

基于改进协作目标外观模型的在线视觉跟踪

宋 涛¹, 李 鸥¹, 刘广怡¹, 崔弘亮²

(1. 信息工程大学信息系统工程学院, 河南郑州 450002; 2. 南京理工大学自动化学院, 江苏南京 210094)

摘 要: 在不受限制的复杂环境中在线跟踪任意类型的感兴趣目标仍是一项极具挑战的难题. 本文在无模型跟踪框架基础上提出一种基于改进协作目标外观模型的在线视觉跟踪方法, 解决了大多数协作模型类跟踪算法在学习阶段无法有效选择正、负样本的问题. 该方法根据人类视觉感知准则将目标边缘信息视为最具区分度的目标特征, 提出边缘判别模型并结合动态模型和检测模块建立二级似然匹配空间, 为生成模型的似然匹配去除了背景干扰; 采用分块策略建立目标生成模型, 为模型引入空间结构信息; 利用 Mean-Shift 计算各子块的最终位置和匹配系数, 并根据子块匹配系数为遮挡处理和模型更新提供依据. 在公开视频序列上同几种流行视觉跟踪算法的对比实验结果证明了本文算法的有效性和优越性.

关键词: 在线视觉跟踪; 协作外观模型; 人类视觉感知; 二级似然匹配空间; 模型更新

中图分类号: TP319 **文献标识码:** A **文章编号:** 0372-2112 (2017)02-0384-10

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.02.017

Online Visual Tracking Based on Improved Collaborative Appearance Model

SONG Tao¹, LI Ou¹, LIU Guang-yi¹, CUI Hong-liang²

(1. Institute of Information System Engineering, Information Engineering University, Zhengzhou, Henan 450002, China;

2. School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China)

Abstract: It is still a very challenging issue to online track arbitrary targets in the unrestricted complex environment. This paper presents an online visual tracking method with improved collaborative appearance model based on model-free framework, solving the problem of most other tracking algorithms with collaborative model, which is unable to effectively select the positive and negative samples. According to the human visual perception rules, object edge information is regarded as the most discriminative feature, on which an edge discriminative appearance model is proposed. In order to remove background interference in likelihood matching space for generative model, a two-stage matching space is put forward via integrating dynamic model, detection module and edge discriminative model. The generative model based on partition strategy is constructed for space and appearance information. The final position and matching coefficient of each sub-block are calculated by mean-shift, as a basis for occlusion handling and model update. Experimental results using challenging public video sequences show the effectiveness and superiority of the proposed method, compared with other state-of-the-art visual tracking approaches.

Key words: online visual tracking; collaborative appearance model; human visual perception; two-stage likelihood matching space; model update

1 引言

目标跟踪是计算机视觉领域的关键底层技术, 其目的是利用视频序列估计感兴趣目标的位置、大小和运动方向等状态, 在视觉导航、智能机器人、自动监控和人机交互等领域有重要的应用价值, 越来越多的跟踪算法被

开发以满足不同的应用需求. 但在实际应用环境中, 由于光照变化、目标遮挡、视角变化及复杂背景等因素引起的感兴趣目标外观变化, 使得构建跟踪精度高、稳定性好的视觉跟踪算法仍是一项极具挑战性的课题.

一个典型的跟踪算法主要包括两个部分^[1]: (1) 目标外观模型, 用于表示感兴趣目标并匹配候选目标模

型;(2)运动模型,预测目标在下一帧内可能出现的位置.由于感兴趣目标的随机运动模式难以定义,其运动模型只能依赖过去已修正状态近似计算或直接利用随机模型(RW, Random-Walk)进行描述.文献[2]利用交互多模结构(IMM, Interacting Multiple Model)为动态模型引入多种运动模式,可以有效处理机动目标,但其计算量较大且多种运动模型需要预先设定,可推广性差.文献[3]在粒子滤波框架中提出双层运动模型来寻找更加可靠的预测点,但该方法无法解决目标不规则运动问题.运动模型在跟踪算法中能够起到辅助作用,可作为搜索匹配策略的可靠初始位置,并为前景目标关联提供依据,但运动模型无法处理目标间歇运动和全遮挡问题,会导致跟踪误差不断累积直到目标丢失.目标外观模型是跟踪算法的核心部分,对跟踪性能起决定性作用;建立鲁棒的外观模型是当前主流跟踪算法重点研究的内容.文献[4~6]分别针对面部、行人和车辆建立专属模型,虽然提高了跟踪精度,但应用范围有限.近些年,不受目标类型限制的无模型跟踪算法得到广泛关注,其外观模型的构建方式主要包括两类,生成型和判别型.其中,生成模型通过建立目标特征模板将跟踪问题转化为在下一帧中寻找与目标模板最相似的候选区域.这类方法易陷入局部极值,当目标外观发生较大变化时,模型失效.文献[7]利用目标颜色信息建立核函数加权的直方图模型,该方法对半遮挡和目标旋转不敏感,但在光照变化干扰下易发生漂移问题且模型缺乏更新机制.文献[8]在稀疏表达框架的基础上建立了多特征融合模型,提高了模型鲁棒性和跟踪精度,但模型比较复杂,计算量大.文献[9]采用局部背景加权的直方图模型去除目标窗口内的背景信息,该算法增强了模型在复杂环境中的区分度,但额外背景信息的引入增加了模型复杂度且无法处理相似背景干扰.判别模型将目标跟踪转化为二值分类问题,通过目标样本和背景样本训练分类器将目标从背景中提取出来.文献[10]在结构约束条件下提出基于在线 P-N (Positive-Negative Constraints) 学习的半监督二值分类器,在模型更新阶段利用未标注数据提高了分类器性能,跟踪过程中能自适应学习目标外观变化.文献[11]在 P-N 学习框架基础上提出基于检测的跟踪方法,在不需要任何先验知识的前提下,能够长时间有效跟踪目标.文献[12]提出多实例学习算法(MIL, Multiple Instances Learning),不同于传统单实例学习,该算法利用目标邻域的多个样本组成训练子集来更新分类器,提高了分类精度.但上述判别跟踪方法都过度依赖学习算法,其通过空间约束或结构约束假设更新阶段标注的训练样本具有高可信度,当训练样本出现误差时,分类器误差逐渐累积导致分类性能迅速恶化.文献[13,

14]结合生成和判别模型提出鲁棒的协作外观模型,通过融合两类模型的优势来提高目标模板可靠性,但该类方法的难点在于如何平衡学习阶段两类模型训练样本集之间的比重;过度依赖正样本易导致模型过拟合,模型推广预测能力和区分度下降,FP(False Positive)误差逐渐增加;反之,过度依赖负样本虽能提高模型区分度,但模型的鲁棒性/泛化能力下降, FN(False Negative)误差逐渐增加.

针对上述问题,本文提出一种基于改进协作目标外观模型的在线视觉跟踪方法,其核心思想:(1)提高生成模型的可辨别性和唯一性,并给出模型的在线更新机制以适应目标外观变化;(2)将判别模型视为生成模型的辅助机制,为生成模型的匹配空间去除背景干扰,且无需正、负样本的学习过程.具体过程:利用动态模型和检测模块建立第一级粗粒度匹配空间,防止目标发生大幅度漂移;在第一级匹配空间内,利用目标边缘像素点提出边缘判别模型,给出第二级细粒度匹配空间,去除了大量背景信息(尤其是复杂的背景纹理),提高了生成模型的匹配精度和效率;采用分块策略建立目标生成模型,为模型引入空间结构信息;利用改进的局部二进制特征(LBP, Local Binary Pattern)构建每个子块的核函数直方图模型,在似然匹配空间利用 Mean-Shift 迭代计算各子块的最终位置和匹配系数;根据子块匹配系数为遮挡处理和模型更新提供依据.在实验分析阶段,本文通过三组对比实验证明了本文算法在复杂应用场景中的优越性和有效性.

2 生成模型和判别模型的问题分析

本文从实际应用角度出发,重点研究在不受限制的复杂环境中在线跟踪任意类型目标的跟踪算法,算法没有任何离线处理部分和先验知识,只能依靠初始帧手动选择或检测得到的有限目标样本和其邻域的背景样本训练目标模型(如图 1(a)所示)并确定目标在下一帧出现的位置.在线视觉跟踪算法可视为小样本情况下的半监督类机器学习问题,根据统计学习理论的经验风险最小归纳原理,因有限样本导致的期望误差无法避免.

$$R(w) \leq R_{\text{emp}}(w) + \varphi(h, n_{\text{train}}) \quad (1)$$

其中, $R(w)$ 表示实际风险,由经验风险 $R_{\text{emp}}(w)$ 和置信界限 $\varphi(h, n_{\text{train}})$ 组成.在有限训练样本 n_{train} 的条件下,初始帧训练得到的目标模型复杂性越高,其置信界限就越大,导致真实风险与经验风险之差越大,即目标模型过拟合.所以,越复杂的训练模型其泛化能力越差,初始帧得到的生成或判别目标模型均无法处理后续帧中不可预知的模型变化,必须依靠合理的模型更新或学习策略在线校正初始模型.如何选择有效的更新样本(正

样本:目标,负样本:背景)成为关键问题.

基于结构性约束的 P-N 学习方法是当前生成模型和判别模型类跟踪方法比较常用的样本选择策略^[10-14],将 t 时刻目标轨迹附近的图像块标注为目标样本 $X^p = \{x; \|(l)(x) - (l)_i^*\| \leq \alpha\}$,并且将目标样本邻域的图像块标注为背景样本 $X^b = \{x; \alpha \leq \|(l)(x) - (l)_i^*\| \leq \beta\}$ (详细步骤见文献[12]). 这类样本选择策略可以不断更新模型,提高模型的可信度,但其缺陷同样明显:(1) 目标位置 $(l)_i^* = (l)(\arg\max_{x \in X^{\text{true}}} p(y|x))$ 必须满足 $\|(l)_i^* - (l)_i^{\text{True}}\| \leq Th$ 的结构约束假设,这在现实应用中很难保证;(2) 在极端情况下,即使满足理想条件 $(l)_i^* = (l)_i^{\text{True}}$,相似背景信息依然会影响样本的可靠性.

从图 1(b)可以看出,当结构约束条件无法满足时, MIL 学习策略所选择的正样本 $X^p = \{(x_1, 1), (x_2, 1) \dots\}$ 中存在大量背景信息,使得校正模型的可信度下降;如图 1(c)所示,即使在理想情况下选择的正样本 $(x_2, 1)$ 依然含有较多背景信息,导致正、负样本的区分度降低. 所以,基于结构约束的样本学习策略无法有效处理复杂环境中相似背景干扰问题. 由上述分析可知,如何在跟踪过程中在线选择有效样本并自适应平衡正、负样本之间的比例是保证生成或判别模型泛化及区分能力的关键,但这在实际应用中很难做到. 本文所提算法避免了正、负样本的选择问题.

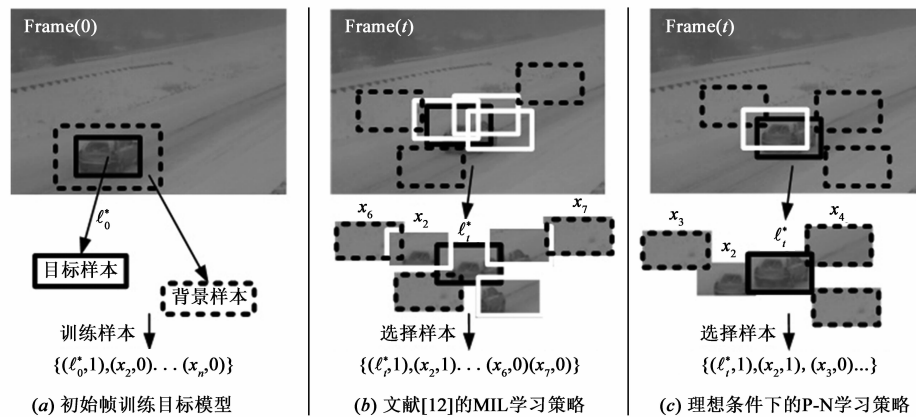


图1 基于结构约束的样本学习方法

3 粗粒度似然匹配空间

3.1 运动模型

目标运动模型可近似表达为过去 K 个已修正状态的线性组合,引入目标运动的平滑机制和限制条件:假设目标速度和方位没有发生突变,且当前帧的状态满足前 K 帧运动规律. 通常认为运动目标在半秒钟内不会发生大的突变,本文按照 25frame/s 的帧率取值 $K = 25/2$.

定义 $(l)_{t-k:t-1} = \{(l)_i\}_{i=t-k}^{t-1}$ 表示目标在前 K 帧中经过校正的位置;按照帧间距离越近相关性越大的原则,令 $\pi_{j=t-1}^{(i=t-k, \dots, t-2)} = \exp\left[-\frac{(i-j)^2}{2\sigma^2}\right]$ 表示第 j 帧与其他帧之间的关联程度,则目标的运动模型如下式所示:

$$\begin{cases} \hat{v}_{t-1} = \left[\frac{\sum_{i=t-k}^{t-2} (l)_{i+1} - (l)_i^2 \cdot \pi_{t-1}^{(i+1)}}{\sum_{i=t-k}^{t-2} \pi_{t-1}^{(i+1)}} \right]^{\frac{1}{2}} \\ (\hat{l})_t = \hat{v}_{t-1} \cdot t + \varepsilon_t \end{cases} \quad (2)$$

其中, $(\hat{l})_t$ 和 \hat{v}_{t-1} 分别表示目标在第 t 帧的预测位置和

速度; ε_t 是随机扰动,由真实值 $(l)_{t-1}$ 和估计值 $(\hat{l})_{t-1}$ 之间的相似度决定,在实际应用中其值可适当增加,以扩大候选目标的搜索范围.

3.2 基于自适应分块的整体判别分类器

为了从背景中提取前景运动目标,本文在背景减除框架基础上结合简单线性迭代聚类算法(SLIC, Simple Linear Iterative Clustering)^[15] 提出自适应分块的整体判别分类器. 首先,利用边缘特征统计分析视频帧复杂度,结合帧分辨率自适应选择合理的分块数量 N_s ;然后将每个子块视为超像素点,并提出相应的前景目标检测方法.

利用式(3)的高斯差分算子(DOG, Difference of Gaussians)对视频帧进行锐化处理,并根据非极大值抑制(NMS, Non-Maximum Suppression)细化高斯差分图像的边缘信息;结合帧分辨率计算所需的 SLIC 分块数目 $N_s = \frac{\text{Resolution}[I(x, y)]}{\text{NMS}[D(x, y, \sigma)] + \varepsilon}$. 从图 2 可以看出,跟踪场景的边缘信息反映了场景复杂程度,边缘信息越少,场景的纹理分布越简单,即场景中存在大量的同构特征(如图 2(a)中的公路),用数量较少的分块就能很好地描述跟踪场景;反之则需要数量较大的分块(如图 2(b)所示).

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \quad (3)$$

本文在背景减除框架基础上提出前景目标二级检测机制,根据 SLIC 自适应分块建立第一级粗粒度检测模块,令 $\mathbf{B}^t = \{\mathbf{B}_1^t(x), \mathbf{B}_2^t(x), \dots, \mathbf{B}_{N_s}^t(x)\}$ 表示一级背景模型,通过与第 $t+1$ 帧进行差分运算得到发生变化

的子块;利用子块内像素点的颜色信息建立第二级细粒度检测模块,令 $\mathbf{B}_i^t(x) = \{\mathbf{b}_1^t(x), \mathbf{b}_2^t(x), \dots, \mathbf{b}_{N_i}^t(x)\}$ 表示第 i 个子块的二级背景模型,其中 $\mathbf{b}_j^t(x)$ 是第 i 个子块内的第 j 个像素点信息,包括颜色信息及邻域二进制信息. 第二级检测模块只处理产生较大变化的子块,捕捉子块内真正发生变化的前景像素点.

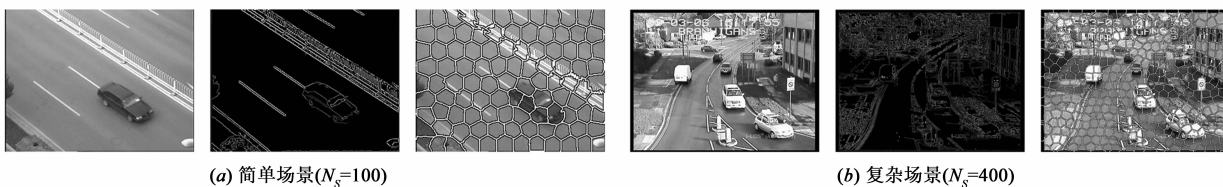


图2 SLIC自适应分块示意图

3.3 第一级粗粒度似然匹配空间

本文将整体判别分类器和运动模型视为跟踪算法的辅助机制,为生成目标模型的似然匹配确定搜索空间. 以 $(\hat{l})_{t+1}$ 为圆心,感兴趣目标的对角线为半径计算目标在第 t 帧可能出现的范围,该范围内的前景目标即为真实目标最有可能存在的置信区域 C_{area} ,如图 3(d) 所示. 在动态背景干扰的复杂环境中,前景检测目标可能包含大量背景像素,因此本文将 C_{area} 定义为粗粒度似然匹配空间,并根据第 4.1 节提出的方法进一步去除匹配空间内的背景干扰.

值得注意的问题,本文提出的整体判别分类器结合运动模型确定匹配空间的方法仅适用于静态摄像头应用环境;而对于动态摄像头应用场景,本文仅根据运

动模型来确定匹配空间,从图 3(b) 可以看出,运动模型确定的搜索范围包含真实目标.

4 构建目标边缘判别模型和基于分块的生成目标模型

4.1 基于判别模型的第二级细粒度似然匹配空间

人类视觉感知准则 (HVP, Human Visual Perception) 认为:人类视觉系统观察的是变化,而不是绝对值,且容易被边界吸引;在不同物体间使用高对比度会使物体更容易区分. 边界是视觉能够将目标从背景中分离出来的最基本要素,如果目标与背景间没有边界,则目标完全融入背景;只要存在目标边界,理论上就可以将目标与背景区分开,如图 4 所示.

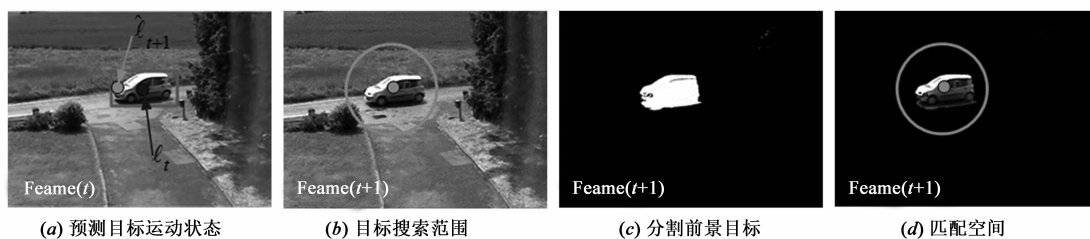
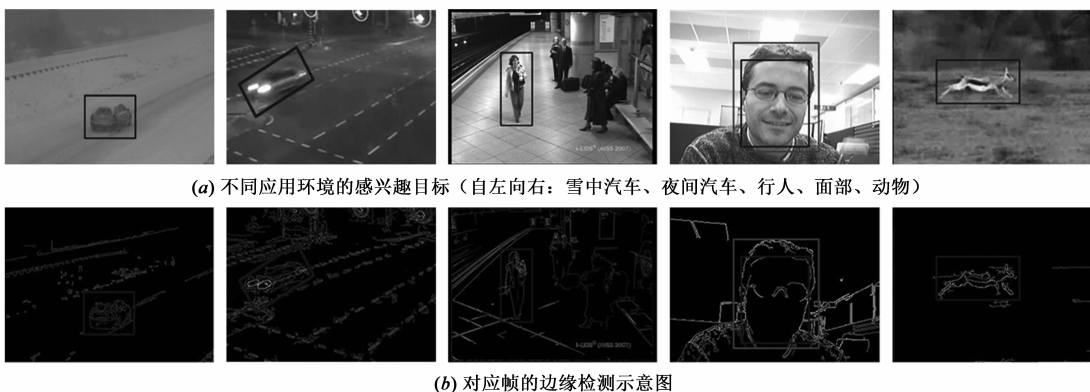
图3 $t+1$ 时刻的粗粒度似然匹配空间

图4 不同目标类型的边界信息

从图 4 可以看出,即使在非常恶劣的跟踪场景中依然存在清晰的目标边界(如图 4 中的汽车目标).不同于文献[16]和[17]等大多数利用轮廓特征描述目标形状进行跟踪的方法,本文将轮廓信息视为目标区分背景的重要线索,即在粗粒度空间 C_{area} 中提取相应的目标轮廓信息,得到更加精确的细粒度匹配空间 R_{area} (如图 5(e) 中的黑色矩形框所示).但在纹理复杂的应用环境中,目标轮廓附近存在大量杂乱无章的边界信息,导致目标轮廓的提取十分困难;本文根据初始帧选择并分割得到的感兴趣目标,提出一种目标边缘判别模型,最大限度去除背景边缘干扰.

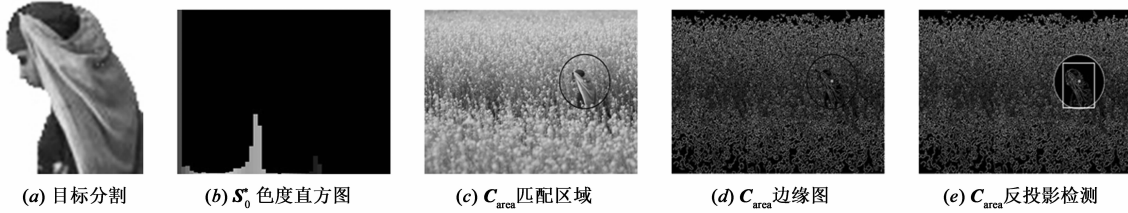


图5 目标边缘判别示意图

$$s_i^* = \operatorname{argmind}(\varphi_i, \varphi_x), \text{ s. t. } \mathbf{I}_x \in \text{Contour}_{\text{in}} \quad (4)$$

$$C_{\text{edge}}^*(x, y) = \sum_{u=1}^m q(u) \cdot \delta[\Gamma(C_{\text{edge}}(x, y)) - u] \quad (5)$$

其中, u 表示量化值索引,将色度空间划分为 m 个 bin 区间; $\Gamma(\cdot)$ 表示像素点色度值量化函数,根据像素点的色度值判断其对应的 bin 区间; $d(\cdot)$ 表示色度 Euclidean 距离,用来度量像素点之间的颜色相似度; $C_{\text{edge}}(x, y)$ 表示 C_{area} 内所有边缘像素点; $C_{\text{edge}}^*(x, y)$ 表示边缘像素点的反投影概率值.

4.2 融入空间结构信息的分块生成模型

为了提高目标生成模型的可辨别性和唯一性,本文采用分块策略建立目标生成模型,为模型引入空间结构信息.本文利用最基本的分块方案,按照目标的几何区域将目标均匀分为 $m \times n$ 个互不重叠的子块 block (i, j) ($1 \leq i \leq m, 1 \leq j \leq n$),简称为 b_i .通过反复试验,本文取 $m = n = 3$,如果子块数过多,会导致各子块间的相似性增加而独立性降低;反之,子块数过少则无法突出目标模型的空间结构信息;而 3×3 的分块方案存在 1 个中心子块 b_0 和 8 个邻域子块 $b_{i=1, \dots, 8}$,邻域子块不但

首先在初始帧利用分割算法得到近似真实目标区域,在目标区域内利用 Canny 算子检测目标的边缘信息 $S_0 = \{s_i\}_{i=1, \dots, N_s}$,其中 $s_i = (\varphi_i, h_i)$ 分别表示第 i 个目标边缘像素点的位置和色度值;然后在目标区域内部根据式(4)计算 s_i 的代替值 $s_i^* = \mathbf{I}_x = (\varphi_x, h_x)$, \mathbf{I}_x 表示目标区域内部与 s_i 距离最近的像素点;利用 $S_0^* = \{s_i^*\}_{i=1, \dots, N_s}$ 建立色度直方图 $q_0 = \{q(u)\}_{u=1, \dots, m}$,在 C_{area} 区域内利用式(5)计算所有边缘像素点的色度直方图反投影概率图并根据阈值去除概率较小的像素点,降低 C_{area} 内的背景边缘干扰(如图 5 所示).

包含目标轮廓信息并可判断目标是否发生部分遮挡.本文利用 Mean-Shift^[7] 跟踪框架建立分块目标模型的似然匹配函数,为了降低光照变化对目标外观的影响,本文将子块像素的颜色信息 \mathbf{I}_x 转化为改进的 LBP_{12,2} 二进制特征 e_x (如图 6 所示),并利用核函数加权直方图进行统计建模 $q_u^{(b)}[(l)_{i-1}^i]$,其中 $(l)_{i-1}^i$ 是子块 b_i 在 $t-1$ 时刻的几何中心位置.

$$q_u^{(b)}[(l)_{i-1}^i] = C_{h_i} \cdot \sum_{i=1}^{N_i} k\left(\left\|\frac{e_0^{(b)} - e_i^{(b)}}{h_b}\right\|^2\right) \cdot \delta[\text{bin}(e_i^{(b)}) - u] \quad (6)$$

在置信区域 C_{area} 内根据 $q_u^{(b)}[(l)_{i-1}^i]$ 的方法计算目标候选分块模型 $p_u^{(b)}[(\hat{l})_i^i]$,利用 Bhattacharyya 系数来度量目标模型 $q_u^{(b)}[(l)_{i-1}^i]$ 和候选模型 $p_u^{(b)}[(\hat{l})_i^i]$ 之间的相似程度:

$$\rho_{\text{MS}}((\hat{l})_i^i) = \rho\{p_u^{(b)}[(\hat{l})_i^i], q_u^{(b)}[(l)_{i-1}^i]\} = \sum_{u=1}^m \sqrt{p_u^{(b)} \cdot q_u^{(b)}} = w_i^{(b)} \quad (7)$$

式(7)中 $w_i^{(b)}$ 反映了候选子块的匹配程度;本文通过阈值 $T_w = 0.5$ 来判断目标被遮挡程度并采用相应的更新方案.

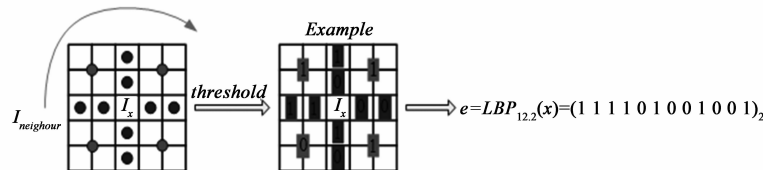


图6 局部二进制模式LBP_{12,2}示意图

5 本文跟踪方法

本文以序贯贝叶斯状态估计和最大后验估计理论为框架,将似然匹配空间、目标生成模型、遮挡处理机制(算法 1 的步骤 5)和模型更新方法(算法 1 的步骤 6)相结合建立视觉跟踪算法,具体如算法 1 所示.

$$\begin{cases} p(X_t | Z_t) \propto p(Z_t | X_t) \cdot \int p(X_t | X_{t-1}) \cdot p(X_{t-1} | Z_{t-1}) d_{X_{t-1}} \\ \hat{X}_t = \underset{X_t}{\operatorname{argmax}} p(X_t | Z_t) \end{cases} \quad (8)$$

算法 1 基于改进协作目标外观模型的在线视觉跟踪算法

步骤 1 初始化:在初始帧手动选择或利用检测模块得到感兴趣目标并分割目标,分别根据式(4)和式(6)计算目标边缘色度直方图 $q_0 = \{q(u)\}_{u=1, \dots, m}$ 和分块目标生成模型 $q_u^{(b_i)}[(l)_0^i]$.

步骤 2 在下一帧 t 时刻,根据第 3.3 节方法计算得到粗粒度似然匹配空间 C_{area}^t .

步骤 3 利用 Canny 算子计算 C_{area}^t 内的边缘信息 $C_{\text{edge}}^t(x, y)$,然后根据式(5)计算边缘像素点对应的直方图反投影概率值 $C_{\text{edge}}^*(x, y)$,令 $C_{\text{edge}}^*(x, y) < T_h$ 的像素点为零,得到细粒度似然匹配空间 R_{area}^t .

步骤 4 在 R_{area}^t 区域内根据式(6)建立候选目标分块模型 $p_u^{(b_i)}[(\hat{l})_t^i]$,并根据式(7)计算各个子块的似然权值 $w_u^{(b_i)}$.

步骤 5 根据 $w_u^{(b_i)}$ 判断目标的受遮挡状态,令 N_w 表示 $w_u^{(b_i)} > T_w$ 的个数;如果 $w_u^{(b_0)} > T_w$,则目标的最终状态 $(l)_t = \sum_{i=0}^{N_w} w_u^{(b_i)} \cdot (\hat{l})_t^i$;如果 $w_u^{(b_0)} < T_w$ 且 $N_w > 1$,则 $(l)_t = \underset{(l)_t^i}{\operatorname{argmax}} w_u^{(b_i)}$;如果 $N_w = 0$,则目标出现全遮挡或丢失,利用第 3 节方法检测目标,直到目标重新出现.

步骤 6 更新权值 $w_u^{(b_i)} > T_w$ 的子块,令 $q_u^{(b_i)}[(l)_t^i] = (1 - \alpha) p_u^{(b_i)}[(l)_t^i] + \alpha q_u^{(b_i)}[(l)_{t-1}^i]$;更新色度直方图 q_t .

表 1 三类代表性算法 Mean-Shift、MIL、SCM 和本文算法的跟踪策略

跟踪算法	外观表达	搜索方式	模型更新	处理速度(帧/秒)
Mean-Shift	全局模板、颜色直方图、生成模型	局部最优	无	66.67
MIL	全局模板、Haar 特征、判别模型	密集采样	有	28.13
SCM	局部模板、稀疏表达、协同模型	粒子滤波	有	0.52
本文算法	局部模板、二进制特征、生成模型	匹配空间采样	有	29.35

6.2 实验结果与性能分析

实验 1 给出了本文算法与 Mean-Shift、MIL、SCM 三类算法在相似背景应用场景“blizzard”和“boats”下的跟踪性能.序列“blizzard”的跟踪难点:感兴趣目标的颜色与背景非常相似,导致目标与背景难以区分;由图 7(a)第一行可以看出,基于颜色直方图统计模型的 Mean-Shift 算法无法处理相似问题,在第 201 帧即完全丢失目标;基于 P-N 学习的 MIL 算法在学习阶段引入大量误判的背景信息,逐渐丢失目标;SCM 算法融合生成和判别模型来表示目标外观,但依然会将大量背景

6 实验结果与分析

6.1 实验说明

为了进一步论证分析所提算法的跟踪性能,本文采用文献[18]提出的视觉跟踪算法通用评估体系,该体系采集了 50 段常用的公开标准视频测试序列,并为每段序列标注了属性及真实目标的位置和尺度,为定性评价跟踪算法提供了更加公正、合理的实验素材;同时,该体系还提供了 29 个当前主流跟踪算法的代码和实验结果,为算法的对比分析提供了便利.为了验证本文算法在处理相似背景干扰和全遮挡方面的优越性,本文与三类具有代表性的跟踪算法 Mean-Shift^[7]、MIL^[12]和 SCM^[13](如表 1 所示)进行了两组对比实验;并通过实验 3 证明了本文算法在其他复杂应用场景下的有效性.所有相关实验都是在 Intel 双核 2.2GHz CPU、2.0GB 内存的 PC 平台上,利用 C/C++、Matlab 和 OpenCV 函数库实现的.

本文通过中心点跟踪误差 d_t 和目标覆盖率 r_t 两个方面来定量评价跟踪算法的有效性.

$$\begin{cases} d_t = \sqrt{(x_{\text{True}}^t - x_{\text{Tracking}}^t)^2 + (y_{\text{True}}^t - y_{\text{Tracking}}^t)^2} \\ r_t = \frac{\text{Area}(R_{\text{True}}^t \cap R_{\text{Tracking}}^t)}{\text{Area}(R_{\text{True}}^t \cup R_{\text{Tracking}}^t)} \end{cases} \quad (9)$$

其中, $(x_{\text{True}}^t, y_{\text{True}}^t)$ 和 $(x_{\text{Tracking}}^t, y_{\text{Tracking}}^t)$ 分别表示 t 时刻目标真实位置和算法得到的估计位置; R_{True}^t 和 R_{Tracking}^t 分别表示 t 时刻目标真实区域和跟踪算法得到的区域.

误判为前景,导致跟踪窗口逐渐发生漂移;本文算法根据二级似然匹配空间确定了目标可能出现的范围并利用目标自身的边缘信息来区分背景(如图 7(a)第二行所示),所以在整个跟踪过程中,本文算法都能有效跟踪目标.序列“boats”的难点:跟踪矩形框在跟踪初始阶段就引入大量背景像素,而且跟踪过程中背景不变,类似于“blizzard”的似然背景应用场景,但背景纹理更加复杂;由图 7(b)第一行可以看出,本文算法的跟踪结果优于其他三类算法;图 7(b)第二行似然匹配空间中去除了背景纹理,为目标模型匹配消除了相似背景干扰.

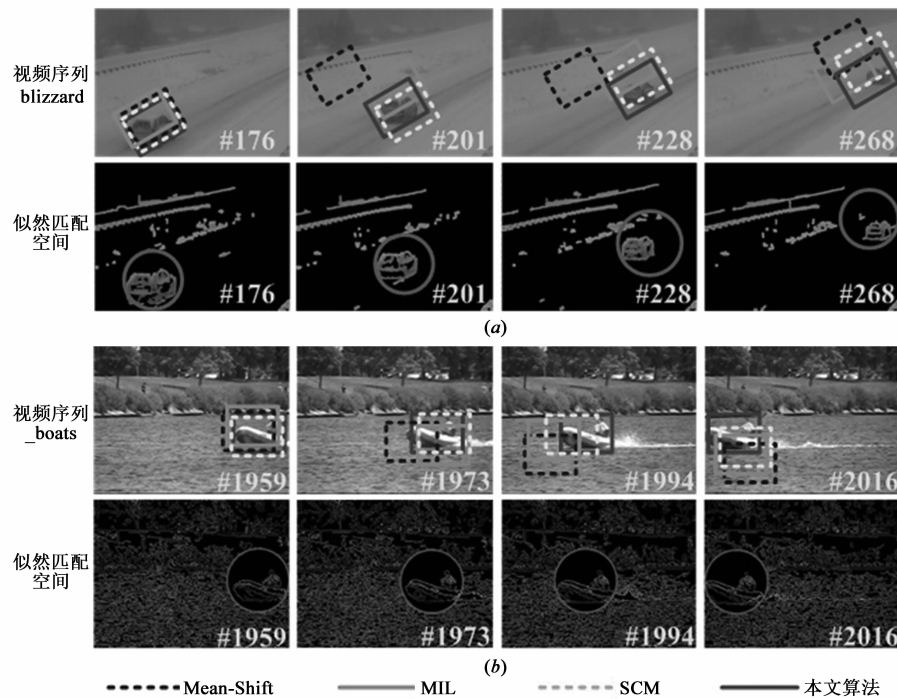


图7 实验1 本文算法、Mean-Shift、MIL和SCM在两类背景相似环境下的跟踪结果
(a) 序列“blizzard”的应用背景纹理比较单一；(b) 序列“boats”的应用背景纹理复杂

从图7可以看出, Mean-Shift 和 MIL 算法在序列“blizzard”和“boats”中都发生了严重的漂移问题, 而 SCM 算法也无法有效跟踪目标, 所以三类跟踪算法均失去了统计分析的意义; 本文算法在序列“blizzard”中的平均中心点误差 $d_i = 5$, 平均目标覆盖率 $r_i = 0.78$; 在序列“boats”中的 $d_i = 7$, $r_i = 0.67$; 说明本文算法在相似背景干扰下有较高的跟踪精度.

遮挡是跟踪中常遇到的问题, Mean-Shift、MIL 和 SCM 三类算法在相关文献中均有处理半遮挡的机制. 因此, 本文通过实验2 给出本文算法和三类算法在全遮挡环境“bungalows”和“subway”下的跟踪性能. 序列“bungalows”中的感兴趣目标于第1526帧发生全遮挡, 从图8(a)可以看出, 只有本文算法和SCM能够按照动

态模型预测的轨迹继续运动, 而 Mean-Shift 和 MIL 已完全丢失目标, 当目标于第1537帧重新出现时本文算法和SCM均能正确捕捉到目标. 序列“subway”中的目标在第3511帧发生全遮挡, 由于目标发生了较大程度地转向, 导致SCM的粒子滤波搜索策略失效, 而本文算法利用检测模块能够检测到重新出现的目标, 从图8(b)可以看出, 只有本文算法能够有效跟踪感兴趣目标.

从图8可以看出, Mean-Shift 和 MIL 算法在序列“bungalows”和“subway”中都丢失了目标, 而SCM算法仅能依靠运动模型处理序列“bungalows”中直线匀速运行的目标, 所以三类跟踪算法均失去了统计分析的意义; 本文算法在序列“bungalows”中的平均中心点误差 $d_i = 6$, 平均目标覆盖率 $r_i = 0.82$; 在序列“subway”中的

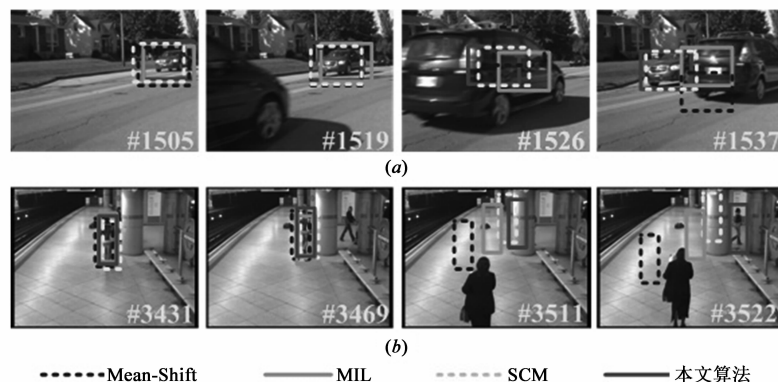


图8 实验2 本文算法、Mean-Shift、MIL和SCM在两类全遮挡环境下的跟踪结果
(a) 序列“bungalows”中的目标直线运动；(b) 序列“subway”中的目标发生转向

$d_t = 8, r_t = 0.64$;说明本文算法在全遮挡条件下依然可以有效跟踪目标.

实验 3 给出了本文算法与三类算法在其他复杂应用环境下的跟踪性能,其中序列“David outdoors”的难点:光照变化、随机运动和复杂的背景纹理;序列“car4”的难点:光照变化、随机运动、对比度低和目标尺度变化(目标逐渐变小).序列“David outdoors”中的感兴趣目标于第 222 帧发生光照变化,使得目标外观产生较大变化,Mean-Shift 算法的初始模型失效,从图 9(a)可以看出,Mean-Shift 算法在第 222 帧已丢失目标;而本文算

法、MIL 和 SCM 则可以正确跟踪目标.序列“car4”属于灰度图像且目标与背景的对比度较低,从图 9(b)可以看出,只有本文算法和 SCM 能够有效跟踪目标,这是由于 SCM 和本文算法都利用局部特征构建外观模型,提高了模型的区分度,能将前景目标从背景中分离出来.

图 10 量化反映了本文算法、Mean-Shift、MIL 和 SCM 在实验 3 中的跟踪性能;从图 10(a)可以看出,在序列“David outdoors”跟踪过程中,本文算法的中心点误差和覆盖率在大部分图像帧上均优于其他三类算法,两个指标的平均值最佳(如表 2 所示),这说明本文

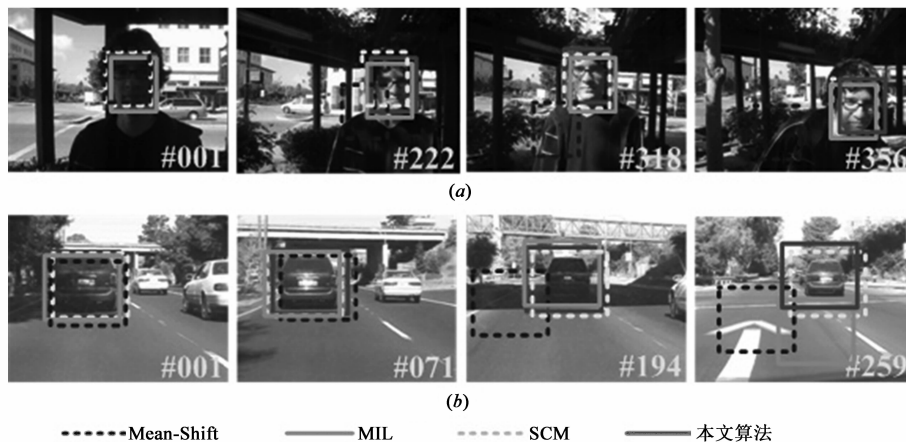


图9 实验3 本文算法、Mean-Shift、MIL和SCM在两类光照变化环境下的跟踪结果
(a) 序列“David outdoors”的背景复杂、对比度高; (b) 序列“car4”的背景单一、对比度低

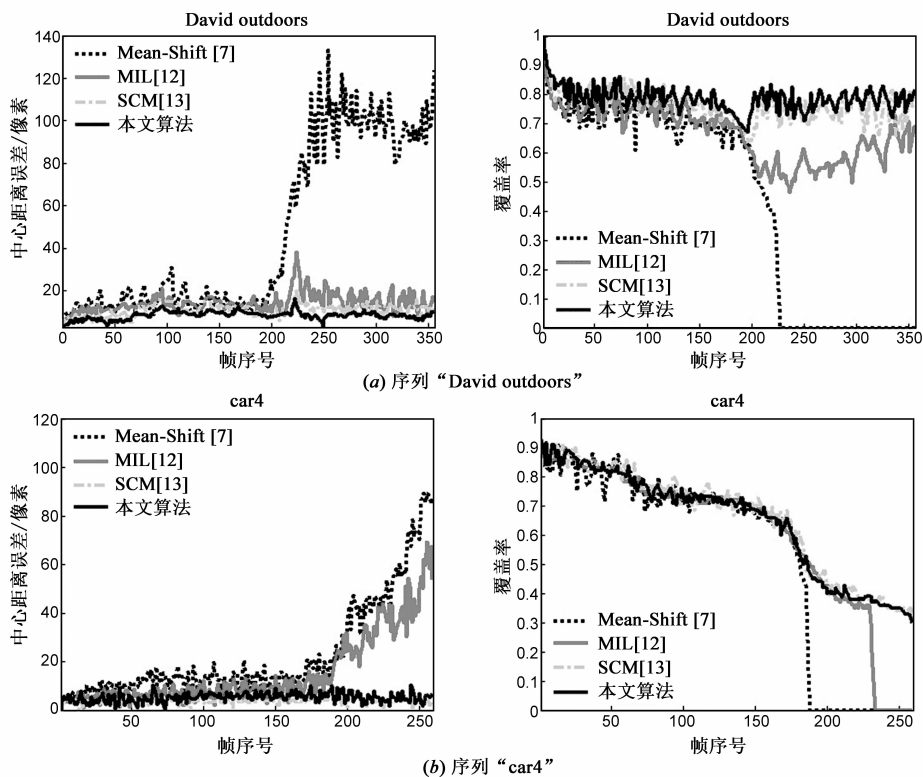


图10 实验3的定量分析图

算法在整个跟踪过程中都能保持较高的跟踪精度和鲁棒性. 从图 10(b)可以看出, 虽然本文算法和 SCM 在序列“car4”的整个跟踪过程中都保持较低的中心点误差, 但其覆盖率逐渐下降, 这是因为当目标逐渐变小时, 两类算法的跟踪窗口保持不变, 导致覆盖率下降; 如表 2 所示, 本文算法的平均指标仅次于 SCM 算法. 一般情况下, 视频流的帧率为 25 帧/秒, 显示一帧图像的时间为 40ms, 因此, 跟踪算法的运行耗时小于 40ms 就可以满足实时跟踪; 本文算法能保持“多项式级的时间复杂度”, 每帧的平均处理时间 $\approx 34.07\text{ms}$, 达到了实时处理的要求.

表 2 实验 3 中各算法的 d_t 、 r_t 指标平均值对比

运动目标 跟踪算法	“David outdoors”		“car4”	
	d_t 平均值	r_t 平均值	d_t 平均值	r_t 平均值
Mean-Shift	67.59	0.37	66.5	0.21
MIL	16.37	0.61	37.3	0.34
SCM	9.08	0.75	3.8	0.67
本文算法	7.72	0.81	4.9	0.59

7 结论

本文在无模型跟踪框架基础上提出一种基于改进协作目标外观模型的在线视觉跟踪算法. 该方法无需正、负样本的训练和学习过程, 将判别模型视为生成模型的辅助机制, 利用目标边缘信息提出基于边缘色度的判别模型, 并结合动态模型和检测模块建立二级似然匹配空间, 为生成模型去除匹配空间中的背景干扰; 采用分块策略和二进局局部特征构建生成模型, 提高了模型的稳健性和唯一性; 根据子块匹配系数处理遮挡并调整模型更新策略. 实验结果表明, 相较于其他三类具有代表性的视觉跟踪算法, 本文算法在相似背景干扰和全遮挡环境中存在明显的优势, 能够长时间有效跟踪目标且平均覆盖率大于 70%; 实验 3 证明了本文算法在其他类型的复杂应用场景下依然可以有效跟踪目标. 在未来工作中, 本文将重点研究解决跟踪窗口随目标大小自适应变化的方法, 进一步提高算法的可靠性.

参考文献

[1] Yilmaz A, Javed O, Shal M. Object tracking: a survey[J]. ACM Computing Surveys, 2006, 38(4): 1-45.
 [2] Xue J, Zheng N, Geng J. Tracking multiple visual targets via particle-based belief propagation[J]. IEEE Transactions on SMC, Part B, 2008, 38(1): 196-209.
 [3] Matej K, Stanislav K. A two-stage dynamic model for visual tracking[J]. IEEE Transactions on SMC, Part B, 2010, 40(6): 1505-1520.

[4] Liu L Q, Zhang L, Liu H R. Toward large-population face identification in unconstrained videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(11): 1874-1884.
 [5] Zhang W C, Shang L F, Antoni B C. A robust likelihood function for 3D human pose tracking[J]. IEEE Transactions on Image Processing, 2014, 23(12): 5374-5389.
 [6] Amirali J, Cai H Y, Zheng J Y. Vehicle detection and tracking in car video based on motion model[J]. IEEE Transaction on Intelligent Transportations Systems, 2011, 12(2): 583-595.
 [7] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
 [8] Zhang X, Hu W, Bao H. Robust head tracking based on multiple cues fusion in the kernel-Bayesian framework[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(7): 1197-1208.
 [9] Wang L F, Yan H P, Wu H Y. Forward-backward mean-shift for visual tracking with local-background-weighted histogram[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(3): 1480-1489.
 [10] Kalal Z, Matas J, Mikolajczyk K. P-N learning: bootstrapping binary classifiers by structural constraints[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. San Francisco: IEEE Press, 2010. 49-56.
 [11] Kalal Z, Mikolajczyk K, Matas J. Tracking-Learning-Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
 [12] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632.
 [13] Zhong W, Lu H, Yang M H. Robust object tracking via sparse collaborative appearance model[J]. IEEE Transactions on Image Processing, 2014, 23(5): 2356-2368.
 [14] Zhang L, Matten D. Preserving structure in model-free tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(4): 756-769.
 [15] Achanta R, Shaji A, Smith K, Lucchi A, Fua P and Susstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2281.
 [16] Chiveron J, Xie X H, Mirmehdi M. Automatic bootstrapping and tracking of object contour[J]. IEEE Transactions on Image Processing, 2012, 21(3): 1231-1245.
 [17] Hu W M, Zhou X, Li W. Active contour-based visual

tracking by integrating color, shapes and motions [J].
IEEE Transactions on Image Processing, 2013, 22 (5) :
1778 - 1792.

[18] Wu Y, Lim J W, Yang M H. Online object tracking: a

benchmark [A]. Proceedings of the 26th IEEE Conference
on Computer Vision and Pattern Recognition [C]. Port-
land: IEEE Press, 2013. 2411 - 2418.

作者简介



宋 涛 男, 1983 年出生于河南焦作, 现为解放军信息工程大学信息工程学院博士研究生, 主要研究方向为计算机视觉和目标跟踪技术.

E-mail: taosong_1983@126.com



李 鸥 男, 1961 年出生于河南郑州, 现为解放军信息工程大学信息工程学院教授、博士生导师, 主要研究方向为无线网络、信息融合和目标跟踪技术.

E-mail: zzliou@126.com