

基于多个相关滤波器的行人跟踪尺度算法

张云洲, 郑瑞, 暴吉宁, 朱尚栋
(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘 要: 外观、尺度变化是行人跟踪的难点,解决行人多尺度跟踪问题是增强算法实用性的关键因素. 在 KCF(kernel correlation filter)算法的基础上,本文采用多个相关滤波器(如头部、臀部)辅助身体躯干滤波器的匹配跟踪. 通过获得图像帧(除第一帧外)与初始帧的行人头部和臀部之间的距离变化率来缩放搜索面积,解决目标定位不准确和时间浪费的问题;通过调整目标框的尺寸,解决目标模板逐渐包括背景特征或者逐渐被局部特征取代的问题. 在 VOT2016 的 18 个有明显尺度变化的行人场景视频序列上进行了测试,实验结果表明所提算法具有更高的跟踪准确率.

关 键 词: 行人多尺度跟踪;相关滤波器相互辅助;KCF(核相关滤波器);搜索范围;跟踪准确率

中图分类号: TP 391.41 **文献标志码:** A **文章编号:** 1005-3026(2019)09-1228-07

Pedestrian Tracking Scale Algorithm Based on Multiple Correlation Filters

ZHANG Yun-zhou, ZHENG Rui, BAO Ji-ning, ZHU Shang-dong
(School of Information Science & Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: ZHENG Rui, E-mail: zhengrui_neu@163.com)

Abstract: Appearance and scale change are the difficulties of pedestrian tracking. To solve the problem of multi-scale pedestrian tracking is the key factor to enhance the practicability of the algorithm. On the basis of KCF(kernel correlation filter) algorithm, this paper uses multiple correlation filters(such as head and hip)to assist the tracking of the body trunk filter. The distance change rate which is obtained by comparing the distance between the pedestrian's head and hip of every image frame(except the first frame)with the initial frame is used to zoom the search area, so as to avoid inaccurate target location and time waste. By adjusting the size of the target's bounding box, the problem of target's template shift caused by gradual change of the target template including background features or local features is solved. The experiment is conducted on eighteen pedestrian scene video sequences with obvious scale changes in VOT2016 dataset, and the experimental results show that the algorithm proposed has higher tracking accuracy.

Key words: pedestrian multi-scale tracking; correlation filters mutually assisted; KCF(kernel correlation filter); search range; tracking accuracy

行人跟踪技术在计算机视觉领域具有重要的研究意义^[1-4]. 解决跟踪过程中的行人图像尺度变化问题是提高跟踪准确率的重要途径. 研究者尝试提出一系列行人跟踪尺度变化问题的理论与算法,主要思路大致包括两种:先目标定位再确定尺度;直接在多尺度的目标图像上进行定位和尺度确定. 前者主要是先确定物体位置、再确定目标尺度,从而精确确定跟踪模板. 例如, Danelljan 等^[5]提出的鲁棒的视觉跟踪尺度估计(DSST)算法,先用平移滤波器确定位置,再以尺度滤波确定目标尺寸. 大多数算法采用了第二种思路,即确定尺度的同时就确定了目标位置. 例如, Li 等^[6]提出了一种具有多特征跟踪器的尺度自适应算法 SAMF,利用 HOG^[7]特征和颜色特征(CN)^[8]的

特征融合及相关滤波器在多个尺度缩放的图像块上进行目标检测,通过最大响应确定目标和尺度;该算法适用于多种运动场景,但速度较慢. Huang^[9]提出在视觉跟踪候选框中启用比例和长宽比适应性,通过将 IoU 超过阈值以上的候选框缩放到合适尺度,然后再进行相关滤波. Li 等^[10]利用集成边界和中心相关滤波器进行长宽比变化的视觉跟踪算法,通过引入边界来处理相关滤波器的宽高比变化,适用于目标的横纵比发生变化时的目标跟踪. 以上算法可以解决一般的尺度变化问题,但是在目标发生大幅度位移并伴随有尺度变化时,很容易导致跟踪失败.

KCF^[11] (kernel correlation filters) 算法是经典的判别式跟踪算法,基于相关滤波跟踪器 MOSSE^[12] 进行改进. KCF 跟踪器通过使用训练样本中的循环结构实现了高效率,还增强了具有内核技巧的常规相关滤波器并支持多通道特性,但跟踪目标的尺度问题仍未解决.

本文基于 KCF 算法提出了多个相关滤波器相互辅助的方法,通过构建多个(头部、臀部、躯干)相关滤波器来辅助跟踪视频中的行人,能够有效地处理目标发生大幅度位移状态下的尺度变化. 当行人完全出现在视频中时,借助跟踪目标的局部信息(头部、臀部)和跟踪目标本身(躯干)的几何信息来辅助目标的多尺度跟踪. 当身体躯干

滤波器跟踪发生偏移时,判断头部和臀部滤波器跟踪的准确性;在准确性较高时,可以由头部或臀部滤波器辅助判定跟踪失误,并利用之前得到的置信度高的跟踪模板生成和替换当前出现跟踪失误的模板. 同时,本文设计了生成模板的感兴趣区域(ROI) 的搜索区域,根据生成模板的匹配效果,当物体发生较大位移时就扩大搜索区域,反之仍然在小范围内进行模板生成. 由此,可以对伴随大幅度位移尺度变化的行人目标进行较好的跟踪. 在 VOT2016^[13] 数据集上进行的测试结果表明,本文算法在伴随大位移的尺度变化时具有更高的跟踪准确率.

1 多个相关滤波器辅助的多尺度跟踪

本文算法利用相关滤波器,并结合跟踪结果的置信度来进行跟踪. 若当前跟踪结果的准确度很高,可以把当前跟踪模板存起来,当后续跟踪结果不够准确(如发生跟踪漂移)时用于重新生成模板,以便于目标能够被继续跟踪到. 如果当前跟踪结果准确度很低,可以用之前跟踪效果好的模板重新生成模板并继续进行跟踪. 同时,使用多个相关滤波器辅助方法进行行人的多尺度跟踪,整体过程如图 1 所示.

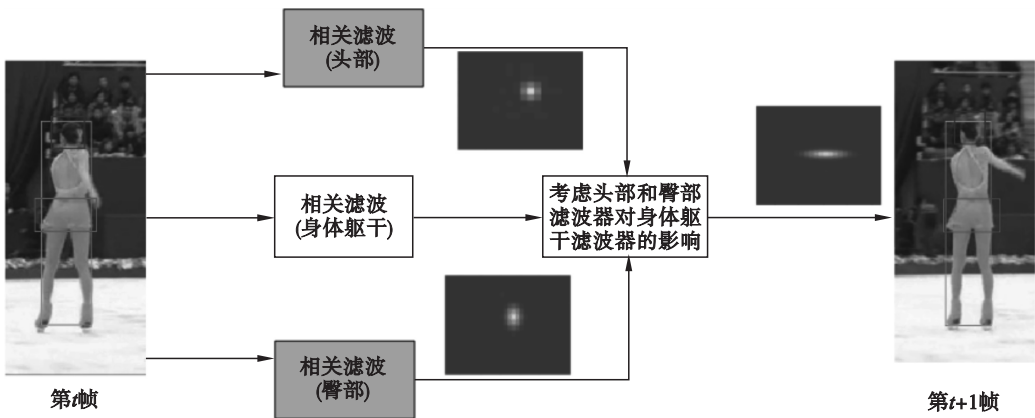


图 1 本文算法示意图
Fig. 1 Schematic diagram of the algorithm

1.1 适应大幅度移位的目标搜索方法

1.1.1 Padding 的范围

在跟踪过程中,为了使跟踪器能够适应在不同场景下的目标跟踪,目标周围一定区域内的信息需要包含进来并进行训练,相应的图像区域称为块(patch). 取模板训练所用的图像面积与跟踪目标的真实面积比值为 $padding$.

当行人目标在视频中发生尺度变化时,使用

相同大小的 $padding$ 进行模板训练是不合理的,当被跟踪的目标尺度变大时且超过训练图像的面积时,目标不会被完全覆盖而造成跟踪失败.

当目标大小占图像大小的比值较小时,目标尺度较小;反之较大. 针对这种情况,本文提出根据目标大小调整 $padding$ 大小的方法,并用于每一视频帧的模板训练的感兴趣区域选择,避免被跟踪目标没有被包含在感兴趣区域内的情况.

在初始帧时,计算人体头部和臀部之间的距离(当人是正面时,头部依旧可以清晰地看到,臀部位置可通过人体的黄金分割比确定,因为要求的变量是人身体的两个位置之间距离的变化,所以较为粗略的位置即可满足条件) d_0 ,在后续帧中计算头部和臀部之间的距离 d_i ,得到尺度变化比例:

$$k = d_0 / d_i. \tag{1}$$

式中, k 是被跟踪目标发生尺度变化的比例系数,则第 i 帧图像中目标搜索区域的面积为被跟踪目标尺寸与 $padding_i$ 的乘积,其中, $padding_i$ 的计算式为

$$padding_i = \frac{padding_0}{k}. \tag{2}$$

式中, $padding_0$ 为初值.

1.1.2 搜索方式

为了更好地跟踪并适应简单的尺度变化,除了预测的目标区域外,KCF 算法还会以目标中心为中心同时匹配一个面积较大和较小的感兴趣区域(ROI),其边长为原目标区域的边长乘以不同比例因子(预测区域边长乘以大于 1 和小于 1 的比例因子分别代表较大和较小面积的边长)进行缩放得到.对 3 个框分别提取特征,判断是否含有目标以及目标的位置,取峰值最大的感兴趣区域作为最终的跟踪结果.

由于运动目标的位移具有方向性,一般地,目标在自身周围较近区域的水平位置上会左右移动,竖直方向会上、下、斜上、斜下移动,然后向周围较远的区域移动.对于这种情况,本文设计了一种新颖的搜索方案.首先在一级搜索区域中将目标进行滑框移动选定 8 个感兴趣区域并训练模板;如果身体模板得到的结果置信度高于 0.75,且同时有头部或者臀部的置信度高于 0.75,则进行下一帧的匹配跟踪;如果效果不佳,再取较远范围内 4 个感兴趣区域进行训练匹配.

如图 2 所示,目标框表示目标本身,一级搜索区域是目标运动大概率发生的区域,即首先提取感兴趣区域的范围;二级搜索区域是运动目标次高概率发生的区域,即后来提取 4 个感兴趣区域的范围.本文采用滑窗的方式在特定范围内确定感兴趣区域,对于被跟踪的目标分别在一级和二级搜索区域内进行感兴趣区域的确定及目标搜索匹配.如果一级搜索区域中的感兴趣区域已经能和模板进行匹配并得到一个较高的峰值,说明运动目标已经找到,则不再进行其他范围的搜索匹配;如果不能匹配,则继续在二级搜索区域搜索,以此类推.该搜索策略在保证跟踪效果的同时也

不会过度降低跟踪速度.

当确定候选框后,以目标框的中心分别向左和向右平移目标框宽度的 20%,提取特征并求出模板后,将三个滤波器预测的位置通过式(4)和式(5)求出置信度后,取置信度最高的滤波器作为最终模板.

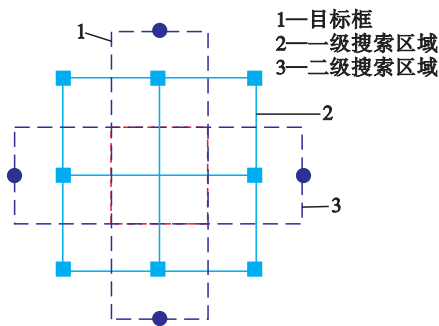


图 2 候选框选定范围
Fig. 2 Candidate frame selection range

1.2 多个相关滤波器相互辅助

基于相关滤波(CF)的跟踪算法在第一帧指定跟踪目标,然后对样本进行训练并得到跟踪目标的相关滤波器,用于后续帧的目标匹配跟踪.当行人完全出现在视频序列中时,本文使用多个相关滤波器相互辅助进行目标跟踪.CF 跟踪算法对行人整体训练一个相关滤波器,本文增加了人体显著部位(头部和臀部)的相关滤波器以协助行人整体的相关滤波器进行行人跟踪.为方便叙述,记 CF 跟踪算法对整个行人的相关滤波器为 B 、行人头部的相关滤波器为 H 、行人臀部的相关滤波器为 A .

本文算法在视频序列的第一帧时需要给出三个部位的真实区域,通过训练得到三个相关滤波器 B, H, A .LMCF^[14]算法利用 Struck^[15]的优质模板保留替换思想和 SVM^[16]分类器进行结果判断,并提出了多峰前向检测策略进行相似背景判断.本文为了让跟踪算法充分利用目标信息,设置 3 个模板存储器 B_{15}, H_{15}, A_{15} 用来存储跟踪效果较好的相关滤波器.存储器的大小设置为 15,因为 15 个目标的先验知识已经足够后续模板更新使用,而且内存和时间的消耗不多.在跟踪过程中,可以将得到的模板匹配结果高于阈值的模板存储起来,若在后续跟踪过程中出现不好结果时可以重新生成模板从而实现较好的目标跟踪.

阈值可根据跟踪算法预测结果的置信度来设定.置信度的计算采用欧氏距离,对于二维平面上的两点 $A(a_1, b_1)$ 和 $B(a_2, b_2)$,其欧式距离的计算公式为

$$d_{AB} = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}. \quad (3)$$

在跟踪过程中,对跟踪模板预测出的目标区域提取特征得到 n 维向量 $\mathbf{T}(v_{i_1}, v_{i_2}, \dots, v_{i_n})$,而跟踪目标在第一帧提取的特征向量为 $\mathbf{T}_r(v_{1_1}, v_{1_2}, \dots, v_{1_n})$,则其欧式距离为 d_{T_r} ,其计算公式为

$$d_{T_r} = \sqrt{\sum_{i=1}^n (v_{1_i} - v_{i_i})^2}. \quad (4)$$

利用欧式距离,可计算得到置信度 conf :

$$\text{conf} = \left| 1 - \frac{d_{T_r}}{\sum_{i=1}^n v_{1_i}} \right|. \quad (5)$$

利用置信度,确定需要的模板 \mathbf{w} ,如式(6)所示:

$$\mathbf{w} = (\mathbf{F} \text{diag}(\hat{x}^* \odot \hat{x}) \mathbf{F}^H + \lambda \mathbf{I})^{-1} \mathbf{X}^H \mathbf{y}. \quad (6)$$

其中: x 是训练图像数据; \mathbf{I} 是单位矩阵; λ 是正则化参数; \hat{x} 是 x 的傅里叶变换结果; \hat{x}^* 是 \hat{x} 的共轭; \mathbf{X} 是所有 x 的向量表达形式; \mathbf{X}^H 是 \mathbf{X} 的厄米特转置; y 是真值 (ground truth); \mathbf{F} 是一个连续矩阵,用于矩阵的对角化转换.由式(7)可以得到对应的输出分布,取输出响应最大值的位置为跟踪目标位置:

$$f(\mathbf{X}) = \mathbf{w}^T \mathbf{X}. \quad (7)$$

本文设定了三个相关滤波器跟踪结果的取舍关系,如表 1 所示.

表 1 相关滤波器跟踪结果判定与模板更新
Table 1 Result determination and template updating of correlation filter tracking

| $\text{conf} \geq$ | B | H | A | 结果判定 | 需生成的模板 |
|--------------------|-----|-----|-----|------|-----------|
| 0.9 | ✓ | | | 跟踪准确 | H, A |
| | ✓ | ✓ | ✓ | 跟踪准确 | 无 |
| | ✓ | ✓ | | 跟踪准确 | A |
| 0.75 | ✓ | | ✓ | 跟踪准确 | H |
| | ✓ | | | 跟踪失败 | B, H, A |

当视频序列中的相关滤波器 B 的置信度高于 0.9 时,认为该跟踪结果就是准确的,并将模板保存到对应位置;与此同时,如果头部和臀部跟踪结果的置信度都很高,且 H, A 滤波器跟踪结果的中心点位置在 B 跟踪结果范围内,则认为 B, H 跟踪是准确的,如果它们的置信度高于阈值,则将这些滤波器模板按置信度值的大小顺序保存到对应的存储器中.如果此时头部和臀部有检测跟踪效果的置信度达不到最低的 0.75,则相应的相关滤波器模板利用之前保存起来的对应位置模板进行当前帧该位置模板的生成,生成规则(以整体滤

波器为例)如式(8)所示,然后继续进行后续跟踪.如果以上的情况都没有发生,则直接进入下一帧检测跟踪.

$$B_i = \text{ratio} \times B_0 + (1 - \text{ratio}) \times \max(B_{15}). \quad (8)$$

若身体整体跟踪置信度在 0.75 ~ 0.9 之间,并与其他滤波器模板置信度在同一区域,相互之间符合几何位置关系(垂直位置大致符合黄金分割比,水平位置目标中心位置间不超过目标宽度的一半),则跟踪正确;否则跟踪失败,需要进行模板生成.若身体整体跟踪置信度不在 0.75 ~ 0.9 之间,但是头部和臀部的置信度值都较为理想,且两者跟踪结果的几何距离小于初始位置的距离 d_0 ,则利用两者之间的位置确定整个身体的位置,否则利用模板存储器进行模板生成.

算法流程如下:

Algorithm: multi-correlation filter tracking
Input:
 n : The total number of image sequences to test
 bef : Confidence of the prediction area being true.
 H_d, H_h, H_a : Image template for body, head and ass.
 d : The distance between head and hip
 gt : Ground truth of the initial target
 pl : Predicted target location
 k : Distance variation coefficient
 padding : Fill between search area and target area
Output:
 T : Template storage container
 Response : target location

for i to n :
 if $i = 1$ then
 $gt \rightarrow H_d, H_h, H_a, d_0$;
 else
 $H_d, H_h, H_a \rightarrow pl, d_i$;
 $d_0, d_i \rightarrow k, \text{padding}$; \longrightarrow Eq. (1)
 $pl \rightarrow H_d, H_h, H_a, \text{bef}$; \longrightarrow Eq. (4)
 if $\text{bef}_b > 0.75$ then
 $H_d \rightarrow T$,
 else
 if $\text{bef}_h > 0.75 \ \&\& \ \text{bef}_a > 0.75$ then
 $H_h, H_a \rightarrow H_h, H_a$
 else
 $T \rightarrow H_h, H_a \longrightarrow$ Eq. (8)
 end if
 end if
 $H_d, H_h, H_a \rightarrow \text{response}$
 end if
end for

2 实验结果及分析

本文提出的算法主要针对完全出现在视野中的行人,在视频序列的第一帧进行行人头部和臀部的标注. VOT 数据集在视频序列的第一帧就进行初始化,适合作为基准测试. 因此,本文选用 VOT 数据集. 设置正则化参数为 10^{-4} ,相关滤波器保存的阈值为 0.9,模板更新的比例值(即 ratio)为 0.7, *padding* 的值为 2,则每一帧的感兴趣区域的面积为初始标注面积的 2 倍的尺度变化比例因子. 本文的硬件平台采用主频 2.10 GHz 的 CPU 和 2 块 8 GB 的 GPU,编程语言采用 C++.

在 VOT 基准测试时,在视频序列的第一帧初始化跟踪器,以精确度、鲁棒性和预期平均重叠度(EAO)来衡量算法的性能. 将本文算法与 KCF 算法、DSST 算法的跟踪效果进行对比分析. 对于 VOT 视频序列的选取,考虑到需要行人完整地出现在视野范围内,在 VOT2016 数据集中选取了 basketball, bmx, iceskater1 等 18 个视频序列. 表 2 列出了在所选取的视频序列上的 VOT2016 实验结果对比,可以看出本文算法的 EAO 为 0.085,超过了 KCF 算法和 DSST 算法. 因此,本文基于 KCF 算法的改进具有较好的效果,且针对多尺度场景效果较好. 在精度上,本文算法远远优于 DSST 和 KCF. 在鲁棒性上,本文算法优于其他算法,有较好的表现,本文较为复杂的候选框选定方法使得本算法具有较好的鲁棒性. 在跟踪速度上,本文算法由于使用 3 个相关滤波器,时间花费较大,所以帧率较另外两个算法低.

表 2 基于 VOT2016 的跟踪算法性能对比
Table 2 Comparison of tracking algorithm performances based on VOT2016

| 性能指标 | DSST | 本算法 | KCF |
|-------------------|-------|--------------|---------------|
| EAO | 0.081 | 0.085 | 0.078 |
| Accuracy – Rank | 2.05 | 1.72 | 2.27 |
| Robustness – Rank | 1.94 | 1.27 | 1.67 |
| Speed/fps | 23.79 | 17.37 | 132.88 |

图 3 展示了 VOT2016 部分视频序列中的本文算法三个跟踪器之间相互辅助的跟踪结果图.

2.1 尺度变化

在 iceskater1 视频序列中,如图 3a 所示(对于视频序列,按照从左到右、从上到下计数为第一幅图、第二幅图等),行人目标发生了下蹲、大幅度的跳跃以及目标整体变大变小等尺度变化等行

为,可能导致跟踪效果变差. 本文算法优化了目标候选框的搜索范围,在偏差发生时能够及时找回跟踪目标并持续跟踪. 在图 3a 的第二幅图中,身体跟踪器的跟踪结果与真实值之间的 IoU 交集面积较小,而此时另外的两个相关滤波器都能准确跟踪;在第三幅图中,3 个跟踪器都能跟踪得很准确,表明跟踪器是有相互辅助作用的,但前提是 3 个跟踪器之间的几何距离符合跟踪目标(人体)本身各个部位之间的几何关系(使得头部滤波器的目标框在躯干滤波器的顶部,臀部滤波器的目标框在躯干滤波器目标位置的黄金分割点位置,且 3 个目标框的目标中心位置在水平距离上不超过目标宽度的一半). 在图 3a 的第五、六幅图中,当行人发生尺度变化,由远及近经过相机时,躯干滤波器使用二阶段搜索方式,行人都可以很好地被跟踪到,表明本文算法较好地解决了行人跟踪过程中的多尺度问题.

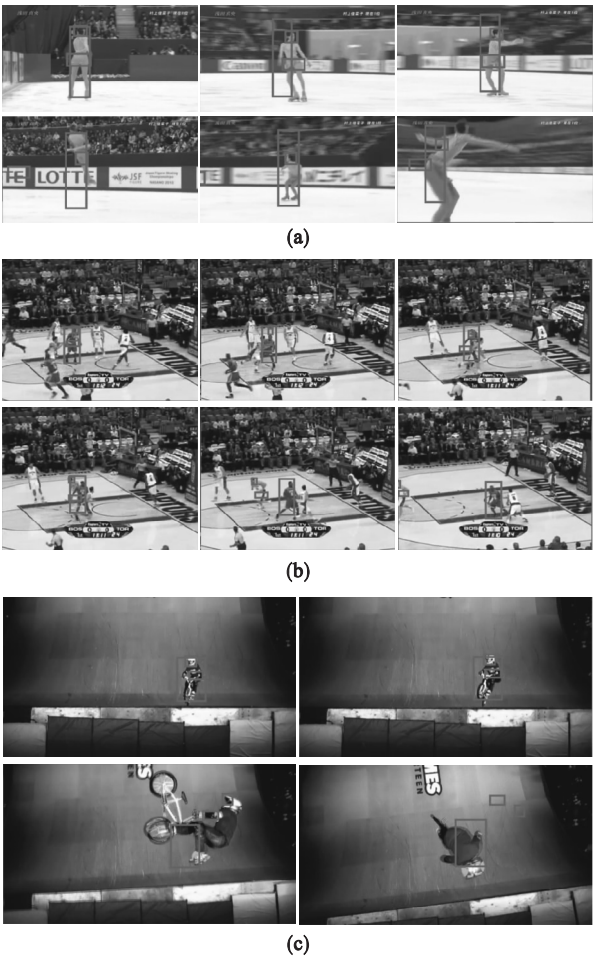


图 3 本文算法对 VOT 部分视频序列跟踪效果
Fig. 3 Tracking effect of the algorithm on the VOT part of the video sequence

(a)—iceskater1 跟踪; (b)—basketball 跟踪;
(c)—bmx 跟踪.

在目标发生较大位移时,本文与改进前的

KCF 算法进行了跟踪效果对比实验. 图 4a 是 KCF 算法的实验效果, 当目标发生跳跃之后, 目标仍然被跟踪到了, 但仅跟踪到了下半身. 当目标发生较大位移时, 如图 4b 所示, 本文算法能迅速跟踪到目标. 在图 4b 中, 本文算法跟踪到的目标比例比图 4a 大, 且定位效果比图 4a 精确.

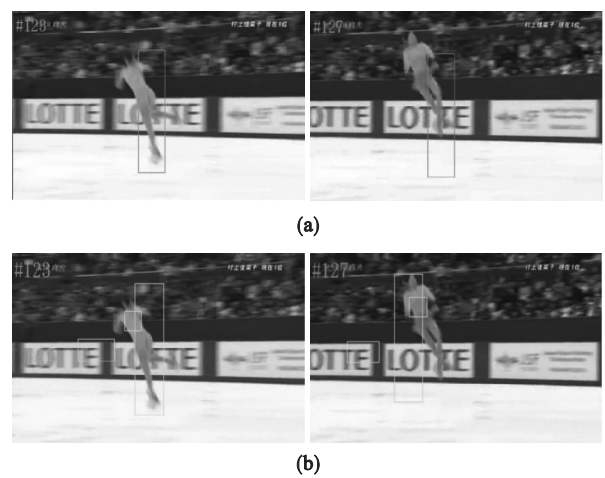


图 4 在跟踪目标发生较大位移时的跟踪效果对比
Fig. 4 Comparison of tracking effects when tracking targets have large displacements
(a)—KCF 算法的跟踪效果; (b)—本文算法的跟踪效果.

2.2 背景干扰

在 basketball 视频序列中, 如图 3b 所示, 从左到右的第一幅图是视频初始时的标注情况, 但是在后几帧中(第二幅图), 没有任何干扰的情况下头部的相关滤波跟踪器已经跟踪失败了, 且在第三幅图中一个非跟踪目标的头部吸引走了头部跟踪器. 对于非跟踪目标的头部与跟踪目标的头部肤色相差很大、而头部跟踪器存在误差的情况, 此时会影响跟踪器的跟踪效果, 但并不影响最终的跟踪效果, 因为本文在跟踪器之间设定了几何关系判定条件, 各个相关滤波器是独立工作和更新的. 图 3c 的 bmx 视频序列中的最后一幅图中, 头部和臀部滤波器并没有实现准确跟踪目标, 但并不影响身体躯干滤波准确跟踪和更新, 验证了本文的思路.

上述实验结果表明, 在背景较为相似时, 本文算法可以避免不必要的背景干扰, 而在无背景干扰时可以利用身体各部位之间的相互辅助作用进行更有效的跟踪.

3 结 语

本文通过构建多个相关滤波器并且彼此间相互辅助进行行人跟踪, 解决行人跟踪的姿态变化

和尺度问题. 通过改变搜索方式, 更精确地选取候选框, 在提高跟踪准确度的同时尽可能不降低跟踪速度, 在准确度与速度之间取得平衡. 另外, 依据行人尺度变化改变感兴趣区域的大小, 可以更准确地确定行人的位置, 也避免了引入不必要的特征和行人目标没有被包含进感兴趣区域的问题. 实验结果表明, 本文和现有的较好的主流跟踪算法相比, 当行人全部出现在视野中, 如果行人发生尺度变化、姿态变化或者较大形变时, 本文算法都表现出了较好的跟踪精度, 但是, 相关滤波器通过一种相互辅助的手段来判定跟踪是否准确, 判定的准则较为粗糙, 导致跟踪的鲁棒性较差, 因此在今后的研究中需进一步提出更完善的滤波器之间的辅助判定条件.

参考文献:

[1] Yang H, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: a review [J]. *Neurocomputing*, 2011, 74 (18): 3823–3831.

[2] 孟煜, 张斌. 检测区域自适应调整的 TLD 多目标跟踪算法 [J]. *东北大学学报 (自然科学版)*, 2017, 38 (2): 214–218.

(Meng Yu, Zhang Bin. A multiple-object tracking algorithm using TLD-based adaptive adjustment of detection areas [J]. *Journal of Northeastern University (Natural Science)*, 2017, 38 (2): 214–218.)

[3] 韩光星, 张轩, 李崇荣. 基于多模型检测的视频监控行人跟踪算法 [J]. *华中科技大学学报 (自然科学版)*, 2016, 44 (11): 43–47.

(Han Guang-xing, Zhang Xuan, Li Chong-rong. Pedestrian tracking algorithm by video surveillance based on ensemble detection [J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2016, 44 (11): 43–47.)

[4] 鲍华, 赵宇宙, 张陈斌, 等. 基于自适应分块表观模型的视觉目标跟踪 [J]. *控制与决策*, 2016, 31 (3): 448–452.

(Bao Hua, Zhao Yu-zhou, Zhang Chen-bin, et al. Visual tracking based on adaptive patches appearance model [J]. *Control and Decision*, 2016, 31 (3): 448–452.)

[5] Danelljan M, Häger G, Khan F S, et al. Accurate scale estimation for robust visual tracking [C]// *British Machine Vision Conference*. Nottingham, 2014: 65. 1–65. 11.

[6] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration [C]// *European Conference on Computer Vision*. Cham, 2014: 254–265.

[7] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// *International Conference on Computer Vision & Pattern Recognition*. San Diegom CA: IEEE Computer Society, 2005: 886–893.

[8] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking [C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus OH: IEEE Computer Society, 2014: 1090–1097.

[9] Huang D F. Enable scale and aspect ratio adaptability in visual tracking with detection proposals [C]// *British Machine Vision Conference*. Swansea, 2015: 185. 1–185. 12.