Eq. (13) and $\boldsymbol{A}$ and $\boldsymbol{b}$ by Eq. (10) and Eq. (11), respectively.
      Evaluate flow $\frac{d\boldsymbol{m}^i_{k|k-1}}{d\lambda}$ by Eq. (9) and $\boldsymbol{m}^i_{k|k-1} = \boldsymbol{m}^i_{k|k-1} + \triangle\lambda \frac{d\boldsymbol{m}^i_{k|k-1}}{d\lambda}$.
    end for
    Update the particle weights by Eq. (12).
  end for
```

[23]. In the visual SMC-PHD filter, the born particles are created randomly in the tracking area but the number of particles is the same as other filters. AV16.3 consists of sequences where speakers are walking and speaking at the same time. Those actions are recorded by three calibrated video cameras at 25 Hz and two circular eight-element microphone arrays at 16 kHz. The audio and video streams are synchronized before running the algorithms. The size of each image frame is 288x360 pixels. All the algorithms are tested with all the three different camera angles of four sequences: Sequences 24, 25, 30 and 45, which correspond to the cases of two and three speakers and are the most challenging sequences in term of movements of the speakers and the number of occlusions.

As in [20], the OSPA metric [28] is employed for measuring the tracking performance. The OSPA is able to evaluate the performance on target number estimation as well as the position estimation, which is suitable for multi-target tracking. A low OSPA implies a better performance. All experiments are run on a computer with Intel i7-3770 CPU with a clock frequency of 3.40 GHz and 8G RAM.

The parameters for the SMC-PHD filter are set as: $p_D = 0.98$, $p_S = 0.99$ and $\sigma_c = 0.1$. The uniform density u is $(360280)^{-1}$ and the number of particles per speaker is 50. The parameters for particle flow are set empirically as: $\triangle\lambda = 0.01$, $\boldsymbol{P} = [5, 5, 1, 1]$ and $d = 30$. The OSPA metric order parameter a is 2.

Due to page limit, we only show part of the results obtained. First, the OSPA results for Sequence 24 camera 1, as an example, is shown in Figure 1. The green dotted line is the OSPA for the visual SMC-PHD fitler, the blue dotted line is the OSPA for the AV-SMC-PHD filter, and the red solid line for the AV-PF-SMC-PHD filter. From frame 400 to frame 600 and from frame 800 to frame 1000, there is no occlusion. At most of time such as from frame 350 to frame 700, OSPA for the AV-PF-SMC-PHD filter is the lowest among the three filters. Compared with audio information, the targets can be more quickly tracked with a lower OSPA as compared with the visual filter, especially when a new target appears, such as from frame 0 to frame 350. However, from frame 750 to frame 800, the error of the AV-PF-SMC-PHD filter is larger than that of the AV-SMC-

PHD filter, since it is the end of the occlusion, and the particles are modified to the wrong direction by the visual information of the occluded speakers in the previous frame.

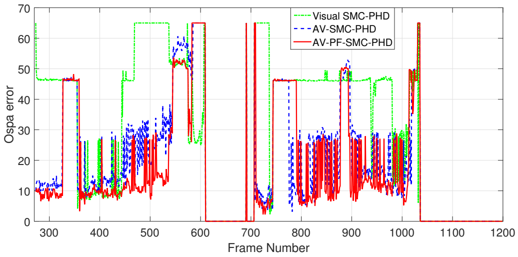

Fig. 1: Performance comparison of the visual SMC-PHD filter, the AV-SMC-PHD filter and the proposed AV-PF-SMC-PHD filters in terms of the OSPA error.

Other sequences are also used in the tests and the results of different methods are given in Table I. The average errors for the visual SMC-PHD filter, the AV-SMC-PHD filter and the AV-PF-SMC-PHD filter are 36.97, 22.75 and 20.14 respectively. With the same number of particles, the visual filter gives a higher OSPA than the audio-visual filters, which means audio information can improve the tracking accuracy of visual SMC-PHD filters. Apart from that, with the particle flow, 12.47% reduction in tracking error has been achieved. However, for the Sequence 24 camera 1, the computational cost has increased from 112 to 703 seconds. The computational cost for the AV-PF-SMC-PHD filter will also increase with the number of particles and targets. For example, the execution time for the Sequence 45 (1034s) is larger than that for Sequence 24 (153s).

5 Conclusion

We have presented a novel AV-PF-SMC-PHD filter for multi-speaker tracking, by adding an adjustment step to smoothly migrate the particles. The proposed algorithm has been tested on the AV16.3 dataset, where the number of speakers varies over time. The experimental results show that the AV filters offer a higher tracking accuracy than the visual filter with the same number of particles. The proposed particle flow method can improve the tracking accuracy over the AV-SMC-PHD filter, with a modest increase in the computational cost.

References

1. Baser, E., Efe, M.: A novel auxiliary particle PHD filter. In: 15th International Conference on Information Fusion. pp. 165–172. IEEE (2012)

		Visual SMC-PHD	AV-SMC-PHD	AV-PF-SMC-PHD
seq24	cam1	30.46	17.71	16.57
	cam2	35.91	19.83	17.04
	cam3	32.61	18.94	16.71
seq25	cam1	34.96	19.13	16.85
	cam2	31.86	18.47	16.72
	cam3	37.15	21.61	18.58
seq30	cam1	39.35	25.22	20.57
	cam2	35.24	19.37	16.92
	cam3	40.21	25.31	20.57
seq45	cam1	43.17	29.46	27.55
	cam2	43.20	29.47	27.55
	cam3	39.52	28.43	26.07
Average		**36.97**	**22.75**	**20.14**

Table 1: Experimental results for the visual SMC-PHD filter, the AV-SMC-PHD filter and the AV-PF-SMC-PHD filter in terms of the OSPA error.

2. Bernardin, K., Gehrig, T., Stiefelhagen, R.: Multi-level particle filter fusion of features and cues for audio-visual person tracking. In: Multimodal Technologies for Perception of Humans, pp. 70–81. Springer (2008)
3. Cevher, V., Sankaranarayanan, A.C., McClellan, J.H., Chellappa, R.: Target tracking using a joint acoustic video system. IEEE Transactions on Multimedia 9(4), 715–727 (2007)
4. Cui, P., Sun, L.F., Wang, F., Yang, S.Q.: Contextual mixture tracking. IEEE Transactions on Multimedia 11(2), 333–341 (2009)
5. Daum, F., Huang, J.: Particle flow for nonlinear filters with log-homotopy. Proceedings of SPIE pp. 696918–1–696918–12 (2008)
6. Daum, F., Huang, J.: Nonlinear filters with log-homotopy. In: Drummond, O.E., Teichgraeber, R.D. (eds.) Optical Engineering + Applications. pp. 669918–669918–15. International Society for Optics and Photonics (2007)
7. Daum, F., Huang, J.: Nonlinear filters with particle flow. In: SPIE Optical Engineering+ Applications. pp. 74450R–1–74450R–9. International Society for Optics and Photonics (2009)
8. Daum, F., Huang, J.: Particle flow for nonlinear filters. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing pp. 5920–5923 (2011)
9. Daum, F., Huang, J.: Renormalization group flow and other ideas inspired by physics for nonlinear filters, bayesian decisions, and transport. In: SPIE Defense+ Security. p. 90910I. International Society for Optics and Photonics (2014)
10. Daum, F., Huang, J.: Renormalization group flow in k-space for nonlinear filters, bayesian decisions and transport. In: 18th International Conference on Information Fusion. pp. 1617–1624. IEEE (2015)
11. Daum, F., Huang, J., Noushin, A.: Exact particle flow for nonlinear filters. In: SPIE Defense, Security, and Sensing. p. 769704 (2010)
12. Daum, F., Huang, J., Noushin, A.: Small curvature particle flow for nonlinear filters. Signal Processing, Sensor Fusion, and Target Recognition XIX 7697(1), 769704 (2010)

13. Daum, F., Huang, J., Noushin, A.: Coulomb's law particle flow for nonlinear filters. In: Drummond, O.E. (ed.) SPIE Optical Engineering+ Applications. pp. 1–15. International Society for Optics and Photonics (2011)
14. Fortmann, T., Bar-Shalom, Y., Scheffe, M.: Sonar tracking of multiple targets using joint probabilistic data association. IEEE Journal of Oceanic Engineering 8(3), 173–184 (1983)
15. Gatica-Perez, D., Lathoud, G., Odobez, J.M., McCowan, I.: Audiovisual probabilistic tracking of multiple speakers in meetings. IEEE Transactions on Audio, Speech, and Language Processing 15(2), 601–616 (2007)
16. Hafner, J., Sawhney, H.S., Equitz, W., Flickner, M., Niblack, W.: Efficient color histogram indexing for quadratic form distance functions. IEEE transactions on pattern analysis and machine intelligence 17(7), 729–736 (1995)
17. Khan, M.A., Ulmke, M.: Non-linear and non-Gaussian state estimation using log-homotopy based particle flow filters. 2014 Workshop on Sensor Data Fusion: Trends, Solutions, Applications, SDF 2014 (2014)
18. Khan, M.A., Ulmke, M.: Non-linear and non-gaussian state estimation using log-homotopy based particle flow filters. In: Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2014. pp. 1–6. IEEE (2014)
19. Kidron, E., Schechner, Y.Y., Elad, M.: Cross-modal localization via sparsity. IEEE Transactions on Signal Processing 55(4), 1390–1404 (2007)
20. Kilic, V., Barnard, M., Wang, W., Hilton, A., Kittler, J.: Mean-shift and sparse sampling based SMC-PHD filtering for audio informed visual speaker tracking. IEEE Transactions on Multimedia (2016)
21. Kılıç, V., Barnard, M., Wang, W., Kittler, J.: Audio assisted robust visual tracking with adaptive particle filtering. IEEE Transactions on Multimedia 17(2), 186–200 (2015)
22. Kong, A., Liu, J.S., Wong, W.H.: Sequential imputations and bayesian missing data problems. Journal of the American Statistical Association 89(425), 278–288 (1994)
23. Lathoud, G., Odobez, J.M., Gatica-Perez, D.: AV16. 3: an audio-visual corpus for speaker localization and tracking. In: International Workshop on Machine Learning for Multimodal Interaction. pp. 182–195. Springer (2004)
24. Li, Y., Zhao, L., Coates, M.: Particle flow for particle filtering. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 3979–3983. IEEE (2016)
25. Liu, Q., Rui, Y., Gupta, A., Cadiz, J.J.: Automating camera management for lecture room environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 442–449. ACM (2001)
26. Maggio, E., Piccardo, E., Regazzoni, C., Cavallaro, A.: Particle PHD filtering for multi-target visual tracking. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing. vol. 1, pp. I–1101. IEEE (2007)
27. Polat, E., Ozden, M.: A nonparametric adaptive tracking algorithm based on multiple feature distributions. IEEE Transactions on Multimedia 8(6), 1156–1163 (2006)
28. Ristic, B., Vo, B.N., Clark, D., Vo, B.T.: A metric for performance evaluation of multi-target tracking algorithms. IEEE Transactions on Signal Processing 59(7), 3452–3457 (2011)
29. Talantzis, F., Constantinides, A.G., Polymenakos, L.C.: Estimation of direction of arrival using information theory. IEEE Signal Processing Letters 12(8), 561–564 (2005)
30. Zhao, L., Wang, J., Li, Y., Coates, M.J.: Gaussian particle flow implementation of PHD filter. In: SPIE Defense+ Security. p. 98420D (2016)