基于 GMM-RBMCDA 的实时监控多目标跟踪方法*

张培尼，穆志纯
（北京科技大学 自动化学院，北京 100083）

摘　要： 为了有效提取视频监控场景中的前景目标信息并准确跟踪目标的状态，提出一种基于混合高斯模型和 Rao-Blackwellized 蒙特卡洛数据关联的视频多目标跟踪方法。该方法根据场景中像素点的特征信息，利用混合高斯模型进行建模，并对前景目标进行检测，使用 Rao-Blackwellized 蒙特卡洛数据关联算法来降低可能的目标交叉及杂波干扰带来的影响，通过设置目标存在和消失参数，实现了实时多目标跟踪。实验结果表明，该方法不仅能对场景中未知目标的个数进行有效估计，而且可以准确地跟踪目标的状态，取得了良好的实际效果。

关键词： 实时监控；多目标跟踪；混合高斯模型；蒙特卡洛数据关联
中图分类号： TP391.41　　**文献标志码：** A　　**文章编号：** 1001-3695(2012)11-4345-05
doi:10.3969/j.issn.1001-3695.2012.11.089

Real-time multi-object tracking method based on GMM-RBMCDA

ZHANG Pei-ni, MU Zhi-chun
(*School of Automation & Electrical Engineering*, *University of Science & Technology Beijing*, *Beijing* 100083, *China*)

Abstract: To extract the foreground objects and track the motions of these moning objects in the video surveillance scenes, this paper presented a GMM and Rao-Blackwellized Monte Carlo data association (GMM-RBMCDA) based on real-time multi-object tracking methood. The algorithm subtracted background adaptively by modeling each pixel as a mixture of Gaussians. And it used the RBMCDA algorithm to deal with the problem of trajectory-crossing and the background noise. Then it realized the real-time multi-object tracking by setting the parameters of objects appearance and disappearance. The experimental results show that the presented method can not only obtain the number of the objects in the video scenes, but also estimate the states of objects more accurately. It also proved the effectiveness of the algorithm.

Key words: real-time surveillance; multi-object tracking; Gaussian mixture model(GMM); Monte Carlo data association

0　引言

视频监控技术在交通管理、楼宇安防、银行服务、医疗监护等众多领域有着广泛的应用。其中，运动目标的检测和跟踪是视频监控系统的重要组成部分，也是目前的一个研究热点。近些年，人们对不同场景下的多目标跟踪进行了深入的研究，融合目标特征提取和跟踪滤波技术，提出了许多解决方案。

为了从复杂的场景中提取感兴趣的目标，目前常用的方法有光流法、帧间差分法和背景减除法。其中：光流法的计算复杂，对硬件要求较高，不适于对场景中的目标进行实时检测；帧间差分法采用相邻帧的差分来检测运动目标，实时性较强，但很难检测出像素变化不明显的点，不适合相对复杂的场景；背景减除法是目前常用的运动目标检测算法之一，其操作简单、实时性较好，一般能够提供较为完整的运动目标信息。文献[1]提出用单个高斯模型对场景建模，但其场景为室内简单的特定场景，因此其模型通常难以满足其他一般的应用场景。文献[2,3]提出的混合高斯背景模型(GMM)，利用混合多个高斯模型对每个像素点的数据序列建模，它能够克服光照变化或树枝摇摆等影响。GMM 是目前应用较为广泛的模型，很多研究者对此模型进行了诸多改进[4]，并应用于复杂背景下运动目标的检测与跟踪。

在提取每帧图像中的目标特征信息后，需要在后继帧中计算出与该特征最匹配的目标位置，并实时估计出目标的运动状态。为此，许多研究者对基于滤波的目标跟踪方法进行了研究。文献[5]假设目标的运动方程和观测方程是线性高斯过程，其后验分布可以由均值和协方差矩阵确定，故直接采用 Kalman 滤波或扩展的 Kalman 滤波进行跟踪。但实际场景通常会受到目标的复杂运动、光照变化以及图像中噪声的影响，这种线性高斯假设往往不成立，因此直接使用 Kalman 滤波或扩展的 Kalman 滤波的跟踪方法容易失败。针对场景中存在的非线性、非高斯、多模态的问题，基于采样原理的粒子滤波(particle filter)方法在视频跟踪领域得到了广泛关注[6]。其核心是蒙特卡洛(Monte Carlo)模拟算法，系统状态的后验密度函数由一组带有权重的离散粒子来表达。文献[7~9]使用了粒子滤波，在单目标跟踪领域，粒子滤波的实效性问题并不突出，但对于多目标跟踪，粒子个数会随着目标数增多呈指数性增长，计算量大，时间开销大。针对此问题，文献[10]提出的基于粒子滤波的多目标跟踪算法，是在跟踪过程中进行了粒子和图像空间的映射，能够对新进入区域和离开区域的目标粒子给予及时建立和删除，可以有效地解决车辆交叉时粒子滤波联合

收稿日期： 2012-04-23；**修回日期：** 2012-05-30　**基金项目：** 国家自然科学基金资助项目(60973064)；北京市自然科学基金资助项目(4102039)；北京市重点学科建设项目(XK100080537)

作者简介： 张培尼(1982-)，女，河南焦作人，博士研究生，主要研究方向为计算机视觉、目标跟踪(zpeini@gmail.com)；穆志纯(1952-)，男，教授，博导，主要研究方向为生物特征识别、图像处理.

分布计算量过大的问题,但该方法的实时性较差。因此,人们研究了优化粒子滤波性能的方法。文献[11]提出了基于马尔可夫(Markov)链蒙特卡洛(MCMC)扩展 Kalman 粒子滤波的目标跟踪算法,利用扩展卡尔曼滤波构造粒子滤波的建议分布函数,使建议分布函数融入最新的观测信息,以便得到更符合真实状态的后验概率分布;并且同时引入 MCMC 方法来解决粒子贫化问题,但由于增加了这一步,计算量较大,实时性相对于经典的粒子滤波较弱。对于优化粒子滤波性能的方法,其中 RBPF(Rao-Blackwellized particle filter)[12]方法根据 Rao-Blackwell 原理,将能够采用系统分析方法处理的环节置于粒子滤波的框架之外(通常采用 Kalman 滤波或扩展 Kalman 滤波),因而减少了用蒙特卡洛采样处理的环节,这无疑会提高跟踪的精度。与此同时,所需要粒子数量也会大大减少,从而节省了计算时间。

本文提出一种基于混合高斯模型和 Rao-Blackwellized 蒙特卡洛数据关联(GMM-RBMCDA)的视频多目标跟踪方法。

1 混合高斯背景模型

运动目标检测是指从视频序列中提取运动目标(与背景存在相对运动的前景目标),其相关数据能够直接影响视频监控和行为分析等后续研究。背景减除法是目前常用的运动目标检测算法之一,此方法操作简单、实时性较好,一般能够提供较为完全的运动目标信息。其中,由 Stauffer 等人提出的 GMM 是目前研究应用比较广泛的方法之一,其核心思想是通过统计像素亮度的变化情况描述多个背景模式,并采用基于回归滤波的在线 EM 近似方法更新该模型的参数。它能够鲁棒地克服光照变化、树枝摇动等造成的影响。

1.1 GMM 的建立

Stauffer 等人提出的方法中,像素点(x_0,y_0)的取值集合为$\{X_1,X_2,\cdots,X_t\}=\{I(x_0,y_0,i)\mid 1\leqslant i\leqslant t\}$,其中 t 为取值时刻,I 为视频序列帧。如果将该像素点的所有历史值用 K 个高斯分布来近似,那么所观察到的当前像素值的概率可表示为

$$P(X_t)=\sum_{i=1}^{K}\omega_{i,t}\times\eta(X_i,\mu_{i,t},\Sigma_{i,t}) \tag{1}$$

其中:X_t 是 t 时刻该点的像素值;K 是混合高斯分布的个数;$\omega_{i,t}$是 t 时刻混合高斯模型中第 i 个高斯模型的权值,且满足条件 $0\leqslant\omega_{i,t}\leqslant 1$,且 $\sum_{i=1}^{K}\omega_{i,t}=1$;$\eta(X_t,\mu_{i,t},\Sigma_{i,t})$是 t 时刻第 i $(i=1,\cdots,K)$个高斯分布,其定义为

$$\eta(X_t,\mu_{i,t},\Sigma_{i,t})=\frac{1}{(2\pi)^{\frac{n}{2}}\mid\Sigma_{i,t}\mid^{\frac{1}{2}}}\times$$
$$\exp\left(-\frac{1}{2}(X_t-\mu_{it})^{\mathrm{T}}\Sigma_{i,t}^{-1}(X_t-\mu_{it})\right) \tag{2}$$

其中:n 是 X_t 的维数;$\mu_{i,t}$是 t 时刻第 i 个高斯模型的均值向量;$\Sigma_{i,t}$为 t 时刻第 i 个高斯模型的协方差矩阵。若满足:

$$\mid X_t-\mu_{i,t-1}\mid\leqslant 2.5\sigma_{i,t-1} \tag{3}$$

则此像素与模型匹配。

1.2 利用 GMM 进行目标检测

1)更新模型参数

$$\begin{cases}\omega_{i,t}=(1-\alpha)\omega_{i,t-1}+\alpha\\ \mu_{i,t}=(1-\rho)\cdot\mu_{i,t-1}+\rho X_t\\ \sigma_{i,t}^2=(1-\rho)\sigma_{i,t-1}^2+\rho(X_t-\mu_{i,t})^{\mathrm{T}}(X_t-\mu_{i,t})\end{cases} \tag{4}$$

其中:ρ 为参数更新率。式(4)表明只有与 X_t 相匹配的高斯分布权值才得到提高,否则通过下式来降低权值:

$$\omega_{i,t}=(1-\alpha)\omega_{i,t-1} \tag{5}$$

因此,可以保证当像素与模型匹配时,其模型权值增大,若像素与模型不匹配,其模型权值变小。

2)进行背景估计

将 K 个高斯模型按照 ω/σ 优先级从大到小排序,满足下式的条件:

$$B=\underset{b}{\operatorname{argmin}}\left(\sum_{i=1}^{b}\omega_{i,t}>T\right) \tag{6}$$

其中:T 为权重阈值。

3)前景分割及目标检测

若满足式(6),那么将前 B 个模型作为背景模型。由此算法的过程可以看出,算法对每个像素用 K 个高斯分布进行建模,并认为当前帧内的像素点在上一帧中为背景像素点,经过算法处理后会将扰动匹配给权值小的高斯分布,经过多帧之后归为背景像素点,这样就可以有效地将一些扰动归为背景,有利于检测出兴趣目标。混合高斯背景模型能够适应场景的缓慢变化,被认为是最好的背景模型之一。本文目标检测的方法就是利用混合高斯背景建模方法进行检测。

2 RBMCDA

2.1 粒子滤波与 Rao-Blackwellized 策略

粒子滤波(particle filter)是一种重要的用于非高斯非线性估计的重要方法,通过赋予每个粒子不同的权重,来替代概率分布函数的积分运算,并通过重采样来剔除一些权重较低的粒子,而复制一些权重较高的粒子[6]。考虑如下的状态空间模型:

$$x_k\sim p(x_k\mid x_{k-1}) \tag{7}$$

$$y_k\sim p(y_k\mid x_k) \tag{8}$$

在第 k 个时间步骤,通过可以直接获得的测量 y_k 来估计状态变量 x_k。根据贝叶斯滤波的原理,需要用重要性分布函数来近似地表示状态的后验分布:

$$p(x_k\mid x_{k-1},y_{1:k}) \tag{9}$$

通常所用到的粒子滤波(bootstrap filter)直接用系统的动态方程 $p(x_k\mid x_{k-1})$来近似代替重要性分布函数,通过重采样策略来应对粒子的衰退现象。这无疑在操作上带来了极大的方便。但为了更好地逼近概率密度函数,需要大量的粒子来进行计算,实时性问题依然是阻碍其应用的最大障碍。

为此,一种被称为 Rao-Blackwellized 粒子滤波(RBPF)的方法被提出用来降低估计的方差[12]。这种方法根据Rao-Blackwell 原理,将可以用系统分析方法处理的环节置于粒子滤波的框架之外,因而减少了用蒙特卡洛采样处理的环节,这无疑会提高算法的精度。与此同时,为达到同样效果,所需要的粒子数量也会大大减少,从而节省了计算时间。具体到目标跟踪领域,就是仅仅使用粒子滤波处理数据关联问题,而目标状态的估计则使用 Kalman 滤波(KF),或用于非线性问题的扩展 Kalman 滤波(EKF)以及无味 Kalman 滤波(UKF)。

2.2 RBMCDA 算法

对于 RBPF,实质上是求解受一个隐含变量 λ_k 影响的条件线性高斯系统。在 Rao-Blackwellized 蒙特卡洛数据关联(RBMCDA)算法中,是将 λ_k 定义为数据关联指示标志(data associ-

ation indicator),多目标跟踪问题则转换为数据关联条件下的目标状态估计问题[12]。当跟踪目标个数已知时,假设为 $T(j=1,\cdots,T)$,杂波测量可以假设均匀分布在整个测量空间(体积为 V):

$$p(y_k|\lambda_k=0)=1/V \tag{10}$$

目标测量

$$p(y_k|x_{k,j},\lambda_k=j) \tag{11}$$

目标运动方程

$$p(x_{k,j}|x_{k-1,j})=N(x_{k,j}|A_{k-1,j}x_{k-1,j},Q_{k-1,j}) \tag{12}$$

其中:$A_{k-1,j}$ 和 $Q_{k-1,j}$ 分别为状态转移矩阵和过程噪声协方差矩阵,$N(\cdot)$ 为高斯分布。目标和杂波先验可以用一个 m 阶的 Markov 链表示[12]:

$$p(c_k|c_{k-1:k-m}) \tag{13}$$

假设选取的粒子数量为 $N_p(i=1,\cdots,N_p)$,目标状态初始值为

$$p(x_{0,j})=\sum_i w^{(i)}N(x_{0,j}|m_{0,j}^{(i)},P_{0,j}^{(i)}) \tag{14}$$

每个粒子 i 在时间步骤 k 时的状态可以表示为

$$\{\lambda_k^{(i)},m_{k,1}^{(i)},\cdots,m_{k,j}^{(i)},\cdots,m_{k,T}^{(i)},P_{k,1}^{(i)},\cdots,P_{k,j}^{(i)},\cdots,P_{k,T}^{(i)},w_k^{(i)}\} \tag{15}$$

其中:$m_{k,j}^{(i)}$ 和 $P_{k,j}^{(i)}$ 分别是目标 j 的均值和方差,它们是独立于数据关联历史 $\lambda_{1:k}^{(i)}$ 的;$w_k^{(i)}$ 是粒子的权重,根据关联的后验概率来进行更新。测量似然函数为

$$p(y_k|\lambda_k,y_{1:k-1},\lambda_{1:k-1}^{(i)})p(\lambda_k|\lambda_{k-1:k-m}^{(i)}) \tag{16}$$

最优重要性分布(后验分布):

$$p(\lambda_k|y_{1:k},\lambda_{1:k-1}^{(i)})\propto p(y_k|\lambda_k,y_{1:k-1},\lambda_{1:k-1}^{(i)})p(\lambda_k|\lambda_{k-1:k-m}^{(i)}) \tag{17}$$

计算 λ_k 的后验分布:

$$\hat{\varphi}^{(i)}=p(y_k|\lambda_k^{(i)}=0,y_{1:k-1},\lambda_{1:k-1}^{(i)})\times p(\lambda_k=0|\lambda_{k-1:k-m}^{(i)}) \tag{18}$$

$$\hat{\varphi}_j^{(i)}=p(y_k|\lambda_k^{(i)}=j,y_{1:k-1},\lambda_{1:k-1}^{(i)})\times p(\lambda_k=j|\lambda_{k-1:k-m}^{(i)}) \tag{19}$$

$$\varphi_j^{(i)}=\frac{\hat{\varphi}_j^{(i)}}{\sum_j^T\hat{\varphi}_j^{(i)}}\qquad j=0,\cdots,T \tag{20}$$

数据关联指示 $\lambda_k^{(i)}(\lambda_k^{(i)}=j)$ 按照概率 $\varphi_j^{(i)}$ 进行采样。

当目标个数未知且随时间变化时,需要建立变化的目标状态空间维数,可以将其视为多假设跟踪(MHT)方法的一种延伸。假设存在一个足够大的固定维数目标个数空间 T_∞,此时,为描述目标的出现和消失(实际为可见和非可见)的概率随机过程模型,需要重新定义 $\lambda_k=\{e_k,c_k\}$,其中 e_k 为可见性指示标志(visibility indicator)。为避免混淆,此处用 c_k 表示数据关联指示变量。此时,数据关联模型为

$$p(c_k|c_{k-1:k-m},e_k) \tag{21}$$

则联合 Markov 链模型:

$$p(e_k,c_k|c_{k-1:k-m},e_{k-1:k-m})=p(c_k|c_{k-1:k-m},e_k)p(e_k|e_{k-1}) \tag{22}$$

其中:$p(e_k|e_{k-1})$ 为生灭(目标出现或消失)过程的先验概率分布模型。可以将可见和非可见目标分为两个集合 ψ_0 和 ψ_1,则所有目标的联合先验分布可以表示为[12]

$$p(x_k|y_{1:k})=\prod_{j\in\psi_1}N(x_{k,j}|m_{k,j},P_{k,j})\times\prod_{j_0\in\psi_0}N(x_{k,j_0}|m_0,P_0) \tag{23}$$

每个粒子 i 在时间步骤 k 时的状态可以表示为

$$\{\lambda_{k-m+1:k}^{(i)},m_{k,1}^{(i)},\cdots,m_{k,j}^{(i)},\cdots,m_{k,T}^{(i)},P_{k,1}^{(i)},\cdots,P_{k,j}^{(i)},\cdots,P_{k,T}^{(i)},w_k^{(i)}\} \tag{24}$$

实际上,在每个时间步骤 k,每个粒子的状态变量中还包含着目标的个数 $T_k^{(i)}$,有测量与目标 j 最后进行数据关联时刻 $\tau_{k,j}^{(i)}$,表示一个目标出现的唯一指示标志 $\kappa_{k,j}^{(i)}$。还需要注意的是,在处理未知目标个数的 RBMCDA 算法中,由于每个目标的初始先验假设为条件独立于 c_k 和 e_k 的,因此在跟踪中仍然保持着相应的独立性。

3　GMM-RBMCDA 视频多目标跟踪

在视频监控中,主要是通过跟踪多个目标的运动,为运动行为分析和理解提供支撑。实际场景中目标个数未知且运动方式多样,还经常伴有外界的临时干扰(如树叶飘落和光线变化等)。因此,将以上介绍的 GMM 方法与 RBMCDA 相结合,有利于更好地处理此类问题。由于场景中的杂波干扰很难在图像处理时被加以区分,RBMCDA 提供了一种处理杂波条件下未知目标个数的跟踪方法,适合目标机动性较小的场合,可以在处理过程中直接应用 Kalman 滤波来预测和估计近匀速运动的状态。当目标运动机动性较强时,则可以考虑加入解决跳变随机过程的交互多模型(IMM)方法。

3.1　基本思想

根据来自一个静态摄像机的图像序列,需要利用 GMM 为背景建立一个合适的描述并检测轮廓。另外,通过对背景模型的连续更新来适应背景光线的逐渐变化过程以及场景中固定物体的变化,这类似于一种在线的 K-means 算法。首先辨别背景中的前景运动目标,对每幅图像中的每一个像素点建立典型值(typical values)的分布,通过参数化的分布函数,可以用较为简洁的方式寻找传感器噪声和重复的背景运动。利用累计背景模型可以算出一个像素点的来源,也就是说如果一个像素点的值不符合背景的分布,则被认为是前景目标;然后按照目标帧间的对应关系,利用 RBMCDA 算法对测量和目标的对应关系进行区分。这个过程无须对目标的个数进行估计,也无须像均值漂移粒子滤波(MSPF)算法[13]那样计算目标之间的 BH 系数来确定最佳配比位置,关联跟踪过程都是自动完成的。由于所使用的最优重要性分布函数(数据关联指示器后验分布)使用了最新的观测信息,并且结合了关联先验和似然函数,因此数据关联的结果将更加准确。对于背景中的连续扰动物体(如风吹树叶、往复运动的物体等)很容易被错误地认做前景目标,这将给跟踪带来麻烦。因此,在使用 RBMCDA 作数据关联时,要充分考虑到局部连续杂波干扰的处理。算法的流程如图 1 所示。

3.2　混合高斯模型目标检测

在进行视频场景中运动物体的跟踪之前,首先利用混合高斯模型检测是否存在运动物体,如果存在则确定被跟踪物体的位置和轮廓。由于这种方法不是仅使用一种特定的模型来为每个像素点建模,而是利用每个高斯模型的持续性和可调性来选择合适的高斯模型组合来处理,处理过程需要用到学习率 α 和背景需要的数据比例 τ。由于背景匹配模型的选取具有很好的灵活性,本方法能够很好地适应光照变化、场景中重复运动的物体干扰、杂波区域的干扰以及固定物体的增加和减少。

然而,由于颜色相对于背景有着明显的变化,慢速运动的物体不会被作为背景而忽略掉。需要说明的是,检测出来的前景像素需要使用文献[14]中的算法进行分组,根据位置、大小、运动方式以及形状等信息确定属于同一个前景目标的信息结合体(connected components)。利用混合高斯模型进行目标检测的流程如图 2 所示。

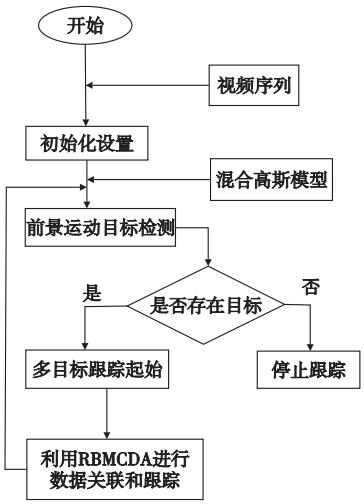

图1 基于GMM-RBMCDA的视频多目标跟踪流程

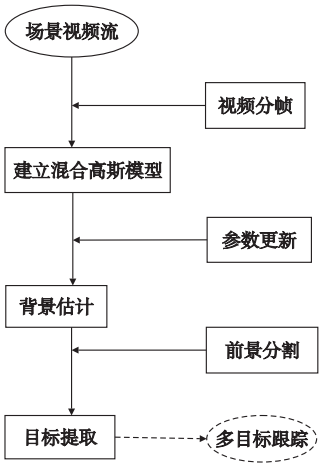

图2 利用混合高斯模型进行目标检测的流程

3.3 基于 RBMCDA 的多目标跟踪

多目标跟踪是指通过目标前一时刻(前一帧)前景目标的状态信息,来预测当前时刻(当前帧)的状态,并结合新得到的测量信息(新的视频序列中提取的前景目标)来进行更为准确的估计。这个过程是使用贝叶斯滤波原理(Kalman 滤波)来实现的。其主要困难在于无法获取测量的来源,特别是在杂波的干扰下,将测量信息分配给正确的目标变得非常困难,数据关联起着至关重要的作用。当被跟踪目标个数未知时,情况将变得更加复杂。

本文中使用 RBMCDA 算法来处理数据关联和跟踪的问题,对每一个粒子中每个目标测量的更新过程依然使用 Kalman 滤波。如果场景中目标的运动都可以被认为是近匀速(nearly constant velocity)运动,针对线性高斯最优估计的 Kalman 滤波器是能够胜任的。而在粒子滤波处理数据关联问题时,通过计算每个目标的边际似然函数,依然可以用最优重要性分布的采样来得到数据关联变量的取值。需要注意的是,在连续两帧前景目标之间建立联系时,需要合理地定义目标的出现和消失控制参数。本文把用一个测量和新目标建立关联事件定义为目标出现,在此之前,目标的状态都被定义为一个常值。在此之后,目标的存在时间为 $t_d \sim p(t_d)$(服从已知的指数或伽马分布),即如果与目标 j 最后一次关联时间为 $\tau_{k,j}$,且在此之前目标出现时的采样时刻为 t_{k-1},则目标在此时刻消失的概率 P_{dis} 为[12]

$$P_{\mathrm{dis}} = P(t_d \in [t_{k-1} - \tau_{k,j}, t_k - \tau_{k,j}] \mid t_d \geqslant t_{k-1} - \tau_{k,j}) \tag{25}$$

这样可以限制存在的目标数量,尤其是在噪声干扰比较大的场合,可以有效地消除随机建立的干扰前景联系。

3.4 算法实现

所提出的算法主要包括参数初始化、GMM 目标运动检测和 RBMCDA 目标跟踪三个过程。

1)初始化参数(混合高斯模型)

预先定义高斯模型,对高斯模型中的控制变量和参数进行初始化。主要包括:

a)K 为选取的高斯模型个数,设定值为 4;

b)α 为学习率(决定高斯分布权重的更新幅度),设定值为 0.015;

c)ρ 为高斯分布方差的学习率,设定值为 0.02;

d)T 为权重阈值(决定背景分布个数),设定值为 0.8;

e)V_{ini} 为方差元素初始值,设置为 2;

f)D_{thr} 为偏差阈值,设定值为 55.31;

g)C_{thr} 为(连通)分量阈值,设定值为 12.5。

2)GMM 目标运动检测

a)对每一帧中的每一个像素进行处理。根据式(3)进行高斯模型匹配。若匹配,将其归入该模型中,并通过式(4)对背景模型参数进行更新;若不匹配,则以该像素建立一个高斯模型,并通过式(5)初始化参数,替代原有模型中最不可能的模型。

b)根据式(6),选择前面几个最有可能的模型作为背景模型,根据背景模型来提取前景目标;如果直到第 k 个分量的权重之和没有超出权重的阈值,则第 $k+1$ 个分量也应属于背景。

c)形态学处理,将得到前景像素进行分组,获得有效的前景目标。

3)RBMCDA 多目标跟踪

a)Kalman 滤波预测。选取粒子个数 $N_p = 100$,根据式(14)设定的初始值,由式(12)对每个粒子 i 中的目标状态进行预测,得到式(24)所示的目标状态。

b)数据关联。根据式(16)计算测量似然函数,再结合式(22)对应的 Markov 链得出的关联先验;根据式(17)计算最优重要性分布,再根据式(18)~(20)进行数据关联采样,根据式(25)计算目标消失的概率,得到数据关联结果。

c)Kalman 滤波更新。根据以上得到的目标数据关联结果,对每一个粒子 i 中相应目标(根据新出现目标指示变量 $\kappa_{k,j}^{(i)}$ 得到存在的目标个数 $T_k^{(i)}$)的状态变量进行更新,计算更新后的目标状态(消失或非关联的目标状态不参与更新过程),归一化粒子权重,进行重采样操作。

4)重复进行步骤 2)3),直到视频结束。

4 实验结果与分析

为了验证本算法的有效性,对三个不同运动场景进行了跟

踪实验。其特点分别为:运动目标个数较少、多个目标存在交互运动、较远视角下的多个运动目标。实验是基于 MATLAB R2009b 平台上,硬件环境是 Intel Core i7,CPU 2.67 GHz,4 GB 内存。实验所采用的视频是使用固定在三脚架上的摄像机进行拍摄的若干个视频序列,帧率为 25 fps,分辨率为 320 × 240。

实验 1　本段视频包含了单个及两个目标的运动,即骑自行车的目标从左边驶入场景,随后另一个目标从左步行出现;骑自行车的目标从场景右边驶出,一辆白色的汽车从右边驶入。此段视频包含的运动较为简单,目标个数少,没有出现目标间的运动交汇。视频格式为 avi,共有 1 170 帧。跟踪实验结果如图 3 所示,目标用黑色方框来表示。其中(a)为混合高斯模型检测出的背景结果;(b) ~ (d)分别表示第 34、108、228 帧(这三帧基本表示了此段视频中所包含目标运动的情况)的实验跟踪结果。

实验 2　和实验 1 是同一场景下拍摄的视频,此段视频包含了多个运动目标,即三辆汽车的交互运动,以及行人的运动。视频格式为 avi,共有 578 帧。跟踪实验结果如图 4 所示,目标用黑色方框来表示。其中(a)为混合高斯模型检测出的背景结果;(b) ~ (d)分别表示第 205、222、242 帧(这三帧基本表示了此段视频中所包含目标运动的情况:两辆汽车从左向右行驶;当黑色汽车驶出场景时,另一辆白色汽车从右驶入场景;两辆汽车平行交互行驶;场景下方的行人运动)的实验跟踪结果。

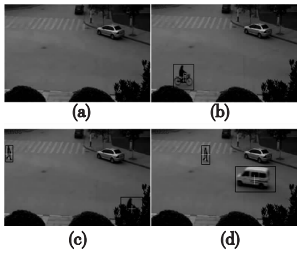

图3　视频1中目标跟踪结果

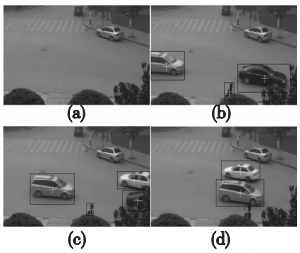

图4　视频2中目标跟踪结果

实验 3　与实验 1、2 的场景不同,远视角场景包含了更多的目标,并且具有目标轮廓较小、位置相对分散和疏密不均的特点。视频格式为 avi,共有 765 帧。跟踪实验结果如图 5 所示(包括三辆汽车和多个行人的运动),目标用黑色方框来表示。其中(a)为混合高斯模型检测出的背景结果;(b) ~ (f)分别表示第 67、335、502、696、762 帧(一辆黄色汽车从左边驶入,从右边驶出;随后一辆黑色汽车从右驶入,从左驶出;接着黑色汽车从左向右行驶;多个行人不规则的运动)的实验跟踪结果。

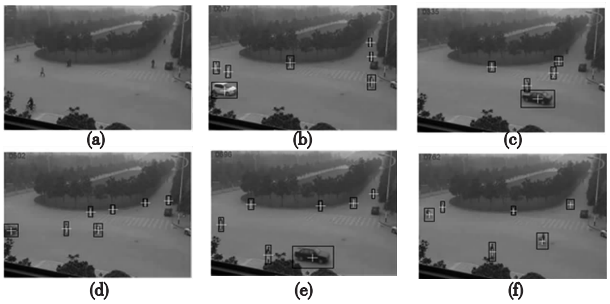

图5　视频3中目标跟踪结果

表 1 记录了处理不同视频段时所需要的运算时间。从实验结果来看,对于处理单一运动目标和多个交互运动目标场景下的目标跟踪问题,本文提出的算法具有一定的实时性。

表 1　实验所需的时间

视频段	视频帧数	每帧处理时间/ms
1	1 170	124
2	578	111
3	765	118

5　结束语

本文提出一种基于 GMM-RBMCDA 的实时监控多目标跟踪方法。该方法利用混合高斯模型作为背景建模方法,并使用形态学方法分割提取出前景目标;在多目标跟踪过程中使用了 Rao-Blackwellized 蒙特卡洛数据关联(RBMCDA)算法。此方法可以处理杂波条件下的跟踪问题,能够有效减少可能的目标交叉及杂波干扰带来的影响,且不需要事先给定目标个数,通过设置目标的存在和消失参数,实现了多个运动目标的自动跟踪。实验结果表明,对于实际场景中所跟踪目标个数无法实现给定的情况,该方法能够较为准确地计算出目标的个数,并且能够同时跟踪多个目标的运动状态,并具有较好的实时性。本文算法存在的一个局限是当场景视角广、目标较小或目标个数增多时,并不能全部给出满意的跟踪效果。另外,当目标之间的距离很近或存在交叉遮挡(或重叠)时,容易误认为是同一个目标。因此,在今后的研究中,还需要借助更多的目标特征,并综合利用模式识别的方法和理论增强前期目标检测预处理的准确性,为后续目标状态估计和跟踪提供更加可靠的保障。

参考文献:

[1] WREN C, AZARBAYEJANI A, DARRELL T, *et al*. Pfinder: real-time tracking of the human body[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780-785.

[2] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking[C]//Proc of the IEEE International Computer Society Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 1999: 246-252.

[3] STAUFFER C, GRIMSON W E L. Learning patterns of activity using real-time tracking[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-757.

[4] 何信华,赵龙. 基于改进高斯混合模型的实时运动目标检测与跟踪[J]. 计算机应用研究, 2010, 27(12): 4768-4771.

[5] PETERFREUND N. Robust tracking with spatio-velocity snakes Kalman filtering approach[C]//Proc of the 6th International Conference on Computer Vision. Washington DC: IEEE Computer Society, 1998: 433-439.

[6] ISARD M, BLAKE A. Contour tracking by stochastic propagation of conditional density[C]//Proc of the Europe Conference on Computer Vision. Berlin: Springer-Verlag, 1996: 343-356.

[7] ISARD M, MACCORMICK J. BraMBLe: a Bayesian multiple-blob tracker[C]//Proc of the IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2001: 34-41.

[8] JEPSON A D, FLEET D J, EI-MARAGHI T F. Robust online appearance models for visual tracking[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1296-1311.

(下转第 4359 页)

量维数为 648 维。将这些高维的特征向量映射到神经元数目为 $W \times W(W=3,\cdots,8)$ 的 LLOM 网络中进行自主学习。实验中,可以对神经元的数目以及维数进行调整测试,已达到最佳的映射。神经元的维数可以为 $H=1,2,3$。例如,如果选择 $W=4,H=2$,那么经训练后,648 维的特征向量就会被映射到 $4\times4\times3$的 LLOM 网络中(每个神经元有一个均值向量和两个正交基向量表示),从而达到高维数据的降维处理。另外,这种映射关系的结果也揭示了高维深度特征向量中隐藏的内在本质特征。

3　实验结果

本文的实验数据来源于斯坦福大学学者 Saxena 采用 3D 扫描设备收集的图像以及对应的参考深度图(ground-truth depths)(http://make3d. stanford. edu),这些图像的深度范围是 1.2 ~81.89 m。为了降低计算复杂度,本文基于 LLOM 的深度识别系统仅包括 80 个 LLOM 网络。其中:被标注为路面(road)的图像块对应 $N_k(k=0,\cdots,16)$ 网络;被标注为草地(grass)的图像块对应 $N_k(k=17,\cdots,41)$ 的网络;被标注为树木(tree)的图像块对应 $N_k(k=42,\cdots,58)$ 的网络;被标注为建筑物(building)的图像块对应 $N_k(k=59,\cdots,80)$ 的网络;而被标注为天空(sky)的图像直接得出深度值为 81.89。

图 3 为深度估计值与其参考深度之间的匹配程度,即估计精度。由于被标注为天空(sky)的图像块深度直接赋值为 81.89,因此其深度估计的精度由语义标注算法来决定。从图 3 中可以看出,深度估计精度不是特别理想,主要原因在于本文将每一类别的语义图像划分在 20 个深度级别以内。这种划分注定了估计结果只是一个近似值,而图 3 的实验数据却是估计值与参考深度值的精确匹配,因此,导致实验结果欠佳。

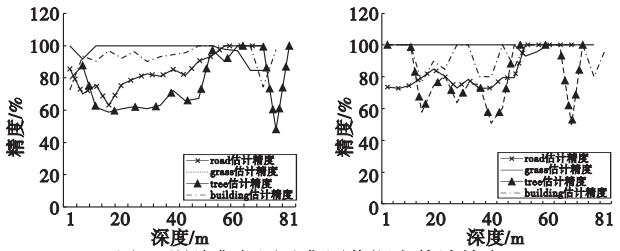

图3　训练集与测试集图像深度估计精度

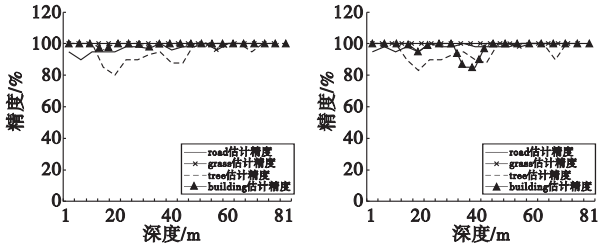

图4　容许0.5 m误差的估计精度

为了更客观地评价本文算法,采用近似匹配策略衡量算法性能,即容许估计值与参考深度值有一定的误差存在。在这种匹配标准的前提下,本文算法估计的深度与参考深度的匹配精度明显提高许多。图 4 为容许 0.5 m 误差的结果。

4　结束语

基于 LLOM 的图像深度估计算法作为一个尝试,为图像深度图估计提供了一种可行的发展空间。实验结果显示,该算法对路面、草地以及建筑物类等这些深度渐进变化的图像块来说,能获得较满意的估计结果。对于深度突变性较大的语义类别图像,如树木等,则需要进一步研究更有效的特征提取方法,这将是本文下一步研究的目标。

参考文献:

[1] SAXENA A, SUN M, ANDREW Y N. Make3D: learning 3D scene structure from a single still image[J]. IEEE Trans on Pattern Analysis and Machine Intelligence,2009,30(5):824-840.

[2] ZHENG Hai-cheng, SHEN Wei, DAI Qiong-hai,*et al*. Learning nonlinear manifolds based on mixtures of localized linear manifolds under a self-organizing framework[J]. Neurocomputing, 2009, 72 (13-15):3318-3330.

[3] 郑慧诚,沈伟. 一种局部化的线性流形自组织映射[J]. 自动化学报,2008,34(10):1298-1304.

[4] EBRAHIMPOUR R, KABIR E, ESTEKY H, *et al*. View independent face recognition with mixture of experts[J]. Neurocomputing, 2008,71(4-6):1103-1107.

[5] XING Hong-jie,HU Bao-gang. An adaptive fuzzy C-means clustering-based mixtures of experts model for unlabeled data classification[J]. Neurocomputing,2008,71(4-6):1008-1021.

[6] GE Liang,ZHU Qing-sheng,FU Si-si. Application of laws' masks to stereo matching[J]. Acta Optica Sinica,2009,29(9):2509-2511.

[7] CARNEIRO G, CHAN A B, MORENO P J,*et al*. Supervised learning of semantic classes for image annotation and retrieval[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29 (3):394-410.

[8] LI Jia, WANG J. Automatic linguistic indexing of pictures by a statistical modeling approach[J]. IEEE Trans on Pattern Analysis and Machine Intelligence,2003,25(9):1075-1088.

[9] 李志欣,施智平. 融合语义主题的图像自动标注[J]. 软件学报,2011,22(4):801-812.

[10] LIU B, GOULD S, KOLLER D. Single image depth estimation from predicted semantic labels[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010:1253-1260.

(上接第 4349 页)

[9] 李景義,王树宗. UPF 算法及其在目标跟踪问题中的应用[J]. 系统仿真学报,2007,19(3):675-677.

[10] GORDON N J, SALMOND D J, SMITH A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation[J]. IEEE Proceedings on Radar and Signal Processing,1993,140(2):107-113.

[11] 王华剑,景占荣,羊彦. 基于改进扩展卡尔曼粒子滤波的目标跟踪算法[J]. 计算机应用研究,2011,28(5):1634-1636,1643.

[12] SARKKA S,VEHTARI A,LAMPINEN J. Rao-Blackwellized particle filter for multiple target tracking[J]. Information Fusion,2007,8 (1):2-15.

[13] 王相海,方玲玲,丛志环. 基于 MSPF 的实时监控多目标跟踪算法研究[J]. 自动化学报,2012,38(1):139-144.

[14] HORN B K P. Robot vision[M]. [S. l.]:MIT Press,1986:66-69, 299-333.