

基于拟蒙特卡洛滤波的说话人跟踪方法

侯代文^{1,2} 殷福亮¹ 陈喆¹

摘要 提出了一种基于拟蒙特卡洛滤波的说话人跟踪方法. 该方法利用拟蒙特卡洛积分技术优化采样粒子在状态空间的分布特性, 降低了滤波过程中的积分误差, 提高了状态估计精度; 同时, 用均值漂移技术使采样粒子向高似然区域移动, 减少了所需采样粒子的数目, 降低了计算需求. 最后, 将所提方法应用于说话人跟踪系统, 提高了说话人位置的跟踪精度. 仿真实验结果验证了本文方法的有效性.

关键词 说话人跟踪, 拟蒙特卡洛滤波, 粒子滤波, 均值漂移, 状态估计
中图分类号 TP391

Quasi-Monte Carlo Filtering for Speaker Tracking

HOU Dai-Wen^{1, 2} YIN Fu-Liang¹ CHEN Zhe¹

Abstract A mean shift quasi-Monte Carlo (MS-QMC) method is proposed for speaker tracking. To explore the state space more efficiently, deterministic samplers are used instead of random draws according to a quasi-Monte Carlo integration rule in the new method. Furthermore, a mean shift procedure is applied to move particles toward the modes of the posterior, leading to a more effective allocation of particles thereupon fewer particles are needed. Simulation results show that compared with the traditional particle filter, both speaker tracking accuracy and convergent rate of the proposed method are improved.

Key words Speaker tracking, quasi-Monte Carlo (QMC) filtering, particle filter, mean shift (MS), state estimation

说话人跟踪是语音信号处理领域的重要课题之一, 它通过麦克风阵列接收说话人的语音信号, 根据语音到达各麦克风的时间延迟或波达方向信息确定说话人的位置. 说话人跟踪技术可以广泛用于电视电话会议系统、视频监控中的摄像头自动导引、远距离说话人识别、计算机人机接口以及机器人导航等场合^[1].

传统的说话人定位方法通过麦克风阵列接收到的语音信号, 计算当前时刻的时间延迟或波达方向, 进而估计出说话人的位置^[2]. 该方法在自由声场条件下, 能够实现对话话人的准确定位, 但在房间混响和环境噪声较强时往往会产生虚声源, 从而导致对话话人位置的错误估计. 近几年提出的状态空间方法通过建立动态方程, 在估计说话人位置时, 不仅利用当前时刻的观测信息, 而且利用当前时刻之前的全部信息, 因而能够滤除观测中具有明显误差的观测量, 从而避免对话话人位置的错误估计. Dvorkind 等^[3] 利用扩展卡尔曼滤波

收稿日期 2008-05-22 收修改稿日期 2008-12-01

Received May 22, 2008; in revised form December 1, 2008

国家自然科学基金 (60772161, 60372082), 高等学校博士学科点专项科研基金 (200801410015) 资助

Supported by National Natural Science Foundation of China (60772161, 60372082) and Specialized Research Fund for the Doctoral Program of Higher Education of China (200801410015)

1. 大连理工大学电子与信息工程学院 大连 116023 2. 海军试验基地 大连 116041

1. School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116023 2. Naval Test Base, Dalian 116041

DOI: 10.3724/SP.J.1004.2009.01016

器跟踪说话人位置,在一定程度上解决了虚声源问题,但在非线性、非高斯观测噪声条件下,使用卡尔曼滤波方法所得估计结果偏差较大.考虑到粒子滤波^[4-5]方法能够有效地处理非线性、非高斯问题,Vermaak等^[6]和Ward等^[7]分别采用粒子滤波方法对说话人进行跟踪,明显地改善了说话人的跟踪效果.

在本质上,粒子滤波方法是利用未知状态空间上的采样点集合来逼近连续的分布函数,因此,采样点位置的选取将直接影响对分布函数描述的准确程度,进而影响粒子滤波方法的性能.在粒子滤波方法中,随机采样过程使获得的粒子具有随意性,这会造成粒子集合在状态空间积聚在一起或形成空洞,从而不能充分描述分布函数.为此,Guo等提出了拟蒙特卡洛(Quasi-Monte Carlo, QMC)滤波方法^[8],该方法以确定性采样代替随机采样,利用拟蒙特卡洛积分中的低偏差序列^[9-10]代替随机采样点集合,使采样粒子最大程度地互相远离,从而提高了粒子滤波的精度.但是,由于拟蒙特卡洛方法的计算复杂度与采样粒子个数的平方成正比,计算量随系统状态维数增加呈指数形式增长,当状态空间维数增加时,拟蒙特卡洛滤波方法会因为计算量太大而失去实用价值.为了降低计算复杂性,本文提出了基于均值漂移(Mean shift, MS)技术^[11]的拟蒙特卡洛(MS-QMC)滤波方法.该方法在每次滤波迭代之后,重新调整各粒子的位置,使之向似然值较大的方向移动,从而提高了采样粒子对分布函数的表示能力,减少了所需粒子的数目,降低了计算需求.最后,将MS-QMC方法应用于说话人跟踪.仿真实验结果表明,本文方法有效提高了说话人位置的估计精度.

1 拟蒙特卡洛滤波方法

1.1 粒子滤波方法

非线性滤波系统的状态方程和观测方程可以表示为

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

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{n}_k \quad (2)$$

其中 $\mathbf{x}_k \in \mathbf{R}^{n_x}$ 表示系统状态, $\mathbf{z}_k \in \mathbf{R}^{n_z}$ 表示观测量, $f(\cdot)$ 为状态转移函数, $h(\cdot)$ 为观测函数, \mathbf{v}_k 和 \mathbf{n}_k 分别是独立同分布的过程噪声和观测噪声.我们的目标是通过观测序列 $\mathbf{z}_{1:k} = \{\mathbf{z}_i, i = 1, \dots, k\}$,递推估计系统状态的后验概率密度函数 $p(\mathbf{x}_k | \mathbf{z}_{1:k})$,从而得到系统状态 \mathbf{x}_k 的最优估计.

贝叶斯估计方法通过以下迭代过程,递推估计系统状态的后验概率密度函数

$$p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1} \quad (3)$$

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \frac{1}{c_k} p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) \quad (4)$$

$$c_k \triangleq \int p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) d\mathbf{x}_k \quad (5)$$

粒子滤波方法,也称序贯蒙特卡洛方法,它通过蒙特卡洛模拟实现状态的贝叶斯递推估计^[4].其核心思想是:用一组随机采样点及其对应的权值表示所需的后验概率密度函数,并利用它们递推估计系统状态.当采样点个数趋于无穷大时,蒙特卡洛模拟的概率密度函数等价于后验概率密度函数,相应的状态估计值接近于最优的贝叶斯估计.

1.2 基于拟蒙特卡洛的粒子滤波方法

粒子滤波方法利用蒙特卡洛技术实现式(3)~(5)的多元积分.但近期的研究发现,完全地随机采样并不能有效利用模拟数据^[12],特别地,采样点往往会发生聚集现象,从而导致采样空间出现空隙,造成对采样空间的搜索不充分.为了获得分布更加均匀的采样序列,人们提出了多种形式的确定性采样方法,通常将这些采样方法称为拟蒙特卡洛方法^[10](如图1和2所示).

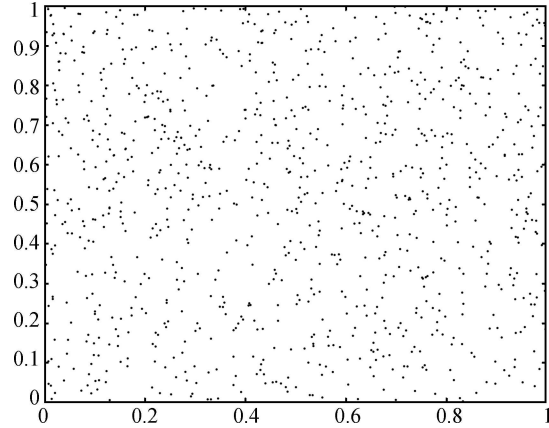


图1 二维空间上1000点的随机采样点集

Fig.1 1000 2-D random sampling set

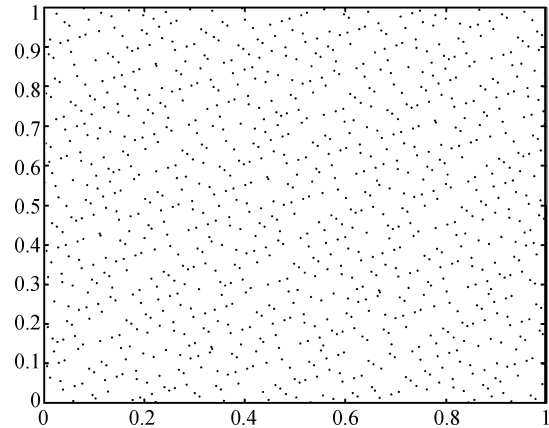


图2 二维空间上1000点的拟蒙特卡洛采样点集

Fig.2 1000 2-D quasi-Monte Carlo sampling set

拟蒙特卡洛采样可以看作是蒙特卡洛采样相对应的确定性采样方法.假定我们要计算积分 $I = \int_{C^d} f(\mathbf{x}) d\mathbf{x}$,经典的蒙特卡洛积分方法通过在 d 维单位超体 $C^d: [0, 1]^d$ 上随机抽取 N 个服从均匀分布的采样点 $\mathbf{x}_i \sim \text{Uniform}(C^d)$ ($i = 1, 2, \dots, n$),得到其近似估计 $\hat{I} = (1/N) \sum_{i=1}^N f(\mathbf{x}_i)$.相应地,拟蒙特卡洛方法也是在 C^d 上抽取 N 个采样点,不同之处在于,这些采样点是通过确定性方法获取的,它们在 d 维空间上尽可能均匀地分布,使彼此之间的距离最大化,以充分描述被积函数 $f(\mathbf{x})$ 在 C^d 上的分布特性.这种通过确定性采样得到且分布较为规整的序列称为低偏差序列.在理论上,对 d 维空间上的函数进行积分,若采样点个数为 N ,则拟蒙特卡洛积分的收敛速度为 $O(N^{-1} \log_d N)$,高于传统蒙特卡洛方法 $O(N^{-\frac{1}{2}})$ 的收敛速度.根据生成方法的不同,低偏差序列可分为Halton序列、Sobol序列和Niederreiter序列

等^[10]. 这里我们以较容易理解的 Halton 序列为例, 说明低偏差序列的生成方法.

设 p 为一质数, 则任意的整数 $K \geq 0$ 都可以表示为以 p 为基的指数描述形式^[10]

$$k = d_0 + d_1p + \cdots + d_jp^j = \sum_{i=0}^j d_i p^i \quad (6)$$

其中, $d_i \in \{0, 1, \cdots, p-1\}$. 这样, 可以得到一维 Halton 序列的第 k 个元素为

$$\phi_p(k) = \frac{d_0}{p} + \frac{d_1}{p^2} + \cdots + \frac{d_j}{p^{j+1}} = \sum_{i=0}^j \frac{d_i}{p^{i+1}} \quad (7)$$

如果以不同的质数 p_1, p_2, \cdots, p_d 为基, 就能得到包含 N 个元素的 d 维 Halton 序列 $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N\}$, 其中 $\mathbf{u}_k = [\phi_{p_1}(k-1), \phi_{p_2}(k-1), \cdots, \phi_{p_d}(k-1)]^T, k = 1, \cdots, N$.

相对于随机采样点集, 低偏差序列在概率空间上分布较规整, 这使得拟蒙特卡洛方法比蒙特卡洛方法收敛速度更快且精度更高. 但在确定性采样方法中, 各采样点之间不再相互独立, 这违背了状态估计的无偏特性^[13], 而且积分精度难以估计. 为此, Owen^[14] 提出了将拟蒙特卡洛序列随机化的方法. 经过随机化处理后的随机拟蒙特卡洛 (Randomization QMC, RQMC) 序列不仅在 C^d 上满足均匀分布, 而且保持了序列的低偏差特性.

由于拟蒙特卡洛序列在积分过程中相对于随机采样点集具有优越性, Guo 等^[8] 提出在粒子滤波方法中, 用随机拟蒙特卡洛序列代替蒙特卡洛采样点集, 以实现系统状态估计的拟蒙特卡洛滤波方法.

在任一时刻 k , 基于拟蒙特卡洛的粒子滤波方法主要包括以下步骤:

- 1) 计算重要性采样函数 $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k})$;
- 2) 从重要性采样函数 $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k})$ 中抽取 N 个拟蒙特卡洛点序列 $\{\mathbf{x}_k^i\}_{i=1}^N$; 并做随机化处理;
- 3) 计算每一个采样粒子的归一化权值 $w_k^i \propto p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k})/\pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k})$.

拟蒙特卡洛方法与粒子滤波方法的唯一区别在于用随机拟蒙特卡洛序列代替了随机采样点集. 从已知分布中抽取低偏差序列的方法参见文献 [15-16], 在 Matlab 中, 我们可以用 `norminv(.)` 函数产生高斯分布下的拟蒙特卡洛序列.

2 基于均值漂移的拟蒙特卡洛滤波方法

在粒子滤波方法中, 由于系统状态的分布范围不易准确界定, 采样粒子处于权值较小区域的概率远大于处于权值较大区域的概率, 因此, 为了满足滤波精度的要求, 人们只能采取增加粒子数的方法, 从而导致计算量加大. 拟蒙特卡洛方法在一定程度上提高了粒子的使用效率, 但相对于基本粒子滤波方法 $O(N)$ 的计算复杂度, 其计算复杂度为 $O(N^2)$. 当采样粒子数目增加时, 由于计算复杂度与粒子个数的平方成正比, 拟蒙特卡洛滤波器会因计算量急剧增大而失去实用价值. 为了降低计算复杂度, 本文提出了基于均值漂移技术的拟蒙特卡洛 (MS-QMC) 方法, 其原理如图 3 和 4 (见下页) 所示. 该方法在每次迭代之后, 没有直接估计系统状态, 而是通过调整采样粒子的位置, 来提高采样粒子对概率密度函数的表示能力, 从而减少了所需采样粒子的数目, 降低了计算量.

均值漂移技术是一种非参数核概率密度函数估计方法^[11], 它首先将离散的概率密度函数做连续化处理, 然后根

据各采样粒子所在位置的梯度信息调整各点位置, 使之向局部极大值方向漂移.

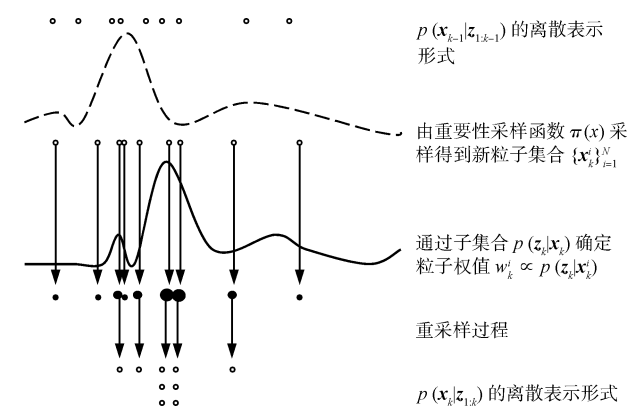


图 3 粒子滤波方法

Fig. 3 Particle filtering method

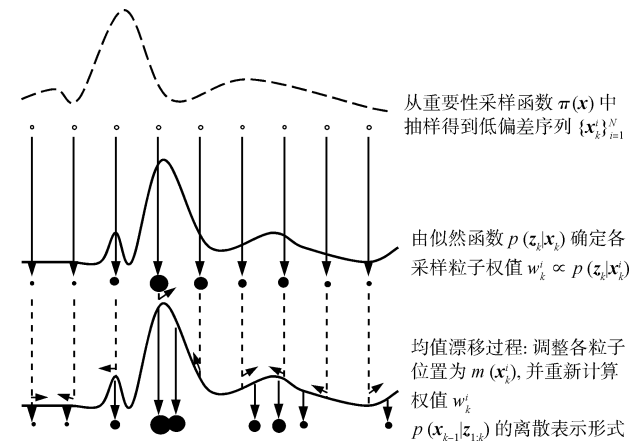


图 4 MS-QMC 方法

Fig. 4 Mean shift quasi-Monte Carlo method

设离散点集为 $\{\mathbf{x}_k^i\}_{i=1}^N$, 其对应权值为 $\{w_k^i\}_{i=1}^N$, 则后验概率密度函数的连续形式可用核函数 K 表示为

$$\hat{p}(\mathbf{x}_k|\mathbf{z}_{1:k}) = \sum_{i=1}^N w_k^i K_\lambda(\mathbf{x}_k - \mathbf{x}_k^i) \quad (8)$$

在实际应用中, 核函数 K 可选取高斯核或者 Epanechnikov 核^[17]. 函数 $K_\lambda(\cdot)$ 是由窗宽 λ 标度的核函数, 即 $K_\lambda(\mathbf{x}_k) = \frac{1}{\lambda^{n_x}} K(\mathbf{x}_k/\lambda)$. 如果定义积分均方误差 (Mean integrated square error, MISE)^[17] 为

$$\text{MISE}(\hat{p}) = \mathbb{E} \left[\int [\hat{p}(\mathbf{x}_k|\mathbf{z}_{1:k}) - p(\mathbf{x}_k|\mathbf{z}_{1:k})]^2 d\mathbf{x} \right] \quad (9)$$

则核函数以及窗宽的选取, 应该使后验概率密度函数与核概率密度函数之差的积分均方误差最小. 在本文中, $K(\mathbf{x})$ 选为高斯核, 即

$$K(\mathbf{x}) = (2\pi)^{-\frac{n_x}{2}} \exp\left(-\frac{\mathbf{x}^T \mathbf{x}}{2}\right) \quad (10)$$

得到核概率密度函数 $\hat{p}(\mathbf{x}_k|\mathbf{z}_{1:k})$ 后, 就可以利用均值漂移技术调整粒子的位置, 使其向后验概率密度函数的局部极大值方向移动. 均值漂移通过把每一个粒子都转移到它的均值位

置 $m(\mathbf{x}_k^i)$ 实现了这一目的. 采样粒子调整后, 其位置为

$$m(\mathbf{x}_k^i) = \frac{\sum_{l=1}^N K_\lambda(\mathbf{x}_k^i - \mathbf{x}_k^l) w_k^l \mathbf{x}_k^l}{\sum_{l=1}^N K_\lambda(\mathbf{x}_k^i - \mathbf{x}_k^l) w_k^l} \quad (11)$$

由于采样粒子的位置发生了改变, 其权值也必须作相应调整才能正确描述其概率密度函数. 在均值漂移方法中, 每一个粒子的权值重新计算为

$$w_k^i = \frac{p(\mathbf{x}_k^i | \mathbf{z}_{1:k})}{q(\mathbf{x}_k^i)} \quad (12)$$

其中 $q(\mathbf{x}_k) = \sum_{l=1}^N K_\lambda(\mathbf{x}_k - \mathbf{x}_k^l)$ 为平衡因子^[18].

Cheng^[11] 从理论上证明, 均值漂移过程将采样粒子沿概率分布的最陡下降方向调整, 是一种局部优化过程. 经过均值漂移处理后, 采样粒子向似然值较大的区域聚集, 处于更加有效的分布位置, 因而能减少所需粒子的数目.

为了应用方便, 我们将上面提出的改进方法综合在一起, 给出了如下 MS-QMC 滤波方法的具体步骤:

在时刻 $k = 1, 2, \dots$

1) 在 $[0, 1)^{n_x}$ 上生成包含 N 个元素的拟蒙特卡洛序列 $\{\mathbf{u}_i\}_{i=1}^N$;

2) 随机化处理: 产生 N 个均匀分布的随机变量 $\mathbf{m}_i \sim \text{Uniform}(C^{n_x})$, $i = 1, \dots, N$, 作运算 $\mathbf{s}_i = (\mathbf{u}_i + \mathbf{m}_i) \bmod 1$, 则 $\{\mathbf{s}_i\}_{i=1}^N$ 为随机拟蒙特卡洛序列, 其中 “mod” 为取模运算;

3) 计算 k 时刻的重要性概率密度函数 $\pi(\mathbf{x}_k)$ 及其对应的分布函数 $P(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \pi(\mathbf{s}) d\mathbf{s}$;

4) 利用变换 $\mathbf{x}_k^i = P^{-1}(\mathbf{s}_i)$, 将随机拟蒙特卡洛序列 $\{\mathbf{s}_i\}_{i=1}^N$ 映射为状态空间上的采样序列 $\{\mathbf{x}_k^i\}_{i=1}^N$;

5) 计算各采样粒子的归一化权值 $\tilde{w}_k^i \propto p(\mathbf{z}_k | \mathbf{x}_k^i) \times p(\mathbf{x}_k^i | \mathbf{z}_{1:k-1}) / \pi(\mathbf{x}_k^i)$, 其中, $p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) \cong \sum_{j=1}^N w_{k-1}^j p(\mathbf{x}_k | \mathbf{x}_{k-1}^j)$;

6) 均值漂移: 计算 $m(\mathbf{x}_k^i) = \sum_{l=1}^N K_\lambda(\mathbf{x}_k^i - \mathbf{x}_k^l) w_k^l \mathbf{x}_k^l / \sum_{l=1}^N K_\lambda(\mathbf{x}_k^i - \mathbf{x}_k^l) w_k^l$, 将粒子 \mathbf{x}_k 调整到其均值位置, 并重新计算权值 $w_k^i = p(\mathbf{x}_k^i | \mathbf{z}_{1:k}) / q(\mathbf{x}_k^i)$.

3 基于 MS-QMC 滤波的说话人跟踪方法

3.1 布朗运动模型

由于说话人运动具有随意性, 难以用一种简单的运动模型描述说话人跟踪问题, 为此, 我们采用适应性较强的布朗运动模型, 其相应的运动方程称为朗之万 (Langevin) 过程^[6].

设说话人状态为 $\mathbf{x}_k = [x_k \ \dot{x}_k \ y_k \ \dot{y}_k]^T$, 其中 $(x_k \ y_k)$ 、 $(\dot{x}_k \ \dot{y}_k)$ 分别表示说话人的位置和速度; 假定声源在 X 、 Y 两个方向上的运动完全独立, 且运动方式的改变是随机的, 即 $E[\mathbf{x}_{k+1}] = \mathbf{x}_k$, 则声源运动可以用布朗运动模型描述. 在 X 轴上, 其状态满足 $\ddot{x} + \beta_x \dot{x} = F_x$, 其中 β_x 是速率常数, F_x 为高斯白噪声. 对应的离散表示形式为

$$\dot{x}_k = a_x \dot{x}_{k-1} + b_x F_x, \quad x_k = x_{k-1} + \Delta T \dot{x}_k \quad (13)$$

其中 $F_x \sim N(0, 1)$, ΔT 是采样周期, $a_x = \exp(-\beta_x \Delta T)$, $b_x = \bar{v}_x \cdot \sqrt{1 - a_x^2}$, \bar{v}_x 是稳态条件下的速率.

由于声源在 X 、 Y 两个方向上运动的独立性, 说话人运动的动态方程可表示为

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{\Gamma} w_k \quad (14)$$

其中 $A = \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$ 为状态转移矩阵, $B = \begin{bmatrix} 1 & \Delta T \\ 0 & a_x \end{bmatrix}$, $\mathbf{\Gamma} = [\Delta T \ 1 \ \Delta T \ 1]^T$, $w_k \sim N(0, b_x^2)$.

3.2 麦克风阵列时间延迟估计及声源定位

在说话人定位系统中, 由于各麦克风所处的位置不同, 语音到达各麦克风的时间存在差异, 这就是麦克风之间的时间延迟. 我们利用广义互相关法^[19], 通过求解两个麦克风接收信号互相关函数的极值点来估计时间延迟.

设声源信号为 $s(k)$, 两个麦克风接收到的信号分别为

$$y_1(k) = s(k - \tau_1) + m_1(k) \quad (15)$$

$$y_2(k) = \alpha s(k - \tau_2) + m_2(k) \quad (16)$$

其中 α 是声源信号的衰减系数, τ_1 和 τ_2 是声源到两麦克风的传播时间, 信号 $s(k)$ 与噪声 $m_1(k)$ 和 $m_2(k)$ 均不相关.

设信号 $y_1(k)$ 和 $y_2(k)$ 的互功率谱为 $\mathbf{S}_{y_1 y_2}(f) = E[\mathbf{Y}_1(f) \mathbf{Y}_2^*(f)]$, 加权函数为 $\varphi(f)$, 广义互功率谱为 $\Psi_{y_1 y_2}(f) = \varphi(f) \mathbf{S}_{y_1 y_2}(f)$, 则广义互相关函数可表示为

$$\Psi_{y_1 y_2}(\tau) = \int_{-\infty}^{+\infty} \varphi(f) \mathbf{S}_{y_1 y_2}(f) e^{j2\pi f \tau} df = \int_{-\infty}^{+\infty} \Psi_{y_1 y_2}(f) e^{j2\pi f \tau} df \quad (17)$$

这样, 两麦克风之间的时间延迟 $\tau = \tau_1 - \tau_2$ 对应于广义相关函数的极值点位置, 即

$$\hat{\tau} = \arg \max_{\tau} \Psi_{y_1 y_2}(\tau) \quad (18)$$

在式 (17) 中, 选取不同的加权函数 $\varphi(f)$, 可以得到不同的时延估计方法. 在相位变换 (Phase transform, PHAT)^[19] 方法中, 加权函数取为 $\varphi(f) = 1/|\mathbf{S}_{y_1 y_2}(f)|$. 该方法通过对互功率谱预白化, 去除了与时延无关的互功率谱幅度信息而仅保留相位特性, 以锐化函数公式 $\Psi_{y_1 y_2}(\tau)$ 的极值, 从而较好地抑制噪声和混响对时延估计的影响. 得到时间延迟估计后, 就可以对说话人进行跟踪. 设在某一时刻 k , 麦克风阵列得到的时延估计为 $\tau_{k,i}$, 说话人位置为 $\mathbf{r}_k = (x_k, y_k)$, 相应的麦克风位置分别为 $\mathbf{r}_{1,i} = (x_{1,i}, y_{1,i})$ 和 $\mathbf{r}_{2,i} = (x_{2,i}, y_{2,i})$, 声音在空气中的传播速度为 c , 则时间延迟 $\tau_{k,i}$ 与说话人位置之间的关系为

$$\tau_{k,i} = c^{-1} \cdot [|\mathbf{r}_{1,i} - \mathbf{r}_k| - |\mathbf{r}_{2,i} - \mathbf{r}_k|], \quad i = 1, 2 \quad (19)$$

即

$$\tau_{k,i} = c^{-1} \cdot [\sqrt{(x_{1,i} - x_k)^2 + (y_{1,i} - y_k)^2} - \sqrt{(x_{2,i} - x_k)^2 + (y_{2,i} - y_k)^2}] \quad (20)$$

对式 (14) 和 (20) 构成的说话人跟踪系统, 利用本文提出的 MS-QMC 滤波方法, 就能够递推估计出系统状态的后验概率密度函数 $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, 进而由其均值得到说话人的位置信息.

4 计算机仿真结果

在说话人跟踪仿真实验中, 我们设计的跟踪方案如图 5 所示. 在 $3\text{m} \times 3\text{m}$ 的房间内, 在 X 、 Y 两个方向上, 分别放置两组包含两个麦克风的线性阵列, 麦克风之间的距离均为 60cm , 说话人沿半圆形轨迹运动. 麦克风接收到

的信号用 IMAGE 模型^[20], 然后加上 30 dB 的高斯白噪声 $T_{60} = 0.39$ s, 对语音信号每隔 32 ms 采集一帧数据, 用 PHAT 方法估计时间延迟.

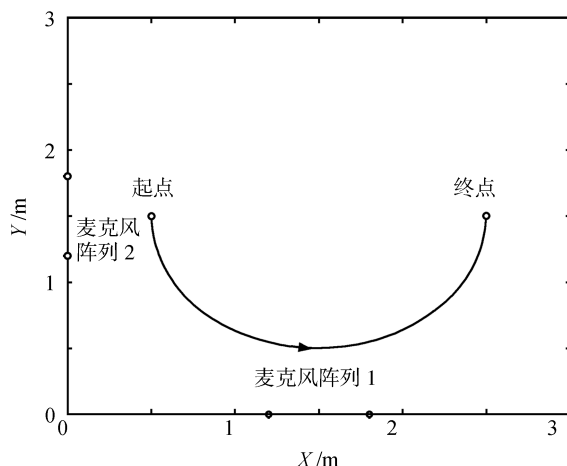


图5 说话人跟踪设计方案
Fig. 5 Speaker tracking scenario

说话人运动模型中各参数值为: $\beta_x = 10 \text{ s}^{-1}$, $\bar{v}_x = 1 \text{ ms}^{-1}$, $\Delta T = 32 \text{ ms}$; 初始状态值为 $\hat{\mathbf{x}}_0 = [2.2 \ 0.1 \ 2.5 \ -0.3]^T$. 为了检验 MS-QMC 方法在实际应用中的性能, 我们将其与基本的粒子滤波方法和拟蒙特卡洛方法进行了仿真实验比较. 在仿真过程中, 各滤波器使用相同的观测输入值和似然函数, 且初始条件统一设定为真值, 以避免初始化对滤波性能的影响. 粒子滤波利用了多项重采样^[4]过程, 拟蒙特卡洛滤波过程中使用辅助的扩展卡尔曼滤波器生成所需的重要性采样函数.

当采样粒子个数 $M = 40$ 时, 经过 30 次蒙特卡洛实验, 得到说话人位置估计的均方根误差随时间变化曲线如图 6 所示. 从图 6 可以看出, 本文所有对于基本粒子滤波方法的改进, 都能有效提高对说话人的跟踪精度. 由于确定性采样能提高滤波精度, 拟蒙特卡洛滤波相对于粒子滤波方法的估计误差降低了约 20%, MS-QMC 比拟蒙特卡洛的精度提高了约 50%, 这是因为均值漂移过程使采样粒子向高似然区域移动, 增强了粒子对后验概率密度函数的表示能力.

各滤波方法的滤波精度和滤波时间随粒子数目的变化如表 1 所示, 滤波时间是在赛扬 2.0 GHz CPU 和 256 M RAM 的个人计算机上, 使用 Matlab 6.5 软件实现滤波过程时统

计得到的. 由表 1 可知, 随着粒子数目的增加, 所有滤波方法的估计误差都会减少, 当粒子数相同时, 粒子滤波、拟蒙特卡洛、MS-QMC 的滤波误差依次减少, 这与图 6 的结果是一致的, 但所需时间依次增加, 即精度的提高是以增加计算量为代价的. 当 CPU 时间基本一致时, 如拟蒙特卡洛滤波 (40 粒子) 和 MS-QMC (29 粒子), MS-QMC 方法的估计精度高于相应的拟蒙特卡洛滤波方法; 当估计误差基本相近时, 如拟蒙特卡洛滤波 (300 粒子) 和 MS-QMC (20 粒子), 拟蒙特卡洛方法每一周期的 CPU 运行时间大大高于对应的 MS-QMC 方法, 即 MS-QMC 方法能够在同样估计精度下, 通过降低所需采样粒子的数目, 减少了计算需求. 从表 1 还可以看出, 随着粒子数目的增加, MS-QMC 相对于拟蒙特卡洛的性能改善程度降低, 这是因为在粒子数较少时, 更适合对其位置分布作优化处理.

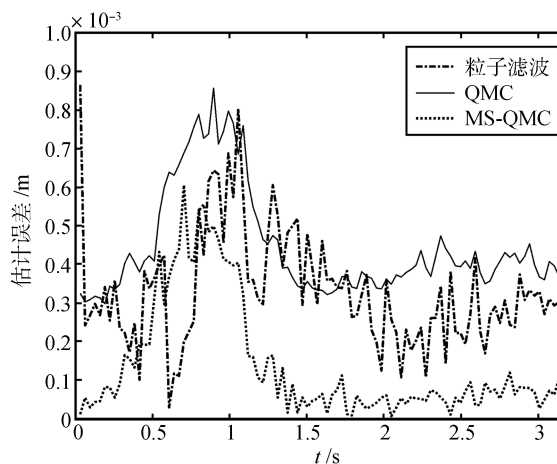


图6 说话人位置估计均方根误差
Fig. 6 RMSE of speaker tracking

在粒子滤波方法中, 有效采样粒子数 (Effective sampling size)^[13] 是衡量粒子衰退程度的一种参数, 它可以等价看作从后验分布中抽取的独立同分布采样点的个数. 在仿真中, 我们对采样粒子个数 $M = 40$, 各滤波方法在不同时刻对应的有效采样粒子数进行了统计, 结果如表 2 所示. 从表 2 可以看出, 随迭代次数的增加, 所有方法中的有效采样粒子数均逐步减少, 但 MS-QMC 方法中的有效采样粒子数总是大于粒子滤波和拟蒙特卡洛方法, 这从另一个方面说明了 MS-QMC 方法的优越性.

表 1 采样粒子数变化时, 不同滤波方法所用时间和估计精度

Table 1 Filtering time and estimation accuracies of different filtering methods with varying number of samples

滤波算法	采样粒子个数	每一运行周期所用时间 (s)	X 轴均方根估计误差 (m)	Y 轴均方根估计误差 (m)
粒子滤波	40	1.579667	0.159939	0.032126
粒子滤波	80	3.002100	0.120170	0.034739
粒子滤波	200	9.972367	0.091070	0.021831
拟蒙特卡洛	20	10.078133	0.053727	0.032087
拟蒙特卡洛	40	38.429700	0.049299	0.029440
拟蒙特卡洛	300	367.124642	0.017849	0.015126
MS-QMC	20	24.341167	0.014169	0.013904
MS-QMC	29	40.663651	0.011764	0.011342
MS-QMC	40	65.056767	0.010339	0.009825

表 2 不同滤波方法在滤波过程中的有效粒子数

Table 2 Effective sample sizes of different filtering methods

迭代次数 k	4	7	15	27	35	45	60	95
粒子滤波	3	3	2	2	2	3	4	3
拟蒙特卡洛	10	5	7	8	7	5	6	9
MS-QMC	24	26	28	18	17	18	11	9

5 结论

本文提出了一种基于均值漂移技术的 MS-QMC 方法。相对于基本的粒子滤波方法, 该方法利用确定性采样方法, 保证了采样粒子分布的均匀性, 使采样粒子对系统状态后验概率密度函数的描述更加充分; 同时, 使用均值漂移过程调整采样粒子的位置, 使之沿梯度方向向高似然区域移动, 从而增加了有效采样粒子的个数, 减少了所需采样粒子的数目; 此外, 该方法在每次迭代之后, 都由重要性采样函数重新生成粒子集合, 这避免了粒子退化现象, 因而不再需要重采样过程。将本文方法应用于对说话人跟踪系统, 利用麦克风阵列接收语音信息, 实现了对说话人的有效跟踪。仿真结果表明, 本文所提方法能够得到更高的跟踪精度和更快的收敛速度。

References

- Potamitis I, Chen H M, Tremoulis G. Tracking of multiple moving speakers with multiple microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 2004, **12**(5): 520–529
- Brandstein M A. A Framework for Speech Source Localization Using Sensor Arrays [Ph.D. dissertation], Brown University, USA, 1995
- Dvorkind T, Gannot S. Speaker localization exploiting spatial-temporal information. In: Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control. Kyoto, Japan: IEEE, 2003. 295–298
- Gordon N J, Salmond D J, Smith A F M. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings on Radar and Signal Processing*, 1993, **140**(2): 107–117
- Liu J S, Chen R. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 1998, **93**(443): 1032–1044
- Vermaak J, Blake A. Nonlinear filtering for speaker tracking in noisy and reverberant environments. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA: IEEE, 2001. 3021–3024
- Ward D B, Lehmann E A, Williamson R C. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *IEEE Transactions on Speech and Audio Processing*, 2003, **11**(6): 826–836
- Guo D, Wang X D. Quasi-Monte Carlo filtering in nonlinear dynamic systems. *IEEE Transactions on Signal Processing*, 2006, **54**(6): 2087–2098
- Fearnhead P. Sequential Monte Carlo Methods in Filter Theory [Ph.D. dissertation], University of Oxford, UK, 1998

- LEcuyer P, Lemieux C. Recent advances in randomized quasi-Monte Carlo methods. *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*. Boston, USA: Springer, 2005. 419–474
- Cheng Y Z. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(8): 790–799
- Morokoff W J, Caflisch R E. Quasi-Monte Carlo integration. *Journal of Computational Physics*, 1995, **122**(2): 218–230
- Liu J S, Chen R, Logvinenko T. A theoretical framework for sequential importance sampling with resampling. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001. 225–246
- Owen A B. Monte Carlo extension of quasi-Monte Carlo. In: Proceedings of the 1998 Winter Simulation Conference. Washington D. C., USA: IEEE, 1998. 571–577
- Marsaglia G, Zaman A, Marsaglia J C W. Rapid evaluation of the inverse of the normal distribution function. *Statistics and Probability Letters*, 1994, **19**(4): 259–266
- Henderson S G, Chiera B A, Cooke R M. Generating “dependent” quasi-random numbers. In: Proceedings of the 2000 Winter Simulation Conference. Orlando, USA: IEEE, 2000. 527–536
- Silverman B W. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall, 1986. 75–94
- Chang C, Ansari R. Kernel particle filter for visual tracking. *IEEE Signal Processing Letters*, 2005, **12**(3): 242–245
- Knapp C, Carter G. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1976, **24**(4): 320–327
- Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 1979, **65**(4): 943–950

侯代文 大连理工大学电子与信息工程学院博士研究生。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。本文通信作者。

E-mail: hodevin@gmail.com

(**HOU Dai-Wen** Ph. D. candidate at the School of Electronic and Information Engineering, Dalian University of Technology. His research interest covers localization and tracking, acoustic signal processing, and array signal processing. Corresponding author of this paper.)

殷福亮 大连理工大学电子与信息工程学院教授。主要研究方向为语音信号处理、图像处理、阵列信号处理、宽带无线通信技术。

E-mail: flyin@dlut.edu.cn

(**YIN Fu-Liang** Professor at the School of Electronic and Information Engineering, Dalian University of Technology. His research interest covers acoustic signal processing, image signal processing, array signal processing, and ultra-banded wireless communication.)

陈喆 大连理工大学电子与信息工程学院副教授。主要研究方向为语音信号处理、阵列信号处理、宽带无线通信技术。

E-mail: eeyin@dlut.edu.cn

(**CHEN Zhe** Associate professor at the School of Electronic and Information Engineering, Dalian University of Technology. His research interest covers acoustic signal processing, array signal processing, and ultra-banded wireless communication.)