Bernoulli Filter for Super-Distance Speech Tracking in Outdoor Environments Using a Circle Microphone Array

Yehui Zhang and Xiang Pan*

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, P. R. China 310012

*Email: panxiang@zju.edu.cn

Abstract. It is a challenge to track super-distance speech in an outdoor environment due to the existence of wind noise and interference sources. After the weak speech has been enhanced by beamforming carried out over a microphone array, Bernoulli filter is utilized to track the speech source. Bernoulli filter is one kind of Bayesian filters under the random finite set (RFS) framework, which can accurately estimate the presence and disappearance of the speech source. The outdoor experimental results have shown that the proposed method can effectively track the interesting speech.

1. Introduction

In outdoor environments, it is difficult to understand the super-distance speech due to the existence of wind noise and interference sources. The term “super distance” refers to the range larger than 30-40m. In general, a microphone array is utilized to receive speech signals, and the signal-to-noise ratio (SNR) of the received speech is further improved by beamforming techniques. In real scenarios, there exists lots of speech signals. Thus, we consider the “tracking before detection” strategy for detection of the interesting speech. In this paper, we focus on implementing the Bernoulli filter for tracking speech.

It is well known that the Kalman filter [1] (KF) is optimal in the linear/Gaussian target tracking problems. Due to the unknown presence time of the interesting speech signal, the KF, the extended Kalman filter [2] (EKF) and unscented Kalman filter [3] (UKF) do not perform well. These trackers do not account for existence of targets when target is absent, and the target cannot be captured when target is present [4]. Bernoulli filter is a kind of Bayesian filters under the random finite set (RFS) framework, suitable for the system which randomly switches on and off. However Bernoulli filter has no analytic solution generally. Bernoulli filters are implemented as particle filters in non-linear/non-Gaussian problems and Gaussian sum filters in linear problems.

For detection of the random signal modelled a randomly switch on or off system, it occurs that there is lots of miss-detections and false alarms. Thus, the Bayesian filter needs the extra resources to determine target trajectory and data association. For example, (a) logic-based track formation is used to represent the number of targets; (b) data association is used to deal with imperfect detection. Bayes filters based on the RFS formulate the optimal Bayes filter for such problems without complex logic operation and data association. For a single target tracking problem, where target randomly appears and disappears, with imperfect detections, Bernoulli filter is a good approximation to the optimal Bayes filter under the RFS framework.
In this paper, we will adopt Bernoulli filters to tracking a super-distant source using the output of beamforming of a circular microphone array. The basic theory of RFS and Bernoulli particle filters will be introduced in Section II. Experimental results are discussed in Section III. Section IV summarizes the paper.

2. Bernoulli filter
First, we discuss the basic theory of Bernoulli filter. The statement equation and measurement equation are described by equation (1) and equation (2), respectively.

\[ x_k = f(x_{k-1}, u_{k-1}) + n_{k-1}, \]

\[ z_k = h(x_k, u_k) + w_k, \]

where \( x_k = [x_k \hat{x}_k \hat{y}_k] \in \mathbb{R}^{2n} \) is the state of the dynamic system; \((x_k, y_k), (\hat{x}_k, \hat{y}_k)\) are position and velocity of target, respectively; \( f: \mathbb{R}^{2n} \times \mathbb{R}^n \rightarrow \mathbb{R}^{2n} \) and \( h: \mathbb{R}^{2n} \times \mathbb{R}^n \rightarrow \mathbb{R}^n \) are some known functions; \( u_k \in \mathbb{R}^n \) is the known control input; \( z_k \in \mathbb{R}^n \) is the measurement; \( n_{k-1} \) and \( w_k \) are independent statement and measurement noise sequences, whose covariances are \( Q_{k-1} \) and \( R_k \), respectively.

Since azimuth of the target is obtained by the beamforming method, thus the measurement equation is defined by

\[ z_k = \arctan\left(\frac{y_k - y_0}{x_k - x_0}\right) + w_k \]

where \((x_0, y_0)\) is the position of the observer. Obviously, it’s a nonlinear question.

2.1. Random finite set
A RFS is a special random variable, whose cardinality is also a random variable and specified by a distribution \( \rho(n) = P(\lvert X \rvert = n), \ n \in \mathbb{N}_0 \). A RFS \( X \) are unordered finite set-values and specified by a set of joint distribution \( p_n(x_1, \ldots, x_n) \), \( x_1, \ldots, x_n \in \mathcal{X} \), which describe the distribution of \( X \) conditioned on cardinality \( n \) [4].

In order to describe the probability density function (PDF) and moments for a RFS, we adopt finite set statistic (FISST), which can be find in [5, 6]. The FISST PDF of a RFS \( X \) is denoted by \( f(X) : \)

\[ f(X) = n! \cdot \rho(n) \cdot p_n(x_1, \ldots, x_n) \]

for \( n \in \mathbb{N}_0 \). It has been proved that \( f(X) \) integrates to one under FISST, satisfying the property of PDF.

Here we analysis two RFSs related with detection of random speech.

Bernoulli RFS: The Bernoulli RFS models \( X \) empty or has one element (with probability \( r \)) with PDF \( p(x) \). And its FISST PDF is given by:

\[ f(X) = \begin{cases} 1-r, & X = \emptyset \\ r \cdot p(x), & X = \{x\} \end{cases} \]

Poisson RFS: The number of this RFS is modelled as Poisson and all elements of \( X \) are independent with each other with PDF \( p(x) \). Its FISST PDF is given by:

\[ f(X) = e^{-\lambda} \prod_{x \in X} \lambda p(x) \]

The multi-object state at time \( k \) can be represented by a RFS \( X_k = \{x_{k,1}, \ldots, x_{k,n_k}\} \). The number of objects \( n_k \) is random and time-varying and so are states. We assume that multi-object dynamic system is a Markov process and its transition density is represented by \( \phi_{k|k-1}(X_k | X_{k-1}) \). Real measurements
of multi-object can always contain false alarms and miss-detections, which can be modelled a RFS $Z_k$. Then $\phi_k = (Z_k \mid X_k)$ is defined to represent the likelihood function of $Z_k$.

Here, we perform the stochastic Bayes filter in RFS framework. Suppose at time $k-1$, we have estimated the posterior FISST PDF $f_{ijk-1}(X_{k-1} \mid Z_{k-1})$ ($Z_{k-1}$ represents the history measurements until time $k-1$) and $Z_k$ is received at time $k$. Then the predict and update step can be achieved using multi-object posterior densities according to the Bayes rule as follows:

$$f_{ijk}(X_k \mid Z_{k:i}) = \int \phi_{ijk}(X_k \mid Z_{k:i}) f_{ijk-1}(X_{k-1} \mid Z_{k-1}) \delta X_{k-1}$$  \hspace{1cm} (7)

$$f_{ijk}(X_k \mid Z_{k:i}) = \frac{\phi_k(Z_k \mid X_k) f_{ijk-1}(X_{k-1} \mid Z_{k-1})}{\int \phi_k(Z_k \mid X) f_{ijk-1}(X \mid Z_{k-1}) \delta X}$$  \hspace{1cm} (8)

It can be seen that the integrals in equation (7) and equation (8) are set integrals, which is intractable.

2.2. Bernoulli particle filter

It is assumed that there is only one target may be present or absent. The Bernoulli RFS is utilized to model the target state $X_k$ at time $k$ and its FISST PDF is given by

$$f(X_k) = \begin{cases} 1-r, & X_k = \emptyset \\ r \cdot p(x_k), & X_k = \{x_k\} \\ 0, & otherwise \end{cases}$$  \hspace{1cm} (9)

the detail derivation of Bernoulli filter can be found in [4] for the following formula.

Suppose that we have estimated the state $X_{k-1}$ at time $k-1$, which can be empty or a singleton. Its FISST PDF $f_{k-1}(X_{k-1})$ is specified by $r_{k-1}$ and $p_{k-1}(x_{k-1})$. When $X_{k-1} = \emptyset$, the new target can be born with probability $p_b$ and state probability density $b_k(x_k)$ or remains absent with probability $1-p_b$. For the Bernoulli Markov process, the conditional function upon $X_{k-1} = \emptyset$ can be described by

$$\phi_{ijk-1}(X_k \mid \emptyset) = \begin{cases} 1-p_b, & X_k = \emptyset \\ p_b \cdot b_k(x_k), & X_k = \{x_k\} \\ 0, & otherwise \end{cases}$$  \hspace{1cm} (10)

When $X_{k-1} = \{x_{k-1}\}$, the target may survive with probability $p_s$ to the next time and its Markov transition probability is $\pi_{ijk}(x_k \mid x_{k-1})$ or dead with probability $1-p_s$. The Bernoulli Markov process, the conditional function upon $X_{k-1} = \{x_{k-1}\}$ can be given by

$$\phi_{ijk}(X_k \mid \{x_{k-1}\}) = \begin{cases} 1-p_s, & X_k = \emptyset \\ p_s \cdot \pi_{ijk}(x_k \mid x_{k-1}), & X_k = \{x_k\} \\ 0, & otherwise \end{cases}$$  \hspace{1cm} (11)

It has been proved that predicted density $f_{ijk-1}(X_k \mid X_{k-1})$ is still a Bernoulli specified by $r_{ijk-1}$ and $p_{ijk-1}(x_k)$, where

$$r_{ijk-1} = p_b (1-r_{k-1}) + p_s r_{k-1}$$  \hspace{1cm} (12)

$$p_{ijk-1}(x_k) = p_b (1-r_{k-1}) b_k(x_k) + p_s r_{k-1} \int \pi_{ijk-1}(x_k \mid x) p_{k-1}(x) dx$$  \hspace{1cm} (13)

The measurements coming from a detector can be represented by a RFS $Z_k = \{z_{k,1}, \ldots, z_{k,nk} \}$ at time $k$. It is a union of two independent RFSs.
\[ Z_k = C_k \cup W_k \]  

where \( C_k \) denotes the false detections (clutter) and \( W_k \) is generated by target. \( W_k \) can be a singleton or empty for a point target here, if detected, which only generates one detection. When \( X_k = \emptyset \), \( W_k \) is an empty set and elements in \( Z_k \) are all clutters. The FISST PDF of clutter detections is

\[ \psi_k(Z_k | \emptyset) = e^{\lambda c(z)} \prod_{z \in Z_k} \lambda c(z) \]  

(15)

When \( X_k \) is not an empty set, target may be detected with probability \( p_d \) or not. When a target has been detected, one element of \( Z_k \) is related to it and \( W_k \) is a singleton. Otherwise \( W_k \) is an empty set. Hence we can write

\[
\eta(W_k \mid \{ x_i \}) = \begin{cases} 1 - p_d, & \text{if } W = \emptyset \\ p_d g(z \mid x_i), & \text{if } W = \{ z \} \end{cases}
\]  

(16)

where \( g(z \mid x_i) \) is the likelihood function of \( z \). Since we do not known which element of \( Z_k \) coming from target, we assume every \( z \) can be target’s measurement while others are clutters. Also target can be miss-detected. When \( X_k = \{ x_i \} \), the FISST PDF is given by:

\[
\psi_k(Z_k \mid \{ x_i \}) = \sum_{W_k \subset Z_k} \eta(W_k \mid \{ x_i \}) \kappa(Z_k \setminus W_k) 
\]  

(17)

when \( \kappa \) is defined by (1.15) and sign \( \setminus \) denotes set-difference operate. Then we put equation (15) and equation (16) into equation (17), which is simplified to

\[
\psi_k(Z_k \mid \{ x_i \}) = \kappa(Z_k) \left[ 1 - p_d + p_d \sum_{z \in Z_k} g(z \mid \{ x_i \}) \frac{\kappa(Z_k \setminus \{ z \})}{\kappa(Z_k)} \right] 
\]  

(18)

Now we are in the position to update the posterior density using Bayesian formula, which is given by

\[

f_{r \mid k}(X_k \mid Z_{ik}) = \frac{\psi_k(Z_k \mid X_k) \cdot f_{r \mid k-1}(X_k \mid Z_{ik-1})}{f_k(Z_k \mid Z_{ik-1})} 
\]  

(19)

The posterior density \( f_{r \mid k}(X_k \mid Z_{ik}) \) is still a Bernoulli RFS and can be specified by \( r_k \) and \( p_k(x_k) \), where

\[
r_k = \frac{1 - \Delta_k}{1 - r_{ik-1} \Delta_k} r_{ik-1} \\
p_k(x_k) = \frac{1 - p_d + p_d \sum_{z \in Z_k} g(z \mid x_k) \lambda c(z)}{1 - \Delta_k} 
\]  

(20)

(21)

where \( \Delta_k = p_d \left( 1 - \sum_{x \in Z_k} g(x \mid x_k) p_{ik-1}(x) dx / \lambda c(z) \right) \).

Since the measurement equation is non-linear, Bernoulli filter has no analytical solution. Therefore, Bernoulli filter is performed as Particle filter in this paper and the pseudocode is given in table 1. The details may be found in [4].
Table 1. Bernoulli particle filter.

1. Input: \( r_{k-1}, \{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{N+B}, Z_k \);
2. Calculate target existence probability according to (1.12);
3. Update survive particles: \( x_{ik-1}^{(i)} \sim \theta_k(x^{(i)}, Z_k), \ i=1,\ldots,N \);
4. Birth particles: \( x_{ik-1}^{(i)} \sim \rho_k(x_{ik-1}^{(i)} | Z_k), \ i=N+1,\ldots,N+B \);
5. Predict weight of particle \( w_k^{(i)} \);
6. For every \( z \in Z_k \), calculate \( I_k(z) \approx \sum_{i=1}^{N+B} w_{ik-1}^{(i)} \cdot g(z \mid x_{ik-1}^{(i)}) \);
7. Calculate \( \Delta_k \approx p_d \left( 1 - \sum_{i \in Z_k} I_k(z) \right) \);
8. Update target existence probability \( r_k = \frac{1 - \Delta_k}{1 - r_{k-1} \Delta_d} r_{k-1} \);
9. Update weight of particle \( \tilde{w}_k^{(i)} \approx \left[ 1 - p_d + p_d \sum_{z \in Z_k} g(z \mid x_{ik-1}^{(i)}) \right] \cdot w_{ik-1}^{(i)}, \ i=1,\ldots,N+B \);
10. Normalization: \( w_k^{(i)} = \frac{\tilde{w}_k^{(i)}}{\sum_{j=1}^{N+B} \tilde{w}_k^{(j)}}, \ i=1,\ldots,N+B \);
11. Resample \( \left\{ x_k^{(i)} \right\}_{i=1}^{N}, \left\{ w_k^{(i)} \right\}_{i=1}^{N+B} \);
12. If \( r_{k-1} > \text{threshold} \) output \( \left\{ x_k^{(i)} \right\}_{i=1}^{N}, \left\{ w_k^{(i)} \right\}_{i=1}^{N+B} \);

3. Experimental results and Analysis

In the outdoor experiments, we utilize a microphone array, a uniform circle array with radius of 0.5m and the number of array elements of 80 as shown in figure 1. The experiment was conducted in the playground of Yuquan Campus of Zhejiang university on 25 November, 2017 and the experimental setup is shown in figure 2. The speech of interest is a segment of the English lecture. The environmental noise is about 50dB. The detail sketch of experimental arrangement is shown in figure 3.

Figure 1. Circle array with 80 elements.
Figure 2. Experimental setup.
Figure 3. Sketch of experimental arrangement

The first element of the circular array is located at $0^\circ$, and the clockwise direction is defined as a negative angle, while the counterclockwise direction is positive angle. In the experiment, the speech source moves from $-45^\circ$ to $45^\circ$ along the straight line as shown in figure 3. The azimuth angle was estimated by the conventional beamforming (CBF) and the estimated direction of arrival (DOA) is marked with black plus in figure 4.

Figure 4. Tracking result with Bernoulli filter

The blue line in figure 4 represents the moving trajectory of player, and its movement process includes three stages of uniform motion and two stages of suspension. Since the speech pauses at 30s
and 45s for about 2 seconds, we cannot estimate the DOA of the speech signal. However, we can use this phenomenon to describe the random appearance and disappearance of the target, which can exactly be handled by Bernoulli filter. And we can observe that the blue line does not appear at the beginning and end of the period, which also represents the birth and death of the target.

The red stars in figure 4 represent the tracking results of Bernoulli filter. It is seen from figure 4 that the Bernoulli filter does not immediately track the target when it first appears, and it takes about four sampling periods to capture the target. After detecting the existence of the target, Bernoulli filter accurately estimates the target's orientation among the clutters. When the speech pauses for the first time at 29s, Bernoulli filter estimates the absence of target and does not output state of target. In the following, the tracker works well. When the target finally disappears, Bernoulli filter estimates the disappearance of the target just delay a sampling period. The red point in figure 5 is the estimation of cardinality, and 0 means the absence of target and 1 means the appearance of target. The estimation of cardinality of target is consistent with the results shown in figure 4.

4. Conclusion
The performance of tracking outdoor super-distant speech can be improved by using a large microphone array. Under the random finite set framework, Bernoulli filter can accurately estimate the state of the target and accurately estimate the appearance and disappearance of the target when the measuring set contains clutter. The outdoor speech trials have verified the effectiveness of Bernoulli filter in the tracking outdoor speech under uncertain detection.

Acknowledgments
The authors would like to thank the financial support from the National Natural Science Foundation of China (No.6151397 and No. 41776108), National Defence Key Laboratory of Funds (No. 614210903010617) and National Key Research and Development Program of China (No.2017YFC0306901).

References
[1] Kalman and R E 1960 A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1), 35-45.
[2] Daum F 2005 Nonlinear filters: beyond the Kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, 20(8), 57-69.
[3] Wan E A and Van Der Merwe R 2001 The unscented Kalman filter. *Kalman filtering and neural networks*, 221-280.
[4] Ristic B, Vo B T and Vo B N 2013 A tutorial on Bernoulli filters: theory, implementation and applications. *IEEE Transactions on Signal Processing* 61(13): 3406-3430.
[5] R Mahler 2007 Statistical Multisource Multitarget Information Fusion. *Norwood, MA, USA: Artech House*.
[6] Vo B T, See C M, Ma N and Ng W T 2012 Multi-sensor joint detection and tracking with the Bernoulli filter. *IEEE Transactions on Aerospace and Electronic Systems*, 48(2), 1385-1402.