

基于贝叶斯理论的分布式多视角目标跟踪算法

冯 巍, 胡 波, 杨 成, 林 青, 杨 涛

(复旦大学电子工程系, 上海 200433)

摘 要: 为了有效解决传统单视角跟踪难于处理的目标遮挡问题, 本文提出了一种分布式多视角目标跟踪算法. 该算法首先基于贝叶斯理论, 为多视角目标跟踪问题建立了分布式数据融合的概率框架; 并利用粒子滤波器对所需后验概率进行近似, 提出了自适应的观测模型和状态转移模型. 各摄像机能够并行化地进行数据采集、处理、融合, 而无需集中式处理单元; 能够有效避免遮挡造成的误差传递, 提高跟踪算法的鲁棒性. 实验证明了本文算法的有效性.

关键词: 分布式多视角目标跟踪; 粒子滤波器; 自适应观测模型; 自适应状态转移模型

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112 (2011) 02-0315-07

A Distributed Multi-View Object Tracking Algorithm Under the Bayesian Framework

FENG Wei, HU Bo, YANG Cheng, LIN Qing, YANG Tao

(Department of Electronic Engineering, Fudan University, Shanghai 200433, China)

Abstract: A distributed multi-view object tracking algorithm is proposed to address the occlusion problem. The Bayesian sequential tracking framework is used to model the multi-view tracking problem and implemented with particle filtering. The centralized computing unit is no longer needed. Image acquisition, processing and data fusion can be performed by each camera in parallel. An adaptive observation model and an adaptive state transition model are also proposed to enable efficient data fusion and robust tracking against various occlusions. Experiments have verified the effectiveness of the algorithm.

Key words: distributed multi-view object tracking; particle filtering; adaptive observation model; adaptive state transition model

1 引言

基于图像序列的目标跟踪是图像处理和计算机视觉领域的研究热点之一, 在视频监控、人机交互、交通控制等领域中有着广泛的应用前景. 其主要任务是在摄像机获取的每帧图像中准确估计出目标的位置信息.

目前国内外发表的目标跟踪算法, 主要针对单个摄像机(单视角)的应用场景. 然而, 单视角跟踪丢失了跟踪场景的深度信息. 在发生遮挡现象时, 常常造成跟踪的失败. 为解决遮挡问题, 文献[1, 2]均提出了可处理局部遮挡的算法, 但对目标被完全遮挡的情况无能为力. 而多摄像机跟踪系统则可以获得跟踪场景多个视角的图像信息, 为解决遮挡问题提供了可能^[3]. 现有的多视角跟踪算法, 从实现结构上而言, 可分为集中式处理和分布式处理两类.

集中式处理是指各摄像机节点将采集到的图像数据首先传输到集中式处理单元, 然后由该单元完成后续

的图像处理和数据融合任务, 给出跟踪结果. 例如, 文献[5]提出了基于特征融合的多摄像机跟踪算法; 文献[6]算法将地平面分割成若干个区域, 利用目标在这些区域上的后验概率来估计目标的位置. 此类算法需要较高的网络带宽传输二维图像数据, 对集中式处理单元的计算能力要求较高, 还存在网络延时等问题, 难以在分布式摄像机网络中实现^[3, 4].

分布式算法则利用分布式智能摄像机系统, 由各摄像机节点并行化地完成图像数据的采集和处理任务, 然后通过网络进行信息融合, 能够弥补集中式处理的不足之处^[4]. 文献[7]使用分布式摄像机扩大了目标跟踪的有效区域, 但难以有效处理遮挡情况. 文献[8]提出了利用粒子滤波器进行多摄像机交互式跟踪的算法; 文献[9]提出了在无线传感器网络中进行分布式跟踪的算法. 以上两种算法均依赖于背景消除技术, 需要预先获得静止背景的信息, 无法应用于背景信息未知的场景. 台湾学者 Lien 等^[10]提出了基于隐马尔科夫过程和粒子

滤波器的协同式目标跟踪算法,可处理运动目标间相互遮挡的情况.但在处理遮挡时,需要首先利用单个摄像头中的数据,对遮挡情况进行分析,然后进行数据融合校正跟踪目标的位置,避免跟踪失败.计算量较大且无法处理目标被静态遮挡物遮挡的情况.

为了有效解决各种遮挡情况,本文首先基于贝叶斯理论对多视角目标跟踪问题进行建模,借助运动目标在不同视角中存在的几何映射关系,建立了分布式融合的概率框架.在此框架下,各摄像机节点能够并行化地进行数据采集、处理,并通过网络完成节点间的数据融合,而无需集中式处理单元.本文进一步使用粒子滤波器近似逼近所需的后验概率,并提出了自适应的观测模型和状态转移模型.其中,自适应观测模型能够有效处理局部遮挡;而在自适应状态转移模型中,每台摄像机的粒子可由前一时刻本摄像机的粒子和其它摄像机的跟踪结果两部分自适应地生成,能够有效避免遮挡造成的误差传递,提高跟踪算法对遮挡的鲁棒性.

2 分布式数据融合的概率框架

2.1 符号表示和问题建模

假设跟踪目标处在 N 台摄像机 (C_1, C_2, \dots, C_N) 的视野中,在 t 时刻摄像机 C_i ($i = 1, 2, \dots, N$) 采集到的图像数据集合表示为 $\mathbf{y}_{t,i}$,跟踪目标在该幅图像中的位置则表示为 $\mathbf{x}_{t,i}$.整个跟踪系统的目标是获得 t 时刻跟踪目标在各摄像机视野中位置的估计值 $\hat{\mathbf{x}}_{t,1}, \hat{\mathbf{x}}_{t,2}, \dots, \hat{\mathbf{x}}_{t,N}$.

如果不考虑摄像机间的数据交互,即每台摄像机只能利用本地获得的图像数据 $\mathbf{y}_{1:t,i}$ ($\mathbf{y}_{1:t,i} = \{\mathbf{y}_{1,i}, \mathbf{y}_{2,i}, \dots, \mathbf{y}_{t,i}\}$,表示从 1 到 t 时刻摄像机 C_i 获得的图像数据集合)对当前视野中目标位置 $\mathbf{x}_{t,i}$ 进行估计,那么多视角跟踪问题就简化为传统的单视角目标跟踪.在贝叶斯理论中,可首先求取后验概率 $p(\mathbf{x}_{t,i} | \mathbf{y}_{1:t,i})$,然后根据最小均方误差或最大后验概率准则,得出 t 时刻跟踪目标位置的估计值 $\hat{\mathbf{x}}_{t,i}$.但是,如果在跟踪过程中发生遮挡情况,那么在采集到的图像 $\mathbf{y}_{t,i}$ 中将含有不完整的目标外观数据,会造成跟踪算法精度的降低.由于贝叶斯方法的递推特性,跟踪结果的误差将逐渐积累,目标位置的估计精度不断降低,并最终导致跟踪失败.

如果各台摄像机间能够进行数据融合,则可以利用跟踪目标各个视角的观测数据对目标位置进行估计,从而为有效解决遮挡提供了可能.一种可能的方案是利用所有摄像机采集到的图像数据 $\mathbf{Y}_{1:t} = \{\mathbf{y}_{1:t,1}, \mathbf{y}_{1:t,2}, \dots, \mathbf{y}_{1:t,N}\}$,求取跟踪目标状态的集合 $\mathbf{X}_t = \{\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, \dots, \mathbf{x}_{t,N}\}$ 的联合后验概率 $p(\mathbf{X}_t | \mathbf{Y}_{1:t})$,进而获得跟踪目标状态集合的估计值.但该方案存在两个难以解决的问题.首先, \mathbf{X}_t 的状态空间是以指数形式随摄像机

的个数 N 增加而增加的,会使问题的求解陷入维数灾难;其次,利用联合后验概率获得目标的状态估计值,要求所有摄像机将采集到的数据传递到集中式处理单元进行数据融合,其缺陷已在简介中介绍.

文献[16]中指出,求取边缘概率可以有效降低联合概率所需的状态空间维度,从而解决维数灾难问题.这里,可以采用另一种方案,即通过求取边缘后验概率 $p(\mathbf{x}_{t,i} | \mathbf{Y}_{1:t,i})$ 获得跟踪目标位置的估计值,以解决维数灾难问题.同时,考虑到实际系统中存在网络延时,假设在估计 $\mathbf{x}_{t,i}$ 时,不能得到同时刻其它摄像机的图像数据,而只能利用本节点的图像数据和其它节点以前时刻的数据,则目标后验概率变为 $p(\mathbf{x}_{t,i} | \mathbf{Y}_{1:t-1,i}, \mathbf{y}_{t,i})$.但可以发现,摄像机节点 C_i 求取边缘概率时,需要将其它节点前面时刻采集到的图像数据传输到该节点,仍无法解决节点间传输数据量大的问题.

为降低节点间的通信开销,可进一步将目标后验概率建模为 $p(\mathbf{x}_{t,i} | \mathbf{y}_{1:t,i}, \mathbf{z}_{1:t-1,i})$.其中, $\mathbf{z}_{1:t-1,i}$ 表示除 C_i 以外其它所有摄像机在 $t-1$ 时刻的估计结果集合,即 $\mathbf{z}_{1:t-1,i} = \{\hat{\mathbf{x}}_{1:t-1,j} | (j = 1, 2, \dots, N) \cap (j \neq i)\}$,同时 $\mathbf{z}_{1:t-1,i} = \{\mathbf{z}_{1,1,i}, \mathbf{z}_{2,1,i}, \dots, \mathbf{z}_{t-1,1,i}\}$.每台摄像机 C_i 首先求取后验概率 $p(\mathbf{x}_{t,i} | \mathbf{y}_{1:t,i}, \mathbf{z}_{1:t-1,i})$,然后利用最小均方误差准则,计算出跟踪目标在该摄像机视野中的位置估计值,即

$$\hat{\mathbf{x}}_{t,i} = \int \mathbf{x}_{t,i} p(\mathbf{x}_{t,i} | \mathbf{y}_{1:t,i}, \mathbf{z}_{1:t-1,i}) d\mathbf{x}_{t,i} \quad (1)$$

这样,节点间只需传递位置的估计值,而无须传递原始的图像数据.上述建模方式有着如下优势:第一,求取每台摄像机视野中目标位置的边缘化后验概率,可以有效解决维数灾难问题.第二,每台摄像机在进行目标位置估计时,考虑了实际系统中可能存在的延时,只需要融合其它摄像机在前面时刻的估计结果,而无须对原始图像数据进行融合,极大程度地降低了摄像机间的通信开销;第三,无需融合中心,各摄像机能够以分布式融合的方式进行目标位置的估计.

2.2 后验概率的递推公式

在本节中,将利用随机变量间的条件独立关系,推导出后验概率 $p(\mathbf{x}_{t,i} | \mathbf{y}_{1:t,i}, \mathbf{z}_{1:t-1,i})$ 的递推表达式.首先使用贝叶斯网络表示后验概率中涉及到的随机变量间的关系.从贝叶斯网络中,可以利用 d-separation 准则^[11]方便地分析出各个变量间的条件独立关系.对于摄像机 C_i 而言, t 时刻采集得到的观测数据 $\mathbf{y}_{t,i}$ 是受到该时刻目标位置 $\mathbf{x}_{t,i}$ 影响的;同时, t 时刻运动目标位置 $\mathbf{x}_{t,i}$ 受到前一时刻目标位置 $\mathbf{x}_{t-1,i}$ 的影响,以上两点与经典贝叶斯滤波的假设是完全一致的.同时,运动目标在各台摄像机视野中的位置需要满足几何映射关系^[12],

所以可以利用其它摄像机前一时刻的跟踪结果集合 $z_{t-1,i}$ 对当前时刻目标位置 $x_{t,i}$ 的状态空间进行约束,即可认为 $z_{t-1,i}$ 也能够对目标位置 $x_{t,i}$ 造成影响. 根据以上假设,最终的动态贝叶斯网络如图 1 所示.

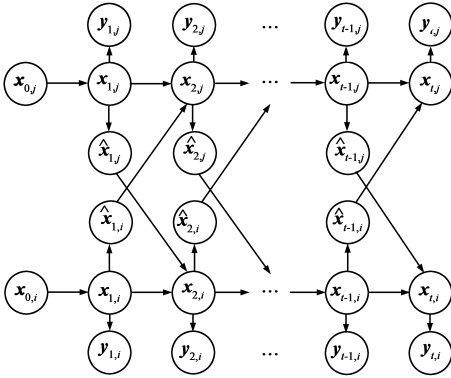


图1 融合其他摄像机前一时刻位置估计值的动态贝叶斯网络

利用图 1 的贝叶斯网络,我们可以得到后验概率的递推公式:

$$p(x_{t,i} | y_{1:t,i}, z_{1:t-1,i}) = p(y_{t,i} | x_{t,i}) \int p(x_{t,i} | x_{t-1,i}, z_{t-1,i}) \cdot p(x_{t-1,i} | y_{1:t-1,i}, z_{1:t-2,i}) dx_{t-1,i} \quad (2)$$

式(2)中, $p(x_{t-1,i} | y_{1:t-1,i}, z_{1:t-2,i})$ 是 $t-1$ 时刻的后验概率. 根据贝叶斯理论, $p(y_{t,i} | x_{t,i})$ 称为观测模型, 而 $p(x_{t,i} | x_{t-1,i}, z_{t-1,i})$ 为状态转移模型. 根据式(2), 利用 $t-1$ 时刻的后验概率、观测模型和状态转移模型, 即可计算出 t 时刻目标位置的后验概率.

2.3 分布式融合算法的示意图

图 2 中显示了分布式跟踪的系统构架. 在此构架下, 各台摄像机能够并行化地进行图像采集和处理. 同时, 利用网络传输各视角中目标位置的估计值, 完成数据融合.

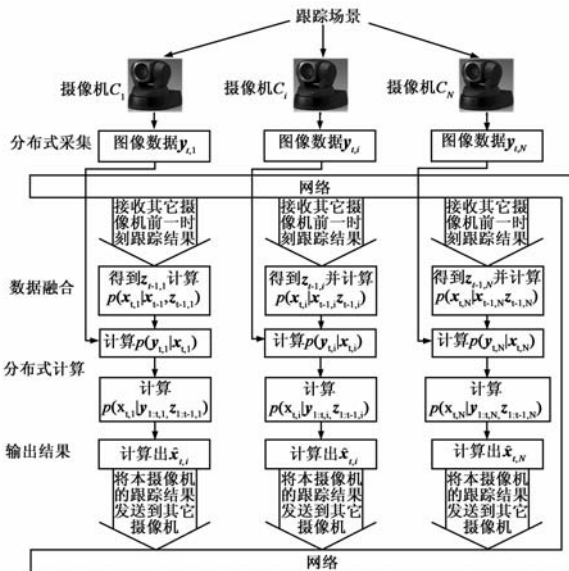


图2 分布式融合算法的直观示意图

在本节中,我们建立了分布式多视角目标跟踪的框架,但并没有讨论如何有效解决遮挡的问题. 在多视角跟踪系统中时常碰到的一个场景是,跟踪目标在一些视角中被遮挡,跟踪精度较低;而在其它视角中没有被遮挡,跟踪精度较高. 进行有效数据融合的目标是利用精度较高的跟踪结果,对其它精度较低的跟踪结果进行补偿,提高跟踪结果的精度;同时,需要避免精度较低的跟踪结果的反向传递,导致其它视角跟踪精度的下降. 下面,将利用粒子滤波器对(2)中所需的后验概率进行近似,并建立自适应的观测模型和状态转移模型,以提高跟踪算法对遮挡的鲁棒性.

3 粒子滤波器实现

3.1 权重递推公式

文献[13]指出,在视频跟踪应用中,受到目标形变和遮挡的影响,观测模型是一个高度非线性的过程. 而粒子滤波算法可以近似逼近所需的后验分布,有效解决观测模型的非线性问题. 针对本文应用,需要为随机向量 $x_{t,i}$ 产生 N_i 个粒子,利用粒子 $\{x_{t,i}^{(m)}\}_{m=1}^{N_i}$ 及其相应的权值 $\{w_{t,i}^{(m)}\}_{m=1}^{N_i}$, 逼近摄像机 C_i 所需的分布 $p(x_{t,i} | y_{1:t,i}, z_{1:t-1,i})$, 即:

$$p(x_{t,i} | y_{1:t,i}, z_{1:t-1,i}) \approx \sum_{m=1}^{N_i} w_{t,i}^{(m)} \delta(x_{t,i} - x_{t,i}^{(m)}) \quad (3)$$

将式(3)带入式(1)中,则每台摄像机视野中目标位置估计值的计算公式为:

$$\hat{x}_{t,i} \approx \sum_{m=1}^{N_i} w_{t,i}^{(m)} x_{t,i}^{(m)}, \quad i = 1, 2, \dots, N \quad (4)$$

若选取状态转移模型 $p(x_{t,i} | x_{t-1,i}, z_{t-1,i})$ 作为粒子生成函数,则粒子 $x_{t,i}^{(m)}$ 权重 $w_{t,i}^{(m)}$ 的递推表达式为:

$$w_{t,i}^{(m)} \propto p(y_{t,i} | x_{t,i}^{(m)}) w_{t-1,i}^{(m)} \quad (5)$$

3.2 自适应观测模型

采用文献[13]中提出的分块色彩模型作为观测模型的基础,即在 t 时刻摄像机 C_i 中的粒子权重 $w_{t,i}^{(m)}$ 计算公式为:

$$w_{t,i}^{(m)} \propto p(y_{t,i} | x_{t,i}^{(m)}) = \exp\left\{\lambda \sum_{j=1}^{J_i} S_c[q_{i,j}^*, q_{i,j}(\mathbf{x}_{t,i}^{(m)})]\right\} \quad (6)$$

其中, λ 为常数 20, J_i 是第 i 台摄像机中跟踪目标的外观模型被划分出的部分数; $q_{i,j}^*$ 表示第 i 台摄像机第 j 部分的外观模型; $q_{i,j}(\mathbf{x}_{t,i}^{(m)})$ 表示 t 时刻第 i 台摄像机中的粒子 $x_{t,i}^{(m)}$ 对应区域的第 j 部分图像; S_c 表示目标外观模型与粒子对应区域图像的 HS 分量直方图的互相关系数,表征了两者间的色彩相似度. 采用此分块模型可以很好地衡量图像的局部特征.

为了进一步提高观测模型对遮挡的鲁棒性,在上述模型中添加了自适应加权策略,即:

$$w_{t,i}^{(m)} \propto \exp\left\{\lambda \sum_{j=1}^{J_i} b_{t,i,j} S_c[q_{i,j}^*, q_{i,j}(\mathbf{x}_{t,i}^{(m)})]\right\} \quad (7)$$

其中, $b_{t,i,j}$ 表示 t 时刻第 i 台摄像机中第 j 部分的权重, 是由前一时间刻相应部分的相似度计算得到的, 即

$$b_{t,i,j} = S_c[q_{i,j}^*, q_{i,j}(\hat{\mathbf{x}}_{t-1,i})] \quad (8)$$

而 t 时刻, 目标外观模型与跟踪估计结果对应区域图像的相似度 (简称相似度) 之和, 则可以大致衡量跟踪算法精度的高低:

$$s_{t,i} = \sum_{j=1}^{J_i} S_c[q_{i,j}^*, q_{i,j}(\hat{\mathbf{x}}_{t,i})] \quad (9)$$

式(9)可理解为, t 时刻摄像机 C_i 视野中, 如果由于遮挡现象造成跟踪结果精度较低, 即跟踪结果对应的区域与目标实际的区域距离很远, 那么跟踪结果对应的图像区域与目标外观模型间的相似度较小, 即 $s_{t,i}$ 较小; 反之, 如果跟踪精度较高, 则 $s_{t,i}$ 较大.

3.3 自适应状态转移模型

在本文的分布式跟踪框架中, 采用 $p(\mathbf{x}_{t,i} | \mathbf{x}_{t-1,i}, \mathbf{z}_{t-1,i})$ 作为粒子生成函数. 这样, 在 t 时刻摄像机 C_i 视野中产生的粒子 $\mathbf{x}_{t,i}^{(m)}$ 不仅可以由前一时间刻状态 $\mathbf{x}_{t-1,i}$ 产生, 还可以由前一时间刻其它摄像机的跟踪结果 $\mathbf{z}_{t-1,i}$ 生成. 假设 t 时刻摄像机 C_i 需要生成的粒子总数为 $N_{t,i}$, 由前一时间刻状态 $\mathbf{x}_{t-1,i}$ 产生的粒子数为 $N_{t,\bar{i}}$, 由其它摄像机 C_j 前一时间刻的跟踪结果生成的粒子数为 $N_{t,\bar{j}}$, 则

$$N_{t,\bar{i}} + \sum_{j=1, j \neq i}^N N_{t,\bar{j}} = N_{t,i} \quad (10)$$

而每部分生成的粒子数与前一时刻的跟踪精度成正比, 在上一节中已经讨论过, 可以用相似度来近似衡量每个摄像机的跟踪精度, 即:

$$N_{t,\bar{j}}/N_{t,\bar{i}} = s_{t-1,j}/s_{t-1,i}, j \neq i \quad (11)$$

通过求解式(10)和式(11), 可以计算出各部分产生的粒子数目. 其物理意义在于: 在 $t-1$ 时刻, 在未发生遮挡现象的摄像机中, 跟踪精度较高, 那么在 t 时刻由其产生的粒子数目应该较多; 在发生遮挡现象的摄像机中, 跟踪精度较低, 那么在 t 时刻由其产生的粒子数目应该较少. 这样, t 时刻各摄像机生成的粒子, 绝大部分是由 $t-1$ 时刻跟踪质量较好的跟踪结果决定的, 只有很小一部分是由跟踪质量较差的跟踪结果产生的, 可以有效减少 $t-1$ 时刻发生遮挡现象的摄像机中的跟踪结果对下一时刻的影响, 避免误差的传递.

下面, 将具体讨论如何由上述的两部分产生粒子. 首先, t 时刻由本摄像机前一时间刻的粒子状态 $\mathbf{x}_{t-1,i}$ 生成 $\mathbf{x}_{t,i}^{(m)}$ 的公式, 可利用目标位置在时间上的连续性^[13], 即

$$\mathbf{x}_{t,i}^{(m)} \sim N(\mathbf{x}_{t-1,i}, \boldsymbol{\Sigma}_{t,i}), m = 1, 2, 3, \dots, N_{t,\bar{i}} \quad (12)$$

其中, $N(\mathbf{x}_{t-1,i}, \boldsymbol{\Sigma}_{t,i})$ 表示均值为 $\mathbf{x}_{t-1,i}$, 协方差矩阵为 $\boldsymbol{\Sigma}_{t,i}$ 的高斯函数.

而使用其它摄像机前一时刻的跟踪结果 $\mathbf{z}_{t-1,i}$ 产生 t 时刻粒子 $\mathbf{x}_{t,i}^{(k)}$ 的情况, 则较复杂. 首先考虑使用 $\mathbf{z}_{t-1,i}$ 中的一个元素 $\hat{\mathbf{x}}_{t-1,j}$ 产生粒子的方法, 由前面的讨论可知, 由 $\hat{\mathbf{x}}_{t-1,j}$ 生成的粒子个数为 $N_{t,\bar{j}}$. 根据目标位置在时间上的连续性, 粒子 $\mathbf{x}_{t,i}^{(k)}$ 在摄像机 C_j 中的对应位置 $\bar{\mathbf{x}}_{t,j}^{(k)}$, 应位于该节点前一时间刻估计结果 $\hat{\mathbf{x}}_{t-1,j}$ 的附近, 即满足 $\bar{\mathbf{x}}_{t,j}^{(k)} \sim N(\hat{\mathbf{x}}_{t-1,j}, \boldsymbol{\Sigma}_{t,j})$, $k = 1, 2, 3, \dots, N_{t,\bar{j}}$ (13) 同时, 因为 $\mathbf{x}_{t,i}^{(k)}$ 和 $\bar{\mathbf{x}}_{t,j}^{(k)}$ 是目标真实位置在摄像机 C_i 和 C_j 上的投影, 所以可利用多视角成像的几何映射关系 (即单应性矩阵^[12]) 将 $\bar{\mathbf{x}}_{t,j}^{(k)}$ 转化为 $\mathbf{x}_{t,i}^{(k)}$, 即

$$\mathbf{x}_{t,i}^{(k)} = \mathbf{H}_{j,i} \bar{\mathbf{x}}_{t,j}^{(k)}, j \neq i \quad (14)$$

其中, $\mathbf{H}_{j,i}$ 为已知的单应性矩阵, 可将摄像机 C_j 中的坐标位置映射为摄像机 C_i 中的坐标位置.

而每台摄像机 t 时刻的粒子数 $N_{t,i}$ 和协方差矩阵 $\boldsymbol{\Sigma}_{t,i}$ 同样可以自适应计算得到. 假设跟踪目标在图像平面 x 轴和 y 轴上的运动是相互独立的, 则跟踪目标状态方程的协方差矩阵 $\boldsymbol{\Sigma}_{t,i}$ 为对角阵, 即 $\boldsymbol{\Sigma}_{t,i} = \text{diag}(\sigma_{t,i,x}^2, \sigma_{t,i,y}^2)$. 假设两个方向上的方差是相同的, 则 $\sigma_{t,i,x}^2 = \sigma_{t,i,y}^2 = \sigma_{t,i}^2$. $\sigma_{t,i}$ 可由下式计算得到:

$$\sigma_{t,i} = \max(\sigma_{\min}, \min(\sigma_{\max}, \sigma_0/s_{t-1,i})) \quad (15)$$

其中, σ_{\min} 、 σ_{\max} 、 σ_0 均为常数, σ_{\min} 为 $\sigma_{t,i}$ 的下限, σ_{\max} 为上限. 而 t 时刻 C_i 的粒子数 $N_{t,i}$ 可由下式计算得出:

$$N_{t,i} = \frac{N_0}{\sigma_0} \cdot \sigma_{t,i} \quad (16)$$

其中, N_0 为常数. 式(16)的物理含义为 t 时刻的协方差矩阵越大, $\mathbf{x}_{t,i}$ 的随机性就越大, 所需的粒子数也越多. 在本文算法中, $\sigma_{\min} = 1$, $\sigma_0 = 1$, $\sigma_{\max} = 2$, $N_0 = 25$.

3.4 算法流程图

整个跟踪算法的具体流程, 如表 1 所示.

表 1 算法流程图

<p>对于每台摄像机 $C_i (i = 1, 2, \dots, N)$ 而言, 初始化和跟踪过程是相同的, 具体流程如下:</p> <p>初始化过程</p> <ol style="list-style-type: none"> (1) 手工指定^[10]摄像机视野中跟踪目标的区域, 并根据文献^[13]中的方法得到外观模型 q_i^*. (2) 根据文献^[12]中的方法, 计算得出其它摄像机与本摄像机间的单应性矩阵 $\mathbf{H}_{j,i} (j = 1, 2, \dots, N \text{ 且 } j \neq i)$. (3) 启动跟踪过程 <p>跟踪过程</p> <ol style="list-style-type: none"> (1) 采集当前时刻的图像数据 $\mathbf{y}_{t,i}$, 并通过网络, 接收其它摄像机前一时刻的跟踪结果 $\{\hat{\mathbf{x}}_{t-1,j}, s_{t-1,j}\} (j = 1, 2, \dots, N \text{ 且 } j \neq i)$. (2) 利用式(10)~式(16), 生成当前时刻的粒子 $\mathbf{x}_{t,i}^{(m)}$. (3) 利用式(7), 更新粒子的权重 $w_{t,i}^{(m)}$. (4) 利用式(4), 计算得出当前时刻目标在摄像机 C_i 视野中的位置 $\hat{\mathbf{x}}_{t,i}$. 同时, 根据式(9), 计算出估计目标区域与目标外观模型间的相似度 $s_{t,i}$. (5) 通过网络, 将当前时刻的跟踪结果 $\{\hat{\mathbf{x}}_{t,i}, s_{t,i}\}$ 发送到其它摄像机, 并重复步骤 1.

4 实验结果

4.1 实验条件

我们使用 12 组多视角拍摄的视频序列进行了实验,均获得了良好的效果.包括 <http://peipa.essex.ac.uk/ipa/pix/pets/PETS2001/>和 <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>网站上提供的双视角视频序列,及 <http://www.cvg.rdg.ac.uk/PETS2009/a.html> 网站上提供的三视角视频序列.由于篇幅所限,只给出三组最具挑战性的视频序列的跟踪结果.

为了更加深入了解算法的性能,本文还实现了两种对比算法.

对比算法 1: 每台摄像机只利用本地采集得到的数据对跟踪目标位置进行估计,而不进行数据融合,即:

$$\hat{\mathbf{x}}_{t,i} = \int \mathbf{x}_{t,i} p(\mathbf{x}_{t,i} | \mathbf{y}_{1:t,i}) d\mathbf{x}_{t,i} \quad (17)$$

通过将本文提出的分布式算法与对比算法 1 进行比较,可以清晰地了解有效数据融合对于算法性能的提高.

对比算法 2: 采用文献[10]提出的算法,该算法同样采用了单应性矩阵作为几何映射关系,而且有一定的分布式特性.通过比较,可以更好地对本文算法的精

度和鲁棒性进行评估.

4.2 抗遮挡性能实验

我们采用文献[15]中提出的 *OTP* (Object Tracking Precision) 参数来定量地衡量算法的跟踪精度.其中, t 时刻跟踪系统的 *OTP* 表示为 OTP_t , 计算公式如下:

$$OTP_t = \frac{1}{N} \sum_{i=1}^N \frac{A_{t,i}^{true} \cap A_{t,i}^{est}}{A_{t,i}^{true} \cup A_{t,i}^{est}} \quad (18)$$

其中, N 表示系统中摄像机的数目, $A_{t,i}^{true}$ 表示 t 时刻目标在第 i 台摄像机视野中的真实区域; $A_{t,i}^{est}$ 则表示由算法估计得出的目标区域. $A_{t,i}^{true} \cap A_{t,i}^{est}$ 表示了两个区域的交集部分的面积, $A_{t,i}^{true} \cup A_{t,i}^{est}$ 表示两个区域的并集部分的面积. OTP_t 计算估计区域与真实区域的平均重合度, 表征了 t 时刻系统中各视角跟踪精度的平均值. 算法的精度越高, 则两个区域的交集面积越大, 并集面积愈小, OTP_t 越大; 反之, 算法的精度越低, 则两个区域的交集面积越小, 并集面积愈大, OTP_t 越小.

图 3 显示了本文算法与对比算法在三组测试序列的处理精度结果. 在 3 组测试序列中本文算法的跟踪精度均是最高. 下面将结合具体的测试序列对图 3 结果做进一步讨论.

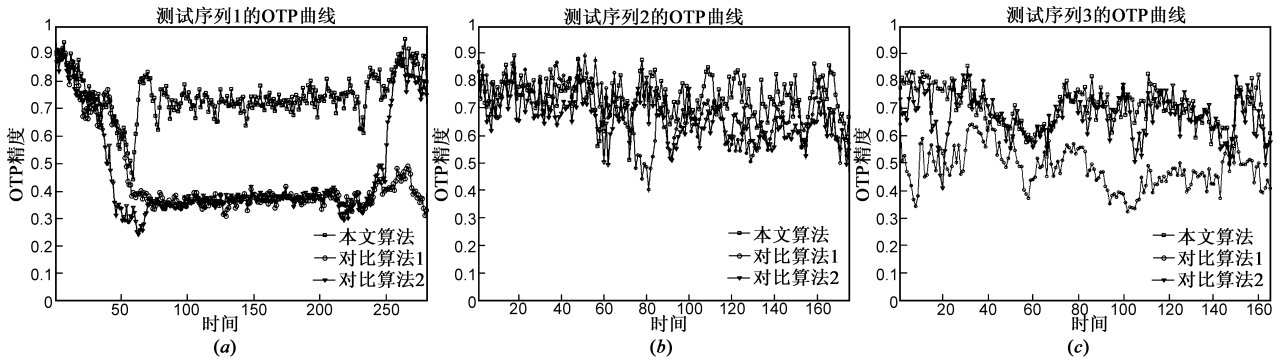


图3 各跟踪算法的跟踪精度曲线图

4.2.1 测试序列 1 该组测试序列包含两个视角,跟踪目标为绿色轿车.两个视角中跟踪目标的外观在不断变化.在视角 2 中,跟踪目标两次被静态遮挡物(即大树)完全遮挡(如图 4(b₂)),且遮挡物与跟踪目标的外观极为相近,给跟踪算法提出了挑战.图 4 中给出了本文算法的跟踪结果,图 4(a₁)~(a₄)、图 4(b₁)~(b₄)分

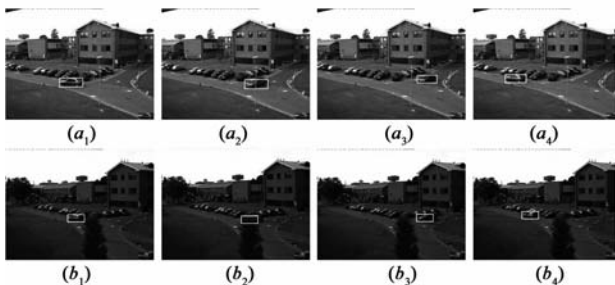


图4 本文算法在测试序列1上的跟踪结果

别对应于两个视角的初始帧和第 58、200、280 帧的图像,每幅图中的白色方框表示了目标所在区域.从图 3(a)中可知,本文算法的跟踪精度远高于对比算法.其原因在于,对比算法 1 只能利用本视角的图像数据,难以处理完全遮挡的情况;对比算法 2 通过对多个目标进行跟踪,可以处理跟踪目标间相互遮挡的情况.但是,在本测试序列中,遮挡物为静止的大树,算法无法对遮挡情况进行分析,所以跟踪精度会大幅度下降.

4.2.2 测试序列 2 图 5 显示了本文算法对第 2 组视频序列的跟踪结果,图 5(a₁)~(a₄)和(b₁)~(b₄)分别对应于两个视角在初始帧和第 77、95 和 167 帧时的跟踪结果.第 2 组视频序列的挑战性最大:在视角 1 和 2 中,跟踪目标同时被其它物体所遮挡.如相同时刻的图 5(a₃)和(b₃)图像,图 5(a₃)中目标被大面积遮挡(超过 50%),只有头部和肩部是可视的;而图 5(b₃)中目标被

完全遮挡.从图 3(b)中可以知道,本文算法的跟踪精度高于对比算法.原因在于,本文算法借助于自适应观测模型,能够有效处理局部遮挡的情况,可在视角 1 中对目标进行精确跟踪;同时,视角 2 利用视角 1 的跟踪结果进行有效融合,可提高目标在完全遮挡状态下的跟踪精度.而对比算法 1 无法处理目标被完全遮挡的情况,对比算法 2 无法处理两视角同时遮挡的情况,所以跟踪精度均较低.

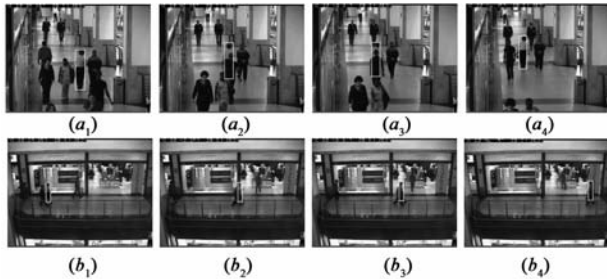


图5 本文算法在测试序列2上的跟踪结果

4.2.3 测试序列 3 在该测试序列包含三个视角,跟踪目标为身着黑色上衣的男人.图 6 中给出了本文算法的跟踪结果,图 6(a₁)~(a₄),(b₁)~(b₄),(c₁)~(c₄)分别对应于三个视角的初始、第 87、144 和 159 帧的图像.在跟踪过程中,跟踪目标不断被其它物体所遮挡,给跟踪造成了较大的困难,如图 6(a₃)中的电线杆、(b₂)和(c₃)图中的人.由图 3(c)中可知,对比算法 1 受到目标遮挡的影响较大,性能是最差的;对比算法 2 可以进行持续跟踪,但跟踪精度仍低于本文算法.

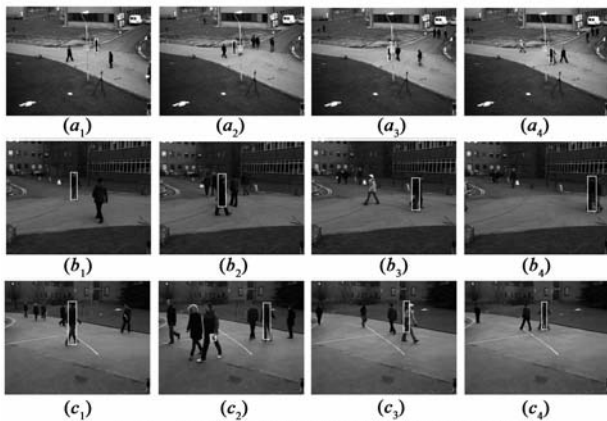


图6 本文算法在测试序列3上的跟踪结果

4.3 抗故障鲁棒性实验

在本实验中,使用序列 3 评估算法的抗故障性能.测试方法是三个视角的摄像机同时开始跟踪,在跟踪过程中的任意时刻,指定 1 或 2 台摄像机节点发生故障,发生故障的时刻和节点都是随机指定的.发生故障的节点不能继续进行跟踪,也不能向其它节点传递结果.我们使用 OTP_t^b 精度衡量故障对于算法的影响,其计算方法如下:

$$OTP_t^b = \frac{1}{\sum_{i=1}^3 B_{ii}} \sum_{i=1}^3 B_{ii} \frac{A_{t,i}^{true} \cap A_{t,i}^{est}}{A_{t,i}^{true} \cup A_{t,i}^{est}} \quad (19)$$

其中, B_{ii} 表示 t 时刻摄像机 C_i 是否发生故障, $B_{ii} = 1$ 为无故障; $B_{ii} = 0$ 表示发生故障. OTP_t^b 表征了系统中某些节点发生故障时,剩余节点平均的跟踪精度.为了消除随机指定故障时刻和节点对结果造成的影响,对于每种故障模式,均重复 100 次实验,取平均值作为最终的 OTP_t^b 精度.

图 7 中显示了各种故障模式下的 OTP_t^b 精度曲线.观察图 7 可知,在本文算法中,任意一台的摄像机节点发生故障时,其它摄像机仍可正常工作,且系统的跟踪精度仅略低于无故障模式.但整个跟踪系统的平均跟踪精度将随着故障机器数目的增加而下降.

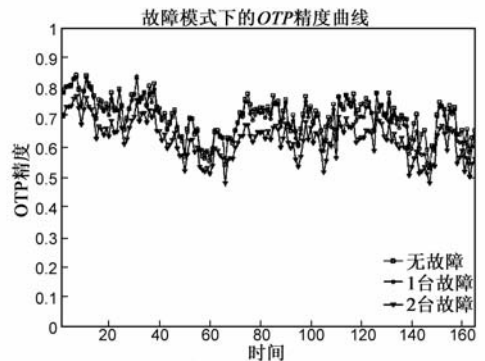


图7 各种故障模式下的 OTP 精度曲线

5 结论

本文借助于贝叶斯理论和粒子滤波器,针对多视角目标跟踪问题建立了分布式融合的概率框架;并在此框架下进一步建立了自适应的观测模型和状态转移模型.主要优点有三点:(1)在遮挡处理方面,自适应观测模型和状态转移模型能够避免遮挡所造成的误差传递,提高了跟踪算法的鲁棒性,且无需复杂的遮挡分析过程;(2)各摄像机节点只需要融合其它摄像机前一时刻的目标位置估计结果,而无需原始的图像数据,可以极大程度降低通信开销;(3)任意摄像机节点出现故障时,其它摄像机仍可继续进行目标跟踪.本文算法可以广泛应用在视频监控、交通管理等领域中.下一步,我们将把算法移植到硬件平台上,实现实时跟踪,以提高算法的应用价值.

参考文献:

- [1] 常发亮,马丽,刘增晓,乔谊正.复杂环境下基于自适应粒子滤波器的目标跟踪[J].电子学报,2006,34(12):2150-2153.

Chang FaLiang, Ma Li, Liu Zengxiao, Qiao Yizheng. Target tracking based on adaptive particle filter under complex back-

- ground[J]. *Acta Electronica Sinica*, 2006, 34(12): 2150 – 2153. (in Chinese)
- [2] 周妍, 胡波, 张建秋. 基于粒子滤波器和风险决策跟踪遮挡目标的方法[J]. *电子学报*, 2007, 35(2): 350 – 353.
Zhou Y, Hu B, Zhang Jianqiu. Tracking occlusive target method based on Bayesian decision theory and particle filter[J]. *Acta Electronica Sinica*, 2007, 35(2): 350 – 353. (in Chinese)
- [3] Rinner B, Wolf W. An introduction to distributed smart cameras [J]. *Proceedings of the IEEE*, 2008, 96(10): 1565 – 1575.
- [4] Sankaranarayanan A C, Veeraraghavan A, Chellappa R. Object detection, tracking and recognition for multiple smart cameras [J]. *Proceedings of the IEEE*, 2008, 96(10): 1606 – 1624.
- [5] Chang T H, Gong S. Tracking multiple people with a multi-camera system[A]. *Proceedings of IEEE Workshop on Multi-Object Tracking*[C]. Vancouver, 2001. 19 – 26.
- [6] Fleuret F, Berclaz J, Lengagne R et al. Multicamera people tracking with a probabilistic occupancy map[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 267 – 282.
- [7] Qin C, Jade A. Tracking human motion in structured environments using a distributed-camera system[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(11): 1241 – 1247.
- [8] Wei D, Jean-Bernard H, Jacques V, et al. Ground-target tracking in multiple cameras using collaborative particle filters and principal axis-based integration[J]. *IPSN Transactions on Computer Vision and Applications*, 2009, 1: 58 – 71.
- [9] Xue W, Sheng W, Daowei B. Distributed visual-target-surveillance system in wireless sensor networks[J]. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 2009, 39(5): 1134 – 1146.
- [10] Lien Kuo-Chin, Huang Chung-Lin. Multiview-based cooperative tracking of multiple human objects[J]. *EURASIP Journal on Image and Video Processing*, 2008, 2: 1 – 13.
- [11] Christopher M B. *Pattern Recognition and Machine Learning* [M]. New York: Springer. 359 – 383.
- [12] Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision* [M]. Cambridge: Cambridge University Press, 2003. 325 – 343.
- [13] Patrick P, Carine H, Jaco V, Michel G. Color-based probabilistic tracking[A]. *Proceedings of the 7th European Conference on Computer Vision* [C]. London: Springer-Verlag, 2002. 661 – 675.
- [14] M Sanjeev A, Simon M, Neil G, et al. A tutorial on particle filters for online nonlinear non-Gaussian Bayesian tracking [J]. *IEEE Transactions on Signal Processing*, 2002, 50(2): 174 – 187.
- [15] R Kasturi, D Goldgof, P Soundararajan, et al. Framework for performance evaluation of face, text, and vehicle detection and tracking in video; Data, metrics, and protocol[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 319 – 336.
- [16] Xu Xinyu, Li Baoxin. Adaptive Rao-Blackwellized Particle Filter and Its Evaluation for Tracking in Surveillance [J]. *IEEE Transactions on Image Processing*, 2007, 16(3): 838 – 849.

作者简介:



冯 巍 男, 1983 年生于辽宁沈阳, 现为复旦大学电子工程系博士研究生, 主要研究方向为视频跟踪.

E-mail: weifeng@fudan.edu.cn



胡 波 男, 1968 年生于江苏常州, 复旦大学教授、博士生导师, 主要研究方向为数字信息处理、数字通信及数字系统设计.

E-mail: bohu@fudan.edu.cn

杨 成 男, 1982 年生于江苏兴化, 现为复旦大学电子工程系博士研究生, 主要研究方向为信号与图像处理.

林 青 男, 1977 年生于山东, 复旦大学讲师, 主要研究方向为图像处理和目标跟踪.

杨 涛 男, 1970 年生于陕西汉中, 复旦大学副教授, 主要从事无线通信理论及信号处理、认知无线电频谱感知等方面的研究.