

一种基于贪心搜索的实时多目标遮挡处理算法

杨涛¹ 李静² 潘泉³ 张艳宁¹

摘要 提出了一种固定摄像机遮挡条件下的多目标跟踪算法, 包括基于区域相关的运动前景分割、基于合并-分裂检测的数据关联和基于贪心搜索的遮挡目标定位三部分. 该算法的主要特点表现在: 1) 将遮挡条件下的目标跟踪问题转化为一个已知目标数量和特征的图像分类问题; 2) 用贪心搜索和积分图算法快速定位遮挡中的目标, 保证了算法的实时性; 3) 对目标数量无约束, 能够处理多目标相互遮挡下的跟踪问题 (发生遮挡的目标数量大于等于 2), 且对目标的遮挡程度和目标运动模式无约束, 具有良好的可扩展性. 采用手工标定的 IBM 多人遮挡数据库的测试结果证明了算法的有效性.

关键词 多目标检测与跟踪, 遮挡处理, 贪心搜索, 智能视频监控

DOI 10.3724/SP.J.1004.2010.00375

A Greedy Searching Algorithm for Multiple Object Tracking and Occlusion Handling

YANG Tao¹ LI Jing² PAN Quan³ ZHANG Yan-Ning¹

Abstract This paper presents a novel real-time multiple object tracking algorithm, which contains three parts: region correlation based foreground segmentation, merging-splitting based data association and greedy searching based occluded object localization. The main characteristics of the proposed algorithm are summarized as follows: 1) the multiple object tracking and occlusion handling problem is successfully changed into an image classification problem with prior knowledge of object number and feature; 2) a highly efficient greedy searching method is presented to meet real-time capability; 3) it has good performance in expansibility, and it has no constraints about the number of occluded objects, the occlusion ratio and the object's motion model. Experiment results with hand labeled IBM database demonstrate that the method is effective and efficient.

Key words Multiple object detection and tracking, occlusion handling, greedy searching, intelligent video surveillance

视频跟踪的任务是找到同一目标在连续两帧中空间位置的对应关系, 它在视频监控、人机交互、视频压缩以及自动视频分析和索引中被广泛应用. 遮挡条件下的多目标跟踪^[1-16] 是视频跟踪中的研究热点和难点之一. 遮挡产生于目标和目标之间、目标与景物之间以及目标的自遮挡, 本文重点讨论多目标跟踪中目标和目标之间的遮挡问题.

目标间的遮挡通常包括以下 5 个阶段: 1) 多目标接近; 2) 多目标合并; 3) 合并目标连续运动; 4) 多目标分离; 5) 多目标远离. 其中, 阶段 1) 和 5) 采用的方法与单目标跟踪相同, 阶段 2) 可以根据发生遮挡前后目标的空间位置进行判断, 而如何在部分或全部遮挡时对每一个目标进行跟踪 (阶段 3) 以及如何在分离时准确标定目标跟踪标号 (阶段 4) 是研究的难点.

粗略地讲, 固定摄像机条件下的遮挡处理算法可分为自上而下和自下而上两类. 前者通过运动分割检测并跟踪前景区域, 然后按照基于合并-分裂的思路进行遮挡处理; 后者则采用基于直接搜索的思路, 在遮挡的全过程始终对特定目标进行跟踪.

1) 基于合并-分裂的遮挡处理算法在遮挡的第一阶段不断对每个目标的特征进行学习和更新, 当遮挡发生时, 算法保存学习好的目标特征, 并开始跟踪合并而成的群目标. 事实上, 这类算法在目标发生遮挡时并不标出每一个目标在群中的位置, 算法关注的问题是当目标发生分离时, 如何判断哪几个目标从群目标中分离. McKenna 等^[13] 和 Bremond 等^[14] 采用外观特征, 如色彩、形状和纹理进行关联. Piater 等^[15] 融合外观特征和运动信息进行遮挡处

收稿日期 2008-10-22 录用日期 2009-01-21
Manuscript received October 22, 2008; accepted January 21, 2009

国家高技术研究发展计划 (863 计划) (2009AA01Z315), 国家自然科学基金 (60903126, 60872145, 60634030), 高等学校科技创新工程重大项目培育资金 (708085) 资助

Supported by National High Technology Research and Development Program of China (863 Program) (2009AA01Z315), National Natural Science Foundation of China (60903126, 60872145, 60634030), and Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (708085)

1. 西北工业大学计算机学院陕西省语音与图像处理重点实验室 西安 710129 2. 西安电子科技大学通信工程学院 西安 710071 3. 西北工业大学自动化学院 西安 710129

1. Shaanxi Key Laboratory of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an 710129 2. School of Telecommunications Engineering, Xidian University, Xi'an 710071 3. School of Automation, Northwestern Polytechnical University, Xi'an 710129

理,并用 Kalman 滤波器估计目标的运动状态,用彩色直方图进行目标匹配。

合并-分裂的思路适用于两个目标的遮挡问题,由于在遮挡过程中将合并目标作为一个整体进行跟踪,因此该类算法对单个目标的遮挡程度、遮挡时间等条件不敏感。当合并目标分裂时,目标在合并前提取的外观特征将被用于数据关联。该类算法的缺点是如果发生遮挡的目标数大于 2,分裂后的单个区域内可包含多个目标,且目标间的相对位置和遮挡程度均未知,此时基于目标外观特征的匹配性能将大幅度下降,从而造成失跟。

2) 基于搜索匹配的遮挡处理算法要求在整个遮挡过程中始终准确地检测和跟踪目标。由于对发生遮挡的目标个数没有限制,因此更适用于处理多目标(数量大于 2)环境中的遮挡问题。该类算法可分为基于局部最优的跟踪和基于全局最优的跟踪。

基于局部最优的跟踪算法包括相关跟踪、Mean shift 跟踪等。该类算法依据某种相似度(如互相关、归一化互相关、误差平方和等)和搜索策略(如序贯搜索、梯度上升等),在目标预测位置的邻域内寻找匹配程度最高的位置作为跟踪结果。这类算法计算量小,实时性强,对目标部分遮挡有一定的适应性,且对目标的运动方程无约束,能在目标做非线性运动下进行跟踪。但是,从本质上讲,局部最优跟踪是将多目标跟踪问题简化为多个单目标跟踪问题,当目标发生严重遮挡时,算法仍然会将当前错误的局部最优匹配结果输出,而且,当目标重新出现时,算法无法自动恢复跟踪。此外,当目标与背景或其他目标特征相似时,这类算法的鲁棒性会下降。

基于全局最优的跟踪不仅考虑单个目标的定位,还充分考虑参与遮挡的其他目标特征,最终的跟踪结果使得整体观测概率最大化。尽管从理论上讲,当遮挡发生时,可以通过遍历目标所有可能的组合得到全局最优解,然而考虑到发生遮挡的目标数量、目标可能的位置组合等因素,穷举的思路无法满足实时性的要求。为此,学者们提出一系列简化方法。Elgammal 等^[16]根据目标的空间排列估计人体的深度信息,算法首先建立人体在遮挡前的色彩和肢体模型(头部、躯干、四肢),遮挡时群目标中每一个像素按照最大概率分配给某一目标,然后用椭圆模型估计该目标的轮廓,并进行目标跟踪。Cucchiara 等^[10]将遮挡目标区域中每一个像素按照最大概率分配给某一目标。Beleznai 等^[11]采用 Mean shift 算法在多目标联合概率空间中进行搜索,并用积分图快速计算 Mean shift 向量。但是,上述算法仍存在以下不足: Cucchiara 等^[10]对每个像素进行独立地判断,损失了目标外观特征的空间信息。此外,由于在进行像素和目标关联时采用确定性的判别

准则,当多个目标色彩近似时,关联的准确性会下降。Beleznai 等^[11]对人体建模时仅在背景减除图像上搜索,损失了色彩信息,而且最终的搜索结果受 Mean shift 初始位置和窗宽的影响较大。

本文将上述两种思路相结合,提出一种新的固定摄像机多目标遮挡处理算法。算法首先采用合并-分裂思路,对多目标间的遮挡进行自动检测,然后提出一种基于全局搜索匹配的遮挡处理算法,将遮挡条件下的目标跟踪问题转化为一个已知目标特征的图像分类问题,并根据目标可视程度,采用贪心搜索算法和积分图对遮挡中的目标进行快速定位。该算法对发生遮挡的目标数量、目标被遮挡程度以及目标运动模式均无约束,实验结果证明了算法的有效性。

本文的后续部分安排如下:第 1 节给出算法的总体介绍;第 2 节给出运动前景提取和合并检测算法;第 3 节具体描述基于贪心搜索的遮挡处理算法;第 4 节给出实验结果;第 5 节是结论。

1 算法流程

算法主要包括 3 部分:1) 背景建模与前景检测;2) 数据关联与合并检测;3) 基于贪心搜索的遮挡目标定位。算法流程如图 1 所示。对于输入图像序列,首先使用背景建模技术进行运动前景检测,得到多个目标运动区域,然后用运动区域的外接矩形做外观模型,进行数据关联和目标间的合并检测。当目标未发生遮挡时,根据数据关联结果进行单目标跟踪,并在线学习和更新目标特征;当有遮挡发生时,对于包含多个目标的合并前景块,采用基于贪心搜索的遮挡处理算法,定位出每个目标在前景块中的具体位置。

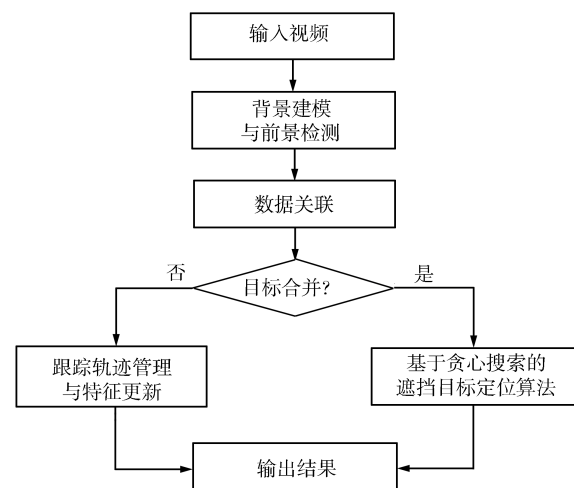


图 1 基于贪心搜索的多目标遮挡处理算法框图

Fig. 1 Greedy searching based multiple object tracking and occlusion handling algorithm

2 运动前景分割与合并检测

2.1 基于区域相关的运动前景分割

背景减除法 (Background subtraction)^[17-22]是目前应用最广泛的运动目标检测算法,其基本思想是从输入图像和背景图像的差异中提取运动前景,由于背景减除法对目标类型、观测角度、距离等因素不敏感,因此在视频监控系统中被广泛采用.背景减除法的设计关键是如何建立有效的背景模型.常见的背景建模算法包括中值滤波、线性预测、单高斯、混合高斯模型^[17]、核密度估计法、局部纹理分布^[18]等.为了进一步消除运动阴影对分割结果的干扰,色彩、边缘、亮度、运动、外观等特征也被学者广泛采用.

本文采用混合高斯模型 (Gaussian mixture model, GMM)^[17]进行背景建模,并用权值最大的高斯分布均值生成背景图像. GMM 算法简述如下:定义像素点灰度值用变量 X 表示,其概率密度函数可用如下 K 个三维高斯函数表示:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \eta \left(X_t, \mu_{i,t}, \sum_{i,t} \right) \quad (1)$$

式中 $\omega_{i,t}$ 为第 i 个高斯分布在 t 时刻的权重,且有 $\sum_{i=1}^K \omega_{i,t} = 1$, $\eta(X_t, \mu_{i,t}, \sum_{i,t})$ 是 X 在 t 时刻的第 i 个高斯分布.

$$\eta(X_t, \mu_{i,t}, \sum_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum_{i,t}|^{\frac{1}{2}}} \times e^{-\frac{1}{2}(X_t - \mu_{i,t})^T \sum_{i,t}^{-1} (X_t - \mu_{i,t})} \quad (2)$$

式中 $i = 1, \dots, K$, n 表示 X_t 的维数,为了降低计算量,通常认为 R, G, B 三个通道相互独立,并有相同的方差,则有 $\sum_{i,t} = \sigma_i^2 I$. 本文将 GMM 中权值最大的几个高斯分布作为动态背景,如果某一像素灰度值不符合所有可能的动态背景分布,则判别为变化前景.此外,为了降低变量 X_t 维数,提高算法的实时性,在预处理时先将彩色图像灰度化,然后用混合高斯对灰度图像进行建模.

由于像素级差分算法对每一个像素孤立地进行分析比较,忽略相邻像素之间的相关性,使得该类方法易受环境的干扰:当前景色彩与背景相似时,该类算法无法检测这部分前景,从而导致漏检率升高;当出现运动阴影时,阴影区的像素灰度会发生变化,算法会将该区域当作运动前景输出,使得虚警率上升.针对上述问题,本文在 GMM 检测结果的基础上,首先遍历整幅图像,估计前景点的空间概率分布.对于像素点 (u, v) ,根据 $(2N+1) \times (2N+1)$ 邻域内前景

点 FT_t 的平均点数 D 的大小提取候选前景点 CF :

$$D(u, v) = \frac{1}{(2N+1)^2} \sum_{y=v-N}^{v+N} \sum_{x=u-N}^{u+N} FT_t(x, y)$$

$$f = \begin{cases} 1, & \text{若 } D(u, v) \geq T_c \\ 0, & \text{否则} \end{cases} \quad (3)$$

实验中,阈值 $T_c = 0.1$. 考虑到阴影造成的图像亮度线性变化,对于每个候选前景点采用对信号强度线性变化不敏感的归一化互相关作为相似性度量,降低运动阴影干扰.令 I 和 B 分别表示输入图像和参考背景,则候选前景点 (u, v) 的 $(2M+1) \times (2M+1)$ 邻域内归一化相关系数的原始计算公式如下:

$$NCC(u, v) = \frac{\sum_{\Delta v=-M}^M \sum_{\Delta u=-M}^M (B(u+\Delta u, v+\Delta v)I(u+\Delta u, v+\Delta v))}{\left(\sqrt{\sum_{\Delta v=-M}^M \sum_{\Delta u=-M}^M B(u+\Delta u, v+\Delta v)^2} \times \sqrt{\sum_{\Delta v=-M}^M \sum_{\Delta u=-M}^M I(u+\Delta u, v+\Delta v)^2} \right)} \quad (4)$$

在计算时,式(4)中的三部分求和用积分图法快速计算.实验表明,该算法可有效地提高算法在相似背景条件下的检测率,减少运动阴影的干扰.

2.2 数据关联与合并检测

用 $T_i, i = 1, \dots, N_1$ 表示被跟踪的历史目标,用 $D_j, j = 1, \dots, N_2$ 表示从当前帧中运动分割出来的观测目标.算法根据历史目标 T 与当前观测 D 的匹配度,采用作者在文献[23]中提出的算法进行数据关联和合并检测.此时需要求解两个矩阵:匹配矩阵 M 和数据关联矩阵 C .

1) 匹配矩阵 M

求解 T 和 D 之间的匹配度可采用多种相似性度量,包括矩形中心的欧氏距离,矩形内部外观、色彩、矩特征的相似性等.由于物体的运动是时空连续的,目标的外接矩形在合并前后会有很大重叠,因此,采用外接矩形重叠区域面积比率作为距离度量,生成匹配矩阵 M .

$$M(i, j) = c \frac{S_{T_i \cap D_j}}{S_{T_i} + S_{D_j}} \quad (5)$$

其中, $S_{T_i \cap D_j}$ 表示历史目标 T_i 和观测目标 D_j 的重叠区域面积, S_{T_i} 和 S_{D_j} 分别表示 T_i 和 D_j 的外接矩形面积,系数 c 是规范化常数,且有 $c = 2$,使得 $M(i, j) \in [0, 1]$. 与欧氏距离相比,该度量综合考虑了目标的空间和形状大小两种特征,因此具有更强的稳定性.

2) 数据关联矩阵 C

获得匹配矩阵 M 后, 根据相似度大小, 通过双向匹配求解目标 T 和观测 D 的关联矩阵 C . 初始时矩阵 C 的所有元素均为 0, 矩阵 C 计算过程如下:

a) 逐行扫描矩阵 M , 对于每个目标 T_i , 找出匹配程度最高的观测所在列 j_{\max} , 并将矩阵 C 的对应位置加 1:

$$j_{\max} = \arg \max_j M(i, j), \quad j = 1, \dots, N_1$$

$$C(i, j_{\max}) = C(i, j_{\max}) + 1 \quad (6)$$

b) 逐列扫描矩阵 M , 对于当前帧的每个观测 D_j , 找出匹配程度最高的目标所在行 j_{\max} , 并将矩阵 C 的对应位置加 1:

$$j_{\max} = \arg \max_i M(i, j), \quad i = 1, \dots, N_2$$

$$C(i_{\max}, j) = C(i_{\max}, j) + 1 \quad (7)$$

c) 扫描关联矩阵 C , 如果存在值为 2 的元素, 将该元素对应的目标和观测相关联, 并将匹配矩阵 M 中相应的行列所有元素赋值为 0, 然后从步骤 a) 重新开始; 如果 C 中所有元素均小于 2, 则中止扫描.

完成关联矩阵 C 的计算后, 历史目标 T 与当前观测 D 将被分为 3 类: 1) 获得观测的历史目标; 2) 未获得观测的历史目标; 3) 未被关联到目标的观测. 对于没有观测的历史目标, 可能的原因包括目标消失、目标被景物的遮挡、相似背景造成的漏检以及目标合并等; 对于没有关联到目标的观测, 可能的原因包括新目标出现、旧目标分裂以及虚警噪声等. 在此, 如果未获得观测的历史目标与本帧某一被关联的观测发生重叠, 则该观测为多个目标合并而成.

基于合并-分裂的遮挡处理算法^[6, 23]将合并目标作为整体跟踪的处理方式, 使得该算法对目标尺寸变化、运动特性、遮挡程度、遮挡时间无要求, 适用于两个目标相互遮挡的应用场合. 但是, 由于将合并目标作为一个整体跟踪, 这种设计决定了该类算法难以解决遮挡过程中的目标定位问题, 当发生遮挡的目标数量大于 2 时, 算法由于难以估计分裂后每个前景块内的目标数, 会造成失跟.

3 基于贪心搜索的多目标遮挡定位

为了克服合并-分裂算法存在的问题, 本节提出一种基于贪心搜索匹配的遮挡处理算法, 将遮挡过程中的目标跟踪问题转化为一个已知目标特征的图像分类问题, 并通过贪心搜索和积分图实时求解最优解. 为了便于后继采用积分图计算观测概率, 降低计算量, 此处仍然采用外接矩形框作为目标外观模型. 通过上节的合并检测算法, 可以得到合并前景块内具体包含哪些目标. 此时, 算法已将遮挡处理问题转化为已知目标模式的图像分类问题.

当多个目标合并时, 基于目标特性的图像分类问题等价于寻找一组最优的位置组合 Z_g^* , 使得观测概率 $P(Z_g|H)$ 最大化. 图 2 给出一组三目标遮挡下的定位示例, 其中 H 代表包含 w 个目标 $\{T_i\}_{i=1, \dots, w}$ 的合并前景块, Z_g 代表目标 $\{T_i\}_{i=1, \dots, w}$ 在合并前景块中的某一种位置组合, 每一个目标的外观模型由合并前的单目标检测跟踪结果给出. 当前景块 H 包含 R 个前景点时, 位置组合 Z_g 的总数为 R^w , 显然, 采用穷举方式寻找最优的位置组 Z_g^* 无法满足实时性的要求.

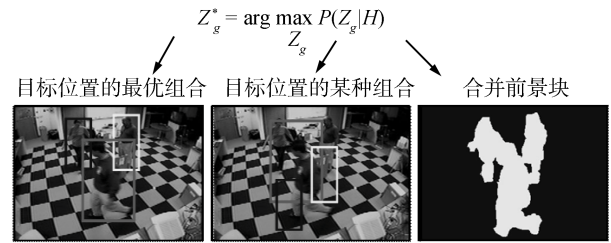


图 2 遮挡前景块中的多目标定位

Fig. 2 Multiple objects location in a merged foreground

为快速寻找最优位置组合 Z_g^* , 定义一个新的变量: 可视度 $\{V_i\}_{i=1, \dots, w}$, 用于描述每个目标的遮挡程度. V_i 越大, 表示该目标被遮挡的程度越轻.

$$V_i = \frac{\text{目标 } T_i \text{ 未被遮挡面积}}{\text{目标 } T_i \text{ 总面积}} \quad (8)$$

假定已经得到前景块内目标的最优位置组合 Z_g^* , 则按照可视度从大到小的顺序对目标进行排序:

$$\begin{aligned} Z_1^* &= \arg \max_i V_i, \quad i \in [1, \dots, w] \\ Z_2^* &= \arg \max_i V_i, \quad i \in [1, \dots, w], V_i \notin \{V_1^*\} \\ &\vdots \\ Z_k^* &= \arg \max_i V_i, \quad i \in [1, \dots, w], \\ &V_i \notin \{V_j^*\}_{j=1, \dots, k-1} \end{aligned} \quad (9)$$

如果目标的可视度 V_i 等于 1, 则观测概率 $P(Z_i|H)$ 达到最大, 随着目标被遮挡程度的加剧, V_i 减少, 此时 $P(Z_i|H)$ 会降低. 基于上述性质, 可知当目标 T_i 在遮挡过程中不发生较大变形或尺度变化时, T_i 的观测概率 $P(Z_i|H)$ 与该目标的可视度 V_i 成正比, 因此, 本算法在遮挡定位时, 并不计算目标可视度的具体大小, 而是根据目标观测概率的相对大小对目标可视度进行排序和定位:

$$\begin{aligned} P(Z_1^*|H) &= \max P(Z_i^*|H) \\ P(Z_2^*|H, Z_1^*) &= \max P(Z_i^*|H, Z_1^*) \end{aligned}$$

$$\begin{aligned} & \vdots \\ P(Z_k^*|H, Z_1^*, \dots, Z_{k-1}^*) = \\ & \max P(Z_i^*|H, Z_1^*, \dots, Z_{k-1}^*) \end{aligned} \quad (10)$$

式中 $i \in [1, \dots, w]$, $Z_i \notin \{Z_j^*\}_{j=1, \dots, k-1}$. 由于目标之间存在相互遮挡, 目标 T_i 的观测概率与 T_j 不独立, 将式 (10) 连乘, 并根据条件概率公式可得:

$$P(Z_1^*|H) \cdots P(Z_w^*|H, Z_1^*, \dots, Z_{w-1}^*) = P(Z_g^*|H) \quad (11)$$

式 (10) 和式 (11) 表明, 使整体观测概率 $P(Z_g^*|H)$ 最大化的最优解 Z_g^* , 可以采用贪心搜索的思想, 通过依次最大化每一目标的观测概率获得. 可视度最高的 Z_1^* 对应的是观测概率最大的目标:

$$Z_1^* = \arg \max_i P(Z_i|H), \quad i = 1, \dots, w \quad (12)$$

在得到 Z_1^* 后, 可通过最大化每个目标的观测概率依次得到最优解 Z_k^* :

$$Z_k^* = \arg \max_i P(Z_i|H, Z_1^*, Z_2^*, \dots, Z_{k-1}^*) \quad (13)$$

其中, $i = 1, \dots, w$, $Z_i \notin \{Z_j^*\}_{j=1, \dots, k-1}$. 为了计算目标的观测概率 $P(Z_i|H, Z_1^*, \dots, Z_{k-1}^*)$, 用目标 T_i 的外接矩形框 Z_i 遍历 H 上所有可能的位置, 并在每一位置计算目标与前景图像的相似度. 遍历结束后将最大相似度作为该目标的观测概率:

$$P(Z_i|H, Z_1^*, \dots, Z_{k-1}^*) = \max_{xc \in H} (P(Z_i|F_{xc})) \quad (14)$$

F_{xc} 表示中心点在 xc 处的外接矩形框 Z_i 覆盖下的前景图像. 概率 $P(Z_i|F_{xc})$ 代表了目标 T_i 与前景 F_{xc} 的相似度. 由于对每一个目标都要遍历整个前景块, 为了保证算法的实时性, 本文采用 F_{xc} 中像素的平均概率计算 $P(Z_i|F_{xc})$:

$$P(Z_i|F_{xc}) = \frac{1}{zw \cdot zh} \sum_{x_k \in F_{xc}} P(T_i|I(x)) \quad (15)$$

式中, $I(x)$ 代表位置在 x_k 处的像素灰度, zw 和 zh 分别表示 Z_i 的宽和高. 像素 $I(x)$ 属于目标 T_i 的概率用 $P(T_i|I(x))$ 表示, $P(T_i|I(x))$ 的和用积分图快速求解. 由于发生遮挡的目标中可能出现色彩近似的情况, 为了锐化相似度函数 $P(T_i|I(x))$, 采用贝叶斯公式, 用前景块 H 内所有尚未进行遮挡定位的目标联合求解 $P(T_i|I(x))$:

$$P(T_i|I(x)) = \frac{P(I(x)|T_i)P(T_i)}{\sum_{s=k}^w P(I(x)|T_s)P(T_s)} \quad (16)$$

式中, $P(I(x)|T_i)$ 用像素 $I(x)$ 在目标 T_i 的颜色直方图中的概率计算.

采用式 (16) 计算概率, 目标间相似色彩对应的权值将被抑制, 同时目标独有颜色对应的权值会得到增强. 该特性可增强对相似目标遮挡定位的鲁棒性和准确性. 但式 (16) 在计算时没有考虑已经完成遮挡定位的集合 Z_1^*, \dots, Z_{k-1}^* 的影响, 此时, 遮挡跟踪中有可能会出现多个目标覆盖在相同的位置, 而一部分前景像素并没有分配给任何目标的情况. 为了解决该问题, 可将 H 中的像素分为两个集合: H 中被集合 Z_1^*, \dots, Z_{k-1}^* 覆盖下的像素集合 X_g^+ 和尚未被覆盖的像素集合 X_g^- , 并将式 (16) 修正为:

$$P(T_i|I(x)) = \begin{cases} \frac{P(I(x)|T_i)P(T_i)}{\sum_{s=k}^w P(I(x)|T_s)P(T_s)}, & x \in X_g^- \\ \frac{P(I(x)|T_i)P(T_i)}{\sum_{s=k}^w P(I(x)|T_s)P(T_s)} \varphi(x), & x \in X_g^+ \end{cases} \quad (17)$$

函数 $\varphi(x) \in [0, 1]$, 其作用是惩罚被覆盖下的像素权值, 降低将一个像素分配给多个目标的概率.

图 3 给出贪心搜索算法流程图. 当有遮挡发生时, 对于包含多个目标的合并前景块 H , 根据合并检测模块和遮挡前的跟踪结果, 可获得前景块内发生遮挡的目标数量、标号、颜色、形状等先验特征. 定位时, 依次遍历所有未被定位的目标, 根据式 (14)、(15) 和 (17) 计算每一个目标 T_i 的观测概率, 并将观测概率最高的目标所在位置作为该目标的定位结果输出, 同时将被该目标覆盖的像素添加到集合 X_g^+ 中, 对合并前景块中的像素集合进行更新. 重复上述过程, 直到获得所有遮挡目标在前景块中的定位结果.

4 算法实现及结果分析

以本文算法为核心, 用 VC++ 开发了一个实时多人跟踪系统. 实验中硬件系统采用 Intel Core 2 CPU 1.66 GHz 处理器, 内存 2 G. 测试数据来自 IBM 室内多人遮挡数据库^[24] 和部分自建数据库, 并选用文献 [25] 中的算法对跟踪结果进行性能评估.

4.1 性能评估指标

按照研究对象不同, 性能评估可分为以下 3 个层次:

1) 基于检测前景与真实前景的像素级性能评估;

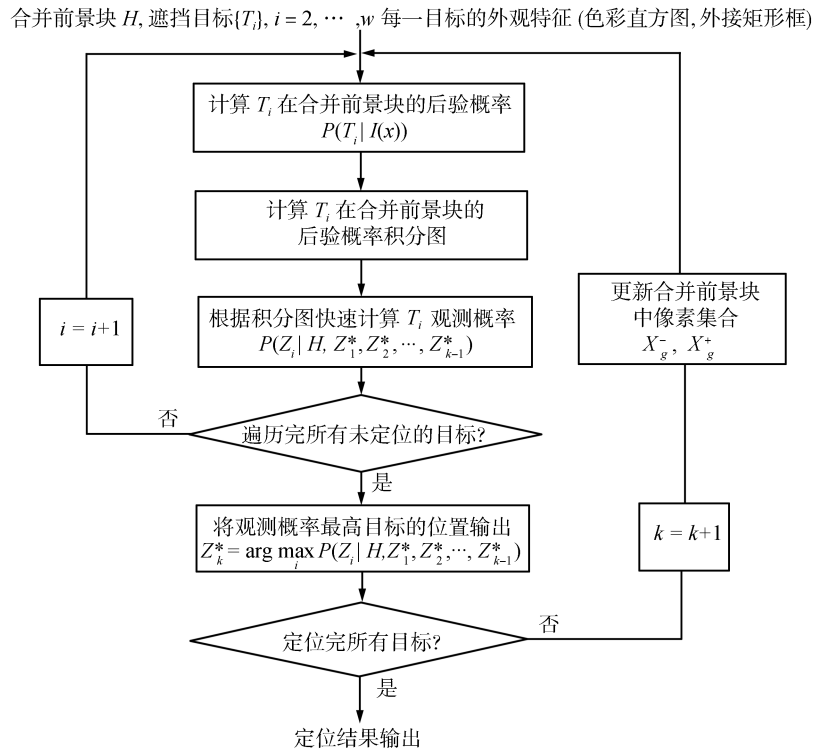


图 3 贪心搜索算法流程图

Fig. 3 Flowchart of greedy searching algorithm

- 2) 基于目标外接矩形框的性能评估;
- 3) 基于目标轨迹的性能评估.

由于像素级评估需要手工精确标定真实前景, 实现困难, 因此本文采用目标外接矩形框的评估方式, 计算每一帧内目标的检测率和虚警率, 并选用 PETS2006 中文文献 [25] 的 9 个指标对算法进行评估, 具体指标如下:

$$\text{Tracker detection rate (TRDR)} = \frac{\text{TP}}{\text{TG}}$$

$$\text{False ala rate (FAR)} = \frac{\text{FP}}{\text{FP}}$$

$$\text{Detection rate (DR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TF}}$$

$$\text{Positive prediction (PP)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Negative prediction (NP)} = \frac{\text{TN}}{\text{FN} + \text{TN}}$$

$$\text{False negative rate (FNR)} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

$$\text{False positive rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (18)$$

其中, TP, TN, FP, FN 表示 True positive, True negative, False positive, False negative 出现的总帧数, TG (Total ground truth) 是视频中含有目标的帧数, TF (Total frame) 是测试视频的总帧数. 基于外接矩形框的评估算法计算简便、直观, 能够有效评价算法在检测率、虚警率方面的性能. 为了定量评估算法处理遮挡的能力, 本文定义一个新的指标: 遮挡错误率 (Occlusion error rate, OER), 用于描述目标标号在遮挡前后发生改变的比率:

$$\text{遮挡错误率} = \frac{\text{遮挡前后目标标号改变总数}}{\text{目标发生遮挡分裂的总数}} \quad (19)$$

4.2 测试数据库

由于视频分析算法的性能与具体应用场合、摄像机参数乃至安装角度均有关系, 且算法自身的参数设置也会对最终分析结果产生重大影响. 因此, 应合理选择与本系统应用条件近似的数据库进行评估.

智能视频监控性能评估的公开数据库包括 PETS、IBM 的室内数据库^[24], 欧盟资助的 CAVIAR 行为识别数据库^[26], 法国政府资助的 ETISEO^[27] 等. 由于 IBM 数据库中包含大量多

人遮挡事件, 因此, 实验中采用该数据库对基于贪心搜索的跟踪算法进行评估.

IBM 提供的目标真实位置采样率较低, 因此, 在实验前先对 10 组 IBM 视频内的运动目标的位置、标号进行手工标定, 生成测试用的真实位置 (Ground truth). 在标定时, 对于每个目标, 用外接矩形框标示出目标所在位置, 并赋予不同的编号, 标定的真实轨迹保存在数据库中. 作者手工标定了 10 组共计 7443 帧图像, 其中含有目标 3085 帧, 发生遮挡总数为 64 次. 图 4 中给出 10 个场景手工标定的目标真实轨迹.



图 4 IBM 室内监控数据库中 10 个场景手工标记的真实轨迹

Fig. 4 Hand labeled ground truth trajectory of IBM indoor surveillance video database

测试时, 先从数据库中同时读取图像文件所在路径和真实轨迹, 然后与算法跟踪结果进行比较, 并按照矩形重叠面积比率判断真实位置和检测位置是否匹配, 统计每一帧的 TP、TN、FP、FN, 最终的评估结果由式 (18) 和 (19) 给出.

4.3 相似背景下的前景检测结果

实验中采用 IBM 室内监控数据库, 从相似背景和运动阴影两方面对第 2.1 节检测算法进行评估, 并与混合高斯背景模型的检测结果比较. 在计算像素邻域间的相关系数时, 采用 9×9 的邻域窗口, 阈值为 0.9. 对 IBM 数据库的实验结果表明, 对输入分辨率为 320×240 的序列, 前景检测的平均处理速度为 15 帧/秒.

相似背景会造成目标的割裂甚至漏检. 图 5 给出一组相似背景条件下的运动前景检测结果. 由于地板格和运动前景目标色彩相近, GMM 漏检了大量前景像素点 (图 5 第 2 行), 场景中同一个运动目标的检测结果会出现割裂, 严重影响后继目标跟踪的稳定性. 本文算法通过候选前景点提取, 能够完整、准确地分割出运动目标, 大幅度提高了对运动前景的检测率, 取得满意的效果 (图 5 第 3 行).

图 6 给出一组场景内同时出现运动阴影和相似背景时的前景检测结果. 当目标在场景中运动时, 在地面、墙面投射了大量阴影. 根据图像照度模型, 阴影区域内的像素灰度会发生比率变化, 这使得用欧氏距离做度量的像素级差分会将该区域当作运动前

景输出, 造成虚警率上升. 从图 6(c) 和 6(d) 中可以看出, 基于像素级差分的 GMM 模型检测结果中包含大量的阴影区域 (图 6(c) 和 6(d) 中远处行人), 此外, 地板和目标灰度相似也造成目标分割结果出现空洞 (图 6(c) 和 6(d) 中近处行人). 图 6(e) 和 6(f) 给出本文算法的检测结果, 通过联合候选景点周围的邻域信息, 将像素级差分扩展为特征级差分, 本算法在增加检测率的同时可减少运动阴影干扰.

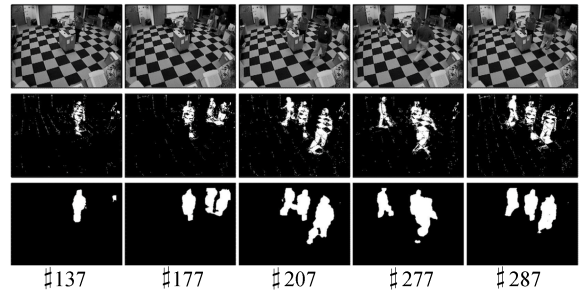


图 5 相似背景条件下的运动前景分割结果 (第 1 行: 输入视频图像; 第 2 行: GMM 检测结果; 第 3 行: 本文检测结果)

Fig. 5 Moving foreground segmentation results under similar background (The first row shows the input video, the second row displays the GMM results, and the third row displays our results)

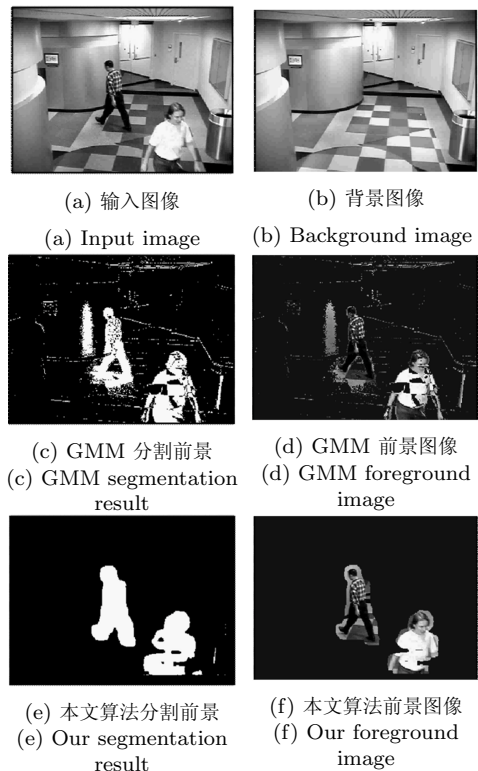


图 6 运动阴影、相似背景条件下的运动前景分割结果

Fig. 6 Moving foreground segmentation results under cast shadow and similar background

4.4 多人遮挡跟踪结果

图 7 中给出 3 组两人交错时的跟踪结果. 当两目标交错时, 基于合并-分裂的跟踪算法^[6, 23]会把目标作为一个整体进行跟踪. 与之不同, 通过贪心搜索, 本算法在遮挡过程中仍能准确地估计目标位置(图 7(a)~(c)中第 2 列和第 3 列). 在图 7(c) 中, 2 号目标在交错时被严重遮挡(图 7(c) 中第 3 列), 目标外观发生较大改变, 此时一些基于局部最优的跟

踪算法, 如相关跟踪、Mean shift 等容易发生失跟. 本算法采用贪心搜索的思路, 首先根据观测概率大小定位出可视度最高的 1 号目标, 然后按照式 (17) 修正合并前景块内像素概率, 并根据 2 号目标的未被遮挡的部分对其进行定位. 该算法对目标的被遮挡程度无约束, 因此能在遮挡过程中连续给出当前两个目标的最优位置组合(图 7(c) 中第 3 列), 而且当被完全遮挡的目标重新出现时, 可恢复对该目标的跟踪(图 7(c) 中第 4 列).

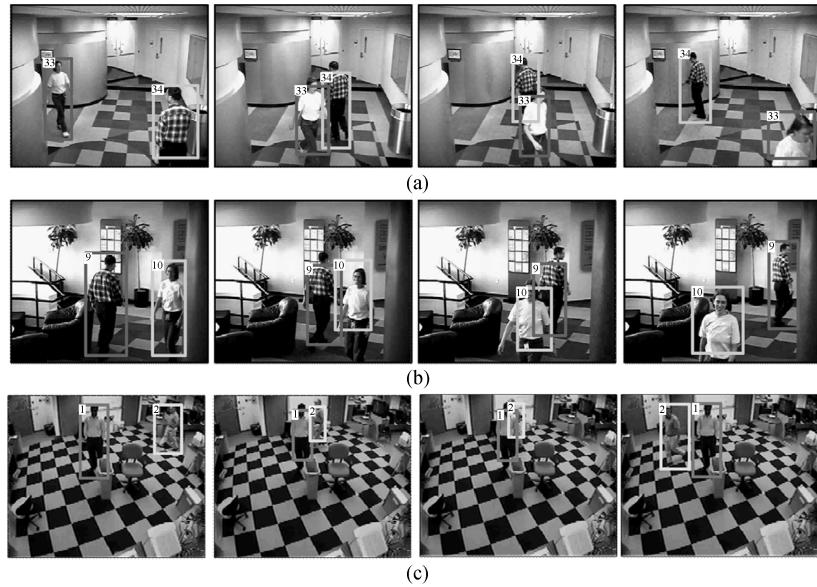


图 7 两人遮挡跟踪结果

Fig. 7 Experiment results of two people tracking through occlusion

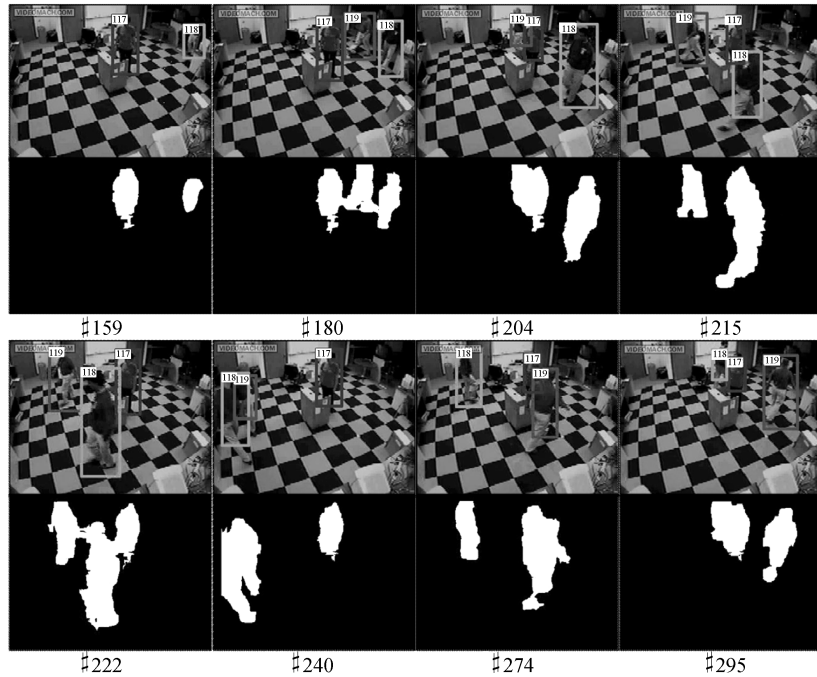


图 8 三人遮挡跟踪结果

Fig. 8 Experiment results of three-people tracking through occlusion

图8给出一组3人交错条件下的跟踪结果. 图中第1行给出输入视频和跟踪结果, 第2行为本文算法的前景检测结果. 由于摄像机视角的因素, 三人在交错时发生互遮挡, 使得检测出的前景块中多次出现同时包含3个目标的情况(图8中#180, #222). 当这个由3个目标组成的合并目标分离时, 由于无法判断每个分裂前景块中的目标组成, 基于合并-分裂的遮挡处理算法往往会发生失跟. 本算法通过贪心搜索, 能够在遮挡过程中准确地定位出3个目标的具体位置(图8中#180, #222), 并在3个目标分离后正确跟踪(图8中#204, #240).

采用文献[25]和式(18)给出的遮挡错误率对算法进行性能评估. 图9给出算法对IBM测试库中的10组视频, 共计7443帧的跟踪轨迹的性能评估结果. 对该数据库, 本算法取得了较高的检测效果: 漏检率(False negative rate, FNR)为0.027(图9), 表明本算法能够有效提高相似背景条件下对目标的检测概率. 在降低虚警方面, 算法检测目标时通过联合像素邻域信息, 减少了运动阴影可能引起的虚警; 同时, 通过采用 N 次检测中出现 M 次相关($M < N$)的跟踪起始逻辑, 减少了随机噪声的干扰. 最终得到的虚警率(False positive rate, FPR)为0.077, 在遮挡处理方面, IBM数据库中频繁出现两目标遮挡和三目标遮挡情况, 10组序列中共发生64次遮挡, 算法正确处理了50次, 遮挡错误率为0.22. 此外, 由于采用贪心搜索和积分图, 使得算法具有良好的实时性, 对该数据库的平均处理速度达到10帧/秒.

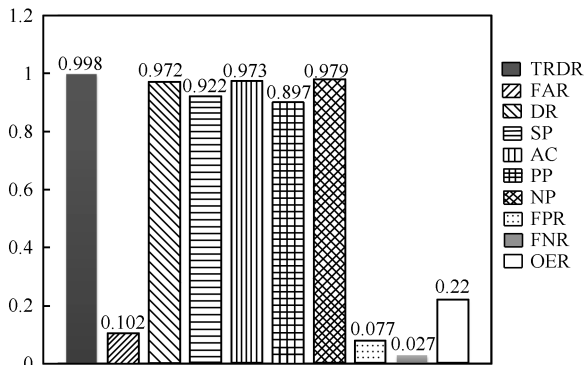


图9 IBM多人遮挡测试库性能评估结果

Fig.9 Performance evaluation results of IBM database

5 结论

针对遮挡条件下的多目标跟踪问题, 提出一种基于贪心搜索的跟踪算法, 将遮挡中的目标跟踪问题转化为已知目标特征和数量的图像分类问题, 拓宽了遮挡处理研究思路, 突破了大多数算法对遮挡时目标数量、运动模式的约束. 同时, 考虑到实时性

的要求, 算法根据目标可视程度, 采用贪心搜索策略和积分图对遮挡目标快速定位. 实验中, 采用手工标定的IBM室内多人遮挡数据库对算法性能进行评估, 结果证明了算法的有效性. 在进一步的研究中, 拟在现有算法的框架下, 采用多特征融合和模糊聚类的思路, 将场景三维模型和目标运动状态引入, 提高遮挡条件下的跟踪算法性能.

References

- Senior A, Hampapur A, Tian Y L, Brown L, Pankanti S, Bolle R. Appearance models for occlusion handling. *Image and Vision Computing*, 2006, **24**(11): 1233–1243
- Cucchiara R, Grana C, Tardini G, Vezzani R. Probabilistic people tracking for occlusion handling. In: *Proceedings of the 17th International Conference on Pattern Recognition*. Cambridge, UK: IEEE, 2004. 132–135
- Nguyen H T, Smeulders A W M. Fast occluded object tracking by a robust appearance filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8): 1099–1104
- Wu Y, Yu T, Hua G. Tracking appearances with occlusions. In: *Proceedings of IEEE Computer Vision and Pattern Recognition*. Wisconsin, USA: IEEE, 2003. 789–795
- Siebel N T, Maybank S J. Fusion of multiple tracking algorithms for robust people tracking. In: *Proceedings of the 7th European Conference on Computer Vision*. Copenhagen, Denmark: Springer, 2002. 373–387
- Huang Y, Essa I. Tracking multiple objects through occlusions. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. San Diego, USA: IEEE, 2005. 1051–1058
- Chang T H, Gong S G, Jong E. Tracking multiple people under occlusion using multiple cameras. In: *Proceedings of British Machine Vision Conference*. Bristol, UK: British Machine Vision Association, 2000. 1–10
- Hu W M, Hu M, Zhou X, Tan T N, Lou J G, Maybank S. Principal axis-based correspondence between multiple cameras for people tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 663–671
- Zhao T, Nevatia R. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9): 1208–1221
- Cucchiara R, Grana C, Tardini G. Track-based and object-based occlusion for people tracking refinement in indoor surveillance. In: *Proceedings of the 2nd International Workshop on Video Surveillance and Sensor Networks*. New York, USA: ACM, 2004. 81–87
- Beleznai C, Schlot T, Wachmann B, Bischof H, Kropatsch W G. Tracking multiple objects in complex scenes. In: *Proceedings of the 26th Workshop of the Austrian Association for Pattern Recognition*. Austria: 2002. 175–182
- Joshi N, Avidan S, Matusik W, Kriegman D J. Synthetic aperture tracking: tracking through occlusions. In: *Proceedings of the 11th IEEE International Conference on Computer Vision*. Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- McKenna S J, Jabri S, Duric Z, Rosenfeld A, Wechsler H. Tracking groups of people. *Computer Vision and Image Understanding*, 2000, **80**(1): 42–56

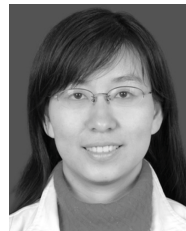
- 14 Bremond F, Thonnat M. Tracking multiple non-rigid objects in video sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, 1998, **8**(5): 585–591
- 15 Piater J H, Crowley J L. Multi-modal tracking of interacting targets using Gaussian approximations. In: Proceedings of the 2nd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Hawaii, USA: IEEE, 2001. 141–147
- 16 Elgammal A E, Davis L S. Probabilistic framework for segmenting people under occlusion. In: Proceedings of the 8th International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 145–152
- 17 Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 747–757
- 18 Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 657–662
- 19 Sheikh Y, Shah M. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(11): 1778–1792
- 20 Wixson L. Detecting salient motion by accumulating directionally-consistent flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 774–780
- 21 Bugeau A, Perez P. Detection and segmentation of moving objects in highly dynamic scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minnesota, USA: IEEE, 2007. 1–8
- 22 Zhu S, Avidan Q, Cheng K T. Learning a sparse, corner-based representation for time-varying background modeling. In: Proceedings of the IEEE Conference on Computer Vision. Beijing, China: IEEE, 2005. 678–685
- 23 Yang T, Pan Q, Li J, Li S Z. Real-time multiple object tracking with occlusion handling in dynamic scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 970–975
- 24 Brown L M, Senior A W, Tian Y L, Connell J, Hampapur A. Performance evaluation of surveillance systems under varying conditions [Online], available: <http://www.research.ibm.com/peoplevision/performanceevaluation.html>, November 3, 2008
- 25 Bashir F, Porikli F. Performance evaluation of object detection and tracking systems. In: Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. New York, USA: IEEE, 2006. 7–14
- 26 CAVIAR test case scenarios [Online], available: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>, January 10, 2009
- 27 Video understanding evaluation [Online], available: <http://www.silogic.fr/etiseo>, November 19, 2008



杨涛 西北工业大学计算机学院博士后。2008 年获得西北工业大学控制理论与控制工程专业博士学位。主要研究方向为视频内容分析、视频目标检测与跟踪、图像配准。本文通信作者。

E-mail: yangtaonwpu@163.com

(**YANG Tao** Postdoctor at the School of Computer Science, Northwestern Polytechnical University. He received his Ph.D. degree in control science and control engineering from the School of Automation, Northwestern Polytechnical University in 2008. His research interest covers video content analysis, visual object detection and tracking, and image registration. Corresponding author of this paper.)



李静 西安电子科技大学通信工程学院讲师。2008 年获得西北工业大学控制理论与控制工程专业博士学位。主要研究方向为图像处理与模式识别、图像配准、文字定位与识别。

E-mail: jinglixid@mail.xidian.edu.cn

(**LI Jing** Lecturer at the School of Telecommunications Engineering, Xidian University. She received her Ph.D. degree in control science and control engineering from the School of Automation, Northwestern Polytechnical University in 2008. Her research interest covers image processing and pattern recognition, image registration, text location and recognition.)



潘泉 西北工业大学教授。主要研究方向为信息融合理论及应用、多目标跟踪、智能信息处理。

E-mail: quanpan@nwpu.edu.cn

(**PAN Quan** Professor at Northwestern Polytechnical University. His research interest covers theory and application of information fusion, multi-target tracking, and intelligence information processing.)



张艳宁 西北工业大学教授。主要研究方向为计算机视觉、图像与视频处理、智能信息处理、模式识别。

E-mail: ynzhang@nwpu.edu.cn

(**ZHANG Yan-Ning** Professor at Northwestern Polytechnical University. Her research interest covers computer vision, image and video signal processing, intelligence information processing, and pattern recognition.)