

文章编号: 0253-2239(2010)06-1721-08

复杂场景下基于条件随机场的视觉目标跟踪

高琳 唐鹏 盛鹏 左航

(四川大学视觉合成图形图像技术国防重点学科实验室, 四川 成都 610065)

摘要 目标模型更新中存在的模型漂移问题,是影响视频跟踪结果的一个重要因素。针对这一难题,提出了一种新的基于前景分割的目标跟踪算法。算法通过引入条件随机场(CRF)模型对跟踪区域和非跟踪区域的时空关系进行建模,实现对图像序列中像素点的标记,标记为跟踪目标或背景,并使用在线学习方法,根据场景的变化调整CRF模型的参数。跟踪过程中,通过对CRF模型的求解,得到最优的标记场和目标像素的置信图像;利用置信图像,结合目标模型的相似性度量定位整个目标;根据目标区域内的标记结果,使用一种选择性采样的方式更新目标模型,从而解决更新中的漂移问题。通过在多个典型的复杂场景中进行实验,验证了该算法的有效性。

关键词 信息处理;视觉目标跟踪;模型漂移;条件随机场;置信图;选择性更新

中图分类号 TP391 文献标识码 A doi: 10.3788/AOS20103006.1721

Visual Object Tracking Based on Conditional Random Field Under Complex Scene

Gao Lin Tang Peng Sheng Peng Zuo Hang

(Key Laboratory of Fundamental Synthetic Vision Graphics and Image Science for National Defense, Sichuan University, Chengdu, Sichuan 610065, China)

Abstract Target model drift is an important factor for visual tracking. To deal with the problem, a novel adaptive tracking algorithm is proposed. The spatial-temporal constraints of the tracking region and non-tracking region in sequence are modeled through a conditional random field (CRF), the parameters of which are updated using online learning method according to the changes of scene. The pixels of all images can be labeled as target or background based on the CRF model. During tracking, the optimal label field and the confidence map are first achieved by resolving the CRF model. Then the similarity measure between the target model and the target candidates, combined with the confidence map, is fed to the mean shift algorithm for the target localization. A selective sampling update strategy is utilized to alleviate the model drift. The experimental results demonstrate the efficiencies of the proposed algorithm in several real sequence testings.

Key words information processing; visual object tracking; model drift; conditional random field; confidence map; selective update

1 引言

目标跟踪是近年计算机视觉领域中备受关注的前沿方向,同时又是智能视频监控、人机交互以及机器人导航等应用领域的核心问题。随着新技术新理论被引入到跟踪问题的研究之中,跟踪算法得到了

长足的发展,然而如何实现复杂场景下的稳健跟踪仍然是困扰该邻域研究的一个主要问题。

实际环境中,由于环境光线变化、背景变化或出现遮挡等原因,跟踪目标的外观也可能发生变化,采用固定目标模型的视频跟踪方法^[1,2]通常不能适应

收稿日期: 2009-06-17; 收到修改稿日期: 2009-09-10

基金项目: 国家自然科学基金(60705005)、国家自然科学基金民航联合基金重点项目(60736046)和教育部博士点基金(20070610031)资助课题。

作者简介: 高琳(1976—),男,博士研究生,主要研究方向为数字图像处理 and 计算机视觉。E-mail: gaolinscu@163.com

导师简介: 游志胜(1945—),男,教授,博士生导师,主要从事图像处理模式识别、计算视觉、智能交通等方面的研究。

E-mail: zsyoun@mail.sc.cininfo.net

这些变化,需要随时对目标模型进行调整和更新。为了实现稳健的跟踪,近些年,自适应方法^[3-5]被用在模型的更新问题上。然而,这些自适应的跟踪算法都存在一个共同的问题,就是模型的漂移问题,即自适应算法总是根据最近时刻的跟踪结果来更新目标模型。由于跟踪结果所对应的目标图像中,或多或少都会包含一些非目标的图像,如背景、遮挡物等,这些错误的信息会带来模型更新的误差。随着跟踪过程的推进,累积的误差导致目标模型逐渐脱离正确的目标。

为了减轻模型漂移对于稳健跟踪的影响,已经提出了很多方法并加以运用。Avidan^[6]将目标表示为一组在线训练的分量分类器以区分目标像素和背景像素,所有的分量分类器通过 Adaboost 方法组成一个总分类器。该方法利用当前的跟踪结果,训练新的分量分类器,并加入到分类器组中,从而实现模型的更新。由于总体中包含了一些旧的分类器,从而在一定程度上减轻了模型漂移问题的影响。Collins 等^[7]用第一帧图像来校正更新的模型,认为目标当前外观与初始跟踪时的外观,始终保持一定的相似性。然而在很多场合下,这种假设并不成立。

Nguyen 等^[8]在更新目标模型时,使用稳健卡尔曼(Kalman)滤波器在时间上对模型进行平滑处理,从而减轻模型漂移。该算法对光照变化和短时间遮挡具有较好的稳健性,但对于长时间的遮挡,模型仍然会产生漂移。

针对模型更新的漂移问题,最直接的解决方法,就是尽量减小非目标观测的影响。为此,提出了一种新的基于前景分割的目标跟踪算法。通过引入条件随机场(CRF)模型对序列的时空关系进行建模,实现对图像序列中像素点的标记,分别标记为跟踪目标和背景。利用标记过程中得到的置信图像估计目标的位置和尺度。在更新目标模型时,根据得到的最优标记场,使用一种选择性采样的更新策略,避免发生模型漂移现象。

2 方法概述

目标跟踪的步骤大致可以分为以下两个阶段:基于 CRF 的像素标记阶段和目标定位及目标模型的自适应更新阶段,总体框图如图 1 所示。

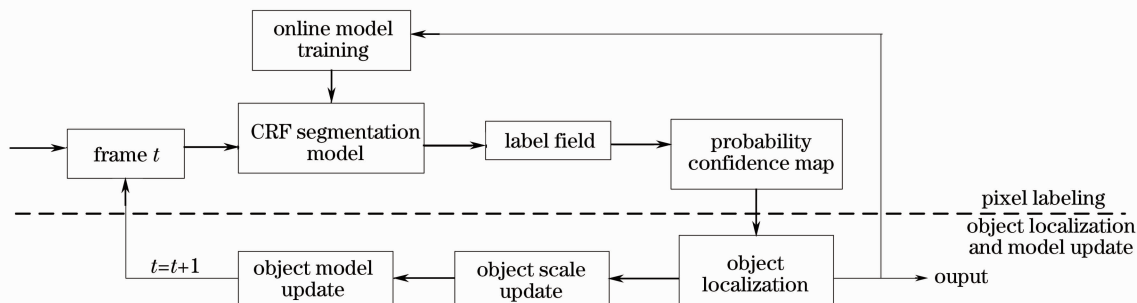


图 1 基于 CRF 的跟踪算法框架图

Fig. 1 Scheme of the tracking algorithm based on CRF

像素标记阶段:利用像素点在图像序列中的时空邻域约束和特征(如光流,颜色),建立一种融合时空上下文信息的 CRF 分割模型。使用一种在线训练方法获取模型参数,使 CRF 模型能够根据场景的变化进行调整。通过对 CRF 模型的求解,获得一个与像素点阵相对应的标记场,其中每个像素被标记为前景目标或是背景。利用标记值的边缘概率分布特点,得到表征目标像素点空间分布的概率置信图像。

目标定位和模型更新阶段:在上一阶段像素级处理的基础上,根据目标的颜色分布模型,在区域中对目标进行定位。结合置信图像,将目标像素点在空间上的概率分布信息,集成到目标模型与候选目标的相似性度量之中。在每帧的跟踪实现后,基于

标记场的结果,对遮挡情况进行判断。如果没有发生遮挡,则对目标模型进行增量更新,增量部分通过对目标区域的选择性采样得到。

3 基于条件随机场的前景分割

CRF 由 Lafferty 等^[9]提出,最初用于自然语言的一维标记处理。相对于隐马尔可夫模型(HMM)等模型而言,CRF 可以选择上下文相关特征,从而解决了标记偏见的问题。可以通过对所有特征进行全局归一化,获得全局的最优值。目前 CRF 的各种改进模型被大量应用于计算机视觉领域,如图像目标识别^[10],图像分割^[11]等,其效果要明显优于使用 HMM 和马尔可夫随机场(MRF)模型的结果。

3.1 融合时空信息 CRF 模型

$\{I_k\}$ 为定义在像素点阵上的一个视频图像序列, k 为帧号, 图像帧 $I_i = \{y_u, u \in V_i\}$, 其中 V_i 为图像帧 I_i 中所有场点(即像素)的集合, y_u 为像素点 u 的观测值。图像序列中的前景背景分割问题, 可以看作是针对于离散场的标记过程。定义随机变量集 $X_i = \{x_u \in L, u \in V_i\}$ 与 I_i 对应, 用于标记图像中的像素, 其中 $L \in \{0, 1\}$ 为分类的标记, 0 为背景, 1 为前景。标记的目标就是为序列中的像素点阵建立一个标记场, 使得标记场在一定的概率准则下达到最优配置。

对于一个静态图像, 标记过程在一个二维的离散场中完成, 利用的是像素在图像空间上的邻域信息, 这称为空域邻域信息。在图像序列中, 沿着时间轴上相邻的对应像素之间也有联系, 这种联系也为标记提供了非常有用的信息, 称为时域邻域信息。如果能够基于实际场景将像素在时间和空间上的邻域信息融合到标记过程中, 则可以进一步提高标记的准确性。为此在模型中同时定义像素在时域和空域上的邻域关系, 如图 2 所示。算法中, 像素在同一帧中拥有 8 pixel 的空域邻域, 在相邻帧上有 25 pixel 的时域邻域。这样就可以在时空域上定义 CRF 分割模型, 不仅优化图像在空间的分布, 还能优化图像序列沿着时间轴的分布。

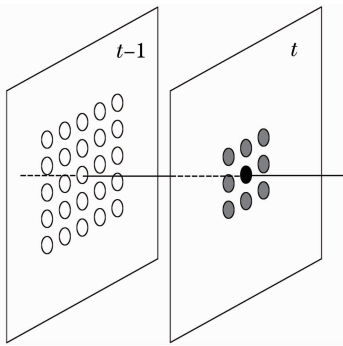


图 2 时空域邻域关系图

Fig. 2 Spatial-temporal neighborhood

给定图像帧 I_t 和 I_{t-1} , CRF 模型直接对标记变量 X_t 的后验概率分布进行建模, 后验概率 $P(X_t | I_{t-1:t}; \Theta)$ 定义为

$$P(X_t | I_{t-1:t}; \Theta) = \frac{1}{Z(I_{t-1:t}, \Theta)} \exp \left\{ \sum_{u \in V_t} [\alpha f_c(x_u, I_t) + \sum_{v \in M_u} \beta g_o(x_u, x_v, I_{t-1:t}) + \sum_{v \in N_u} \gamma g_s(x_u, x_v, I_{t-1:t})] \right\}, \quad (1)$$

式中 I_t, I_{t-1} 为当前帧和前一帧的图像观测, $\Theta = \{\alpha, \beta, \gamma\}$ 为模型的参数, f_c 为单变量势函数, 描述的是标记变量在给定图像观测的条件下所对应的强度, M_u, N_u 分别为像素点 u 在时域和空域上的邻域, g_o 和 g_s 为双变量交互势函数, 分别描述变量 x_u 与邻域 M_u, N_u 中的变量之间形成的强度, $Z(I_{t-1:t}, \Theta)$ 为分配函数, 用于函数的归一化处理。

单变量势函数仅利用图像对像素的标记进行判别, 而不考虑该像素的邻域约束。令 $q = \{q_i\}_{i=1,2,\dots,m}$ 和 $h = \{h_i\}_{i=1,2,\dots,m}$ 分别为目标和背景的颜色直方图, 势函数 $f_c(x_u, I_t)$ 定义为目标和背景的色彩分布对数似然比

$$f_c(x_u, I_t) = \left[\ln \left(\frac{q_{y(u)}}{h_{y(u)}} \right) \right] x_u. \quad (2)$$

利用上一帧的跟踪结果来计算背景颜色直方图, 即假定 $t-1$ 时刻的目标矩形区域为 R_{t-1} , 在 R_{t-1} 与其所对应的搜索矩形之间的环状区域内, 计算背景直方图 h 。

考虑到图像序列中前景背景通常存在着不同的运动方式, 在时域上, 利用光流特征来区分前景和背景像素。前景运动具有连续性, 前景像素的光流在时域邻域内近似相同。相比而言, 背景通常以一种随机的方式运动, 背景像素的光流也呈现出随机性。根据这个特点, 文中定义势函数 $g_o(x_u, x_v, I_{t-1:t})$ 为

$$g_o(x_u, x_v, I_{t-1:t}) = \frac{1}{\|\vec{OF}_u - \vec{OF}_v\|^2 + \tau} x_u x_v, \quad (3)$$

式中 \vec{OF}_i 表示像素点 i 的光流向量, $\|\cdot\|$ 为向量间的欧式距离, τ 为一个较小的常数。

空域邻域势函数描述的是同一帧图像上邻居像素之间的相互约束关系, 为了简化计算, 使用一种独立于图像观测的空间连通性约束定义势函数

$$g_s(x_u, x_v, I_{t-1:t}) = \delta(x_u, x_v), \quad (4)$$

式中 $\delta(\cdot)$ 为克罗内克(Kronecker Delta)函数。

3.2 模型参数学习

相对于生成模型, 如 MRF, HMM 等, 判别模型 CRF 具有参数学习简单快捷的特点。由于生成模型需要对状态变量的先验分布进行描述, 因而需要大量的训练数据进行参数学习。相比之下, 判别模型更关注对数据类型的判别而较少考虑数据的分布形式, 所以只需要很少的训练数据。这有利于设计在线的训练方法, 以适合解决实际跟踪问题。

使用先前的分割结果, 将目标区域内的标记值作为训练数据, 对模型参数进行迭代更新, 使参数可以按照目标和背景的变化而变化。给定单个训练数

据 $\{\mathbf{X}_t, \mathbf{I}_t\}$, 利用条件极大似然准则训练该条件判别模型

$$\Theta^* = \arg \max_{\Theta} \ln P(\mathbf{X}_t | \mathbf{I}_{t-1,t}; \Theta). \quad (5)$$

采用随机梯度增长算法^[12] 搜索参数 Θ 的最优解, 参数 $\theta_j \in \Theta$ 的学习规则如下:

$$\Delta \theta_j = \eta \frac{\partial}{\partial \theta_j} \ln P(\mathbf{X}_t | \mathbf{I}_{t-1,t}; \Theta), \quad (6)$$

式中 η 为学习速率。例如, 假定第 n 次迭代后的参数为 Θ^n , 时域势函数 $g_o(x_u, x_v, \mathbf{I}_{t-1,t})$ 的系数 β^{n+1} 则按照下式迭代更新

$$\begin{aligned} \beta^{n+1} &= \beta^n + \eta^n \frac{\partial}{\partial \beta} \ln P(\mathbf{X}_t | \mathbf{I}_{t-1,t}; \Theta^n) = \beta^n + \\ &\eta^n \left[\sum_{u \in V_t} \sum_{v \in M_u} g_o(x_u, x_v, \mathbf{I}_{t-1,t}) - \frac{\partial}{\partial \beta} \ln Z(\mathbf{I}_{t-1,t}, \Theta^n) \right]. \end{aligned} \quad (7)$$

由于对数分配函数 $\ln Z(\mathbf{I}_{t-1,t}, \Theta^n)$ 是指数族的累积生成函数^[12], 则 $\ln Z(\mathbf{I}_{t-1,t}, \Theta^n)$ 关于 β 的一阶偏导为

$$\begin{aligned} \frac{\partial}{\partial \beta} \ln Z(\mathbf{I}_{t-1,t}, \Theta^n) &= \\ E_{P(\mathbf{X}_t | \mathbf{I}_{t-1,t}, \Theta^n)} \left[\sum_{u \in V_t} \sum_{v \in M_u} g_o(x_u, x_v, \mathbf{I}_{t-1,t}) \right], \end{aligned} \quad (8)$$

式中 $E[\cdot]$ 为分布 $P(\mathbf{X}_t | \mathbf{I}_{t-1,t}; \Theta^n)$ 下的期望值。

将(8)式代入(7)式得到

$$\begin{aligned} \beta^{n+1} &= \beta^n + \eta^n \left\{ \sum_{u \in V_t} \sum_{v \in M_u} g_o(x_u, x_v, \mathbf{I}_{t-1,t}) - \right. \\ &E_{P(\mathbf{X}_t | \mathbf{I}_{t-1,t}, \Theta^n)} \left[\sum_{u \in V_t} \sum_{v \in M_u} g_o(x_u, x_v, \mathbf{I}_{t-1,t}) \right] \left. \right\}, \end{aligned} \quad (9)$$

式中的后验分布 $P(\mathbf{X}_t | \mathbf{I}_{t-1,t}; \Theta^n)$ 使用 LBP(Loop Belief Propagation) 算法^[13] 求解得到。

3.3 标记场的估计和置信图的生成

给定模型参数和图像帧 $\mathbf{I}_{t-1}, \mathbf{I}_t$, 利用极大后验边缘准则求取概率上最优的标记场, 即每个标记变量 x_u 的值通过下式来计算

$$x_u = \arg \max_{x_u} P(x_u | \mathbf{I}_{t-1,t}; \Theta), \quad (10)$$

式中 $P(x_u | \mathbf{I}_{t-1,t}; \Theta)$ 同样用 LBP 算法计算得到。

通过求取最优的标记场, 在像素级上实现了目标和背景的分, 然而二值的标记场不包含像素属于目标或是背景的概率信息。考虑到图像序列中存在着大量的随机性, 目标像素在空间分布上的概率信息更有助于后续的区域级目标定位, 根据得到的后验分布 $P(x_u | \mathbf{I}_{t-1,t}; \Theta)$ 生成置信图像。置信图像中像素点 u 的值等于变量 x_u 标记为前景的概率, 即 $w_i(u) = P(x_u = 1 | \mathbf{I}_{t-1,t}; \Theta)$ 。

4 运动目标跟踪

4.1 相似性度量 and 目标定位

采用颜色直方图作为描述跟踪目标的目标模型。考虑到目标的运动速度, 认为目标在相邻帧的位置变化不会超过某个范围。根据这个假设, 在跟踪时, 使用矩形框来对目标区域进行覆盖, 并且在目标矩形的外围, 选取一个环绕目标的较大矩形(实验中, 其尺寸为目标矩形的长和宽各放大 10%), 作为目标在下一时刻的搜索区域。

根据当前的目标模型, 目标定位是在前一时刻目标位置所对应的搜索区域中, 寻找使相似性最大的候选目标。令当前的目标颜色直方图为 $\mathbf{q} = \{q_i\}_{i=1,2,\dots,m}$, 中心位置为 y 的候选颜色直方图为 $\mathbf{p} = \{p_i\}_{i=1,2,\dots,m}$, $\rho(\mathbf{p}(y)) = \rho[\mathbf{p}(y), \mathbf{q}]$ 为相似性度量函数。采用 Bhattacharyya 系数作为颜色分布 \mathbf{p} 和 \mathbf{q} 之间的相似性度量函数, 有

$$\rho(\mathbf{p}(y)) = \rho[\mathbf{p}(y), \mathbf{q}] = \sum_{i=1}^m \sqrt{p_i(y) q_i}. \quad (11)$$

由于颜色直方图仅仅描述了目标的全局颜色分布, 而忽略了目标在图像空间中的分布, 当目标处在易混淆的背景中, 目标和背景的颜色分布相似时, (11)式的相似性函数不能很好地区分目标和背景区域。为了提高相似性度量的性能, 利用上节得到的置信图像, 定义改进的相似性函数

$$\tilde{\rho}(y) = \rho(\mathbf{p}(y)) \frac{1}{wh} \sum_{u \in R_c} w_i(u), \quad (12)$$

式中 R_c 为候选目标的矩形区域, w, h 为 R_c 的宽和高。相似性函数 $\tilde{\rho}(y)$ 包含了两个层次上的度量信息: 一是来自区域级的直方图匹配, 另一个是来自像素级的置信度求和。通过对这两种层次信息的集成, 提高了目标定位的可靠性。

定义了相似性度量函数后, 采用均值漂移算法寻找目标在搜索区域中位置。在对每个候选区域计算相似度时, 如果直接计算像素置信度的和, 会大大增加跟踪算法的计算复杂度。为了降低算法的计算复杂度, 计算出概率权值图在搜索区域的积分图像^[14], 此后, 对任何矩形区域 R_c , 计算 $\sum_{u \in R_c} w_i(u)$ 的值只需要执行三个加操作, 效率非常高。

4.2 目标尺度更新

跟踪过程中, 由于目标与摄像机的相对运动, 目标在图像上的尺度会发生变化, 如果目标矩形的尺寸固定不变, 将不能保证对目标区域的有效覆盖。借鉴文献^[15]的思想, 根据尺度空间理论, 在尺度维上搜索目

标的最佳尺度,目标矩形则按照最佳尺度改变大小。

当得到了目标当前的中心位置后,在对应的置信图像上计算目标的尺度。方法是分别使用不同尺度下的高斯差分(DOG)算子与置信图像进行卷积滤波,滤波响应的局部极值所对应的尺度即为最佳尺度。假设当前目标的中心位置为 x_t ,目标尺度的计算可以转化为下面的优化问题:

$$\sigma_t = \arg \max_{\sigma} \sum_{u \in R_S} d(u - x_t, \sigma) w(u), \quad (13)$$

式中 R_S 为搜索区域, $d(x, \sigma)$ 为 DOG 函数,其形式如下:

$$d(x, \sigma) = \frac{1}{\sqrt{2\pi}k\sigma} \exp[-x^2/2\pi(k\sigma)^2] - \frac{1}{\sqrt{2\pi}\sigma} \exp(-x^2/2\pi\sigma^2), \quad (14)$$

式中 k 为比例常数,选取 $k=1.6$ 。

文献[15]使用 Epanichnikov 核函数实现尺度维上的核密度估计,然后利用均值漂移算法搜索最佳尺度,由于其计算量较大,在此做一些简化。考虑到前后两帧的目标在尺度上的变化不会太大,因此仅在前一时刻的尺度 σ_{t-1} 附近,对 5 个不同的尺度 $\sigma_{t-1} + \Delta\sigma$ ($\Delta\sigma$ 分别为 $0, \pm 0.1\sigma_{t-1}, \pm 0.15\sigma_{t-1}$) 进行比较,选择 DOG 滤波响应最大的尺度作为目标的尺度,目标矩形的长和宽则分别按照比例 σ_t/σ_{t-1} 进行缩放。

4.3 目标模型更新

在遮挡情况下,系统不能获得完整的目标观测,对于基于颜色分布的目标模型,不能提供可靠的更新信息。为此,设计目标模型只在非遮挡情况下更新,遮挡时模型则保持不变。遮挡判断的方法是,根据像素标记的结果,如果目标矩形中标记为前景的像素占整个目标矩形的比例 P_r 小于阈值 T_h , $T_h \in [0, 1]$ (实验中 T_h 取值为 0.45),则认为遮挡发生。

为了解决模型更新时的漂移问题,根据获得的标记场,仅选择标记为前景的像素参与模型的更新。假设跟踪结果是以 y_0 为中心的矩形区域,计算该区域中前景像素的颜色分布 $q^{cur} = \{q_i^{cur}\}_{i=1,2,\dots,m}$

$$q_i^{cur}(y_0) = C \sum_{j=1}^l k\left(\frac{y_0 - y_j}{h_0}\right) \delta\{[b(y) - i](x_j - 1)\}, \quad (15)$$

式中 C 为归一化系数, k 为空间加权的核函数, x_j 为像素点 j 的标记值。考虑到前景分割和目标定位的过程中,都会存在着一些误差,如果直接以 q^{cur} 替代目标模型 q 作为新的目标模型,当出现较大的误差

时,会由于更新过快而使得模型难以恢复到正确的状态。因此,以一种增量的方式对目标模型进行更新。总而言之,模型的更新公式为

$$q_{i,t} = \begin{cases} \lambda q_{i,t-1} + (1-\lambda)q_i^{cur}, & P_r > T_h \\ q_{i,t-1}, & \text{otherwise} \end{cases} \quad (16)$$

式中权值 λ 为更新系数,用于控制模型更新的速度,取值范围 $\lambda \in [0, 1]$ 。

5 实验结果

为了检验算法的有效性,对多个真实视频序列进行了实验。目标的初始区域为手动选择,目标的颜色直方图在 HSV 色彩空间中构建,整个颜色空间离散化为 $8 \times 8 \times 4$ 个量化级。算法使用 Visual C++ 和 OpenCV 实现,硬件平台为: Pentium 4, 3.0 GHz, 内存为 1 GB, 程序的处理速度约为 12 frame/s。

图 3 为机场停机坪场景下的车辆跟踪实验结果。该序列中,跟踪目标的外观和尺度发生了较大的变化,这增加了跟踪的难度。为了对比算法的性能,以文献[7]的算法(简称 Collins 算法)作为比较,在序列中采用相同的初始条件进行目标跟踪。Collins 算法的跟踪结果如图 3(a) 所示。在第 155 帧附近,车辆开始转弯,其姿态的改变造成外观发生变化,此时目标矩形逐渐偏离了目标,这是由于该算法总是利用初始目标来校正模型。当目标外观发生较大变化时,这种校正不再有效,从而产生了模型漂移。图 3(b) 显示的是本文算法的跟踪结果,图 3(c) 显示的是与图 3(b) 各帧相对应的置信图像。由于利用前景分割的结果更新目标模型[图 3(b) 中右下角的小图像为目标搜索区域中的分割结果],减少了背景对模型的影响,因此尽管目标的外观和尺度均发生了较大变化,仍然抑制住了模型漂移。跟踪结果表明,这一算法能够准确地跟踪上目标。

通过手动得到车辆运动的真实轨迹后,将两种方法在水平和垂直方向上跟踪得到的轨迹与真实值进行比较,结果如图 4 所示。从中 4 可以看出,在第 155 帧附近,目标姿态变化时,Collins 算法的跟踪结果开始产生偏离,水平方向的偏离较为严重,垂直方向的偏离也很明显。本文算法的结果始终接近真实情况。

图 5 为部分遮挡情况下的实验结果,帧数分别为 55, 109, 222, 253。该序列中,被跟踪的目标被另一个行人部分遮挡,并且遮挡的状态保持了一段时间。实验分别采用了文献[7]中的方法(Collins 算

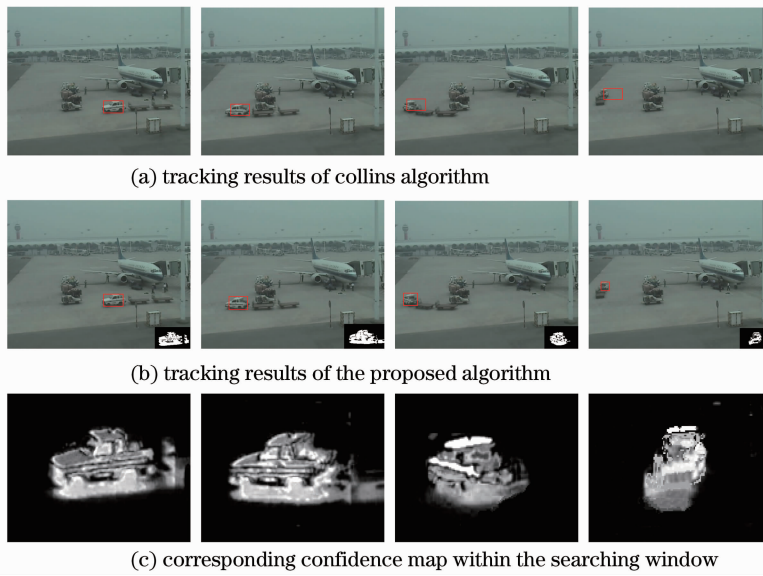


图 3 姿态变化下的车辆跟踪(帧数为:76,155,191,279)

Fig. 3 Tracking a vehicle under pose variation. (Frames: 76,155,191,279)

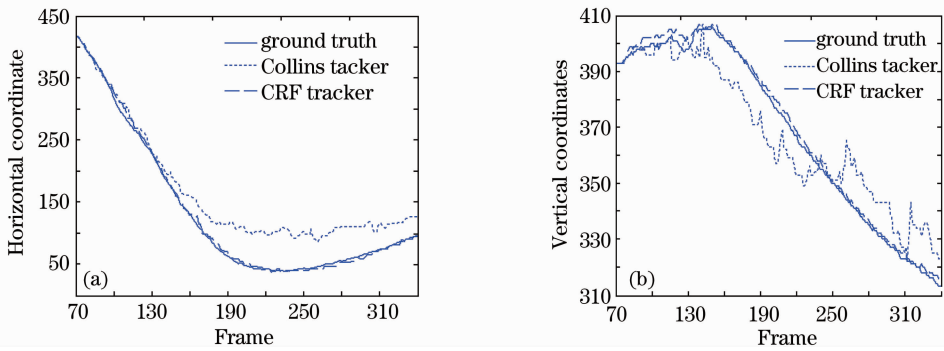


图 4 (a)CRF 算法与(b)Collins 算法得到的跟踪轨迹比较

Fig. 4 Comparison of the tracking trajectories obtained by (a) CRF algorithm and (b) Collins algorithm

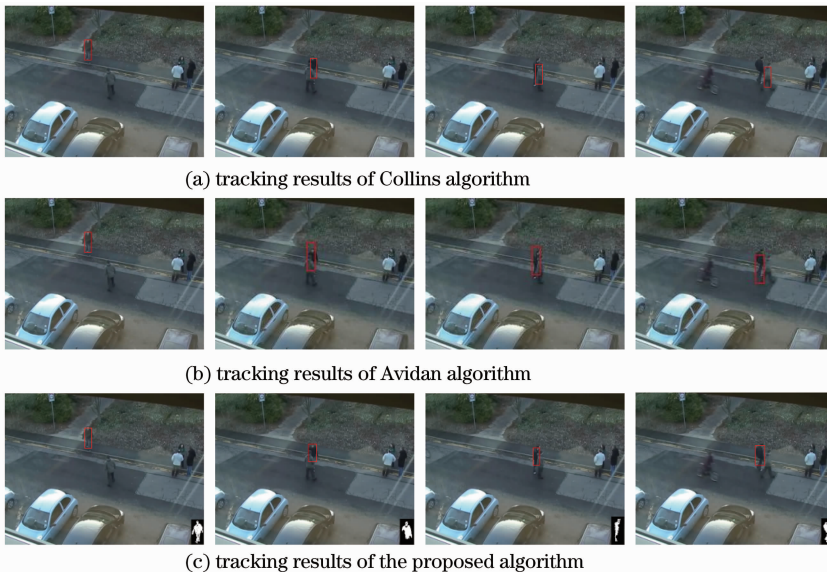


图 5 部分遮挡下的行人跟踪

Fig. 5 Tracking a pedestrian under partial occlusion

法)、文献[6]中的方法(Avidan 算法)以及本文方法。从图 5 可以看到,Collins 算法在遮挡情况下跟踪效果差,在第 253 帧遮挡结束时,错误地跟随上遮挡的行人。Avidan 算法虽然跟上了目标,但由于较长时间的遮挡造成了目标模型的漂移,导致出现较大的跟踪误差。本文算法利用前景分割的结果来对遮挡进行判断和处理,表现出了较好的效果。

图 6 显示了 3 种算法在背景混淆干扰下的行人跟踪结果,帧数分别为 135,169,196,249。行人的衣服与背景的颜色很相近,很容易造成目标与背景

相互混淆。跟踪结果显示,Collins 算法不能准确地定位目标,很快就跟踪失败。Avidan 算法在易混淆的背景下,跟踪变得不稳定,并在第 249 帧丢失了跟踪目标,这是由于在 Avidan 算法中,目标模型的分类器只是根据上一帧目标和背景的特征差异进行训练和更新,而没有考虑到特征在时空域上的邻居约束关系,因而在目标和背景相似时,分类器的性能出现下降。本文算法在该序列中,能够有效地区分前景和背景,因而能够稳定地跟踪到目标。

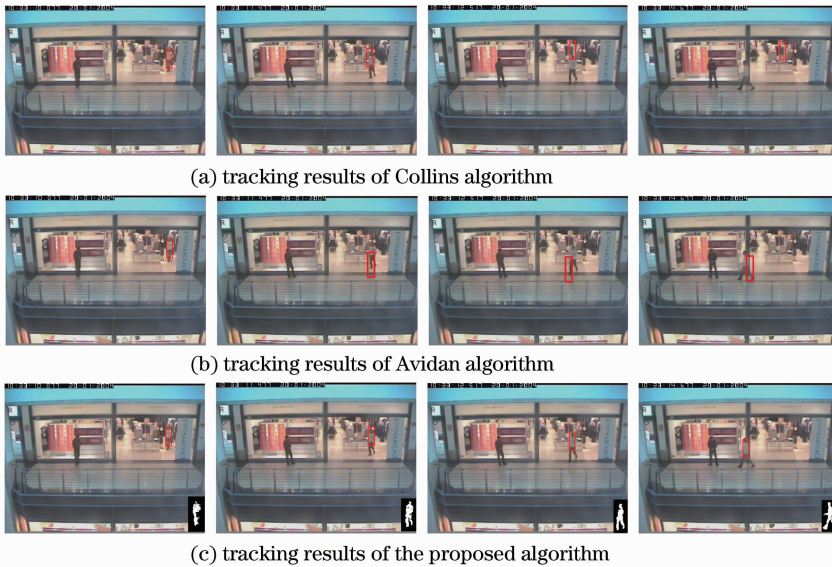


图 6 混淆背景下的行人跟踪

Fig. 6 Tracking a pedestrian under background clutters

6 结 论

针对视觉目标跟踪中存在的目标模型漂移问题,提出了一种基于前景分割的目标跟踪算法。算法通过引入条件随机场模型对跟踪区域和非跟踪区域的时空关系进行建模,实现对图像序列中像素点的标记,分别标记为跟踪目标或背景,并使用在线学习方法,根据场景的变化调整 CRF 模型的参数。在跟踪过程中,基于前景分割的结果,结合目标模型的相似性度量定位整个目标的位置和尺度。目标模型更新时,根据得到的标记场,使用一种选择性采样的方式进行更新,减少了非目标观测的影响,从而解决更新中的漂移问题。多个典型场景下的实验表明该算法是有效的。

参 考 文 献

1 D. Comaniciu, V. Ramesh, P. Meer. Real-time tracking of non-rigid objects using mean shift [C]. Proc. IEEE Conf on Computer Vision and Pattern Recognition(CVPR), 2000, 2: 142~149

2 A. Elgammal, R. Duraiswami, L. Davis. Probability tracking in joint feature-spatial spaces[C] Proc. IEEE Conf. on Computer Vision and Pattern Recognition(CVPR), 2003, 781~788

3 Jia Guimin, Wang Xiangjun, Zhang Shihai. Target tracking algorithm based on adaptive template update in complex background[J]. *Acta Optica Sinica*, 2009, 29(3): 659~663

贾桂敏, 王向军, 张世海. 复杂背景下基于自适应模板更新的目标跟踪算法[J]. *光学学报*, 2009, 29(3): 659~663

4 Risheng Han, Zhongliang Jing, Yuanxiang Li. Kernel based visual tracking with scale invariant features [J]. *Chin. Opt. Lett.*, 2008, 6(3): 168~171

5 Guan Zhiqiang, Chen Qian, Qian Weixian *et al.*. Infrared target tracking algorithm based on algorithm fusion[J]. *Acta Optica Sinica*, 2008, 28(5): 860~865

管志强, 陈 钱, 钱惟贤 等. 一种基于算法融合的红外目标跟踪方法[J]. *光学学报*, 2008, 28(5): 860~865

6 S. Avidan. Ensemble tracking [J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007, 29(2): 261~271

7 R. T. Collins, Y. Liu, M. Leordeanu. Online selection of discriminative tracking features [J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2005, 27(10): 1631~1643

8 H. T. Nguyen, A. W. M. Smeulders. Fast occluded object tracking by a robust appearance filter[J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2004, 26(8): 1099~1104

9 J. Lafferty, A. McCallum, F. Pereira. Conditional random

- fields; probabilistic models for segmenting and labeling sequence data [C]. Proc. IEEE Int. Conf. Machine Learning, 2001. 282~289
- 10 S. Kumar, M. Hebert. Discriminative random fields: a discriminative framework for contextual interaction in classification [C]. Proc. IEEE Int. Conf. Computer Vision, 2003. 1150~1159
- 11 X. Ren, C. Fowlkes, J. Malik. Cue integration for figure/ground labeling [C]. Proc. Neural. Information Processing Systems (NIPS), 2005. 1121~1128
- 12 S. V. N. Vishwanathan, N. N. Schraudolph, M. W. Schmidt *et al.*. Accelerated training of conditional. random fields with stochastic gradient methods [C]. Proc. Int. Conf. Machine Learning, 2006. 969~976
- 13 Y. Weiss. Correctness of local probability propagation in graphical models with loops [J]. *Neural Computation*, 2000, **12**(1): 1~41
- 14 P. Viola, M. Jones. Robust real-time face detection [J]. *International J. Computer Vision*, 2004, **57**(2): 137~154
- 15 R. Collins. Mean-shift blob tracking through scale space [C]. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2003, **2**: 234~240