

基于 HSV 颜色特征和贡献度重构的行人跟踪

刘梦飞¹, 付小雁^{1,2}, 尚媛园^{1,3}, 丁辉^{1,4}

¹首都师范大学信息工程学院, 北京 100048;

²首都师范大学电子系统可靠性技术北京市重点实验室, 北京 100048;

³成像技术北京市高精尖创新中心, 北京 100048;

⁴首都师范大学高可靠嵌入式系统技术北京市工程技术研究中心, 北京 100048

摘要 在光照、背景变化、遮挡、噪声、快速运动等复杂环境下, 准确地实现行人跟踪一直是富有挑战性的任务。针对这些问题, 提出基于 HSV 颜色特征和贡献度重构的行人跟踪算法。在粒子滤波的框架内, 从 HSV 空间提取目标的混合颜色特征生成目标模板集, 依据不同区域对跟踪结果的影响对区域进行贡献度分配, 并将其引入到一个自适应的正则化模型中, 将具有最小重构误差的区域判定为待跟踪目标。为了增强算法的稳健性, 跟踪过程中对模板进行实时更新。在 OTB 100 个序列上进行测试, 本文算法得到跟踪结果的平均中心误差和跟踪成功率两项指标分别为 0.6624 pixel 和 0.4153, 优于同类其他算法。实验结果表明, 该算法能够在复杂的视频场景中实现对行人的连续跟踪, 且稳健性较好, 有利于在实际系统中的实现。

关键词 图像处理; 目标跟踪; 混合颜色特征; 贡献度重构; 稀疏表示; 粒子滤波

中图分类号 TP391 **文献标识码** A

doi: 10.3788/LOP54.091004

Pedestrian Tracking Based on HSV Color Features and Reconstruction by Contributions

Liu Mengfei¹, Fu Xiaoyan^{1,2}, Shang Yuanyuan^{1,3}, Ding Hui^{1,4}

¹ College of Information Engineering, Capital Normal University, Beijing 100048, China;

² Beijing Key Laboratory of Electronic System Reliability Technology, Capital Normal University, Beijing 100048, China;

³ Beijing Advanced Innovation Center for Imaging Technology, Beijing 100048, China;

⁴ Beijing Engineering Research Center of High Reliable Embedded System, Capital Normal University, Beijing 100048, China

Abstract It is a challenging task to track pedestrian accurately in complicated environment such as illumination, background variation, occlusion, noise and fast motion. Aiming at these problems, the tracing algorithm based on HSV color features and reconstruction by contributions is proposed. The proposed algorithm extracts the mixed color features of target in HSV space to generate the target template set within the particle filter framework. According to the influence of different regions on the tracking results, the contribution of the region is distributed. And it is introduced into the adaptive regularization model, and the region with the minimum reconstruction error is determined as the target to be tracked. In order to be more robust, the templates are updated in real time during the tracking progress. The average center error of tracking results and tracking success rate of 100 sequences tested in OTB are 0.6624 pixel and 0.4153, respectively, and the proposed algorithm has better performance than others.

收稿日期: 2017-02-14; **收到修改稿日期:** 2017-03-16

基金项目: 国家自然科学基金(61303104, 61203238, 11178017, 61473015)、北京市自然科学基金(4132014, 4162017)、北京市优秀人才资助项目(2016000020124G088)

作者简介: 刘梦飞(1994—), 女, 硕士研究生, 主要从事目标跟踪与计算机视觉方面的研究。E-mail: lmf_getbetter@sina.com

导师简介: 付小雁(1979—), 女, 博士, 副教授, 主要从事目标跟踪、计算机视觉与图像处理等方面的研究。

E-mail: fuxiaosg@163.com(通信联系人)

Experimental results show that the proposed algorithm can realize the continuous tracking for pedestrian in complex video scenes and is beneficial to be realized in the practice system with better robustness.

Key words image processing; target tracking; mixed color features; reconstruction by contributions; sparse representation; particle filter

OCIS codes 100.4999; 100.4993; 000.3860

1 引言

在计算机视觉中,行人跟踪广泛应用于智能监控、智能交通管理、车辆导航等领域,因此,研究行人跟踪对于建立智能化城市具有重要意义。在过去的二十多年,视觉跟踪已经取得了很大的突破。但是,由于实际采集的视频场景复杂多样、行人在运动过程中自身形态的变化或者受到遮挡等,行人跟踪过程中依然存在跟踪精度不高以及跟踪不连续的问题。

行人跟踪是目标跟踪的一种,目前目标跟踪算法主要包括判别式模型跟踪和生成式模型跟踪算法。在判别式模型跟踪算法中,跟踪问题被视为一个二元的分类问题,即首先锁定物体所在的局部区域,对目标区域的像素构建一个特征向量,并训练一个自适应的分类器来区分当前像素所在区域是否为所要追踪的目标^[1]。常见的判别式模型算法有全局模式搜寻^[2]、在线多例学习^[3]、稀疏贝叶斯^[4]等。生成式模型跟踪算法依赖于物体的外观模型,即通过某种方法得到目标外观模型在图像中最匹配的位置和大小,匹配的标准如候选区与目标模板的欧氏距离、相似度等。常见的生成式模型跟踪算法为改进的粒子滤波^[5]、均值漂移(MS)跟踪^[6]、增量跟踪^[7]和压缩感知^[8]等。对比这两类跟踪算法可知,判别式模型比生成式模型简单,比较容易训练,但其缺点是仅能指出所要搜索的区域是或者不是目标,而不能反映训练数据本身的特性;生成式模型反映的信息更加丰富,比判别式模型更灵活,能用于数据不完整的情况,但其学习和计算的过程比较复杂。

近些年来,利用稀疏表示进行目标跟踪的方法受到很多学者的青睐。2009年,Mei等^[9]首次将稀疏表示引入到目标跟踪领域,该方法以粒子滤波为框架,将目标的灰度信息进行编码后生成目标模板集,然后联合目标模板集和琐碎模板集建立稀疏表示模型,将跟踪问题转化成了L1正则化问题。其中,琐碎模板集被用来线性表示目标的遮挡和图像中的噪声。这种方法使用内插点的方法对L1正则化问题进行求解,计算量较大,难以实现实时跟踪。2013年,Mei等^[10]使用最小误差边界算法对其进行了优化,对粒子进行了筛选,减少了L1范数每帧的运算次数,提高了运算速度。2012年,Zhang等^[11]提出压缩跟踪(CT),利用稀疏测量矩阵对正负样本的多尺度图像特征进行降维,不仅降低了复杂度,还提高了跟踪速度,但是这种算法的代价是损失了Mei等^[9-10]算法的跟踪精度。同年,Bao等^[12]在其基础上利用加速近端梯度(APG)法求解L1正则化问题,然而,APG算法仅提高了处理速度,而对跟踪的准确度没有太大的提高。2016年,Yang等^[13]使用正交的目标模板集作为目标的外观模型,然后求出与模板具有最小残差的样本作为跟踪结果。在计算L1正则化问题中使用最小平方法,这种计算方法虽然减少了运算量,但跟踪的准确度低于APG算法。2015年,Henriques等^[14]提出核相关滤波(KCF)算法,利用循环矩阵进行快速傅里叶变换,通过相关滤波的方法进行目标跟踪,取得了不错的效果,速度优势为其一大特点,但不能很好地处理前后帧目标位移过大以及目标尺度变化较大的问题。

由于跟踪目标的多样以及跟踪场景的复杂性,上述算法对行人这种特定目标的跟踪效果依然有待提高。作为目标跟踪中的一类,行人跟踪有着独特的特点:行人出现在视觉中心区域的概率比较大,四周则为背景区。在建立目标的外观模型时,原始的L1跟踪算法没有考虑到不同区域对跟踪结果的贡献度也不同,背景的变化会导致模型发生变化,从而造成跟踪结果出现偏差。在进行模板更新的过程中,行人可能会受到遮挡,如果此时错误地对模板进行更新则会导致跟踪结果出现漂移,因此建立一个能够准确捕捉行人在复杂背景中的运动状态的自适应正则化模型显得尤为重要。L1跟踪算法仅使用目标的灰度信息作为稀疏表示的目标模板,若出现较长时间的光照变化等干扰时,会出现跟踪不稳定现象。

针对上述问题,本文提出了基于HSV颜色特征和贡献度重构的行人跟踪算法。该算法以粒子滤波为框架,从原始图像中提取HSV空间的混合颜色特征作为稀疏表示的目标模板,结合贡献度分配,对有着较大视觉中心区域分配较大的权重,四周则分配较小的权重,以此应对背景发生较大变化的情况。最后将其引入到一个自适应的正则化模型,通过求解该模型与目标模板的最小残差得到跟踪结果。

2 传统的 L1 跟踪算法

基于 L1 正则化的跟踪算法采用粒子滤波作为框架,并结合稀疏表示理论获得目标的外观模型,通过求取图像与目标模板的最小重构误差获得目标的位置。

2.1 粒子滤波

在 t 时刻,使用 \mathbf{x}_t 描述目标的状态,它由 6 个仿射变换参数构成,分别代表目标的形状和位置。粒子滤波将跟踪过程转化为 $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ 概率估计问题,是预测下一帧目标状态的重要工具。粒子滤波过程由预测和更新两个步骤组成,方程为

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{z}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{z}_{t-1}) d\mathbf{x}_{t-1}, \quad (1)$$

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1})}{p(\mathbf{z}_t | \mathbf{z}_{1:t-1})}, \quad (2)$$

式中 $\mathbf{z}_{1:t-1} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{t-1}\}$ 是从第 1 帧到第 $t-1$ 帧目标的观测值, t 时刻的目标信息由最大后验概率得到,即

$$\mathbf{x}_t^* = \operatorname{argmax}_x p(\mathbf{x} | \mathbf{z}_{1:t}), \quad (3)$$

t 时刻粒子样本状态集 $\mathbf{S}_t = [\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^n]$ 是通过重要性采样分布来获得的,粒子的权重更新方式为

$$w_t^i \propto w_{t-1}^i \frac{p(\mathbf{z}_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i)}{\prod (\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{x}_{1:t-1})}, \quad (4)$$

这里需要说明的是,对权值需要进行归一化处理。

基于重要性采样分布的粒子滤波算法中存在样本退化的问题,即在滤波过程中经过几次迭代,仅有一个或几个样本的权值较大。随着无效采样粒子数的增加,粒子集无法表示样本实际的后验概率分布,造成整体估计性能下降。该问题一般通过重采样(舍去权值小的粒子并复制权值大的粒子)解决,即

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) = \sum_{i=1}^N \frac{n_i}{N} \delta(\mathbf{x}_t - \mathbf{x}_t^i), \quad (5)$$

式中 n_i 是粒子 \mathbf{x}_t 产生新的粒子 \mathbf{x}_t^i 时被复制的次数, N 为粒子总数, $\delta(x)$ 是冲激函数。经过重采样以后,所有的粒子视为独立同分布,权值均为 $1/N$ 。且重采样前后粒子总数不变,从而避免粒子退化现象。

2.2 稀疏表示理论

稀疏表示的目的是在一个过完备字典中寻找对目标的最佳表示,通过目标模板集和琐碎模板集的线性组合来实现^[15],其中稀疏性主要体现在琐碎模板集及其对应的加权系数上。

建立稀疏表示模型首先需要获得目标模板集 \mathbf{T} 和琐碎模板集 \mathbf{I} 。设 t 时刻的目标模板集 $\mathbf{T}_t = \{\mathbf{t}_t^1, \mathbf{t}_t^2, \dots, \mathbf{t}_t^i, \dots, \mathbf{t}_t^n\} \in \mathbf{R}^{d \times n}$,该模板集包含了 n 个模板, d 为每一个模板向量的维度。第一个模板集 \mathbf{T}_1 是通过手动截取目标区域并进行平移旋转产生的。琐碎模板集 \mathbf{I} 是一个单位矩阵,被用于线性表示目标遮挡部分以及图像噪声。通过目标模板集 \mathbf{T} 和琐碎模板集 \mathbf{I} 的线性组合可得到目标的候选集 \mathbf{O}_t ,即

$$\begin{cases} \mathbf{O}_t = \{\mathbf{a}_t^1, \mathbf{y}_t^2, \dots, \mathbf{y}_t^n\} \in \mathbf{R}^{d \times n} \\ \mathbf{y}_t^i = \mathbf{T}_t \mathbf{a}_t^i + \mathbf{I} \mathbf{a}_t^i, \forall \mathbf{y}_t^i \in \mathbf{O}_t \end{cases}, \quad (6)$$

式中 $\mathbf{a}^i = \{\mathbf{a}_t^i, \mathbf{a}_t^i\}$ 是模板系数,上标 i 表示第 i 个粒子, \mathbf{a}_t^i 和 \mathbf{a}_t^i 分别对应目标模板集 \mathbf{T} 和琐碎模板集 \mathbf{I} 对应的加权系数。当不存在遮挡时,候选区的外观模型 \mathbf{y}_t^i 可以由目标模板集来有效表示,这时 \mathbf{a}_t^i 的稀疏性较强;当存在较大遮挡时, \mathbf{a}_t^i 中包含较多的非零元素用于补偿。因此,较好的跟踪结果应该使得琐碎模板集的加权系数达到最大的稀疏性,这意味着求取跟踪结果的过程可以转化为求解 \mathbf{y} ,即

$$\min \frac{1}{2} \|\mathbf{y} - \mathbf{A} \mathbf{a}\| + \lambda \|\mathbf{a}\|_1, \quad (7)$$

式中 λ 是对稀疏系数大小的限定,一般取 0.01;模板集 $\mathbf{A} = (\mathbf{T}, \mathbf{I})$,目标模板集 \mathbf{T} 由动态模板和静态模板组成,在第 1 帧中手动选取目标作为固定模板,其他 9 个模板均为动态模板,在跟踪过程中被实时更新^[16-17]。

2.3 模型的求解

在求解 L1 范数最小化问题中,使用内插点的算法运算起来比较慢^[9,18],Mei 等^[10]提出的最小误差边界

算法,对粒子进行了筛选,可以减少计算量。该算法将下式作为基础,即

$$\|\mathbf{T}_t \mathbf{a} - \mathbf{y}\| \geq \|\mathbf{T}_t \hat{\mathbf{a}} - \mathbf{y}\|_2^2, \forall \mathbf{a} \in \mathbf{R}^N, \quad (8)$$

式中 $\hat{\mathbf{a}} = \operatorname{argmin} \|\mathbf{T}_t - \mathbf{y}\|_2^2$, 对于任意一个样本,观测相似度 $p(\mathbf{z}_t | \mathbf{x}_t^i)$ 有着上边界,即

$$p(\mathbf{z}_t | \mathbf{x}_t^i) = \frac{1}{F} \exp(-\alpha \|\mathbf{y}_t^i - \mathbf{T}_t \mathbf{a}_T^i\|_2^2), \quad (9)$$

$$p(\mathbf{z}_t | \mathbf{x}_t^i) \leq \frac{1}{F} \exp(-\alpha \|\mathbf{y}_t^i - \mathbf{T}_t \hat{\mathbf{a}}\|_2^2) \stackrel{\text{def}}{=} q(\mathbf{z}_t | \mathbf{x}_t^i), \quad (10)$$

式中 $q(\mathbf{z}_t | \mathbf{x}_t^i)$ 是 \mathbf{x}_t^i 的概率上边界,可证明,如果

$$q(\mathbf{z}_t | \mathbf{x}_t) < \frac{1}{2N} \sum_{j=1}^{i-1} p(\mathbf{z}_t | \mathbf{x}_t^j), \quad (11)$$

则样本 \mathbf{x}_t^i 不会出现坍塌退化,因此不需要经过重采样步骤,从而在一定程度上降低运算复杂度。

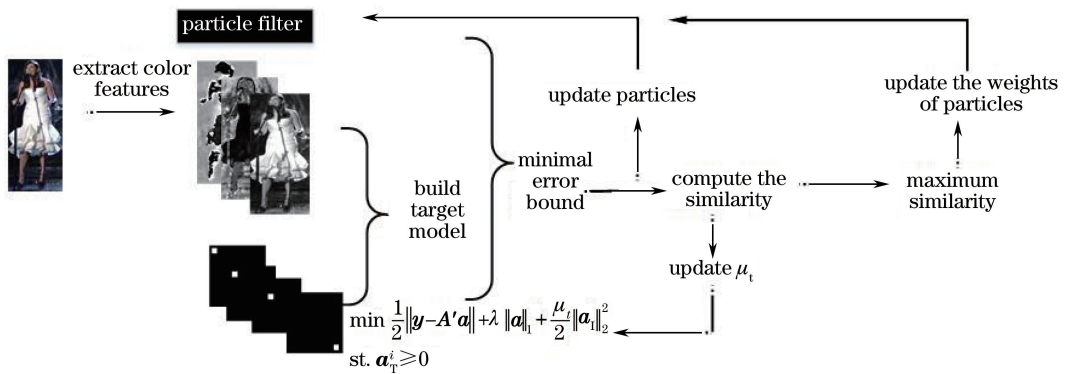


图1 实验算法框图

Fig. 1 Diagram of experimental algorithm

3 改进的行人跟踪算法

传统的 L1 跟踪器提取目标的灰度信息作为稀疏表示的目标模板,当出现光照及背景变化时,跟踪结果往往会出现漂移现象。行人跟踪的目标通常出现在视觉中心区域,而原始的 L1 正则化跟踪算法在建立目标外观模型时没有考虑到各区域对跟踪结果的贡献度不同,而是对整个行人及周围区域的灰度值进行统一编码来建立目标模板,当背景区域变化过大时跟踪效果会出现不稳定现象。另外,L1 跟踪用琐碎模板集线性表示遮挡部分,这种处理方式也可能将遮挡部分表示成目标,从而增加跟踪误差。因此提出更加兼顾跟踪精确性和稳健性的行人跟踪算法。

3.1 算法框架

提出一种基于 HSV 颜色特征和贡献度重构的行人跟踪算法,该算法的整体框架如图 1 所示。预处理阶段,在 HSV 空间提取一个具有混合颜色特征的目标模板,归一化处理之后将该模板进行高斯扰动仿射参数生成模板集。跟踪阶段,利用粒子滤波估计下一刻的状态(候选区),根据行人跟踪中不同区域对跟踪结果的不同影响,对候选区进行贡献度分配,构造出一个新的样本特征,然后联合目标模板集和琐碎模板集进行稀疏表示,重构出该样本的外观模型。最后求出与目标特征具有最小重构误差的外观模型,将该模型对应的样本确定为待跟踪的目标。为了提高算法的稳健性,在跟踪过程中对模板进行实时更新。

3.2 混合颜色特征提取

在行人检测和跟踪中,方向梯度直方图(HOG)^[19]特征因其对图像几何和光学的形变不敏感性而得以广泛应用,但是其损失了颜色这一基本特征。作为人类感知的基本信息之一,颜色特征在目标跟踪和模式识别方面已经引起了研究者的广泛关注^[20]。最近研究表明,提取不同颜色空间中的信息作为外观模型比使用传统灰度信息的跟踪效果更好^[21]。

图像的颜色模型表示方法有很多种,如 RGB(R 为红色分量、 G 为绿色分量、 B 为蓝色分量),HSV(H

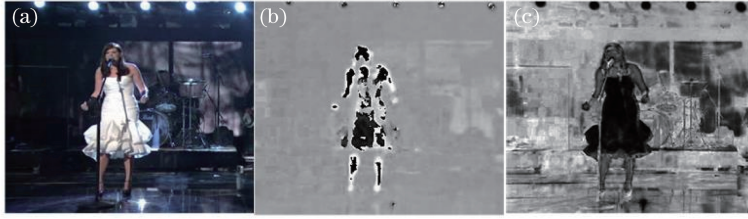


图2 颜色特征提取效果。(a)原图;(b)色调分量;(c)饱和度

Fig. 2 Extraction effects of color features. (a) Original image; (b) hue component; (c) saturability

为色调、 S 为饱和度、 V 为亮度)等。RGB 空间的三个分量随着光照的改变而改变,且两个相近的颜色其 R 、 G 、 B 值可能相差很大,同时 RGB 空间对目标的颜色描述相对复杂,各分量之间冗余信息多;而 HSV 空间能够独立感知各分量的变化,较好地反映人眼对色彩的感知和鉴别能力。并且,监控视频或录像过程中的光照条件各异,利用 H 和 S 分量可以减少光照变化的影响,对目标跟踪具有十分重要的意义。因此在 HSV 空间提取 H 和 S 分量的混合特征,如图 2 所示。由 RGB 空间到 HSV 空间的转换关系为

$$V \leftarrow \max(R, G, B), \quad (12)$$

$$S \leftarrow \begin{cases} \frac{V - \min(R, G, B)}{V}, & \text{if } V \neq 0 \\ 0, & \text{otherwise} \end{cases}, \quad (13)$$

$$H \leftarrow \begin{cases} 60(G - B)/[V - \min(R, G, B)], & \text{if } V = R \\ 120 + 60(B - R)/[V - \min(R, G, B)], & \text{if } V = G \\ 240 + 60(R - G)/[V - \min(R, G, B)], & \text{if } V = B \end{cases}, \quad (14)$$

$$H \leftarrow H + 360, \quad \text{if } H < 0$$

式中分量通常规定为 $0 \leq V \leq 1, 0 \leq S \leq 1, 0 \leq H \leq 360$ 。

转换成 HSV 空间后,生成图像的混合颜色特征为 L_i^i ,为了提高跟踪效果,实验中在保持色度信息的同时对饱和度进行均衡化处理。均衡化是图像增强常见的方法,在计算机视觉中有着非常重要的作用。通过对图像进行均衡化处理可以提高算法的适用性,如对低照度图像、运动模糊图像以及雾霾天气图像的识别问题。最终的特征提取公式为

$$\begin{cases} L_i^i = \gamma \cdot H_i^i + \beta \cdot S_i^i \\ \text{st. } \gamma + \beta = 1 \end{cases}, \quad (15)$$

式中 γ 和 β 分别表示对色调和饱和度特征所取的权值, $\gamma \in (0.5, 1)$ 。

3.3 按贡献度重构的稀疏表示模型

本文使用核函数来进行贡献度分配。近几年来,核函数广泛地应用于数据的概率密度估计,不同的核函数反映了数据的不同分布,常见的核函数有均匀核、多项式核、高斯核、Epanechnikov 核等^[21]。结合行人跟踪的特点,本文引入 Epanechnikov 核来实现贡献度分配。该核函数的形式为

$$K_E(\mathbf{X}) = \begin{cases} c \cdot \left(1 - \frac{\|\mathbf{X} - \mathbf{X}_0\|_2^2}{h}\right), & \text{if } \mathbf{X} \in D, \\ 0, & \text{otherwise} \end{cases}, \quad (16)$$

式中 \mathbf{X} 为混合特征中像素点的坐标, \mathbf{X}_0 是目标区域中心点的坐标, h 表示目标所在矩形框中对角线的长度, c 为归一化常数, D 为目标所在矩形框。

贡献度函数整体呈山峰型,函数值从中心向四周递减。对于行人跟踪而言,目标出现在视觉中心区域的概率比较大,四周则为背景区。这与函数 $K_E(\mathbf{X})$ 的峰值趋势较吻合,因此利用函数 $K_E(\mathbf{X})$ 对行人的混合颜色特征进行贡献度分配:将较大概率是前景的中心区域分配较大的权重,四周则分配较小的权重,权值分布如图 3 所示。利用该函数对 2.2 节提取到的目标混合颜色特征进行处理,即

$$Q_i^i = L_i^i \cdot K_E(\mathbf{X}), \quad (17)$$

式中 L_i^i 是第 i 个粒子对应的目标混合颜色特征, Q_i^i 是对其按照贡献度加权之后的目标模板。将 Q_i^i 作为新

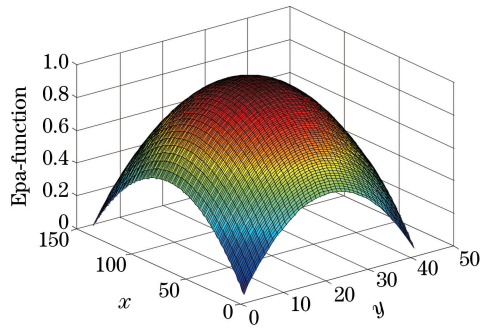


图3 权重分布图

Fig. 3 Distribution of weights

的模板特征,生成目标模板集,然后联合琐碎模板集可以重构目标的外观模型。虽然琐碎模板集可以有效表示图像中的噪声或者遮挡区域,但其本身也有表示目标部分的可能。这种情况可能导致跟踪结果的不稳定,因此引入遮挡参数来修正 L1 正则化模型^[12]。这时,求解稀疏表示模型就可以转化为带非负约束下的 L1 范数相关最小化问题,即

$$\begin{cases} \min \frac{1}{2} \|\mathbf{y} - \mathbf{A}'\mathbf{a}\| + \lambda \|\mathbf{a}\|_1 + \frac{\mu_t}{2} \|\mathbf{a}_1\|_2^2, \\ \mathbf{A}' = [\mathbf{K}_E \cdot \mathbf{T}, \mathbf{I}] \end{cases}, \quad (18)$$

式中 $\mathbf{K}_E \cdot \mathbf{T}$ 是按贡献度进行分配之后的目标模板,参数 μ_t 用来控制琐碎模板的能量。实验中, μ_t 视 t 时刻目标的遮挡情况而自动赋值:将稀疏表示中琐碎模板集对应的系数向量转化为二维平面,将该平面二值化,若其白色区域的面积大于 30% 则认为有遮挡,此时 $\mu_t = 0$; 否则,将其赋值为一个预设的常数^[10]。采用文献^[12]中的加速梯度算法来求解(18)式。步骤如下:

1) 初始化值 $\alpha_0 = \alpha_{-1} = 0, t_0 = t_{-1} = 1$ 。

2) 令 k 为循环变量,反复迭代直到收敛,即:

Step1 初始化值 $\alpha_0 = \alpha_{-1} = 0, t_0 = t_{-1} = 1$ 。

Step2 令 k 为循环变量,反复迭代直到 α_{k+1} 收敛。

$$\beta_{k+1} := \alpha_k + \frac{t_{k-1} - 1}{t_k} (\alpha_k - \alpha_{k-1}), \quad (19)$$

$$\alpha_{k+1} := \operatorname{argmin}_a \frac{L}{2} \|\mathbf{a} - \beta_{k+1} + \frac{\nabla F(\beta_{k+1})}{L}\|_2^2 + G(\mathbf{a}), \quad (20)$$

$$t_{k+1} := \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \quad (21)$$

式中 L 为利普西茨常数, β_k 为中间向量, t 为步长,最后收敛的 α_k 即所求。

4 实验与结果分析

为了验证本文算法的有效性和实用性,在 OTB 上进行了测试。这个测试集包含了 98 个测试视频,其中, Jogging、Skating2 的视频库中各自包含两个目标对象,因此总共 100 个测试序列。

本文算法分别与改进前 APG 算法、经典 MS 算法、CT 算法以及目前较为流行的 KCF 算法进行了比较实验。其中,对比算法使用作者提供的代码,参数为默认参数^[22]。以上算法的运行环境均为 Intel(R) Core (TM) 的计算机 (CPU i5-3470、3.20 GHz、内存 4 GB、32 位 Windows 操作系统),均采用 Matlab2014a 平台实现。

4.1 实验过程

实验中取 $\alpha = 0.6, \mu_t = 0.5$ 。实验过程如下。

1) 输入模版集 \mathbf{T}_t 。将 t 时刻的图像从 RGB 颜色空间转换到 HSV 空间后,第 1 帧目标的模板集为手动

选取及向四周平移产生的目标。

2) 生成样本集 S_t 。样本集由粒子滤波产生,得到 600 个高斯分布的粒子,每个粒子对应的样本代表 t 时刻目标的状态。

3) 由样本状态集生成目标候选集 O_t 。使用贡献度函数 $K_E(x)$ 对目标模板进行处理,之后求得具有最小重构误差的候选区,其对应的样本就是目标的状态。为了减少粒子滤波中重采样的次数,加入了(9)式中的判断条件。步骤为:

①对(10)式中的 q 进行降序排列,当满足 q 小于或等于预设值时进行下一步;该预设值的取值可参考(11)式;

②利用 APG 算法[(19)~(21)式]解决(18)式的范数相关最小化问题;

③计算(9)式中的相似度,直到其达到最大,同时更新步骤 3)中步骤①中的预设值。

4) 输出跟踪结果。根据(3)式得到此刻的目标信息,并进行遮挡检测,更新仿射系数集和模板集。

4.2 结果分析

对算法的跟踪结果分别进行定性和定量分析。定性分析直接从视觉观察上判断目标跟踪的准确程度,依赖于人眼的视觉特性;定量分析从跟踪误差和跟踪成功率来量化跟踪结果,更具说服力。在定量分析中,本次实验的跟踪效果判别采用两种跟踪基准:1) 跟踪误差,即跟踪的目标的中心点与手工标定的准置之间的欧氏距离,误差率为跟踪误差大于所选距离阈值的帧出现的频率;2) 目标跟踪成功率,首先引入重叠率的概念,即所有的视频中跟踪区域和目标真实区域的交集与这两个区域并集的比值。成功率是重叠率大于给定阈值(比如 0.5)的成功帧的比率。在 OTB 中 100 个序列上的测试结果如图 4 所示。

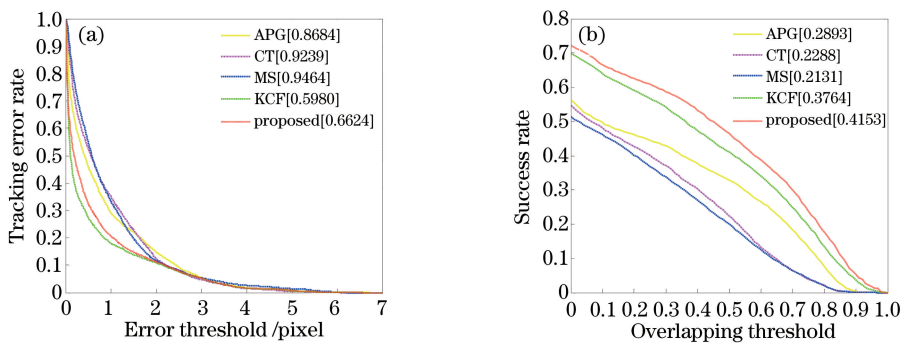


图 4 100 个序列的(a)跟踪误差和(b)跟踪成功率

Fig. 4 (a) Tracking error and (b) tracking success rate for 100 sequences

从跟踪效果来看,本文算法的成功率高于同类其他算法,具有一定的优势。但是,其跟踪误差并不是最小的。原因是:1) 本文算法利用的是目标的全局信息,当目标发生大规模的位移等变化时对中心点的定位精确度有待提高;2) 跟踪误差作为衡量跟踪效果的标准,其本身就有局限性,即主要衡量的是目标是否能够被定位到,不能反映跟踪的具体性能,比如跟踪区域的契合度等。

为了更加直观和详细地对不同算法的结果进行比较,选取 7 个行人跟踪库进行详细分析,这些跟踪库包含了绝大多数可能出现的干扰情况,其干扰属性分布统计如图 5 所示,OPR 为离面旋转,IPR 为面内旋转,BC 为背景相似,FM 为快速运动,MB 为运动模糊,DEF 为变形,OCC 为遮挡,SV 为尺度变化,IV 为光照变化。

4.2.1 定性分析

图 6(a)中 Couple 库是两个在室外采集的行人,其中包含了光照变化以及背景相似等情况。在第 118 帧出现了光照变化,APG、CT、KCF 算法仅利用目标的灰度信息,因此跟踪出现错误;本文算法利用 HSV 空间特征,该特征对光照不敏感,能较好地保持对目标的跟踪;MS 算法利用目标的 RGB 三通道亮度直方图,但在第 17、105 等帧出现较大的跟踪偏差,这是背景 RGB 特征和行人的 RGB 特征相似所致。

图 6(b)为 Human7 库的实验结果,在这个序列中主要存在光照、行人自身姿态变化以及快速运动等问题。在第 119 帧中存在模糊干扰,从而造成 APG、CT、KCF、MS 算法均出现了较大的跟踪误差,在第 244 帧

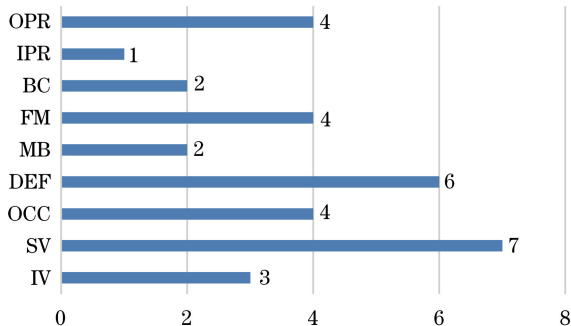


图 5 干扰属性分布

Fig. 5 Attribute distribution of interferences

中,出现了目标部分遮挡现象,APG 算法和本文算法能够对目标进行有效的跟踪,其他三种算法均出现严重偏移。本文算法能一直保持稳定有效地跟踪,进一步的证明了对图 6(a)中不同算法分析的正确性。

在图 6(c) Singer1 库中,5 种算法都能够将目标区域检测出来,CT、MS、KCF 算法跟踪结果和真实目标的契合度较低,这是由于算法不具有尺度不变性,不能根据目标大小及时做出相应的调整,而 APG 算法和本文算法的跟踪结果较好,后者的调整更明显一些,尺度契合更好。

图 6(d) Walking 库是行人慢慢远离摄像头的过程,整个序列中,行人的移动较为缓慢,中间出现了几秒钟被电线桩遮挡的画面。在前 300 帧中,5 种算法都能跟踪上目标,但是 300 帧之后由于背景特征与目标相似,MS 算法跟踪结果开始出现漂移;而随着行人距离摄像头越来越远,目标在视野中开始变小,由于 MS、CT、KCF 算法对尺度比较敏感,因此对目标的跟踪契合度较低,而本文算法对目标跟踪的精度较高。

较大遮挡的行人跟踪效果如图 6(e) 中 Woman 库所示。观察可知,该库中的第 80 帧左右出现明显的光照变化,由于原始 APG 算法和 CT 算法提取的特征对光照敏感,且模板更新具有滞后性,因此无法持续跟踪到目标。MS 算法没有利用目标的局部特征,在行人被多次遮挡之后出现跟不上的情况。KCF 算法能够保持较高的跟踪精度。总体来看,本文算法相比于 APG、MS、CT 算法跟踪效果较好,但是跟踪过程中也会出现偏移,对于较大的遮挡情况处理效果依然有待改进。



图 6 不同算法的跟踪结果。(a) Couple; (b) human7; (c) singer1; (d) walking; (e) woman; (f) crossing; (g) skater

Fig. 6 Tracking results of different algorithms. (a) Couple; (b) human7; (c) singer1;

(d) walking; (e) woman; (f) crossing; (g) skater

图 6(f) Crossing 库包含了背景相似、部分遮挡以及光照变化等情况。从整体跟踪结果来看,第 30 帧很多算法的结果出现了偏移,这是背景与目标的灰度特征相似所致。而 KCF 算法和本文算法能够对目标保持稳定的跟踪。其次,目标的真实轮廓存在尺度变化,因此本文算法跟踪契合度优于 KCF 算法。

包含较大变形的跟踪效果如图 6(g)中 Skater 库所示。在目标发生较大变形时,5 种算法的跟踪结果和真实目标的契合度较低,这是由于目标特征不能适应较大的空间变化所致。如第 37 帧,目标的胳膊和腿部发生较大变化,而 APG 算法和本文算法的模板没能及时更新,导致跟踪出现偏移;MS 算法和 CT 算法的结果依然有待提高。总体来看,KCF 算法利用多通道数据融合算法并融入了分类器,取得了不错的效果。

4.2.2 定量分析

将目前行人跟踪中广泛使用的 HOG 特征与本文提出的混合颜色特征分别进行实验,得到二者的跟踪误差对比如图 7 所示。该误差对比图是在 Couple 库上所测,该序列在第 20 帧时出现光照变化。HOG 特征的跟踪结果开始出现偏移,并且随着时间的延长,偏移量越来越大,而颜色混合特征在光照变化以及背景相似等复杂情况下,对跟踪目标具有一定的优势。

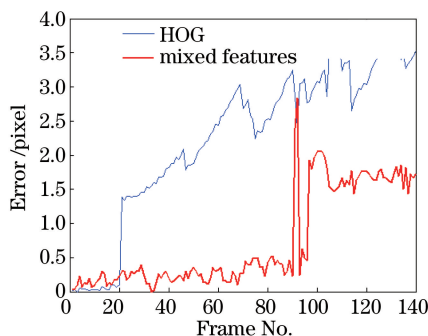


图 7 HOG 特征和混合颜色特征跟踪误差

Fig. 7 Tracking error for HOG and mixed color features

5 种算法的平均跟踪误差如表 1 所示。由表 1 可知,本文算法在 7 种测试库中跟踪中心点的平均误差最小。但并不是对于每一个测试库都有最小的跟踪误差。结合图 6 来看,当目标存在尺度变化或变形时,虽然算法能够定位到目标的中心,但往往失真严重,所以仅使用跟踪误差无法准确衡量跟踪结果的好坏。为了使跟踪结果更具有代表性,统计了不同阈值下的成功率曲线面积(AUC)^[22]。由表 2 可知,在 Couple、Singer、Walking、Woman、Crossing 库中,本文算法的 AUC 均为最大,反映了本文算法跟踪的稳定性和持续性。

表 1 5 种算法的跟踪误差比较

Table 1 Comparison of tracking error for five algorithms

Algorithm	Database							Average error /pixel
	Couple	Human7	Singer1	Walking	Woman	Crossing	Skater	
APG	1.0329	0.7480	0.1721	0.0684	0.3208	0.1176	0.9279	0.4840
CT	0.9724	1.0952	0.6890	0.1280	1.5586	0.2797	0.8842	0.8010
MS	0.5827	1.7546	0.5919	0.1966	1.8432	0.5744	1.8154	1.0513
KCF	1.4786	1.2231	0.5164	0.0665	0.2574	0.0966	0.6776	0.6166
Proposed	0.7436	0.3966	0.1456	0.0888	0.6986	0.0752	0.8864	0.4335

表 2 5 种算法的 AUC 值比较

Table 2 Comparison of AUC for five algorithms

Algorithm	Database							Average AUC
	Couple	Human7	Singer1	Walking	Woman	Crossing	Skater	
APG	0.4531	0.4546	0.3153	0.6191	0.1619	0.7070	0.4766	0.4554
CT	0.4812	0.2786	0.3520	0.5477	0.1322	0.6597	0.5731	0.4321
MS	0.2868	0.2451	0.4081	0.4102	0.1293	0.4432	0.4289	0.3359
KCF	0.2069	0.2868	0.3579	0.5384	0.4352	0.7084	0.6136	0.4496
Proposed	0.5786	0.4877	0.6334	0.6483	0.1897	0.7148	0.5564	0.5441

5 结 论

针对行人跟踪的特点以及行人易受到光照、背景变化及遮挡等影响,提出了一种基于 HSV 颜色特征和贡献度重构的行人跟踪算法。该算法以粒子滤波为框架,考虑到色度饱和度信息对光强不敏感,将原始的灰度特征转化成 HSV 混合颜色特征。在跟踪过程中,由于不同区域对跟踪结果的贡献度不同,对行人的特征进行贡献度分配,对较大概率为行人的中心区域分配较大权重,则周围区域权重较小,然后联合琐碎模板集求出与模板具有最小重构误差的外观模型。这种方法充分利用了空间信息,对光照和背景变化的行人的跟踪稳健性较好。另外,根据遮挡系数的大小,判断目标是否受到遮挡,自动选择正则化模型,使得模型对于遮挡也有一定的适应性。实验表明了本文算法的有效性。然而,本文算法对于遮挡较大的行人视频集的处理效果依然有待改进。下一步的计划是寻找对于较大遮挡状态不敏感的特征,将此特征作为建立外观模型的模板,从而进一步提高算法的精度。

参 考 文 献

- [1] Avidan S. Ensemble tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261-271.
- [2] Zhao Z Y, Collins R T. Object tracking and detection after occlusion via numerical hybrid local and global mode-seeking [C]. Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008: 1-8.
- [3] Babenko B, Yang M, Belongie S. Visual tracking with online multiple instance learning[C]. Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009: 983-990.
- [4] Williams O, Blake A, Cipolla R. Sparse Bayesian learning for efficient visual tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1292-1304.
- [5] Zhang Yueqiang, Su Ang, Liu Haibo, *et al.* Three dimensional rigid object pose tracking and optimization based on multilevel line representation and M-estimation[J]. Acta Optica Sinica, 2015, 35(1): 0115003.
张跃强, 苏 昂, 刘海波, 等. 基于多级直线表述和 M-估计的三维目标位姿跟踪优化算法[J]. 光学学报. 2015, 35(1): 0115003.
- [6] Zhao Y J, Zhang B, Zhang X L. Mean shift blob tracking with target model adaptive update[C]. Proceedings of Chinese Control Conference, Nanjing, China, 2014: 4831-4835.
- [7] Ross D, Lim J, Lin R, *et al.* Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1): 125-141.
- [8] Liu Wei, Zhao Wenjie, Li Cheng. An online learning visual tracking method based on compressive sensing[J]. Acta Optica Sinica, 2015, 35(9): 0915001.
刘 威, 赵文杰, 李 成. 一种基于压缩感知的在线学习跟踪算法[J]. 光学学报, 2015, 35(9): 0915001.
- [9] Mei X, Ling H B. Robust visual tracking using L1 minimization[C]. International Conference of Computer Vision, 2009: 1436-1433.
- [10] Mei X, Ling H B, Wu Y, *et al.* Minimum error bounded efficient L1 tracker with occlusion detection[J]. IEEE Transactions on image processing, 2013, 22(7): 2661-2675.
- [11] Zhang K H, Zhang L, Yang M H. Real-time compressive tracking[C]. European Conference on Computer Vision, 2012: 864-877.
- [12] Bao C G, Wu Y. Real time robust tracker using accelerate proximal gradient approach[C]. Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012: 1830-1837.
- [13] Yang X, Wang M, Zhang L M, *et al.* An efficient tracking system by orthogonalized templates[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5): 3187-3197.
- [14] Henriques J F, Caseiro R, Martins P. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [15] Mei X, Ling H B, Jacobs D W. Sparse representation of cast shadows via L1 regularized least squares[C]. International Conference of Computer Vision, 2009: 583-590.
- [16] Kaneko T, Hori O. Feature selection for reliable tracking using template matching[C]. Proceedings of the 2006 IEEE

- Computer Society Conference on Computer Vision and Pattern Recognition, 2003: 796-802.
- [17] Matthews I, Ishikawa T, Baker S. The template update problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 810-815.
- [18] Kim S J, Koh K, Lustig M, *et al.* A interior-point method for large-scale L1 regularized least squares[J]. IEEE Journal on Selected Topics in Signal Processing, 2007, 1(4): 606-617.
- [19] Li Shuangshuang, Zhao Gaopeng, Wang Jianyu. Distractor-aware object tracking based on multi-feature and scale-adaptive[J]. Acta Optica Sinica, 2017, 37(5): 0515005.
李双双, 赵高鹏, 王建宇. 基于特征融合和尺度自适应的干扰感知目标跟踪[J]. 光学学报, 2017, 37(5): 0515005.
- [20] Villegas M, Paredes R, Juan A, *et al.* Face verification on color images using local features[C]. Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008: 1-6.
- [21] Mehta R, Zhu. R. Blue or red? Exploring the effect of color on cognitive task performances[J]. Science, 2009, 323(5918): 1226-1229.
- [22] Wu Y, Lim J, Yang M H. Online object tracking a benchmark[C]. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.