

一种新的基于在线采样的多物体跟踪方法

朱 琳 周 杰 宋靖雁

(清华大学自动化系 北京 100084)

摘 要 跟踪多个运动物体,尤其是在遮挡过程中跟踪多个运动物体,是计算机视觉领域一个重要但具有挑战性的问题. 该文提出了一种新的在线采样、更新学习和分类的跟踪框架来处理多物体跟踪问题. 首先,对遮挡发生前若干帧的各物体进行块采样,作为训练样本进行在线分类器设计. 各帧的物体区域也在线进行块采样,并用这些分类器来进行分类标号. 如果遮挡没有发生,一些新的训练样本被添加用来更新分类器. 当遮挡发生时,根据标号结果,前景区域被分割成多个目标物体. 和以往方法相比,新方法不依赖于一些假设条件,如场景深度信息、物体的先验模型(比如形状、种类、区域内颜色各向同性、运动规律等),具有更好的适应能力. 实验结果验证了该文方法的稳定性和有效性.

关键词 遮挡;多目标跟踪;在线采样

中图法分类号 TP391

Multiple Objects Tracking Based on Online Sampling

ZHU Lin ZHOU Jie SONG Jing-Yan

(Department of Automation, Tsinghua University, Beijing 100084)

Abstract Multiple objects tracking, especially multiple objects tracking with occlusion, is an important and challenging problem in computer vision. This paper proposes a novel on-line sample based framework to deal with this problem. First, training samples are obtained from each isolated object before occlusion occurs. A simple sample based scheme is used to produce a set of classifiers. The objects in the following frames are labeled using the classifiers. If no occlusion occurs in the frame, the classifiers will be updated by adding new training samples. When there is occlusion in the present frame, according to the labeled results, the foreground region can be segmented into multiple objects. Compared with previous works, this approach needs neither the depth information of the scene nor the prior models of objects such as color blobs, the type of objects (person or vehicle), velocity assumption. Experiments show that the proposed approach can work robustly under the more complex conditions in which the above assumptions may be unreliable.

Keywords occlusion; multiple objects tracking; online sampling

1 引 言

视觉监控系统中,对运动物体进行持续稳定的

跟踪,有助于物体的识别、行为的分析以及高层事件的理解^[1-5]. 目前很多跟踪系统能很好地跟踪孤立运动物体,遇到遮挡时则很容易发生跟踪错误. 多个运动物体的分割和跟踪不仅对视频监控很重要,也有

利于视频索引、视频归档和检索系统^[6]. 因此,跟踪多个运动物体,尤其是遮挡情况下分割和定位多个运动物体,是一个非常具有挑战性的问题.

许多跟踪框架利用下面两类途径来跟踪物体:一类是特征对应的途径,即通过前后帧同一物体的特征相互对应来跟踪物体;另一种是状态预测的途径,即根据物体历史上的状态变化规律预测下一帧的状态来跟踪物体. 从特征对应角度来看,一些方法采用全局特征描述,很难独立解决遮挡情况下的问题,比如,基于物体全局特征(大小、颜色、形状、速度或者质心)以及基于物体主动轮廓的跟踪方法. 另有一些方法是基于运动物体的局部特征(如特征点),能部分解决遮挡情况下的问题,但确定特征点属于哪一个物体是一个难题^[7]. 基于物体 2D/3D 空间模型的方法^[3,6,8-17],能处理部分遮挡情况下的行人定位. 但这种方法依赖于物体的类别(行人)以及姿态形状(直立)先验信息,而且,复杂模型的参数估计的计算量很大. 从状态预测角度看,许多跟踪框架常常假设已知物体运动模型或帧间位移小^[6,12,16-22],所以基于滤波器的模型被广泛用于遮挡过程中的目标定位. 但当目标(如人)存在随意性运动或帧间位移较大时,此类假设不合理,可能造成跟踪错误^[23-24]. 总体上看,在这些框架中,一些先验信息往往被采用,如物体种类(用来构建特定的 2D 或 3D 模型)、运动速度假设、地平面信息或场景深度信息等^[8-9,25-26].

从跟踪结果的关注层面上看,一般分为两个层面:轨迹层面和具体区域层面. 前者,主要关注的是目标质心的轨迹,遮挡过程中关注多个目标的定位,而不关注于对各目标所属区域进行具体划分. 上面提到的绝大多数属于轨迹层面上的跟踪. 而后者,重

点关注遮挡过程中多个目标的具体区域的区分,不仅对多个目标进行定位,而且进行区域分割^[6,15-16]. 为了对多个目标进行定位和分割,这些方法都对跟踪目标——行人建立了表观模型. 文献[27]将人(胳膊、头、脚、上衣和裤子)描述成各个连接的彩色的区域,用来跟踪场景中的单人,但没有涉及多人遮挡时的情况. 而文献[6]借助这一思想,着重解决多人遮挡过程中的跟踪以及区域分割问题,将直立姿态的行人描述成一组垂直排列的区域,分别表达人体的主要部分(头部、上身和下身),然后用最大相似分类器对前景区域的每个像素进行分类,从而实现对区域进行分割. 文献[15]将直立的行人描述为头、上身、下身以及皮肤区域的模型,假设各区域内颜色特征各向同性,对相互靠近的行人各部分区域进行区分,但没有对于严重遮挡情况进行研究. 文献[16]用方框前景区域的颜色特征以及空间特征来对直立的行人进行建模,在假设行人总面积、姿态以及位置变化不大的前提下,给出了多个目标区域分割的结果. 这些多目标分割定位方法不需要已知地平面的深度信息,但都依赖于直立姿态假设以及帧间位移小的假设,而这些假设在许多应用场合下是不成立的. 而且这种方法计算前景区域所有像素的最大相似度也需要很大的计算量(为了减少计算量,Hu 等人^[17]对文献[6]作了一些简化,对多目标进行定位而不分割区域,但仍依赖于先验的垂直椭圆人模型以及空间上的直立姿态假设). 当物体类型不确定(多种未知类型物体)时,以上这些方法也不适用.

针对单目固定摄像头条件下的多目标跟踪,特别是遮挡过程中的多目标跟踪,我们对最新的各种方法进行了分析和比较,详见表 1.

表 1 单目固定摄像头条件下,在相互遮挡过程中对多个目标进行跟踪的方法分析与比较

作者	年份	目标类别	先验信息	目标模型	目标特征	分类元素	跟踪层次	跟踪结果	可以处理的遮挡程度
Haritaoglu ^[3]	2000	人	直立的行人	2D 轮廓	形状,纹理	像素	轨迹	定位	严重
Oliver ^[14]	2000	人	直立的行人,帧间运动规律	2D 方框区域	颜色,运动	像素	轨迹	定位	轻微
Elgammal ^[6]	2001	人	直立的行人,区域颜色各向同性,帧间位移小	2D 多区域(头/上身/下身模型)	颜色,运动	像素	具体区域	定位+分割	严重
Senior ^[10-11,21]	2006	不限	帧间位移小,形变不大	颜色模型	颜色,形状,运动	像素	轨迹	定位	严重
Zhao ^[8-9]	2004	人	摄像头模型,平坦地面,直立的行人,帧间位移小	3D 椭球体	形状,运动	像素	轨迹	定位	轻微
Hu ^[17]	2004	人	直立的行人,帧间位移小	2D 椭圆型模型	颜色,运动	像素	轨迹	定位	严重
Conte ^[20]	2006	不限	帧间位移小	多尺度的区域-图模型	大小,颜色,运动	像素,区域	轨迹	定位	严重
Park ^[15]	2006	人	直立的行人,区域颜色各向同性	2D 多区域(头/上身/下身)	颜色	像素	具体区域	定位+分割	轻微(相似颜色的人严重遮挡时不能区分)

(续 表)									
作者	年份	目标类别	先验信息	目标模型	目标特征	分类元素	跟踪层次	跟踪结果	可以处理的遮挡程度
Xu M ^[22]	2006	不限	帧间位移小	方框	场景信息	像素	轨迹	定位	严重
Xu L Q ^[26]	2006	人	地平面可见	颜色直方图	颜色,空间深度信息	像素	轨迹	定位	严重
Wang ^[16]	2007	人	直立的行人,遮挡过程形变不大,区域颜色各向同性,帧间位移小	2D 方框	颜色,运动	像素	具体区域	定位+分割	严重
本文方法	—	不限	帧间纹理变化不大	局部块模型	纹理	局部块	具体区域	定位+分割	严重

在本文中,我们提出一种新的基于在线采样的方法来解决遮挡过程中多目标区域的分割与定位问题.它能在物体模型、物体姿态、地平面信息都未知以及运动不规则的情况下快速、稳定地对多个目标进行定位和区域分割.新方法对遮挡前若干帧物体区域进行在线块采样,作为分类的训练样本,然后在线学习,来对遮挡时的前景区域的块采样进行标号分类.与前人工作相比,我们的方法有以下特点:

(1)基于块采样的表观描述,在体现局部纹理信息的同时,大大简化了计算量.文献[6,15-16]中的方法都是对每个像素点进行分类,计算量很大.而且,在文献[6,15]中假设一个区域内的所有像素都具有统一的概率分布,文献[16]中认为像素和周围像素颜色相一致,没有考虑像素间的纹理信息.

(2)不依赖于物体运动假设,可以用于帧间运动不规则或者帧间位移较大的情况.文献[6,16]中的方法是假设帧间位移小,用前一帧的位置作为当前帧的位置初始来判断分类,当帧间位置差异较大并且物体间距很小时,可能产生累计误差,甚至造成错误.新方法不依赖于这种假设,而且可以和传统方法相结合,在具体条件下减少由于速度估计误差造成的跟踪错误.

(3)不依赖于物体类型的先验假设,能同时跟踪多个类型未知的物体(行人和车辆),也适用于姿态变化或者局部区域消失在场景外的情况.而文献[6,15-16]中的方法假设跟踪目标都是直立姿态的人,并且假设整个跟踪过程中行人身体各部分都在场景视野范围内.这样就不能适用于其它类型物体的跟踪,当行人不是直立姿态或者局部区域消失在视野之外时也容易出现错误.

本文第 2 节详细解释我们方法.2.1 节介绍系统框架.从 2.2 节~2.7 节,分别介绍各个步骤:样本提取、在线学习、标号分类、遮挡判断、在线更新训练样本和后处理;第 3 节列出了实验结果;第 4 节给出结论.

2 方法描述

2.1 系统框架

在一些复杂监控情况下,跟踪前并不知道地平面信息,也不知道跟踪目标的形状或者运动规律的模型.这时,无法在跟踪前对物体空间模型和速度规律进行合理假设.为了对遮挡过程中的多物体区域进行更鲁棒的分割和定位,我们提出了一个新的基于在线采样的方法,不限定目标类别和运动规律,仅利用遮挡发生前若干帧的表观信息,来解决这个问题.系统流程图见图 1.

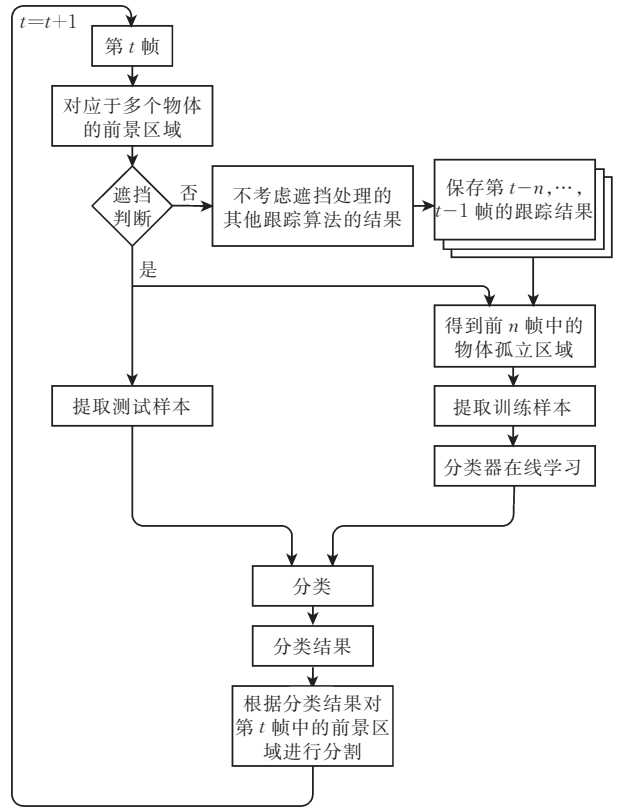


图 1 跟踪处理的系统流程图

对于多个运动物体的跟踪,许多方法能在没有遮挡的情况对多个运动物体进行很好的跟踪,而且能判断物体间遮挡的开始和结束.所以,遮挡发生前

的多个孤立物体的前景区域可以得到. 在系统流程中, 遮挡处理算法的输入信息是运动物体对应的前景区域, 这些区域是用背景差分算法^[27]得到的.

遮挡发生前的物体信息被保留用来在线更新训练样本. 当一些孤立的对应于多个运动物体的前景区域合并为一个时, 认为遮挡发生. 我们重点需要解决的问题是在遮挡过程中将前景区域正确地分割成若干物体. 本文希望利用“在线采集样本-在线训练分类器-在线分类”的思路, 来解决遮挡过程中多目标定位和区域分割问题.

基于目标样本进行类别识别的分类方法已广泛应用于物体类别识别与检测中, 但在在线跟踪问题中, 有以下几点不同: (1) 由于是在线采样, 包含完整物体的样本远远小于需要的数量; (2) 不能区分同类的两个目标, 因而也就不能确定帧间的对应关系. 所以, 传统的样本学习分类方法不适用于在线跟踪.

新方法将前若干帧的物体区域进行同样大小的块采样来得到带标号的样本, 主要基于以下几点考虑: (1) 物体的局部特征在遮挡情况下仍有可能可见; (2) 虽然在全局上物体可能存在很大的形变, 但短时间内同一物体各个局部的纹理具有一致性. 这种样本和传统的基于样本训练分类器的方法有很大不同: (1) 同一物体的所有采样块带都有相同的标号, 虽然可能这些块分别代表了这个物体的不同部位(比如人的头部和脚部); (2) 不同物体的采样块带有不同的标号, 虽然可能这些块都是同类部位(比如都是人的脸部).

考虑到以下 3 个方面: 需要在线训练分类器, 并在线分类; 训练样本的数目是很小的; 样本的特殊性, 经过实验, 我们选择 K 近邻分类器, 对当前帧前景区域得到的样本块进行物体标号的分类. 然后, 对分类结果进行后处理, 并将前景区域分割给多个物体. 通过将区域分块, 新方法能对遮挡过程中的多目标进行定位和区域分割, 甚至可以伴随着严重的形

变、剧烈的姿态变化以及速度的不规律. 显然, 在一些具体情形下, 我们的方法也可以结合其它信息来得出更准确的分割和跟踪结果.

2.2 提取样本

假设在遮挡发生前孤立物体被很好地跟踪, 前景区域都标号正确. 每个孤立的物体最近 10~20 帧的信息被保留并不断更新. 遮挡前 n 帧的物体区域被用来得到训练样本. 这些样本块通过一定大小的正方形区域在前景区域上水平和垂直方向的平移来获得. 如果窗口中的前景区域面积小于窗口面积的一半, 这个样本将被丢弃. 采样窗口的宽度是 L 个像素, 平移间距一般是 $L/2$ 个像素. 窗口的大小 L 一般在 15~25 像素之间, 帧数 n 一般在 5~10 帧之间, 至少 3 帧. 这些参数是在遮挡发生前根据场景中涉及遮挡的物体区域确定的, 用来得到每个物体不少于 100 个的样本块. 如果最小的物体区域不能提取到多于 10 个的块, 窗口的大小将变小, 而且帧数将增加. 有时, 跟踪的物体的面积有很大差异, 在遮挡涉及的多个物体中, 遮挡前总面积最小的那个物体的像素总数设为 S , 则窗口的宽度

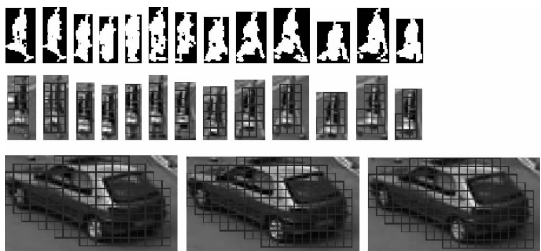
$$L = \begin{cases} \sqrt{S/5}, & \sqrt{S/5} < 15 \\ 15, & 15 \leq \sqrt{S/5} < 25. \\ 25, & 25 \leq \sqrt{S/5} \end{cases}$$

比如, 一个人和一辆车相比非常小, 如图 2(a)所示(数据来源于 PETS2001 的数据集 1(摄像头 1)). 较小的物体(人)的区域只能提取到小于 10 个的块, 所以针对人的帧数 n 应该较大($n > 10$). 而对于车辆的区域每帧都能提取接近 100 个块, 如图 2(b)所示. 对于车辆来说 3 帧足够了.

这样, 对于每个物体, 都有至少 3 帧并且总数超过 100 个的样本块, 被用来作为监督学习的训练样本. 如果采样窗口的宽度设为 L , 每个采样的像素数是 $P = L \times L$. 将这 P 个像素转化为一个向量 $\mathbf{V}_{\text{train}} = [V_1, V_2, \dots, V_p]$.



(a) 一个行人从车辆边经过



(b) 采样结果(第 1 行: 第 540~552 帧的行人的模板. 中间行: 第 540~552 帧的行人的采样. 最下行: 第 550~552 帧的车辆的采样)

2.3 在线学习

我们将两个物体的所有采样块都看作是训练样本,没有考虑属于一个物体的采样块之间的差异.这和传统的训练样本不同.对于属于同一物体(比如人)的两块,可能一个来自头部,一个来自腿部,而这两块的标号是相同的.有了这些带标号的训练样本,一些监督学习的方法能产生一组分类器.考虑到在线学习分类以及样本的特殊性,我们采用 K 近邻的方法.由于剧烈的姿态变化或者由于遮挡产生的形变,当前帧物体的整体表现观信息可能与前几帧有较大的差异,但是当它们被分为小块后,训练样本和测试样本还是有很高相似性的.从实验结果可以看到, K 近邻方法能在一些复杂情况下取得好的结果.

2.4 标号分类

当新的一帧来到时(这里我们仅考虑两个物体的情况),如果遮挡判断模块已经判断物体进入遮挡,对应于两个物体的前景区域需要被划分.利用前

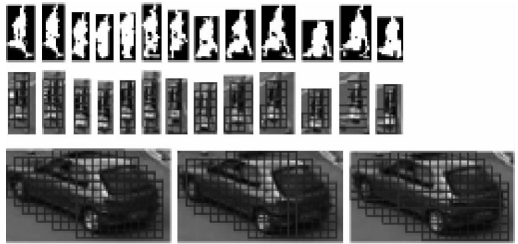
面提到的采样方法,从当前帧的前景区域中得到的采样块作为测试样本.这些样本都被转换成向量的形式 $V_{test}=[V_1,V_2,\cdots,V_p]$. K 近邻分类器的相似度评价函数为 $f=\|V_{train}-V_{test}\|$. 所有的测试样本块通过分类器得到属于各物体标号的隶属度.隶属度被用来描述样本在多大程度上属于某一类,用 A 表示.如果种类的个数是两个, $A=\{(a_1,a_2)|a_1+a_2=1,a_1\geq 0,a_2\geq 0\}$, a_1 是属于第 i 类的可能性.如果用 $K=3$ 近邻分类器, $a_i\in\{0,0.333,0.666,1\}$. 如果 $A=(1,0)$,样本被判断为第 1 类,它的中心被填为浅色;如果 $A=(0,1)$,样本被判断为属于第 2 类,它的中心被填为深色.如果 $A=(0.333,0.666)$ 或者 $A=(0.666,0.333)$,样本的类别不能准确确定下来,这种样本的中心填为黑色.

2.5 遮挡判断

已有的大多数跟踪算法,可以对遮挡发生和结束进行判断.利用这些算法,当一些对应于不同物体



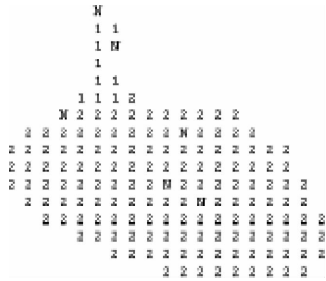
(a) 一个行人从车辆边经过(第540帧)



(b) 采样结果. 第一行: 第540~552帧的行人的模板(中间行: 第540~552帧的行人的采样.最下行: 第550~552帧的车辆采样)



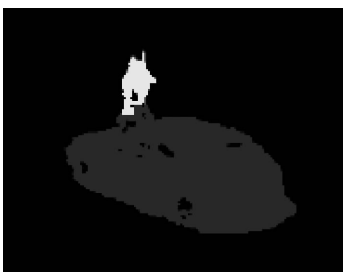
(c) 第560帧的分类结果,所有测试样本的中心被填上不同颜色.浅色代表被分为一类;深色被分为另一类;黑色表示样本块的类别不确定



(d) 对应于(c),每个样本的数字代表类别,N代表不确定



(e) 第560帧的块标号后处理结果,所有测试样本的中心点都填充了不同颜色.浅色点表示该样本块分到物体 1 的类别;深色点表示该样本被分到了物体 2 的类别.注意,这里面已经没有不确定的块



(f) 根据(e)得到的第560帧前景区域分割结果



(g) 根据(f)得到的各物体的区域范围

的前景区域合并为一个区域时,认为遮挡发生;当对应于多个物体的前景区域分裂为多个区域时,认为遮挡结束.我们的方法同这些方法相结合,不仅可以在遮挡发生整个过程中持续稳定地跟踪物体,还可以避免由于前景提取造成的遮挡发生与结束的误判.在我们的系统中,如果当前帧同时满足以下两个条件,认为遮挡发生:(1)多个物体的孤立区域合并为一个区域;(2)超过 20% 的标号结果和不进行遮挡处理的算法的标号结果不同.反之,如果当前帧同时满足以下两个条件,认为遮挡结束:(1)对应于多个物体的前景区域分裂为多个区域;(2)每个物体超过 80% 的标号结果都在同一个区域内.另外,根据分类结果,当遮挡结束,可以得到每个新区域和物体标号的对应关系.

2.6 训练样本的在线更新

在遮挡前,系统始终保留最近 m 帧的物体区域跟踪结果, m 可以是 10~20 之间,这些结果可以从不处理遮挡的其它跟踪算法得到的.如果在当前帧上遮挡没有发生,当前帧物体的区域被保留下来,最早帧的区域信息将被丢掉.这样在线更新,如果遮挡发生,训练样本是从遮挡发生前最近的一些帧里提取.接着,由这些在线的训练样本训练得到分类器.注意,只有几百个从先前帧得到的训练样本被用来训练分类器.所以,这种基于在线采样的分类器训练计算量的开销并不很大.

2.7 后处理

分类结束后,有一些测试样本块的标号还没有确定,所以还需要后处理步骤.我们根据周围点的标

号结果,确定所有块的标号.对于黑色点(不确定的点),如果在 8 个邻域中的点中,深色或浅色的点的数目超过邻域中有效点数的一半,该块被重新标上深色或浅色.另一方面,对于浅色和深色的点,如果和它周围的大多数点都不同,它的颜色会重新标,来对整体进行平滑处理.最后,前景区域根据块的标号进行填充.根据这个结果,我们能得到许多判断,比如,遮挡深度顺序、遮挡程度以及前景区域的对应关系.在图 3 中,列出了以上框架中每一步的结果(数据来自 PETS2001 的数据集 1 (摄像头 1)).

3 实验结果

为了验证新方法在各种复杂情况下的稳定性以及有效性,实验数据选择各种有代表性的情况,有的来源于 PETS 公共数据库^①,有些是自拍的典型数据.下面列的实验结果中,室外有 4 段数据,室内有两段,重点显示多个运动物体在发生遮挡过程中的跟踪结果.一些实验图像序列是灰度图,另一些是彩色图,彩色数据主要用来和其它方法进行比较.前景区域的标号结果用不同颜色(灰度)的点表示(浅色表示属于标号为 1 的目标物体,深色表示属于标号为 2 的目标物体).

图 4 显示了两个人发生遮挡过程中的跟踪结果.注意,他们的颜色很相似,依靠颜色信息很难区分.利用前后帧局部纹理的相似性,我们的方法能成功跟踪两人.遮挡发生在第 364~385 帧之间.



图 4 最上面的是遮挡发生前的一帧,其它的是实验结果:从左到右依次对应于第 364,373,379,385 帧

实验结果显示,我们的方法能够较好地同时跟踪不同类型的物体.图5显示了当一个人和一辆车发生严重遮挡过程的关键帧,数据来自 PETS 公共数据库^①.遮挡发生在第 552 帧~587 帧之间.这时,基于先验模型的人的跟踪算法^[6,15-16]可能不太适用.在一些情况下,物体运动速度很不规律或者由于跳帧导致帧间位移很不规律,此时基于速度预测

的跟踪方法可能会在遮挡发生过程出现误差.利用遮挡前的信息,我们的方法可以对较长时间间隔之后的图像中物体进行较稳定的跟踪.而且,新方法不需要场景深度信息,在地平面不可见的情况也能很稳定地跟踪.图6显示了3个人在严重遮挡过程中的跟踪关键帧.3人遮挡发生在第 67 帧~112 帧之间.此数据中遮挡持续了大概 3s,而且地面不可见.



图5 实验结果:从左到右帧号分别是第 553,563,573,581



图6 实验结果:从左到右分别对应于第 50,80,98,112 帧

新方法能在长时间完全遮挡后稳定地跟踪不规律运动的物体.图7显示了两个人先后被树完全遮挡,其中第一个人被树遮挡后停留片刻,此时第二个人也从树后经过,第二个人先从树后面走出来,接

着第一个人按原路返回这段情景的跟踪关键帧.第一人被树完全遮挡是在第 122~296 帧之间,第二人被树完全遮挡是在遮挡发生在第 253~309 帧之间.

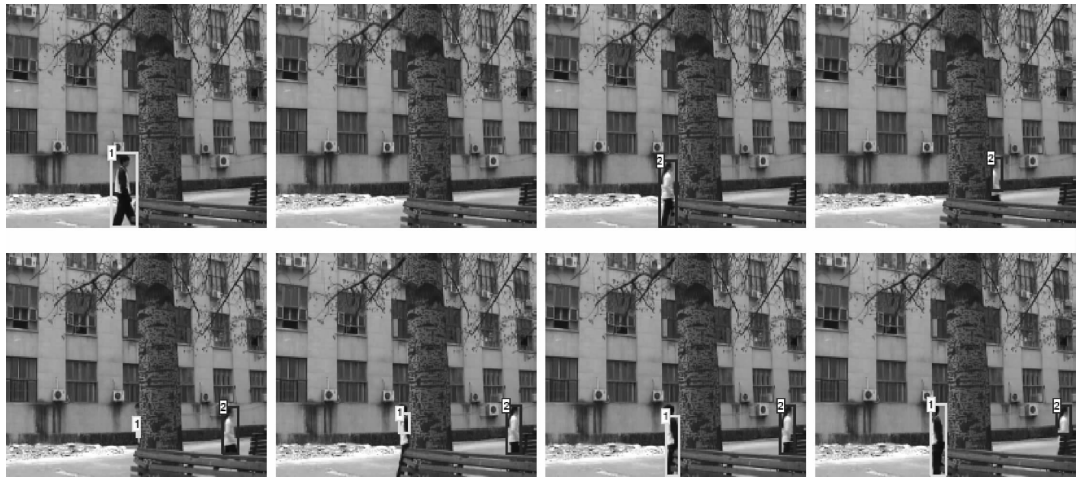


图7 实验结果:第一行从左到右分别对应于第 121,200,252,277 帧,第二行从左到右分别对应于第 296,304,306,309 帧

我们针对的问题是在遮挡情况下对多目标区域进行分割和跟踪,不仅仅是轨迹层面上的定位,还包括具体区域层面的分割.在我们所看到的最新的方法中,与之类似的有文献^[6,15-16]中的方法,参见表1.而其中文献^[15]没有具体涉及严重遮挡情况.文献^[16]重点在背景建模上,对于遮挡情况下的分

割,描述很简单,参数阈值的确定方法不明确,因而难以复现.同时,该方法依赖于以下假设:(1)限于直立的行人;(2)在整个遮挡过程中,行人姿态、在场景中的总面积以及帧间绝对位置变化不大.在这两方面,和文献^[6]类似.因此,我们着重于将新方法同文献^[6]的工作进行了比较实验,如图8、图9所示.

图 8 显示的是两个人上半身有类似的概率密度而纹理不同,发生遮挡后,我们的方法可以很好的工作,而文献[6]中的方法会跟踪失败.图 9 显示的是

在遮挡过程中,人们突然蹲下或者其它不同于直立姿态,这时,我们的方法能工作得很好,而文献[6]中的方法往往会出现错误.

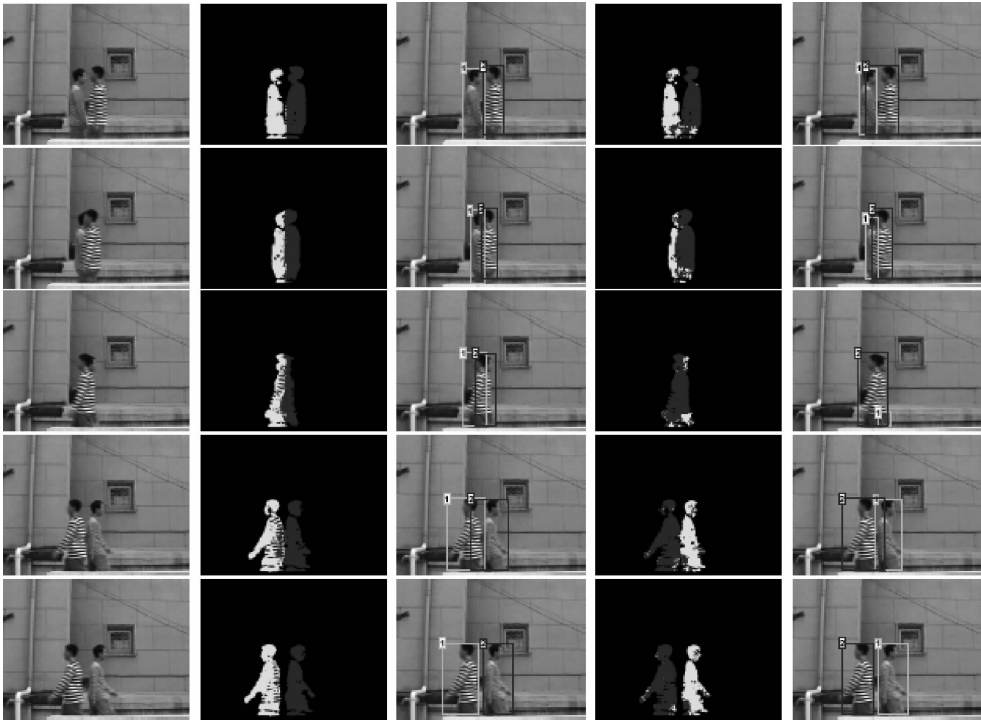


图 8 结果比较 1(左列:原始图像,中间两列:采用参考文献[6]中方法的结果,右边两列:采用新方法的结果,从上到下的帧号依次是第 27,28,29,30,31 帧)



图 9 结果比较 2(左列:原始图像,中间两列:采用参考文献[6]中方法的结果,右边两列:采用新方法的结果,从上到下的帧号依次是第 34,36,38,40,42 帧)

通过对实验结果的分析,与前人方法特别是在严重遮挡情况下对区域进行分割的方法^[6,16]相比,本文方法有如下特点:

(1)如图5所示,我们方法能稳定地同时跟踪多个未知类别的物体,不依赖于先验物体模型,而在文献[6,16]中,跟踪物体的先验模型是前提假设之一。

(2)我们的方法是基于块采样的分类方法,能区分局部的纹理信息,而基于区域颜色各向同性^[6]或相邻点颜色一致^[16]的分类方法不能区分纹理上的差异。

(3)文献[6,16]中的方法假设物体是保持直立姿态的,文献[16]还假设遮挡过程中,物体姿态以及场景内面积变化不大。这对于真实场景过于苛刻。而我们的方法没有这样的姿态假设,而且块采样对于姿态变化比较鲁棒,可以适用于遮挡过程中物体部分区域消失在场景外的情况(见图9)。

(4)我们的方法不依赖于速度平滑的假设,能在物体运动不规律或者帧间位移大的情况下正常工作,而文献[6,16]中的方法有一个假设是目标在连续两帧间的位移不大。

(5)我们没有采用基于概率的方法,也不是对每个像素进行分类,而是对块进行分类,从而大大减少了计算量,避免了在文献[6,16]中对像素分类高昂的计算消耗。

4 结 论

在本文中,我们提出了一种新的基于在线样本的分类方法来对遮挡过程中的多目标进行跟踪。训练样本是从遮挡发生前的物体区域提取的,基于这些在线更新的训练样本在线训练得到分类器。遮挡发生时,对应于多物体的前景区域也同样被分成若干块,通过分类器得到分类标号。本文方法的贡献是:(1)它不需要跟踪前的先验假设,新方法既不需要场景的假设条件,也不需要跟踪前物体的任何先验信息,比如,基于类型的物体空间结构模型、物体速度规律假设。(2)它采用基于在线采样,利用 K 近邻分类器,对前景样本块进行在线分类,避免了统计方法的复杂性和高计算量。实验表明新方法能在一些复杂的情况下对严重遮挡过程的物体进行正确的跟踪,比如,巨大的形变、地平面未知或者物体的类别未知。(3)在特定条件下,它能很容易地和其它方法相结合,来提高跟踪准确性。

未来的工作包括在遮挡过程中在线更新训练样本以及和其它方法相结合来提高分类的准确性。

参 考 文 献

- [1] Amer A, Dubois E, Mitiche A. Real-time system for high-level video representation: Application to video surveillance//Proceedings of the Conference on Visual Communication and Image Processing (VCIP). Santa Clara, USA, 2003; Proceedings of the SPIE International, 2003, 5022: 530-541
- [2] Collins R, Lipton A, Kanade T. A system for video surveillance and monitoring//Proceedings of the American Nuclear Society (ANS) 8th International Topical Meeting on Robotics and Remote Systems. 1999
- [3] Harwood D, Haritaoglu I, Davis L S. W4: Real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809-830
- [4] Lv F, Kang J, Nevatia R, Cohen I, Medioni G. Automatic tracking and labeling of human activities in a video sequence//Proceedings of the 6th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS04). Prague, Czech Republic, 2004.
- [5] Stauffer C, Grimson W. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-757
- [6] Elgammal A M, Davis L S. Probabilistic framework for segmenting people under occlusion//Proceedings of the 8th IEEE International Conference on Computer Vision, IEEE Computer Soc. Los Alamitos, 2001, II: 145-152
- [7] Zang Q, Klette R. Object classification and tracking in video surveillance//Proceeding of the Computer Analysis of Images and Patterns. Lecture Notes in Computer Science 2756. Berlin: Springer, 2003: 198-205
- [8] Zhao T, Nevatia R. Tracking multiple humans in crowded environment//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Soc. Washington DC, 2004, II: 406-413
- [9] Zhao T, Nevatia R. Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2004, 26(9): 1208-1221
- [10] Senior A, Hampapur A, Tian Y L, Brown L, Pankanti S, Bolle R. Appearance models for occlusion handling//Proceedings of the 2nd IEEE Workshop on Performance Evaluation of Tracking and Surveillance (PETS). Kauai, Hawaii, USA, 2001
- [11] Senior A. Tracking with probabilistic appearance models//Proceedings of the ECCV Workshop on Performance Evaluation of Tracking and Surveillance Systems. 2002: 48-55
- [12] Babu R V, Perez P, Boutheimy P. Robust tracking with motion estimation and local kernel-based color modeling. Image and Vision Computing, 2007, 25(8): 1205-1216
- [13] Park S, Trivedi M. Multi-person interaction and activity analysis: A synergistic track-and body-level analysis framework. Machine Vision and Applications, 2007, 18(3/4):

- 151-166
- [14] Oliver N M, Rosario B, Pentland A P. A Bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 831-843
 - [15] Park S, Aggarwal J K. Simultaneous tracking of multiple body parts of interacting persons. *Computer Vision and Image Understanding*, 2006, 102(1): 1-21
 - [16] Wang H Z, Suter D. A consensus-based method for tracking: Modelling background scenario and foreground appearance. *Pattern Recognition*, 2007, 40(3): 1091-1105
 - [17] Hu M, Hu W M, Tan T N. Tracking people through occlusions//*Proceedings of the 17th International Conference on Pattern Recognition*. Los Alamitos, 2004, 2: 724-727
 - [18] Rosales R, Sclaroff S. Improved tracking of multiple humans with trajectory prediction and occlusion modeling//*Proceedings of the IEEE Workshop on Interpretation of Visual Motion*. Santa Barbara, CA, 1998
 - [19] Weng S K, Kuo C M, Tu S K. Video object tracking using adaptive Kalman filter. *Journal of Visual Communication and Image Representation*, 2006, 7(6): 1190-1208
 - [20] Conte D, Foggia P, Jolion JM, Vento M. A graph-based, multi-resolution algorithm for tracking objects in presence of occlusions. *Pattern Recognition*, 2006, 39(4): 562-572
 - [21] Senior A, Hampapur A, Tian YL et al. Appearance models for occlusion handling. *Image and Vision Computing*, 2006, 24(11): 1233-1243
 - [22] Xu M, Ellis T. Augmented tracking with incomplete observation and probabilistic reasoning. *Image and Vision Computing*, 2006, 24(11): 1202-1217
 - [23] Hotter M, Thoma R. Image segmentation based on object oriented mapping parameter estimation. *Signal Processing*, 1988, 15(3): 315-334
 - [24] Lim K, Das A, Chong M. Estimation of occlusion and dense motion fields in a bidirectional Bayesian framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 712-718
 - [25] Pang C C, Lam W L W, Yung N H C. A novel algorithm for the handling of occlusion for visual traffic surveillance//*Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*. Bellingham, WA, USA, 2003, 5014: 437-447
 - [26] Xu L Q, Puig P. Multi-level attention model for tracking and segmentation of objects under complex occlusion. *Bt Technology Journal*, 2006, 2(24): 180-185
 - [27] Wren C R, Azarbayejani A, Darrell T, Pentland A P. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 1997, 19(7): 780-785



ZHU Lin, born in 1978, Ph.D. candidate. Her research interests include computer vision, machine learning, video surveillance.

ZHOU Jie, born in 1968, professor, Ph.D. supervisor. His research interests include pattern recognition, image processing, computer vision, multimedia processing and information service science and technology.

SONG Jing-Yan, born in 1964, professor, Ph.D. supervisor. His research interests include intelligent control, intelligent transportation system.

Background

In video surveillance systems, accurate and real-time track for multiple objects is greatly helpful to object recognition, activity analysis and high-level event understanding. Segmentation and tracking multiple objects is important not only for visual surveillance, but also for other video analysis application such as video indexing, video archival and retrieval system. Many tracking systems now can track the isolated objects well, but may lose the objects for a period of time or even miss the objects through occlusion. So, multiple objects tracking, especially multiple objects tracking with occlusion, is an important and challenging problem.

Recently, many researches aim to handle the problem of tracking multiple objects though occlusion by using some prior assumptions, such as, the deep information of the scene known, the kind of objects known (to build the 2D or 3D model), motion regularity. But in real complex situations,

the prior assumptions may be unreasonable, which may cause tracking failure. However human being can stably track objects with occlusion by online learning the information of the objects, even if the objects' kind, motion regularity and the scene depth information are unknown.

Inspired by the online learning approach of human vision, the authors propose a novel online sample based framework to track multiple objects with occlusion. First, training samples are online obtained from each isolated object before occlusion to produce a set of classifiers. Then the foreground region through occlusion can be segmented into multiple objects by the classifiers' labelling. The authors test the proposed framework in many indoor/outdoor data, and the results show that the new approach can track multiple objects with occlusion fast and stably even when the objects' kind, motion uniform degree and the scene depth information are unknown.