

基于数值积分卡尔曼-概率假设密度滤波的多说话人跟踪方法

陈里铭¹ 陈 喆¹ 殷福亮¹ 侯代文²

(1. 大连理工大学信息与通信工程学院 大连 116023;
2. 91439 部队 460 所 大连 116041)

摘 要: 针对多说话人跟踪的非线性系统模型, 提出了一种基于数值积分卡尔曼-概率假设密度滤波的多说话人跟踪方法。该方法采用麦克风阵列的时间延迟估计作为观测数据, 利用具有三次代数精度的球面-径向数值积分准则计算非线性系统贝叶斯滤波器中的多维积分, 通过数值积分卡尔曼滤波和概率假设密度滤波对后验多说话人状态的一阶统计量进行估计, 并通过递推更新得到说话人状态信息, 实现非线性高斯系统的多说话人跟踪。该方法无需求解非线性系统函数的雅可比矩阵, 且计算量较小。仿真实验分析了检测概率、虚警点数目、采样周期、信噪比以及混响时间变化时跟踪算法的性能。实验结果表明, 该方法降低了系统模型非线性对滤波算法的影响, 增强了跟踪算法的鲁棒性, 提高了说话人状态和数目的估计精度。

关键词: 多说话人跟踪; 概率假设密度滤波; 卡尔曼滤波; 非线性滤波

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1003-0530(2012)09-1209-10

Cubature Kalman-Probability Hypothesis Density Filter for Multiple Speakers Tracking

CHEN Li-ming¹ CHEN Zhe¹ YIN Fu-liang¹ HOU Dai-wen²

(1. School of Information and Communication Engineering, Dalian University of Technology, Dalian 116023, China; 2. Institute 460 Unit 91439, PLA, Dalian 116041, China)

Abstract: Aiming at nonlinear system model in multiple speakers tracking, a cubature Kalman-probability hypothesis density filter for multiple speakers tracking is proposed in this paper. Time difference of arrival for microphone array is taken as measurements, third-degree spherical-radial rule is utilized to compute the multidimensional integral in Bayesian filter of nonlinear system in proposed method, cubature Kalman filter and probability hypothesis density filter is applied to estimate first-order statistical moment of posterior multiple speakers states, and finally multiple speakers tracking of nonlinear Gaussian system is realized while the speakers' states are extracted by recursive updating. Compared with some filters in multiple speakers tracking, the proposed method has several advantages. Calculating Jacobian matrix of nonlinear system function, which is usually hard to be done, is no longer necessary in proposed filter and computational complexity also goes down. Simulation experiments have been taken to analyze the performance of proposed method when detection probability, false speakers' number, sampling period, speech-signal-to-noise ratio and reverberation time varies. Simulation results show that the proposed method reduces the impact on the performance of filtering algorithm from nonlinear system model, enhances the robustness of the algorithm, and improves estimation accuracy of multiple speakers' number and states.

Key words: Multiple speakers tracking; Probability hypothesis density filter; Kalman filter; Nonlinear filter

1 引言

多说话人跟踪是利用语音信息对应用环境中存在的多个说话人的位置状态进行估计。目前,多说话人跟踪技术在多媒体会议系统、视频监控、人机接口、机器人等领域得到了广泛应用^[1-3]。

自1970年代起,人们相继提出了一些经典的声源定位方法,如波束形成方法、高分辨率谱估计方法、时延估计方法等^[4-5],这些方法已成功应用在说话人定位领域。但受房间混响等因素影响,这些声源定位方法容易产生虚声源,从而导致说话人位置估计错误。针对这一问题,Vermaak等提出了基于贝叶斯框架的声源跟踪方法^[6],该方法首先建立说话人运动的动态方程,然后对观测值进行滤波,以减小房间混响对跟踪结果的影响。在实际应用中,由于说话人数目通常都是未知且随时间变化,Ma等提出了基于随机有限集理论的多说话人跟踪方法^[7],该方法用序贯蒙特卡罗技术实现贝叶斯随机有限集滤波器,将其用于时间延迟估计的滤波,并通过设定说话人标签进行数据关联,实现说话人跟踪。该方法在说话人出现和消失的时间未知,或者说话人数目发生变化时,仍然可以进行有效跟踪,但其计算量较大。

基于随机有限集理论的多说话人跟踪系统是非线性的,因此可以从非线性滤波角度解决多说话人跟踪问题。已有的适用于非线性系统的多目标跟踪算法可分为两大类:全局方法和局部方法^[8]。其中全局方法无需对后验密度函数的形式进行假设,但计算量大,如高斯和概率假设密度(Gaussian-Sum Probability Hypothesis Density, GSPHD)滤波^[9]、序贯蒙特卡罗概率假设密度滤波(Sequential Monte Carlo-Probability Hypothesis Density, SMC-PHD)^[10]等。相对全局方法,局部方法需要假设后验密度函数具有某种先验形式,但滤波器的设计简单,计算量小,如扩展卡尔曼概率假设密度(Extended Kalman Probability Hypothesis Density, EK-PHD)滤波^[11]、无轨迹卡尔曼概率假设密度(Unscented Kalman Probability Hypothesis Density, UK-PHD)滤波^[11]和数值积分卡尔曼概率假设密度(Cubature Kalman Probability Hypothesis Density, CK-PHD)滤波^[12]。EK-PHD滤波方法是使用较多的非线性滤波方法,但该方法对系统线性化时,需要求解非线性函数的雅可比(Jacobian)矩阵,且仅具有一阶精度,通常会带来较大的线性

化误差^[13]。UK-PHD滤波方法提高了滤波精度,但当引入的尺度参数小于零时,可能导致更新后的方差阵为非正定阵,从而影响系统的稳定性^[8,13]。CK-PHD滤波方法用球面-径向数值积分准则得到数值积分节点,并对概率假设密度滤波中的多维积分进行数值计算,它在保持计算量较小的同时,能够得到优于EK-PHD的滤波效果,是一种有效的非线性滤波方法。

针对多说话人跟踪的非线性系统模型,本文将CK-PHD滤波器应用于多说话人跟踪中,提出了一种基于数值积分卡尔曼-概率假设密度滤波的多说话人跟踪方法。该方法利用麦克风阵列的时延估计作为观测数据,利用具有三次代数精确度的球面-径向数值积分准则计算概率假设密度滤波器中的多维积分,利用数值积分卡尔曼滤波和概率假设密度滤波来估计表示后验多目标状态一阶统计量的高斯成分,并通过递推更新高斯成分的参数得到说话人的状态,从而实现非线性系统的多说话人跟踪。该方法降低了系统模型非线性对滤波精度的影响,增强了非线性滤波算法的鲁棒性,提高了说话人位置和数目的估计精度。仿真实验结果验证了本文方法的有效性。

2 数值积分卡尔曼滤波

数值积分卡尔曼滤波(Cubature Kalman Filter, CK-F)是一种非线性滤波方法,该方法基于球面-径向数值积分准则,利用数值积分节点,对贝叶斯滤波器中的积分进行数值计算,以得到高斯假设下贝叶斯滤波器的最佳近似^[8]。

设目标状态向量为 $\mathbf{x}_k \in \mathbb{R}^{n_x}$,目标观测向量为 $\mathbf{z}_k \in \mathbb{R}^{n_z}$,则非线性动态系统模型为^[14]

$$\begin{cases} \mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{v}_{k-1} \\ \mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k \end{cases} \quad (1)$$

其中 f_k 和 g_k 均为非线性函数, $\mathbf{u}_k \in \mathbb{R}^{n_u}$ 为系统输入, \mathbf{v}_{k-1} 和 \mathbf{w}_k 均为独立同分布的高斯白噪声,且分别满足 $\mathbf{v}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$, $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ 。

在数值积分卡尔曼滤波方法中,假设非线性系统(1)的贝叶斯滤波器中预测密度和似然密度均服从高斯分布,则贝叶斯滤波器的核心是计算服从高斯分布的加权积分,且所有积分形式均为非线性函数 \times 高斯分布密度函数^[8]。

设向量 $\mathbf{e}_p = (0, \dots, 1, \dots, 0)^T \in \mathbb{R}^{n_x}$,其中1在向量

的第 p 个位置, $\mathbf{I} \in \mathbb{R}^{n_x}$ 为单位矩阵, 利用具有三次代数精确度的球面-径向数值积分准则, 计算标准正态分布加权积分为^[8]

$$I_N(f) = \int_{\mathbb{R}^{n_x}} f(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{I}) d\mathbf{x} \approx \sum_{p=1}^m w_p f(\boldsymbol{\xi}_p) \quad (2)$$

其中

$$\begin{cases} \boldsymbol{\xi}_p = \sqrt{m/2} [1]_p, & p=1, 2, \dots, m=2n_x \\ w_p = 1/m \end{cases} \quad (3)$$

这里

$$[1]_p = \begin{cases} \mathbf{e}_p, & p=1, 2, \dots, n_x \\ -\mathbf{e}_{p-n_x}, & p=n_x+1, n_x+2, \dots, m \end{cases} \quad (4)$$

由式(2), 利用数值积分节点集合 $\{\boldsymbol{\xi}_p, w_p\}$ 对贝叶斯滤波方法中的积分进行数值计算, 得到数值积分卡尔曼滤波器为

$$\mathbf{X}_{p,k|k-1}^* = f(\mathbf{S}_{k-1} \boldsymbol{\xi}_p + \hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) \quad (5)$$

$$\hat{\mathbf{x}}_{k|k-1} = \frac{1}{m} \sum_{p=1}^m \mathbf{X}_{p,k|k-1}^* \quad (6)$$

$$\mathbf{P}_{k|k-1} = \frac{1}{m} \sum_{p=1}^m \mathbf{X}_{p,k|k-1}^* \mathbf{X}_{p,k|k-1}^{*\top} - \hat{\mathbf{x}}_{k|k-1} \hat{\mathbf{x}}_{k|k-1}^\top + \mathbf{Q}_{k-1} \quad (7)$$

$$\mathbf{Z}_{p,k|k-1} = h(\mathbf{S}_{k|k-1} \boldsymbol{\xi}_p + \hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k) \quad (8)$$

$$\hat{\mathbf{z}}_{k|k-1} = \frac{1}{m} \sum_{p=1}^m \mathbf{Z}_{p,k|k-1} \quad (9)$$

$$\mathbf{P}_{zz,k|k-1} = \frac{1}{m} \sum_{p=1}^m \mathbf{Z}_{p,k|k-1} \mathbf{Z}_{p,k|k-1}^\top - \hat{\mathbf{z}}_{k|k-1} \hat{\mathbf{z}}_{k|k-1}^\top + \mathbf{R}_k \quad (10)$$

$$\begin{aligned} \mathbf{P}_{xz,k|k-1} &= \frac{1}{m} \sum_{p=1}^m (\mathbf{S}_{k|k-1} \boldsymbol{\xi}_p \\ &+ \hat{\mathbf{x}}_{k|k-1}) \mathbf{Z}_{p,k|k-1}^\top - \hat{\mathbf{x}}_{k|k-1} \hat{\mathbf{z}}_{k|k-1}^\top \end{aligned} \quad (11)$$

$$\mathbf{W}_k = \mathbf{P}_{xz,k|k-1} \mathbf{P}_{zz,k|k-1}^{-1} \quad (12)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \quad (13)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{P}_{zz,k|k-1} \mathbf{W}_k^\top \quad (14)$$

其中 $\mathbf{P}_{k-1} = \mathbf{S}_{k-1} \mathbf{S}_{k-1}^\top$, $\mathbf{P}_{k|k-1} = \mathbf{S}_{k|k-1} \mathbf{S}_{k|k-1}^\top$ 。

在式(5)-(14)的数值积分卡尔曼滤波器中, 对非线性系统进行状态估计时不要求解雅克比矩阵, 且滤波过程中不需设定尺度参数, 从而提高了滤波系统的稳定性。

3 数值积分卡尔曼-概率假设密度滤波算法

概率假设密度 (Probability Hypothesis Density, PH-

D) 滤波在贝叶斯框架下以递推形式传递目标的概率假设密度, 以实现为目标状态和目标数目的有效估计。数值积分卡尔曼-概率假设密度 (CK-PHD) 滤波是 PHD 滤波的一种实现形式, 其主要优点是滤波过程无需求解雅克比矩阵以及设定尺度参数, 且计算量较小。该滤波器以式(1)作为系统模型, 其实现主要有三个步骤: PHD 递推、高斯成分修剪和状态提取^[12]。其中 PHD 递推是该滤波器的关键步骤, 它包含预测和更新两部分。

由于目标的出现和消失是随机的, k 时刻的预测概率假设密度 $\nu_{k|k-1}$ 表示为目标继续存在、自发产生新目标和分离得到新目标三种情况的概率假设密度之和^[11], 即

$$\nu_{k|k-1}(\mathbf{x}) = \nu_{s,k|k-1}(\mathbf{x}) + \nu_{\beta,k|k-1}(\mathbf{x}) + \gamma_k(\mathbf{x}) \quad (15)$$

其中 $\nu_{s,k|k-1}(\mathbf{x})$, $\nu_{\beta,k|k-1}(\mathbf{x})$ 和 $\gamma_k(\mathbf{x})$ 可分别用高斯混合形式表示, 即

$$\nu_{s,k|k-1}(\mathbf{x}) = \sum_{j=1}^{J_{k-1}} w_{s,k|k-1}^{(j)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{s,k|k-1}^{(j)}, \mathbf{P}_{s,k|k-1}^{(j)}) \quad (16)$$

$$\nu_{\beta,k|k-1}(\mathbf{x}) = \sum_{j=1}^{J_{k-1}} \sum_{l=1}^{J_{\beta,k}} w_{\beta,k|k-1}^{(j,l)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{\beta,k|k-1}^{(j,l)}, \mathbf{P}_{\beta,k|k-1}^{(j,l)}) \quad (17)$$

$$\gamma_k(\mathbf{x}) = \sum_{j=1}^{J_{\gamma,k}} w_{\gamma,k}^{(j)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{\gamma,k}^{(j)}, \mathbf{P}_{\gamma,k}^{(j)}) \quad (18)$$

其中 $\nu_{\beta,k|k-1}(\mathbf{x})$ 可由目标分离前的状态得到, 具体形式参见文献[11], $\gamma_k(\mathbf{x})$ 的各参数作为滤波器输入参数可提前设定, $\nu_{s,k|k-1}(\mathbf{x})$ 可利用球面-径向数值积分准则对多维积分进行数值计算, 得到高斯成分的参数。

设 $k-1$ 时刻概率假设密度 ν_{k-1} 中高斯成分的权值、均值和协方差矩阵分别为 $w_{k-1}^{(j)}$ 、 $\mathbf{m}_{k-1}^{(j)}$ 和 $\mathbf{P}_{k-1}^{(j)}$, $j=1, \dots, J_{k-1}$, 且 $\mathbf{P}_{k-1}^{(j)} = \mathbf{S}_{k-1}^{(j)} [\mathbf{S}_{k-1}^{(j)}]^\top$, 则目标继续存在的预测概率假设密度 $\nu_{s,k|k-1}(\mathbf{x})$ 中高斯成分的权值、均值和协方差分别为

$$w_{s,k|k-1}^{(j)} = p_{s,k} w_{k-1}^{(j)} \quad (19)$$

$$\mathbf{X}_{p,k|k-1}^* = f(\mathbf{S}_{k-1}^{(j)} \boldsymbol{\xi}_p + \mathbf{m}_{k-1}^{(j)}, \mathbf{u}_{k-1}) \quad (20)$$

$$\mathbf{m}_{s,k|k-1}^{(j)} = \frac{1}{m} \sum_{p=1}^m \mathbf{X}_{p,k|k-1}^* \quad (21)$$

$$\begin{aligned} \mathbf{P}_{s,k|k-1}^{(j)} &= \frac{1}{m} \sum_{p=1}^m \mathbf{X}_{p,k|k-1}^* \mathbf{X}_{p,k|k-1}^{*\top} \\ &- \mathbf{m}_{s,k|k-1}^{(j)} \mathbf{m}_{s,k|k-1}^{(j)\top} + \mathbf{Q}_{k-1} \end{aligned} \quad (22)$$

其中 $p_{s,k}$ 为 $k-1$ 时刻存在的目标在 k 时刻继续存在的

概率, ξ_p 具体形式如式(3)所示。

根据式(19)-(22), 就可以用高斯分布加权近似表示目标继续存在的预测概率假设密度 $\nu_{s,k|k-1}(\mathbf{x})$, 如式(16)所示。

设 k 时刻预测概率假设密度 $\nu_{k|k-1}$ 中高斯成分个数为 $J_{k|k-1}$, 由式(15)-(22)有

$$\nu_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}) \quad (23)$$

在 CK-PHD 递推过程的更新中, k 时刻的概率假设密度 ν_k 同样可写成高斯和形式, 即

$$\nu_k(\mathbf{x}) = (1 - p_{D,k})\nu_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z} \in Z_k} \nu_{D,k}(\mathbf{x}; \mathbf{z}) \quad (24)$$

$$\nu_{D,k}(\mathbf{x}; \mathbf{z}) = \sum_{j=1}^{J_{k|k-1}} w_k^{(j)}(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k}^{(j)}(\mathbf{z}), \mathbf{P}_{k|k}^{(j)}) \quad (25)$$

其中 $p_{D,k}$ 为 k 时刻的检测概率, Z_k 为观测值集合。

设 k 时刻概率假设密度 ν_k 的高斯成分个数为 J_k , 由式(23)-(25)有

$$\nu_k(\mathbf{x}) = \sum_{i=1}^{J_k} w_k^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_k^{(i)}, \mathbf{P}_k^{(i)}) \quad (26)$$

设 k 时刻噪声随机有限集的概率假设密度为 $\kappa_k(\mathbf{z})$, 则式(26)中高斯成分的权值 $w_k^{(j)}$ 、均值 $\mathbf{m}_{k|k}^{(j)}$ 和协方差矩阵 $\mathbf{P}_{k|k}^{(j)}$ 的递推公式为

$$\mathbf{Z}_{p,k|k-1}^{(j)} = h(\mathbf{S}_{k|k-1}^{(j)} \xi_p + \mathbf{m}_{k|k-1}^{(j)}, \mathbf{u}_k) \quad (27)$$

$$\mathbf{z}_{k|k-1}^{(j)} = \frac{1}{m} \sum_{p=1}^m \mathbf{Z}_{p,k|k-1}^{(j)} \quad (28)$$

$$\mathbf{P}_{zz,k|k-1}^{(j)} = \frac{1}{m} \sum_{p=1}^m \mathbf{Z}_{p,k|k-1}^{(j)} [\mathbf{Z}_{p,k|k-1}^{(j)}]^\top - \mathbf{z}_{k|k-1}^{(j)} [\mathbf{z}_{k|k-1}^{(j)}]^\top + \mathbf{R}_k \quad (29)$$

$$\mathbf{P}_{xz,k|k-1}^{(j)} = \frac{1}{m} \sum_{p=1}^m (\mathbf{S}_{k|k-1}^{(j)} \xi_p + \mathbf{m}_{k|k-1}^{(j)}) [\mathbf{Z}_{p,k|k-1}^{(j)}]^\top - \mathbf{m}_{k|k-1}^{(j)} [\mathbf{z}_{k|k-1}^{(j)}]^\top \quad (30)$$

$$w_k^{(j)}(\mathbf{z}) = \frac{p_{D,k} W_{k|k-1}^{(j)} q_k^{(j)}(\mathbf{z})}{\kappa_k(\mathbf{z}) + p_{D,k} \sum_{l=1}^{J_{k|k-1}} W_{k|k-1}^{(l)} q_k^{(l)}(\mathbf{z})} \quad (31)$$

$$q_k^{(j)}(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{z}_{k|k-1}^{(j)}, \mathbf{P}_{zz,k|k-1}^{(j)}) \quad (32)$$

$$\mathbf{W}_k^{(j)} = \mathbf{P}_{xz,k|k-1}^{(j)} [\mathbf{P}_{zz,k|k-1}^{(j)}]^{-1} \quad (33)$$

$$\mathbf{m}_{k|k}^{(j)}(\mathbf{z}) = \mathbf{m}_{k|k-1}^{(j)} + \mathbf{W}_k^{(j)}(\mathbf{z} - \mathbf{z}_{k|k-1}^{(j)}) \quad (34)$$

$$\mathbf{P}_{k|k}^{(j)} = \mathbf{P}_{k|k-1}^{(j)} - \mathbf{W}_k^{(j)} \mathbf{P}_{zz,k|k-1}^{(j)} [\mathbf{W}_k^{(j)}]^\top \quad (35)$$

其中 $\mathbf{S}_{k|k-1}^{(j)}$ 为协方差阵 $\mathbf{P}_{k|k-1}^{(j)}$ 的 Cholesky 展开, 即 $\mathbf{P}_{k|k-1}^{(j)} =$

$$\mathbf{S}_{k|k-1}^{(j)} [\mathbf{S}_{k|k-1}^{(j)}]^\top.$$

上述 CK-PHD 滤波算法的 PHD 递推过程的具体步骤总结如下:

输入: 观测集 Z_k 和 $\{w_{k-1}^{(i)}, \mathbf{m}_{k-1}^{(i)}, \mathbf{P}_{k-1}^{(i)}\}_{i=1}^{J_{k-1}}$

(1) 求目标自发产生预测概率假设密度 $\gamma_k(\mathbf{x})$ 的参数

$i=0$

for $j=1, \dots, J_{\gamma,k}$

$i:=i+1, \{w_{k|k-1}^{(i)}, \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}\} = \{w_{\gamma,k}^{(j)}, \mathbf{m}_{\gamma,k}^{(j)}, \mathbf{P}_{\gamma,k}^{(j)}\}$

end

(2) 求分离得到目标预测概率假设密度 $\nu_{\beta,k|k-1}(\mathbf{x})$ 的参数

for $j=1, \dots, J_{\beta,k}$

for $l=1, \dots, J_{k-1}$

$i:=i+1,$

$\{w_{k|k-1}^{(i)}, \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}\} = \{w_{\beta,k|k-1}^{(j,l)}, \mathbf{m}_{\beta,k|k-1}^{(j,l)},$

$\mathbf{P}_{\beta,k|k-1}^{(j,l)}\}$

end

end

(3) 求目标继续存在预测概率假设密度 $\nu_{s,k|k-1}(\mathbf{x})$ 的参数(见式(19)-(22))

for $j=1, \dots, J_{k-1}$

$i:=i+1$, 求 $\mathbf{X}_{p,k|k-1}^{*(j)}, \{w_{k|k-1}^{(i)}, \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}\} = \{w_{S,k|k-1}^{(j)}, \mathbf{m}_{S,k|k-1}^{(j)}, \mathbf{P}_{S,k|k-1}^{(j)}\}$

end

$J_{k|k-1} = i$

(4) 求概率假设密度 $\nu_k(\mathbf{x})$ 的参数(见式(24)-(35))

for $j=1, \dots, J_{k|k-1}$

求 $\mathbf{Z}_{p,k|k-1}^{(j)}, \mathbf{z}_{k|k-1}^{(j)}, \mathbf{P}_{zz,k|k-1}^{(j)}, \mathbf{P}_{xz,k|k-1}^{(j)}, \mathbf{q}_k^{(j)}(\mathbf{z}), \mathbf{W}_k^{(j)}, \{w_k^{(j)}, \mathbf{m}_k^{(j)}, \mathbf{P}_k^{(j)}\} = \{(1-p_{D,k})w_{k|k-1}^{(j)}, \mathbf{P}_{k|k-1}^{(j)}\}$

end

$l=0$

for $\mathbf{z} \in Z_k$

$l:=l+1$

for $j=1, \dots, J_{k|k-1}$

$\{w_k^{(j)}, \mathbf{m}_k^{(j)}, \mathbf{P}_k^{(j)}\} = \{w_{k|k-1}^{(j)}, \mathbf{m}_{k|k-1}^{(j)}, \mathbf{P}_{k|k-1}^{(j)}\}$

$q_k^{(j)}(\mathbf{z}), \mathbf{m}_{k|k}^{(j)}(\mathbf{z}), \mathbf{P}_{k|k}^{(j)}$

end

$$w_k^{(U_{k|k-1}+j)} = \frac{w_k^{(U_{k|k-1}+j)}}{\kappa_k(\mathbf{z}) + \sum_{i=1}^{J_{k|k-1}} w_k^{(U_{k|k-1}+i)}},$$

$$j = 1, \dots, J_{k|k-1}$$

end

$$J_k = IJ_{k|k-1} + J_{k|k-1}$$

$$\text{输出: } \{ \mathbf{w}_k^{(i)}, \mathbf{m}_k^{(i)}, \mathbf{P}_k^{(i)} \}_{i=1}^{J_k}$$

在 CK-PHD 滤波的实现过程中,由递推得到的概率假设密度中高斯成分数量增加很快,导致计算复杂度迅速增加,因此需要对高斯成分进行修剪。修剪算法的主要思想是先剔除权值较小的高斯成分,然后在余下的高斯成分中对距离较近的高斯成分进行合并。

用修剪后的高斯成分可以获得多目标的状态,这也是 CK-PHD 滤波的最后一步。因为高斯成分的均值是概率假设密度 ν_k 的局部极大值,所以从均值中可以获取目标状态。为了去除 ν_k 中峰值过低极大值点,只有权值超过 0.5 的成分才用于状态获取。

从上述的 CK-PHD 滤波器中可以看出,在概率假设密度的递推过程中无需求解雅克比矩阵,且在保持计算量较小的同时,提高了滤波器的精度和系统的稳定性。

4 基于 CK-PHD 的多说话人跟踪算法

基于贝叶斯框架的多说话人跟踪系统利用状态模型对说话人的运动进行描述。考虑说话人运动的机动性,本文采用广泛使用的朗之万(Langevin)模型建立说话人状态方程^[6]。

设说话人在 k 时刻的状态 $\mathbf{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k]^T$, 其中 x_k 和 y_k 分别表示 x 和 y 方向的位置, \dot{x}_k 和 \dot{y}_k 分别表示 x 和 y 方向的速度。假设声源在 x 和 y 方向上的运动是独立同分布的^[6],则在 x 轴上,状态 \mathbf{x}_k 满足

$$\dot{x}_k = a_x \dot{x}_{k-1} + b_x F_{x_k}, \quad x_k = x_{k-1} + T \dot{x}_k \quad (36)$$

其中 $F_{x_k} \sim \mathcal{N}(0, 1)$, T 为采样周期, $a_x = \exp(-\beta_x T)$, β_x 为速率常数, $b_x = \bar{v}_x \sqrt{1 - a_x^2}$, \bar{v}_x 为稳态条件下的均方根速率,则说话人跟踪系统模型的状态方程为

$$\mathbf{x}_k = \mathbf{A} \mathbf{x}_{k-1} + \mathbf{v}_k \quad (37)$$

其中 $\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$ 为状态转移矩阵, $\mathbf{B} = \begin{bmatrix} 1 & T a_x \\ 0 & a_x \end{bmatrix}$, 过程

噪声 \mathbf{v}_k 满足 $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q})$, 且协方差阵 $\mathbf{Q} = \text{diag}[b_x^2 T^2, b_x^2, b_x^2 T^2, b_x^2]$ 。

由于语音信号到达不同位置麦克风的时间存在差异,利用应用场景内摆放的麦克风阵列,可得到一组麦克风对的时间延迟。设应用场景中有 M 对麦克风,其位置为 $\mathbf{u}_i^{(m)} = (x_i^{(m)}, y_i^{(m)})$, $\mathbf{u}_j^{(m)} = (x_j^{(m)}, y_j^{(m)})$, 且 $m = 1, \dots, M, k$ 时刻说话人的位置为 $\alpha_k = (x_k, y_k)$, 声音在空气中的传播速度为 c , 则 k 时刻第 m 对麦克风的时间延迟与说话人位置满足

$$\tau_k^{(m)} = \frac{1}{c} \left(\|\alpha_k - \mathbf{u}_i^{(m)}\| - \|\alpha_k - \mathbf{u}_j^{(m)}\| \right), m = 1, \dots, M \quad (38)$$

其中 $\|\cdot\|$ 为向量范数。

说话人跟踪系统的观测方程为

$$z_k^{(m)} = \frac{1}{c} \left(\sqrt{(x_k - x_i^{(m)})^2 + (y_k - y_i^{(m)})^2} - \sqrt{(x_k - x_j^{(m)})^2 + (y_k - y_j^{(m)})^2} \right) + w_k^{(m)}, m = 1, \dots, M \quad (39)$$

其中 $w_k^{(m)}$ 为观测噪声,且满足 $w_k^{(m)} \sim \mathcal{N}(0, R)$ 。

广义互相关(Generalized Cross-Correlation, GCC)法是广泛使用的时延估计方法^[15],该方法通过计算麦克风对接收信号的互相关函数的极值点来估计时间延迟。设声源信号为 $s(k)$, 第 m 对麦克风 i 和 j 处的背景噪声分别为 $n_i^{(m)}(k)$ 和 $n_j^{(m)}(k)$, $m = 1, \dots, M$, 则麦克风接收的信号为

$$y_i^{(m)}(k) = s(k) + n_i^{(m)}(k) \quad (40)$$

$$y_j^{(m)}(k) = s(k - \tau_k^{(m)}) + n_j^{(m)}(k) \quad (41)$$

设信号 $y_i^{(m)}(k)$ 和 $y_j^{(m)}(k)$ 的互功率谱为 $S_{y_i y_j}^{(m)}(\omega)$, 则广义互相关函数为

$$R^{(m)}(\tau_k^{(m)}) = \int_{-\infty}^{\infty} \Phi(\omega) S_{y_i y_j}^{(m)}(\omega) e^{j\omega\tau} d\omega \quad (42)$$

其中 $\Phi(\omega)$ 为加权函数。求解目标函数 $R^{(m)}(\tau_k^{(m)})$ 的极值估计时间延迟

$$\hat{\tau}_k^{(m)} = \arg \max_{\tau_k^{(m)} \in [-\tau_{\max}^{(m)}, \tau_{\max}^{(m)}]} R^{(m)}(\tau_k^{(m)}) \quad (43)$$

其中 $\tau_{\max}^{(m)}$ 为时间延迟的最大值,且 $\tau_{\max}^{(m)} = \frac{1}{c} \|\mathbf{u}_i^{(m)} - \mathbf{u}_j^{(m)}\|$ 。

由此,将广义互相关法得到的时间延迟估计作为观测值,在说话人数目和说话时间未知的情况下,针对多说话人跟踪系统模型式(37)和式(39),采用 CK-

PHD 滤波方法,可得到多说话人的状态估计,即利用 $k-1$ 时刻说话人的位置和速度信息以及 k 时刻麦克风阵列的时间延迟估计值,估计 k 时刻说话人的位置和说话人数目,以实现多说话人跟踪。该算法没有对不同说话人信息进行标记,也没有采用数据关联算法,因此无法分辨说话人,但该算法减少了观测值中的虚警信息,估计了多个说话人的位置和数目,是后续说话人分辨算法的基础。文献[7]给出了一种多说话人分辨算法,这里不再赘述。

为了分析基于 CK-PHD 的多说话人跟踪滤波器的算法复杂度,本文以式(37)和(39)系统模型为例,将文献[11]提出的 EK-PHD 滤波器和 UK-PHD 滤波器应用在多说话人跟踪系统中,将本文方法与这两种方法的计算复杂度进行了比较。设概率假设密度 ν_{k-1} 和自产生新目标的概率假设密度 γ_k 的高斯成分数目分别为 J_{k-1} 和 $J_{\gamma,k}$,分离得到新目标的数目为 $J_{\beta,k}$,说话人状态向量的维数为 n_x ,麦克风的对数为 M ,观测集元素的个数为 $|Z_k|$,为了方便比较,以一次 PHD 递推为例,表 1 给出了三种滤波器计算复杂度的比较结果,其中 $J=J_{\gamma,k}+J_{k-1}J_{\beta,k}$ 。从表 1 可以看出,CK-PHD 的计算复杂度低于 UK-PHD,与 EK-PHD 相比,两者的加减乘运算复杂度相近,但 CK-PHD 的开方运算复杂度更低。

表 1 EK-PHD 滤波器、UK-PHD 滤波器和 CK-PHD 滤波器的计算复杂度比较

Tab. 1 Computational complexity of EK-PHD, UK-PHD and CK-PHD filters

	加减乘运算	开方运算
EK-PHD	$2n_x^3 J_{k-1} + J(M^3 + Z_k n_x^2 M^2)$	$2J(3M + Z_k)$
UK-PHD	$8n_x^3 J_{k-1} + J(M^3 + 2 Z_k n_x M^2)$	$2J(M + Z_k)$
CK-PHD	$5n_x^3 J_{k-1} + J(M^3 + 2 Z_k n_x M^2)$	$2J(M + Z_k)$

5 计算机仿真与结果讨论

仿真实验场景如图 1 所示,房间大小为 $3\text{m} \times 3\text{m} \times 3\text{m}$,说话人运动轨迹采用经典的半圆形设置^[6],麦克风位置和说话人运动轨迹如图 1 所示。在实验中共放置了 4 对麦克风,每对麦克风的间距为 0.6m ,具体位置分别为 $(1.4, 0)$ 、 $(2, 0)$ 、 $(0, 1.4)$ 、 $(0, 2)$ 、 $(1.4, 3)$ 、 $(2, 3)$ 、 $(3, 1.4)$ 、 $(3, 2)$ 。为了方便分析结果,给出说话人数目和说话时间的真实值。房间内有三个说话人,其中说话人 1 和 3 在不同位置同时说话,对应帧数

为 $k=1 \sim 100$,说话人 2 出现语音信息的帧数为 $k=10 \sim 90$,说话人数目变化情况如图 2 所示。实验采用的语音数据为单通道 16bit 的女性英语发音,其采样率为 32kHz ,利用 IMAGE 模型^[16]生成房间冲激响应来得到麦克风接收信号,每帧数据长度为 32ms ,信噪比 SNR 为 25dB ,混响时间 T_{60} 为 0.2s 。实验观测数据为应用 GCC 方法求得的时间延迟估计,且说话人的数目和说话时间未知。

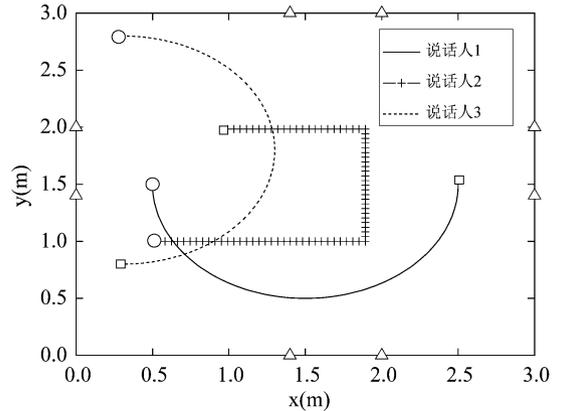


图 1 说话人运动轨迹图(○-说话人起始位置, □-说话人终止位置, △-麦克风位置)

Fig. 1 Speakers trajectories (○-initial point of speakers, □-final point of speakers, △-locations of microphones)

在系统模型(37)和(39)中,状态方程中参数设置为 $\beta_x = 10\text{s}^{-1}$, $T = 32\text{ms}$, $\bar{\nu}_x = 1\text{ms}^{-1}$,观测噪声的协方差阵 $R = \sigma^2$, $\sigma = 125\mu\text{s}$ 。

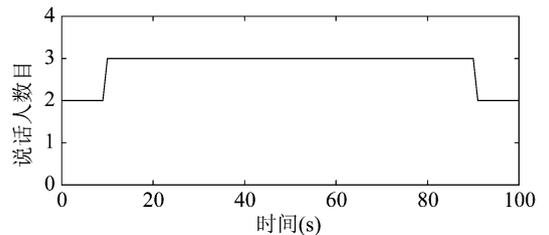


图 2 说话人数目随时间变化图

Fig. 2 The number of speakers with varying time

跟踪场景中每个说话人的存活概率 $p_{s,k} = 0.99$, CK-PHD 滤波器的初始高斯成分 $\{w_k^{(i)}, m_k^{(i)}, P_k^{(i)}\}_{i=1}^{J_0}$ 的参数为 $J_0 = 2$, $w_0^{(1)} = w_0^{(2)} = 0.1$, $m_0^{(1)} = [0.5, 0, 1.5, 0.98]^T$, $m_0^{(2)} = [0.3, 0.98, 2.8, 0]^T$, $P_0^{(1)} = P_0^{(2)} = \text{diag}([10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}]^T)$ 。目标自发产生的概率假设密度 $\gamma_k(x) = w_r^{(i)} \mathcal{N}(x; m_r^{(i)}, P_r^{(i)})$ 的参数为 $w_\gamma^{(1)} = 0.1$, $m_\gamma^{(1)} = [0.5, 0, 1, 0.98]^T$, $P_\gamma^{(1)} = P_0^{(1)}$ 。

跟踪场景中每个说话人的检测概率 $p_{D,k} = 0.98$, 每对麦克风对应的观测值中, 虚警随机有限集服从泊松分布, 其概率假设密度 $\kappa_k(\mathbf{z}) \triangleq \lambda_c Vu(\mathbf{z})$, 其中 $\lambda_c = 7$ 为每对麦克风虚警时间延迟估计的平均数目, $u(\cdot)$ 为虚警点服从均匀分布的概率密度函数, 由于时间延迟估计取值区间为 $[-\tau_{\max}^{(m)}, \tau_{\max}^{(m)}]$, 则 $V = 2\tau_{\max}^{(m)}$ 为取值区间长度。仿真实验中, 高斯成分的修剪阈值为 10^{-5} , 合并阈值为 4, 状态提取阈值为 0.5。

为了有效地评价本文提出的说话人跟踪滤波方法的性能, 实验中分别采用最优子模式分配 (Optimal Subpattern Assignment, OSPA)^[17] 和平均相对误差^[11] 两个准则对跟踪性能进行评价。设说话人状态的真值为 $X = \{\mathbf{x}_1, \dots, \mathbf{x}_{|X|}\}$, 说话人状态的估计值为 $\hat{X} = \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{|\hat{X}|}\}, \{1, 2, \dots, |\hat{X}|\}$ 中集合长度为 $|X|$ 的所有排列的集合为 $\Pi_{|\hat{X}|}$, 参数为 n 的截断距离为 $d^{(n)}(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \min(n, d(\mathbf{x}, \hat{\mathbf{x}}))$ ^[17], 则当 $|X| \leq |\hat{X}|$ 时, p 阶 OSPA 定义为

$$\bar{d}_p^{(n)}(X, \hat{X}) = \left(\frac{1}{|\hat{X}|} \left(\min_{\pi \in \Pi_{|\hat{X}|}} \sum_{i=1}^{|\hat{X}|} d^{(n)}(\mathbf{x}_i, \hat{\mathbf{x}}_{\pi(i)})^p + (|\hat{X}| - |X|)n^p \right) \right)^{1/p} \quad (44)$$

当 $|X| > |\hat{X}|$ 时, $\bar{d}_p^{(n)}(X, \hat{X}) = \bar{d}_p^{(n)}(\hat{X}, X)$ 。

为了评价多说话人跟踪滤波器对说话人数目估计准确性的影响, 平均相对误差 (ARE) 定义为

$$\text{ARE} = E \left\{ \frac{||\hat{X}| - |X||}{|X|} \right\} \quad (45)$$

为了考察说话人跟踪滤波器的性能, 在仿真实验中将 EK-PHD 滤波器和 UK-PHD 滤波器应用在多说话人跟踪系统中, 并将本文方法与这两种方法进行了比较, 其中本文方法的跟踪结果如图 3 所示, 可以看到, 本文方法基本反映了说话人的运动轨迹。对三种滤波器分别进行 1000 次蒙特卡罗模拟仿真实验, OSPA 结果和平均相对误差如图 4 所示, 其中二阶 OSPA 评价准则参数 $n = 1.5m$ 。从图 4 可以看出, 当 $k < 5$ 时, 三种滤波器的 OSPA 结果基本相同; $k > 5$ 后, CK-PHD 滤波结果优于 EK-PHD 和 UK-PHD。当 $k > 10$ 时, 由于说话人数目变化, 三种滤波器的平均相

对误差增大, 但 CK-PHD 滤波器对说话人数目的估计更准确。

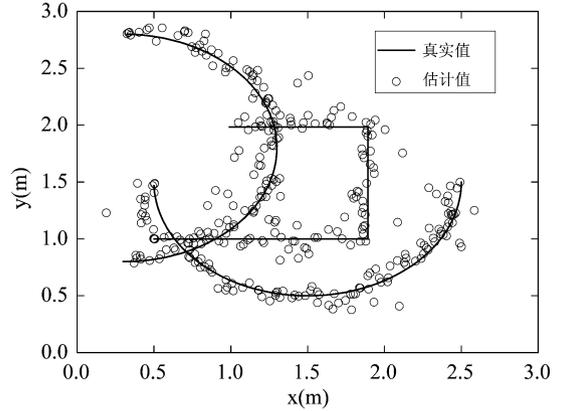
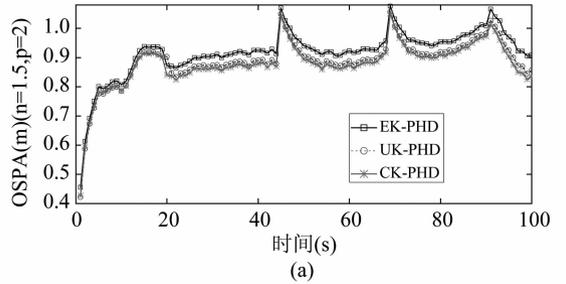
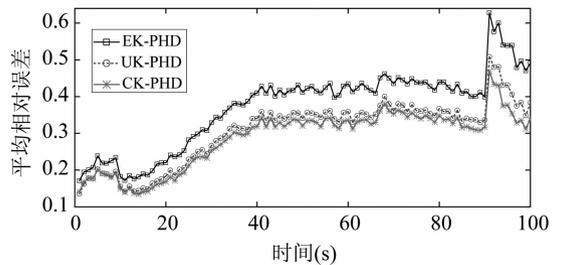


图 3 CK-PHD 滤波器的多说话人位置估计结果图

Fig. 3 The estimates of speakers' position of CK-PHD filter



(a)



(b)

图 4 EK-PHD 滤波器、UK-PHD 滤波器和 CK-PHD 滤波器的 OSPA 和平均相对误差比较

Fig. 4 OSPA and average relative errors comparison of EK-PHD, UK-PHD and CK-PHD filters

为了分析说话人跟踪滤波器的鲁棒性, 在说话人检测概率 $p_{D,k}$ 、每对麦克风虚警的平均数目 λ_c 、采样周期 T 、麦克风接收信号信噪比 SNR 和混响时间 T_{60} 变化情况下, 本文分别对三种滤波器的性能进行了比较。在 $p_{D,k}$ 变化且 $\lambda_c = 7, T = 32ms, SNR = 25dB, T_{60} = 0.2s$ 时, EK-PHD、UK-PHD 和 CK-PHD 的 OSPA 结果和平均相对误差对比如图 5 所示。从图 5 可以看出, 当 $p_{D,k} = 1$ 时, 场景内的说话人全部检测成功, 不存在漏检, 三

种滤波器的滤波结果较 $p_{D,k} < 1$ 时更准确, OSPA 结果和平均相对误差较小, 随着 $p_{D,k}$ 减小, 漏检增多, 滤波器对说话人状态和数目的估计误差增大, 但 CK-PHD 滤波器的滤波结果更准确, 受 $p_{D,k}$ 变化的影响更小。在 λ_c 变化且 $p_{D,k} = 0.98, T = 32\text{ms}, \text{SNR} = 25\text{dB}, T_{60} = 0.2\text{s}$ 时, 三种滤波器的性能如图 6 所示。从图 6 可以看出, 当 $\lambda_c = 0$ 时, 检测区域内不存在虚警点, 三个滤波器对目标数目的估计较 $\lambda_c > 0$ 时更准确, 随着 λ_c 增大, 虚警点增多, 滤波器的 OSPA 结果和平均相对误差变大, 但与 EK-PHD 和 UK-PHD 相比, CK-PHD 的性能曲线变化缓慢, OSPA 结果较小, 对说话人数目的估计更准确。在采样周期 T 变化且 $p_{D,k} = 0.98, \lambda_c = 7, \text{SNR} = 25\text{dB}, T_{60} = 0.2\text{s}$ 时, 三种滤波器的性能对比如图 7 所示。从图 7 可以看出, T 取值越大, 说话人在一个采样间隔内状态的变化越大, 对滤波结果影响越大, 所以三个滤波器的平均相对误差随 T 增加而增大, 但 CK-PHD 滤波器相对另两种滤波器具有更小的 OSPA 值和对说话人数目更小的估计误差。在信噪比 SNR 变化且 $p_{D,k} = 0.98, \lambda_c = 7, T = 32\text{ms}, T_{60} = 0.2\text{s}$ 时, 三种滤波器的性能比较如表 2 和表 3 所示。从表 2 和表 3 可以看出, SNR 越大, 滤波效果越好, 当 SNR 为 10dB 和 15dB 时, 信噪比较小, 三种滤波器的 OSPA 结果和平均相对误差基本相同, 但 SNR 为 20dB、25dB 和 30dB 时, 相比其他两种滤波器, CK-PHD 的 OSPA 结果更小, 对说话人数目的估计误差更小。在混响时间 T_{60} 变化且 $p_{D,k} = 0.98, \lambda_c = 7, T = 32\text{ms}, \text{SNR} = 25\text{dB}$ 时, 三种滤波器的性能对比如表 4 和表 5 所示。从表 4 和表 5 可以看出, T_{60} 增大时, 说话人跟踪精度明显降低, 这是因为混响时间增大导致麦克风阵列的时间延迟估计的准确性降低, 使得滤波器对说话人位置和数目的估计误差增大, 但与另两种滤波器相比, CK-PHD 滤波器的 OSPA 和平均相对误差更小, 跟踪结果更准确。

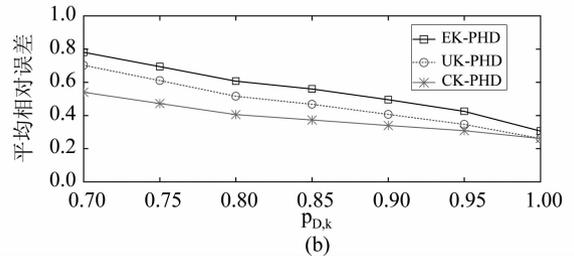
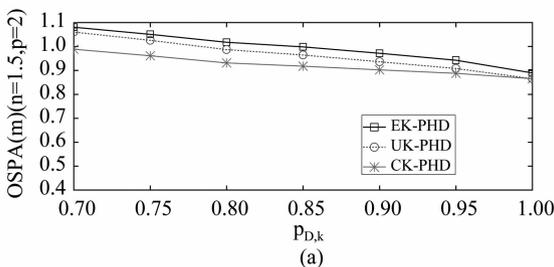


图5 EK-PHD 滤波器、UK-PHD 滤波器和 CK-PHD 滤波器的 OSPA 和平均相对误差随 $p_{D,k}$ 变化图
Fig. 5 OSPA and average relative errors comparison of EK-PHD, UK-PHD and CK-PHD filters with varying $p_{D,k}$

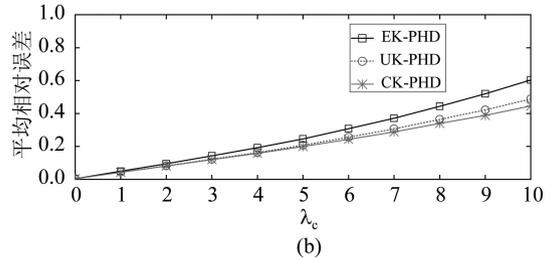
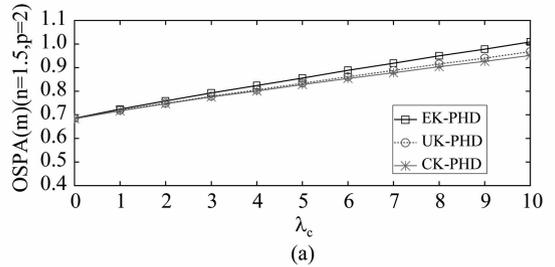


图6 EK-PHD 滤波器、UK-PHD 滤波器和 CK-PHD 滤波器的 OSPA 和平均相对误差随 λ_c 变化图
Fig. 6 OSPA and average relative errors comparison of EK-PHD, UK-PHD and CK-PHD filters with varying λ_c

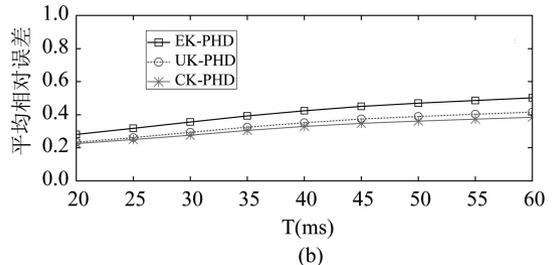
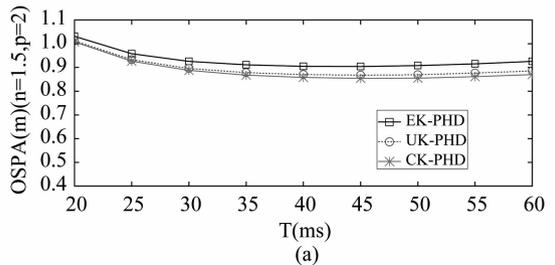


图7 EK-PHD 滤波器、UK-PHD 滤波器和 CK-PHD 滤波器的 OSPA 和平均相对误差随 T 变化图
Fig. 7 OSPA and average relative errors comparison of EK-PHD, UK-PHD and CK-PHD filters with varying T

表 2 EK-PHD 滤波器、UK-PHD 滤波器
和 CK-PHD 滤波器的 OSPA 随 SNR 变化表

Tab. 2 OSPA comparison of EK-PHD,
UK-PHD and CK-PHD filters with varying SNR

	10dB	15dB	20dB	25dB	30dB
EK-PHD	1.4349	1.4346	1.3476	0.8883	0.5870
UK-PHD	1.4350	1.4344	1.3463	0.8613	0.5810
CK-PHD	1.4351	1.4345	1.3450	0.8540	0.5799

表 3 EK-PHD 滤波器、UK-PHD 滤波器
和 CK-PHD 滤波器的平均相对误差随 SNR 变化表

Tab. 3 Average relative errors comparison of
EK-PHD, UK-PHD and CK-PHD filters with varying SNR

	10dB	15dB	20dB	25dB	30dB
EK-PHD	2.5430	2.5505	2.2175	0.3073	0.0329
UK-PHD	2.5437	2.5507	2.1980	0.2554	0.0277
CK-PHD	2.5433	2.5507	2.1781	0.2435	0.0275

表 4 EK-PHD 滤波器、UK-PHD 滤波器
和 CK-PHD 滤波器的 OSPA 随 T_{60} 变化表

Tab. 4 OSPA comparison of EK-PHD, UK-PHD
and CK-PHD filters with varying T_{60}

	0.2s	0.25s	0.3s	0.35s	0.4s
EK-PHD	0.8883	1.3066	1.3318	1.3581	1.3623
UK-PHD	0.8613	1.2993	1.3285	1.3578	1.3620
CK-PHD	0.8540	1.2967	1.3266	1.3569	1.3613

表 5 EK-PHD 滤波器、UK-PHD 滤波器
和 CK-PHD 滤波器的平均相对误差随 T_{60} 变化表

Tab. 5 Average relative errors comparison of EK-PHD,
UK-PHD and CK-PHD filters with varying T_{60}

	0.2s	0.25s	0.3s	0.35s	0.4s
EK-PHD	0.3073	1.9300	2.1128	2.2955	2.3315
UK-PHD	0.2554	1.8473	2.0677	2.2896	2.3247
CK-PHD	0.2435	1.8215	2.0427	2.2753	2.3132

图 3 ~ 图 7 以及表 2 ~ 表 5 的仿真结果表明,对多说话人跟踪问题,CK-PHD 滤波效果优于 EK-PHD 和 UK-PHD,且当检测概率变化、场景中虚警点数目增多、采样间隔变化、信噪比不同以及混响时间变大时,CK-PHD 滤波器仍可以保持良好的跟踪性能。

6 结束语

针对多说话人跟踪问题,本文提出了基于数值积分卡尔曼-概率假设密度滤波的多说话人跟踪方法,该

方法适用于说话人出现和消失时间不确定且说话人数目未知情况下的多说话人跟踪系统。该方法无需求解非线性系统函数的雅克比矩阵,且计算量较小。仿真结果表明,本文方法降低了系统模型非线性对滤波结果的影响,提高了说话人数目和说话人状态的估计精度,具有良好的鲁棒性,有效地解决了非线性高斯系统的多说话人跟踪问题。

参考文献

- [1] Kapralos B, Jenkin M R M, Milios E. Audiovisual localization of multiple speakers in a video teleconferencing setting [J]. International Journal of Imaging Systems and Technology, 2003, 13(1): 95-105.
- [2] Gatica-Perez D, Lathoud G, Odobez J-M, etc. Audiovisual probabilistic tracking of multiple speakers in meetings [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(2): 601-616.
- [3] Wang Q H, Ivanov T, Aarabi P. Acoustic robot navigation using distributed microphone arrays [J]. Information Fusion, 2004, 5: 131-140.
- [4] Brandstein M. A framework for speech source localization using sensor arrays [D]. Providence: Massachusetts Institute of Technology, Brown University, 1995.
- [5] Potamitis I, Chen H, Tremoulis G. Tracking of multiple moving speakers with microphone arrays [J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(5): 520-529.
- [6] Vermaak J, Blake A. Nonlinear filtering for speaker tracking in noisy and reverberant environments [C]. IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, Utah, USA: IEEE, 2001. 3021-3024.
- [7] Ma W-K, Vo B-N, Singh S S, etc. Tracking an unknown time-varying number of speakers using TDOA measurements: a random finite set approach [J]. IEEE Transactions on Signal Processing, 2006, 54(9): 3291-3304.
- [8] Arasaratnam I, Haykin S. Cubature Kalman filter [J]. IEEE Transactions on Automatic Control, 2009, 54(6): 1254-1269.
- [9] Yin J, Zhang J, Zhuang Z. Gaussian sum PHD filtering algorithm for nonlinear non-Gaussian models [J]. Chinese Journal of Aeronautics, 2008, 21: 341-351.
- [10] Vo B-N, Singh S. Sequential Monte Carlo methods for mul-

titarget filtering with random finite sets [J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(4):1224-1245.

- [11] Vo B-N, Ma W. The Gaussian mixture probability hypothesis density filter [J]. IEEE Transactions on Signal Processing, 2006, 54(11):4091-4104.
- [12] Macagnano D, de Aberu G T F. Multitarget tracking with the cubature Kalman probability hypothesis density filter [C]. Conference Record of the Forth Fourth Asilomar Conference on Signals, Systems and Computers. Pacific Grove, CA, USA: IEEE, 2010. 1455-1459.
- [13] Julier S J, Uhlmann J K. Unscented filtering and nonlinear estimation [J]. Proceedings of IEEE, 2004, 92(3):401-422.
- [14] Li X R, Jilkov V P. Survey of maneuvering target tracking. Part I: dynamic models [J]. IEEE Transactions on Aerospace and Electronic Systems, 2003, 39(4):1333-1364.
- [15] Gustafsson T, Rao B, Trivedi M. Source localization in reverberant environments: Modeling and statistical analysis [J]. IEEE Transactions on Speech Audio Process, 2003, 11(6):791-803.
- [16] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics [J]. Journal of the Acoustical Society of America, 1979, 65(4):943-950.
- [17] Schuhmacher D, Vo B-T, Vo B-N. A consistent metric for performance evaluation of multi-object filters [J].

IEEE Transactions on Signal Processing, 2008, 56(8):3447-3457.

作者简介



陈里铭(1981-),女,黑龙江齐齐哈尔人,大连理工大学博士生,主要研究方向为目标跟踪和视频处理。

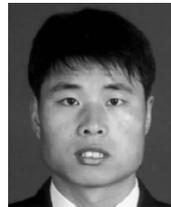
E-mail: chenliming1127@163.com



陈喆(1975-),男,黑龙江泰来人,博士,大连理工大学副教授,主要研究方向为语音处理、图像处理和宽带无线通信技术。E-mail: zhechen@dlut.edu.cn



殷福亮(1962-),男,辽宁抚顺人,大连理工大学教授、博士生导师,主要研究方向为语音处理、图像处理和宽带无线通信技术。E-mail: flyin@dlut.edu.cn



侯代文(1972-),男,山东嘉祥人,博士,主要研究方向为跟踪与定位技术、语音信号处理、阵列信号处理等。

E-mail: hodevin@yahoo.com.cn