

采用时空上下文的抗遮挡实时目标跟踪

陈法领^{1,2,3,4,5}, 丁庆海^{1,6}, 罗海波^{1,2,4,5}, 惠斌^{1,2,4,5}, 常铮^{1,2,4,5}, 刘云鹏^{1,2,4,5}

1. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;
2. 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169;
3. 中国科学院大学, 北京 100049;
4. 中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;
5. 辽宁省图像处理与视觉计算重点实验室, 辽宁 沈阳 110016;
6. 北京航天恒星科技有限公司, 北京 100086)

摘要: 针对目标跟踪算法在光照变化、背景干扰、目标形变及遮挡时出现的跟踪稳定性下降甚至失败的问题, 提出了一种采用时空上下文的抗遮挡实时目标跟踪算法。首先, 在时空上下文模型框架下采用自适应降维的颜色特征构建目标外观模型, 提高算法在复杂场景中对目标的辨别能力; 然后, 联合置信图响应的峰值和峰值旁瓣比对目标跟踪的状态进行评估; 接着, 利用目标模板之间相关系数的变化进一步判断目标是否被严重遮挡; 最后, 当目标跟踪出现波动时, 降低目标模型更新速度, 并通过 Kalman 滤波修正目标位置, 当目标被严重遮挡时, 则根据 Kalman 滤波预测目标位置, 同时停止更新目标模型, 在脱离遮挡后重新捕获目标并进行跟踪。选取了 36 组具有多种挑战因素的彩色视频序列测试算法的跟踪性能, 并与其他表现优异的目标跟踪算法进行了对比分析。实验结果表明, 所提算法具有较强的抗遮挡能力, 并且在光照变化、背景干扰和目标形变等不利因素影响下仍具有较好的跟踪鲁棒性, 同时能够满足目标跟踪的实时性要求。

关键词: 目标跟踪; 时空上下文; 目标跟踪置信度; 遮挡判别; Kalman 滤波

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20200105

Anti-occlusion real time target tracking algorithm employing spatio-temporal context

Chen Faling^{1,2,3,4,5}, Ding Qinghai^{1,6}, Luo Haibo^{1,2,4,5}, Hui Bin^{1,2,4,5}, Chang Zheng^{1,2,4,5}, Liu Yunpeng^{1,2,4,5}

1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China;
4. Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China;
5. The Key Lab of Image Understanding and Computer Vision, Shenyang 110016, China;
6. Space Star Technology Co., LTD, Beijing 100086, China)

Abstract: An anti-occlusion real time target tracking algorithm employing spatio-temporal context was proposed to solve the problems of tracking instability or even failure, which were caused by illumination variation, background clutters, target deformation or occlusion. Firstly, in the framework of spatio-temporal context model, the adaptive dimensionality reduced color features were adopted to describe the target to promote the distinguish ability in complex scene. Secondly, the peak and the peak-to-sidelobe ratio of confidence map

收稿日期: 2020-04-17; 修订日期: 2020-05-18

基金项目: 国家重点研发计划 (2018YFC080192); 中国科学院国防科技创新重点基金 (Y8K4160401)

response were combined to evaluate the target tracking status. Then, occlusion was discriminated by the correlation coefficient between target templates. Finally, when the target tracking status fluctuated, the update speed of target model was reduced, and the target coordinates were corrected by the Kalman filter. When the target was occluded seriously, the target coordinates was predicted according to the Kalman filter, and the target model was stopped to update for recapturing and tracking the target again after occlusion released. 36 color sequences with multiple challenging attributes were selected to evaluate the performance of the proposed algorithm, and it was compared with other excellent target tracking algorithms. The experimental results demonstrated that this algorithm has strong anti-occlusion ability, and improved the robustness of target tracking effectively under the influence of disturbance factors such as illumination variation, background clutters and target deformation. Meanwhile, it met the real time requirement of target tracking.

Key words: target tracking; spatio-temporal context; target tracking confidence; occlusion discrimination; Kalman filter

0 引言

目标跟踪是计算机视觉领域的重要研究课题,它在智能监控、自动驾驶、人机交互、军事侦查和机器人导航等领域应用广泛。目标跟踪技术虽已取得很大进步,但在复杂环境中特别是在非受控条件下,目标跟踪容易受到光照变化、背景干扰、目标形变或遮挡等因素影响,导致实现稳定可靠的目标跟踪仍面临许多困难^[1]。

基于核相关滤波 (Kernel Correlation Filter, KCF)^[2]的目标跟踪算法由于具有较快的运行速度和较高的跟踪精度,近年来受到广泛关注。参考文献 [3] 利用多个分块核相关滤波器解决跟踪中的目标形变、尺度变化和遮挡问题;房胜男等^[4]将多种特征融合以提高目标跟踪性能。另外,随着深度学习在图像分类和目标识别等领域的成功应用,目前也出现了许多基于深度学习的目标跟踪算法: Song 等^[5]在目标跟踪中首先引入了生成对抗网络解决正负样本的比例失衡问题; Bhat 等^[6]利用卷积神经网络提取的深层特征和浅层特征的互补性提高目标跟踪的鲁棒性和准确性; Gao 等^[7]在孪生网络基础上,通过增加分类网络对目标和背景进行更精细分类,同时增加候选缩放网络适应目标尺度变化。但由于深度学习算法的计算开销较大,目前在嵌入式系统中还很难实现实时的目标跟踪。

此外,上述目标跟踪算法均没有考虑目标与其周围邻域的空间上下文关系, Zhang 等^[8]根据生物视觉模型的中心注意机制提出了时空上下文 (Spatio-Temporal Context, STC) 目标跟踪算法,将目标跟踪转化为估计目标出现在候选区域内各点的概率问

题。它充分利用了目标周围邻域的空间上下文信息,具有较好的跟踪性能,同时由于它的主要计算过程是在频域,跟踪速度可达每秒数百帧。在此基础上,参考文献 [9] 采用多个时空上下文模型和贝叶斯卡尔曼滤波解决目标被长时遮挡的模型更新问题;翟静梅等^[10]利用特征点进行目标的辅助定位并实时调整时空上下文模型的更新策略;郭文等^[11]通过对历史帧的目标信息进行低秩表达来解决跟踪失败后的目标重定位问题;王向军等^[12]将跟踪过程分为连续跟踪和目标检测两个环节解决长时跟踪问题。

尽管上述改进方法都取得了一定的效果,但还有很大的提升空间,文中主要针对 STC 算法在复杂场景中对目标的辨别能力差以及在目标跟踪过程缺乏对跟踪状态的评估问题,提出了一种采用时空上下文的抗遮挡实时目标跟踪算法。首先,在提取目标及其上下文区域的颜色名 (Color Name, CN) 特征后,对其进行自适应降维并以此构建目标的外观模型;然后,利用置信图响应的峰值和峰值旁瓣比对目标跟踪状态进行评估;接着,根据跟踪过程中目标模板的相关系数的变化进一步对目标遮挡情况进行判别;最后,在目标跟踪置信度较高时进行正常跟踪,反之,则降低目标模型更新速度并通过 Kalman 滤波修正目标位置,并在目标被严重遮挡时停止更新目标模型同时利用 Kalman 滤波预测目标位置。选取了 36 组具有多种挑战因素的彩色视频序列测试文中算法的跟踪性能,并与其他性能优异的算法进行了对比分析。

1 时空上下文模型

时空上下文模型将目标跟踪转化为估计目标在

候选区域内各点出现的概率问题,其最大概率值对应的坐标点即为目标位置^[8]。目标位置似然的置信图定义如公式(1)所示:

$$m(x) = P(x|o) \quad (1)$$

式中: x 为目标中心位置; o 为包含目标的候选区域。当前帧中,假设目标中心坐标为 x^* ,那么它的上下文特征集合可以定义为 $X^c = \{c(z) = (I(z), z) | z \in \Omega_c(x^*)\}$,其中 $I(z)$ 为位置 z 处的图像特征, $\Omega_c(x^*)$ 是围绕目标中心的局部上下文区域。根据贝叶斯估计理论可将公式(1)转换为:

$$m(x) = P(x|o) = \sum_{c(z) \in X^c} P(x, c(z)|o) = \sum_{c(z) \in X^c} P(x|c(z), o)P(c(z)|o) \quad (2)$$

式中: $P(x|c(z), o)$ 为空间上下文模型,表示了目标与其周围上下文区域的空间关系; $P(c(z)|o)$ 为上下文先验模型,表示了上下文区域各点的先验概率,它的具体形式为:

$$P(c(z)|o) = I(z)\omega_\sigma(z - x^*) \quad (3)$$

式中: $I(z)$ 为位置 z 处的图像特征; $\omega_\sigma(z - x^*) = ae^{-\frac{\|z-x^*\|^2}{\sigma^2}}$ 为加权函数,其中 a 是归一化常数, σ 为尺度参数。

置信图模型 $m(x)$ 的具体定义如公式(4),它描述了目标在其周围上下文区域中出现的概率

$$m(x) = P(x|o) = be^{-\|x-x^*\|^\beta} \quad (4)$$

式中: b 为归一化常数; α 为尺度参数; β 为置信图模型的形状控制参数。

空间上下文模型 $P(x|c(z), o)$ 表示的是目标与其上下文区域的空间关系,为避免场景中相似目标对跟踪的影响,定义了一个径向非对称的矢量函数,如公式(5)所示:

$$P(x|c(z), o) = h^{sc}(x - z) \quad (5)$$

综合上述公式(2)~公式(5),可得:

$$m(x) = be^{-\|x-x^*\|^\beta} = \sum_{c(z) \in X^c} h^{sc}(x - z)I(z)\omega_\sigma(z - x^*) = h^{sc}(x) \otimes (I(x)\omega_\sigma(x - x^*)) \quad (6)$$

式中: \otimes 表示卷积运算。对公式(6)两边分别进行傅里叶变换可得:

$$\mathcal{F}(be^{-\|x-x^*\|^\beta}) = \mathcal{F}(h^{sc}(x)) \odot \mathcal{F}(I(x)\omega_\sigma(x - x^*)) \quad (7)$$

这里 \mathcal{F} 表示傅里叶变换, \odot 代表点乘运算,由此可

以获得空间上下文模型的表示形式为:

$$h^{sc}(x) = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(be^{-\|x-x^*\|^\beta})}{\mathcal{F}(I(x)\omega_\sigma(x - x^*))} \right) \quad (8)$$

式中: \mathcal{F}^{-1} 表示傅里叶逆变换。

在得到目标第 t 帧的空间上下文模型 $h_t^{sc}(x)$ 后,目标在第 $t+1$ 帧的时空上下文模型为:

$$H_{t+1}^{sc}(x) = (1 - \rho)H_t^{sc}(x) + \rho h_t^{sc}(x) \quad (9)$$

式中: $H_{t+1}^{sc}(x)$ 和 $H_t^{sc}(x)$ 分别为目标在第 t 和 $t+1$ 帧的时空上下文模型; ρ 为更新速率,其值越大表示时空上下文模型的更新速度越快。

目标在第 $t+1$ 帧的位置 x_{t+1}^* 则可以通过寻找它在第 $t+1$ 帧置信图中的最大值得到,即:

$$x_{t+1}^* = \arg \max_{x \in \Omega_c(x_t^*)} m_{t+1}(x) \quad (10)$$

其中,

$$m_{t+1}(x) = \mathcal{F}^{-1}(\mathcal{F}(H_{t+1}^{sc}(x)) \odot \mathcal{F}(I_{t+1}(x)\omega_{\sigma_t}(x - x_t^*))) \quad (11)$$

通过上述公式(9)和公式(11)的不断更新过程即可完成目标跟踪任务。

2 文中算法

2.1 基于颜色特征的目标定位

颜色特征是从全局角度对目标特征进行了描述,它在目标发生形状、姿态及旋转等变化时不敏感,被广泛应用于目标跟踪。相对于 RGB、HSV 和 LAB 特征,利用 CN 特征可以获得更好的跟踪性能^[13]。由于目标跟踪算法的运行速度与其采用的特征维数密切相关,为提高算法的运行速度,保证目标跟踪的实时性,通过自适应降维技术将原始的 11 维 CN 特征降维,在保留重要信息的同时减少 CN 特征维数。

自适应降维技术的核心是主成分分析,它通过公式(13)所示的最小化代价函数 η_t^{total} 寻找满足 $B_p^T B_p = I$ 的 $D_1 \times D_2$ 维投影矩阵 B_p (D_1 和 D_2 分别是降维前后的特征维数)进行降维。

$$\eta_t^{\text{total}} = \gamma_t \eta_t^{\text{data}} + \sum_{j=1}^{t-1} \gamma_j \eta_j^{\text{smooth}} \quad (12)$$

式中: $\eta_t^{\text{data}} = \frac{1}{MN} \sum_{m,n} \|I_{t,c} - B_p B_p^T I_{t,c}\|^2$ 是取决于第 t 帧的重构误差; $\eta_j^{\text{smooth}} = \sum_{k=1}^{D_2} \lambda_j^{(k)} \|b_j^{(k)} - B_p B_p^T b_j^{(k)}\|^2$ 是和第 j 帧相关的平滑项,它对应的平滑系数为 γ_j ; $b_j^{(k)}$ 为投影矩

阵 B_p 中的基本向量, 它的权重由 $\lambda_j^{(k)}$ 决定, $I_{t,c}$ 为降维前的第 t 帧 CN 特征矩阵; M 和 N 为该矩阵的大小。

在求解得到投影矩阵后, 即可获得降维后的第 t 帧 CN 特征 $\tilde{I}_{t,c} = B_p^T I_{t,c} = \{\tilde{I}_{t,c1}, \tilde{I}_{t,c2}\}$, 它由两个相互独立的二维特征 $\tilde{I}_{t,c1}$ 和 $\tilde{I}_{t,c2}$ 组成。当采用降维后的 CN 特征描述目标外观模型时可以将公式 (11) 转换为如下形式:

$$x_{t+1}^* = \arg \max_{x \in \Omega_c(x_t^*)} m_{t+1}(x) = \arg \max_{x \in \Omega_c(x_t^*)} \mathcal{F}^{-1}(\mathcal{F}(H_{t+1,c1}^{stc}(x)) \odot \mathcal{F}(\tilde{I}_{t+1,c1}(x) \omega_{\sigma_t}(x - x_t^*)) + \mathcal{F}(H_{t+1,c2}^{stc}(x)) \odot \mathcal{F}(\tilde{I}_{t+1,c2}(x) \omega_{\sigma_t}(x - x_t^*))) \quad (13)$$

式中: $\tilde{I}_{t+1,c1}(x)$ 和 $\tilde{I}_{t+1,c2}(x)$ 为第 $t+1$ 帧时自适应降维后的两个相互独立的二维 CN 特征; $H_{t+1,c1}^{stc}(x)$ 和 $H_{t+1,c2}^{stc}(x)$ 则是它们在第 $t+1$ 帧所对应的时空上下文模型, 分别由第 t 帧的时空上下文模型和空间上下文模型加权得到, 如公式 (14) 所示:

$$\begin{aligned} H_{t+1,c1}^{stc}(x) &= (1 - \rho)H_{t,c1}^{stc}(x) + \rho h_{t,c1}^{sc}(x) \\ H_{t+1,c2}^{stc}(x) &= (1 - \rho)H_{t,c2}^{stc}(x) + \rho h_{t,c2}^{sc}(x) \end{aligned} \quad (14)$$

通过上述公式 (13) 和公式 (14) 即完成了基于自适应降维 CN 特征的目标定位。

2.2 目标跟踪置信度评估

在时空上下文模型中, 通常将置信图响应的峰值

坐标作为当前帧的目标位置, 当目标没受到干扰时得到的置信图响应是不存在波动的单峰, 跟踪效果良好。但在跟踪过程中会不可避免地受到光照变化、遮挡或者目标形变等干扰, 图 1 为 Coke 序列在不同条件下的置信图响应对比, 该序列存在较剧烈的光照变化和遮挡干扰。

从图 1 可以看出, 在稳定跟踪 (图 1(a) 所示)、光照变化 (图 1(b) 所示) 和目标被遮挡 (图 1(c) 所示) 3 种情况下的置信图响应存在较大差异: 不仅响应峰值在光照变化和被遮挡时出现明显下降, 而且在光照变化时置信图响应的旁瓣较高, 主峰不够明显, 同时在目标被遮挡时, 置信图表现出“多峰”形式, 此时容易导致错误跟踪。置信图响应的峰值旁瓣比 (Peak-to-Sidelobe Rate, PSR) 可以用来评估目标跟踪的状态以及受干扰的严重程度, 当前帧中置信图响应在峰值位置 x 处的 PSR 定义为:

$$PSR_i(x) = \frac{\max_i(x) - \mu_i(x)}{\sigma_i(x)} \quad (15)$$

式中: $\max_i(x)$ 为该置信图响应的峰值; $\mu_i(x)$ 和 $\sigma_i(x)$ 分别为它的均值和标准差。

如图 1 所示, 当目标未受外界干扰或者未被遮挡时, 它的置信图响应是波动较小的尖锐单峰, 对应的

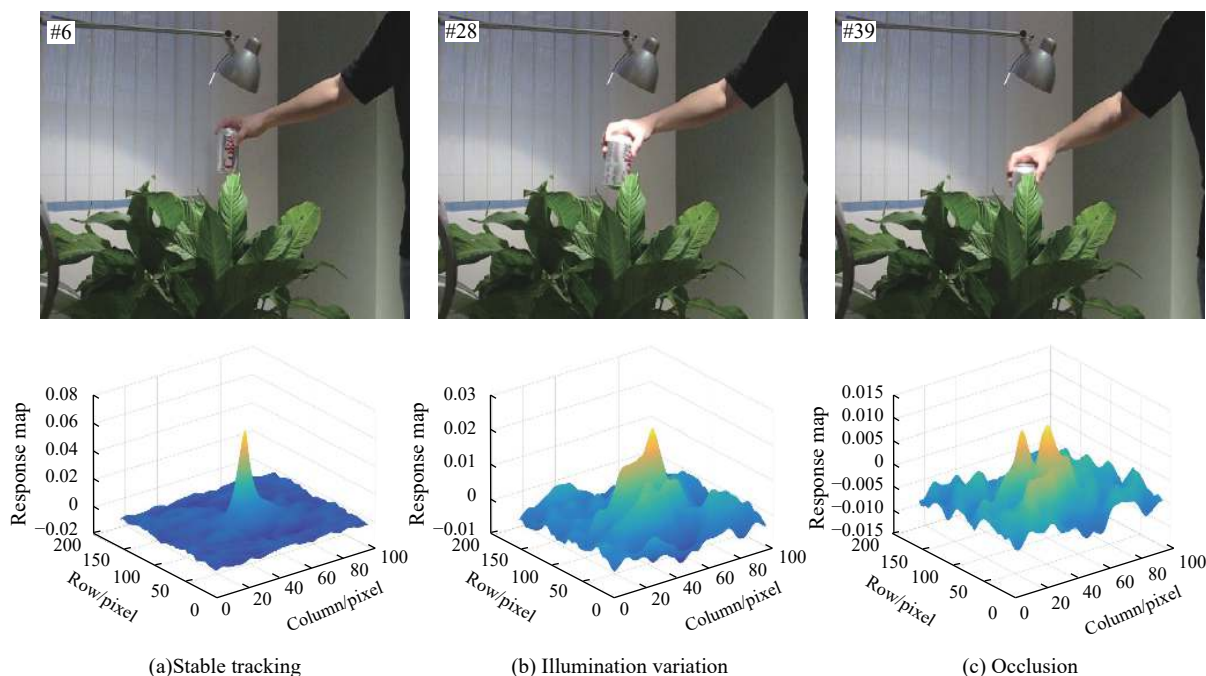


图 1 在稳定跟踪和存在干扰情况下的置信图响应对比

Fig.1 Comparison of responses under stable tracking and interference conditions

峰值和 PSR 值较大;反之,置信图响应则会出现较大波动甚至出现“多峰”,对应的峰值和 PSR 值也较小。综上分析,选择联合置信图响应的峰值 $\max_i(x)$ 和 $PSR_i(x)$ 对当前帧的目标跟踪置信度进行评估。考虑到目标的自身变化也会对置信图响应产生影响,为了对目标跟踪置信度的评估更加可靠,还参考了历史帧的响应峰值 $\max_i(x)$ 和峰值旁瓣比 $PSR_i(x)$,只有在当前帧的 $\max_i(x)$ 值和 $PSR_i(x)$ 值均以一定比例超过了历史帧的 $\max_i(x)$ 和 $PSR_i(x)$ 均值时,才认为当前帧目标跟踪置信度较高,即:

$$\begin{cases} \max_i(x) > \lambda_1 \frac{1}{n} \sum_{i=t-n}^{t-1} \max_i(x), t \geq n \\ PSR_i(x) > \lambda_2 \frac{1}{n} \sum_{i=t-n}^{t-1} PSR_i(x), t \geq n \end{cases} \quad (16)$$

式中: $\max_i(x)$ 和 $PSR_i(x)$ 、 $\max_i(x)$ 和 $PSR_i(x)$ 分别为当前帧和历史帧响应的峰值和 PSR 值; n 为参考的历史帧数; λ_1 和 λ_2 为调节系数。当目标跟踪的置信度较高时,正常跟踪,否则将公式 (14) 的目标模型更新速度降为 $\rho/2$ 。

2.3 遮挡判别

在目标跟踪中,目标被场景中的其他物体遮挡是非常普遍的,如何有效地处理目标被遮挡问题,尤其是严重遮挡甚至完全遮挡对提升目标跟踪性能至关重要。在上一节根据置信图响应的峰值和 PSR 值完成了对目标跟踪置信度的评估后,还需进一步对目标是否被严重遮挡进行判断。

当目标被其他物体严重遮挡时,从“目标”区域获得的“目标模板”将会发生很大变化,这种变化可以通过目标被遮挡前后的目标模板的相关系数的变化来体现。通过公式 (17) 计算两帧目标模板的去均值归一化相关系数,进而判断目标是否被严重遮挡。

$$Corr(I_t, I_{t-1}) = \frac{\sum_{m,n} (I_t - \bar{I}_t)(I_{t-1} - \bar{I}_{t-1})}{\sqrt{\sum_{m,n} (I_t - \bar{I}_t)^2 \sum_{m,n} (I_{t-1} - \bar{I}_{t-1})^2}} \quad (17)$$

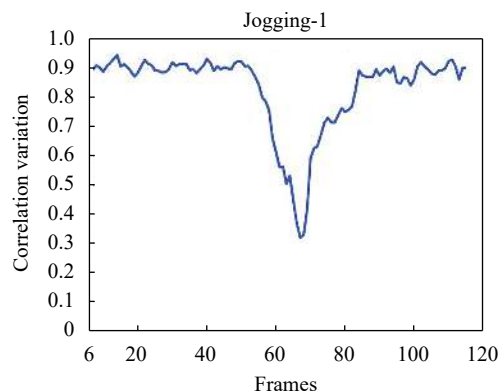
式中: I_t 和 I_{t-1} 分别表示第 t 帧和第 $t-1$ 帧的目标模板; \bar{I}_t 和 \bar{I}_{t-1} 表示这两个目标模板的均值。

图 2(a)、(b) 和 (c) 给出了 Jogging1 序列中目标被遮挡过程:目标在第 63 帧被轻微遮挡,随后目标被遮挡面积逐渐增加并在第 73 帧被完全遮挡,最后在第

83 帧脱离遮挡;图 2(d) 所示为该序列当前帧目标模板与第 $t-6$ 帧目标模板的相关系数变化曲线。可以看出,当目标未被遮挡时,第 $t-6$ 帧与当前帧的目标模板变化不大,它们的相关系数在小范围内波动;当目标的被遮挡面积逐渐增大时,此时目标模板与未被遮挡时目标模板的差异增加,它们的相关系数呈下降趋势;当目标被完全遮挡时,此时“目标模板”相对于未被遮挡时的目标模板变化最大,它们的相关系数达到了极小值。



(a) Occlusion slight (b) Occlusion serious (c) Occlusion release



(d) Correlation variation

图 2 目标被遮挡过程及其相关系数变化

Fig.2 Procedure of target occlusion and its correlation variation

综上,利用当前帧与历史帧的目标模板的相关系数的变化,可以判断目标是否被严重遮挡。为提高遮挡判别的可靠性,保存了 6 帧历史帧的目标模板信息,并依次计算第 $t-5$ 帧、第 $t-4$ 帧、第 $t-3$ 帧、第 $t-2$ 帧、第 $t-1$ 帧及当前帧与第 $t-6$ 帧目标模板的相关系数,如果这 6 个相关系数呈下降趋势且最大值小于设定阈值 $Corr_{TH1}$,则表示目标被严重遮挡,这时将停止更新目标模型,以便在脱离遮挡后重新捕获目标并进行跟踪,同时预测目标位置。

2.4 Kalman 滤波

在目标跟踪中,由于目标在相邻帧的间隔通常为毫秒量级,在如此短的时间内可以认为目标的状态保持稳定,因此可以通过 Kalman 滤波对目标在短时间内的状态进行有效估计^[14],用于在目标跟踪出现波动

时修正目标位置,并在目标被严重遮挡时进行目标位置预测。

假设目标在 t 至 $t + \Delta t$ 这段时间内在 X 方向和 Y 方向上分别服从速度为 v_x 和 v_y 的匀速运动,则目标在第 t 帧的状态方程和观测方程可以分别由公式 (18) 和公式 (19) 表示:

$$X_t = AX_{t-1} + W_{t-1} \quad (18)$$

$$Z_t = HX_t + V_t \quad (19)$$

式中: $X_t = [x(t), y(t), v_x(t), v_y(t)]^T$ 和 $Z_t = [x(t), y(t)]^T$ 分别表示目标在第 t 帧的状态向量和观测向量; $x(t)$ 和 $y(t)$ 是目标在第 t 帧的中心坐标; $v_x(t)$ 和 $v_y(t)$ 是

目标在第 t 帧的速度分量; $A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ 和 $H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ 分别是状态矩阵和观测矩阵; Δt 为时间间隔; W_{t-1} 和 V_t 分别是均值为零的状态误差和观测误差。

2.5 算法流程

算法的具体工作流程如下:

Step1: 初始化。利用起始帧给出的目标信息完成时空上下文模型及相关参数的初始化。

Step2: 位置估计。依次读取序列中的图像数据,提取目标及其上下文区域 CN 特征并进行自适应特征降维;根据公式 (13) 计算当前帧目标位置置信图,完成目标位置初始估计。

Step3: 置信度评估。根据公式 (16) 计算当前帧目标跟踪置信度,若置信度较高,则正常跟踪,执行 step7,否则执行 step4。

Step4: 严重遮挡判别。根据公式 (17) 分别计算当前帧与历史帧目标模板的相关系数,如果这些相关系数值呈下降趋势,并且最大值小于设定阈值 $Corr_{TH1}$,说明目标被严重遮挡,执行 step6;否则,说明仅出现跟踪波动但未被严重遮挡,执行 step5。

Step5: 位置修正。通过 Kalman 滤波修正初始估计的目标位置,并将目标模型的更新速度降为 $\rho/2$,执行 step7。

Step6: 位置预测。停止更新目标模型,根据 Kalman 滤波预测目标位置,并以预测的位置为中心重新搜索目标,并计算当前帧的“目标模板”与被遮挡前的目标模板之间的相关系数,如果大于设定阈值

$Corr_{TH2}$,说明目标脱离遮挡,执行 step2;否则继续执行 step6。

Step7: 更新目标位置和模型,并判断跟踪是否完成,未完成则继续执行 step2。

3 实验结果与分析

3.1 实验平台与测试条件

实现算法的计算机硬件配置为: CPU 为 Intel(R) Core(TM) i5-2500, 主频 3.30 GHz, 内存 8 GB; 软件开发平台为 Matlab R2016b。选取 OTB 数据集^[15] 中 36 组具有多种挑战因素的彩色序列测试算法跟踪性能,实验采用的是一次性通过评估 (One Pass Evaluation, OPE) 方法和参考文献 [15] 中的中心位置误差 (Center Location Error, CLE)、距离精度 (Distance Precision, DP)、重叠精度 (Overlap Precision, OP) 和每秒跟踪帧数 (Frames Per Second, FPS) 4 种性能评价指标。

算法的具体参数设置为: 公式 (4) 中 $\alpha = 2.25$, $\beta = 1$, 公式 (14) 中 $\rho = 0.075$, 公式 (16) 中用于置信度评估的历史帧数 n 设置为 6, 调节系数分别设置为 $\lambda_1 = 0.6$, $\lambda_2 = 0.8$, 用于遮挡判别和脱离遮挡的相关系数阈值分别设置为 $Corr_{TH1} = 0.82$, $Corr_{TH2} = 0.90$ 。

3.2 跟踪性能分析

3.2.1 定量分析

为验证算法的跟踪性能,首先在 3 种不同配置模式下对算法性能进行了测试,其中 Proposed (Only CN₂) 是仅使用自适应降维的 CN 特征,既没有置信度评估,也没有遮挡判别和 Kalman 滤波策略; Proposed (CN₁₀) 则表示使用未降维的原始 CN 特征,其余设置与文中算法一致,表 1 为在 OTB 数据集中 3 种配置模式下的目标跟踪性能对比。

由表 1 可知,文中算法性能较 Proposed (Only CN₂)

表 1 算法在不同配置模式下的跟踪性能评估

Tab.1 Performance evaluation of proposed algorithm in different configuration modes

Tracker	CLE	DP	OP	FPS
Proposed(only CN ₂)	17.6	63.3%	59.4%	89.2
Proposed	12.3	71.2%	64.6%	84.8
Proposed(CN ₁₀)	10.2	73.5%	67.4%	46.6

有较大提高,并且在 DP 和 OP 两种指标上与 Proposed (CN₁₀) 算法相差不多,但平均速度约是它的 2 倍。由此可以看出,文中算法在提升目标跟踪性能的同时,实时性也得到很大提高,降低了目标跟踪算法对硬件资源的要求,有利于在资源受限的嵌入式系统中实现实时的目标跟踪。

为进一步定量评估算法的跟踪性能,在 36 组彩色视频序列中对比了 32 种算法,这些算法包括了文中算法、STC 算法^[8]、CN 算法^[13],以及在 OTB 数据集中表现出色的 29 种算法,如 Struck 算法、SCM 算法、TLD 算法、ASLA 算法和 CT 算法等。

图 3 为这些算法在 24 组具有遮挡属性的彩色视频序列中的 DP 和 OP 指标对比结果,为简化显示,只列出了每种指标排名前 10 的算法。在图 3(a)中,在

CLE 阈值为 20 像素时,文中算法的 DP 指标为 0.708,相对于排名第 2 的 SCM 算法提高了 18.8%,相对于原 STC 算法提高了 19.6%;在图 3(b)中,在 OS 阈值为 0.5 时,文中算法的 OP 指标为 0.622,相对于排名第 2 的 SCM 算法提高了 9.9%,相对于原 STC 算法提高了 18.3%。可以看出,相对于原始的 STC 算法和其他对比算法,文中算法具有较强的抗遮挡能力。

图 4 为这些算法在所有 36 组彩色序列中的 DP 和 OP 指标对比结果,同样的,也只列出了每种指标排名前 10 的算法。在图 4 中,文中算法的 DP 和 OP 指标相对于第 2 名的原 STC 算法分别提高了 16.5% 和 9.7%。

表 2 为文中算法与其他 5 种跟踪性能较好的目标跟踪算法在测试视频序列中的平均跟踪速度对比。

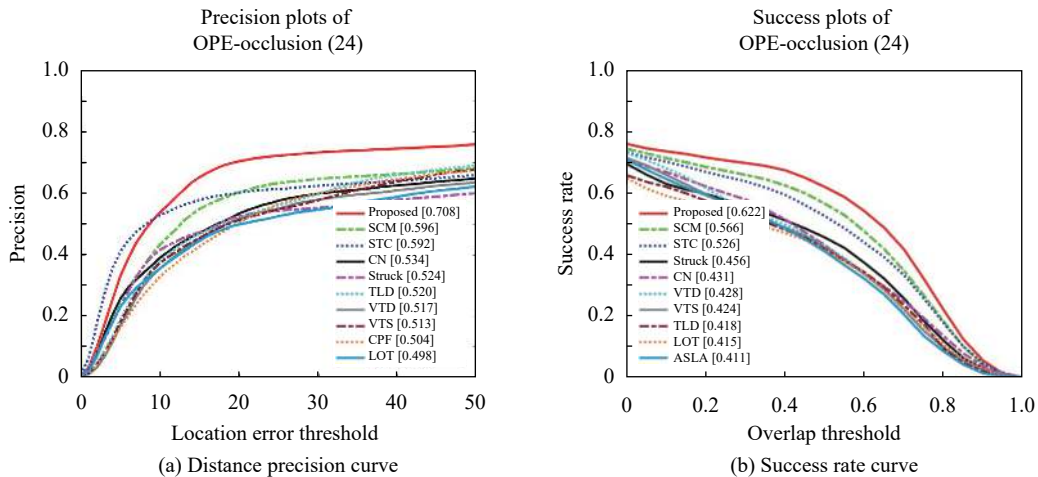


图 3 不同跟踪算法在遮挡属性下的距离精度和成功率

Fig.3 Distance precision and the success rate of different algorithms under occlusion attribute

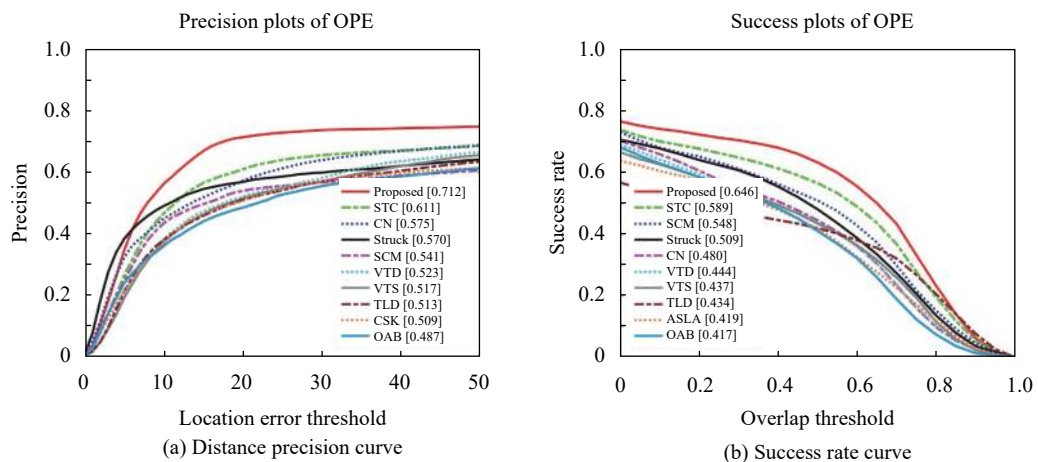


图 4 不同跟踪算法的综合距离精度和成功率

Fig.4 Distance precision and the success rate of different tracking algorithms

表 2 6 种不同算法的平均跟踪速度

Tab.2 Average tracking speed of six different algorithms

Tracker	Proposed	STC	CN	Struck	SCM	TLD
FPS	84.8	270.2	98.6	8.6	0.62	19.8

由表 2 可知,虽然 STC 算法和 CN 算法的平均跟踪速度都高于文中算法,但是其整体性能均与文中算法存在较大差距,其余 3 种算法则都不满足目标跟踪的实时性要求。

通过上述定量分析可以看出文中算法具有较强的抗遮挡能力,同时其整体跟踪性能也优于其他对比算法,并具有较高的实时性。

3.2.2 定性分析

为了更直观地对文中算法的跟踪性能进行评估,分别在图 5~图 10 中给出了文中算法与其他 5 种对比算法(具体包括 STC 算法、CN 算法、Struck 算法、SCM 算法和 TLD 算法)在测试序列 Coke、Jogging1、Human4、Shaking、Bolt 和 David3 中的实际跟踪结果。

图 5 为 Coke 序列跟踪结果,该序列的主要挑战包括光照变化和遮挡。目标在第 36 帧时逐渐被遮挡,CN 算法、SCM 算法和 TLD 算法的跟踪框在第 40 帧时包含了较多干扰信息,并且在第 42 帧时 CN 算法和 TLD 算法都跟踪到了错误目标;在后续的第

44 帧 Struck 算法跟踪框也包含了较多的干扰信息,而文中算法和 STC 算法由于充分利用了空间上下文信息,可以较好地跟踪目标,表现出一定的抗遮挡能力。

图 6 为 Jogging1 序列跟踪结果,该序列中目标先被轻微遮挡,接着目标的被遮挡面积逐渐增加直至被完全遮挡。图 6(d) 中,在目标脱离遮挡后,只有文中算法和 TLD 算法可再次跟踪目标,其他算法都出现了跟踪失败的现象。这是由于文中算法通过置信图响应的峰值和 PSR 值对目标跟踪的状态进行了评估,并利用目标模板的相关系数变化进一步分析了目标被遮挡程度,在目标被严重遮挡时通过 Kalman 滤波预测了目标位置,如图 6(c) 所示,同时停止目标模型更新,使得在脱离遮挡后可以再次捕获目标;而 TLD 算法则是通过在线检测机制实现了目标丢失后的重定位。

图 7 为 Human4 序列跟踪结果,该序列的主要挑战包括光照变化、目标形变及遮挡,特别是目标在运动过程中出现了多次严重遮挡。如图 7(a),在第 90 帧目标还未被遮挡时 TLD 算法已丢失目标,目标在第 126 帧时被轻微遮挡,这时所有算法都还能有效跟踪目标,后续随着目标被遮挡面积逐渐加大,其他算法跟踪框都停留在了遮挡物上,如图 7(c) 所示,但文中算法仍能跟踪目标,并在目标被多次严重遮挡后还能

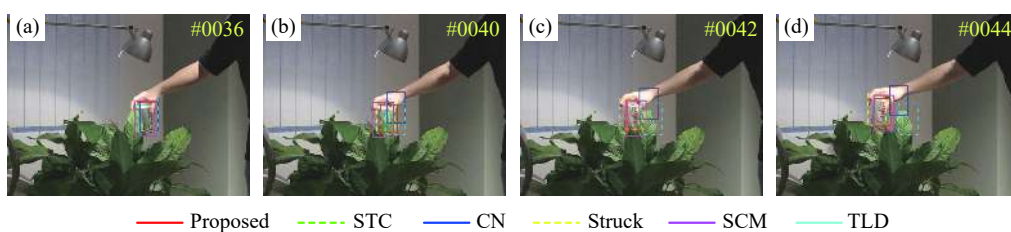


图 5 文中算法与其他 5 种对比算法在 Coke 序列中的跟踪结果

Fig.5 Tracking results of the proposed algorithm and other five comparison algorithms on Coke sequence

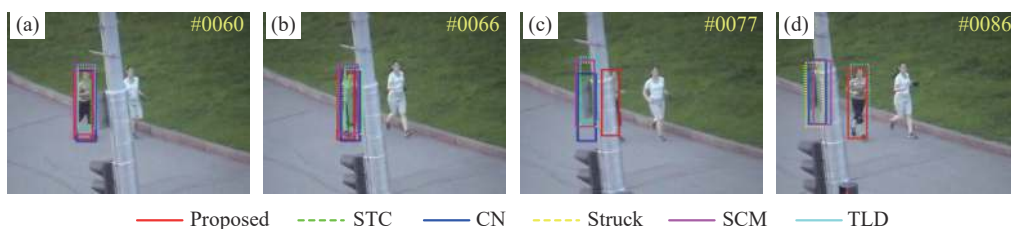


图 6 文中算法与其他 5 种对比算法在 Jogging1 序列中的跟踪结果

Fig.6 Tracking results of the proposed algorithm and other five comparison algorithms on Jogging1 sequence

稳定跟踪目标,进一步验证了文中算法的抗遮挡能力。

图 8 为 Shaking 序列跟踪结果,该序列中目标与周围背景的分度度较低,比较容易受到背景干扰,同时还存在光照变化。Struck 算法在第 30 帧时出现跟踪漂移,CN 算法和 TLD 算法的跟踪框在第 78 帧包含了较多背景信息,这种现象在第 106 帧进一步加剧,在第 266 帧时 SCM 算法也出现了跟踪漂移,文中算法和 STC 算法由于充分利用了目标周围的空间上下文信息,较好地适应了背景干扰,在整个过程中始

终稳定地跟踪目标。

图 9 为 Bolt 序列跟踪结果,该序列的主要挑战是目标在快速运动时的外观变化。如图 9(a),Struck 算法在第 5 帧就已经跟踪在错误目标上,STC 算法和 SCM 算法在第 30 帧时丢失了目标,在第 120 帧 TLD 算法也出现了跟踪失败,但文中算法和 CN 算法由于都采用了 CN 特征,对目标形变具有较好的适应性,能够一直稳定地跟踪目标。

图 10 为 David3 序列跟踪结果,该序列的主要挑

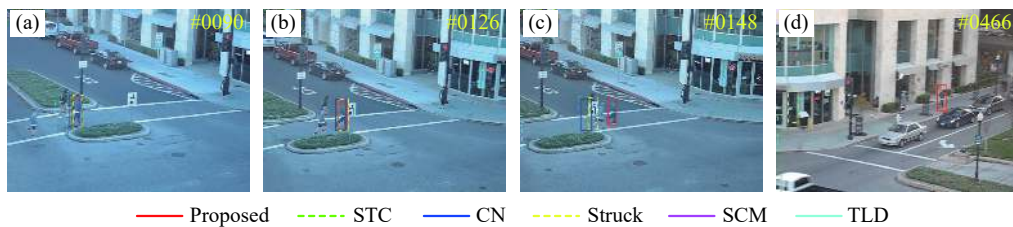


图 7 文中算法与其他 5 种对比算法在 Human4 序列中的跟踪结果

Fig.7 Tracking results of the proposed algorithm and other five comparison algorithms on Human4 sequence

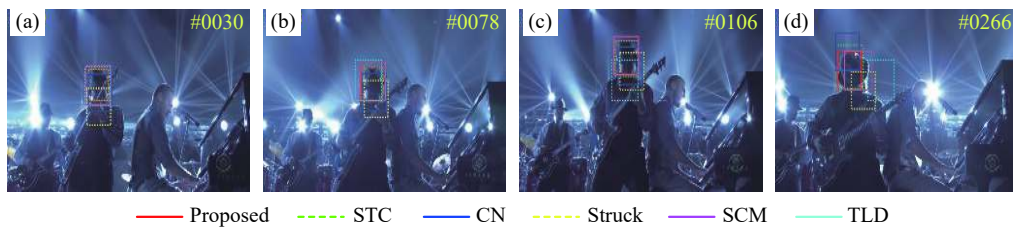


图 8 文中算法与其他 5 种对比算法在 Shaking 序列中的跟踪结果

Fig.8 Tracking results of the proposed algorithm and other five comparison algorithms on Shaking sequence

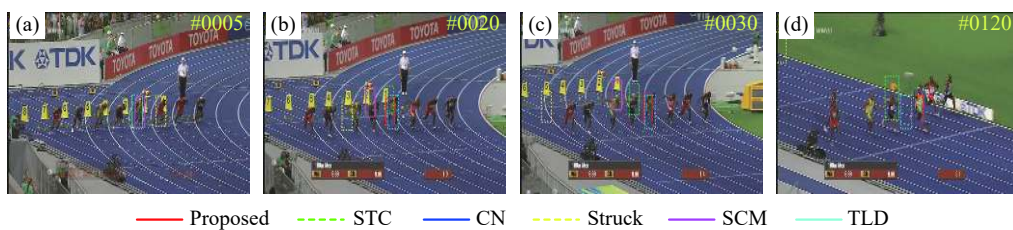


图 9 文中算法与其他 5 种对比算法在 Bolt 序列中的跟踪结果

Fig.9 Tracking results of the proposed algorithm and other five comparison algorithms on Bolt sequence

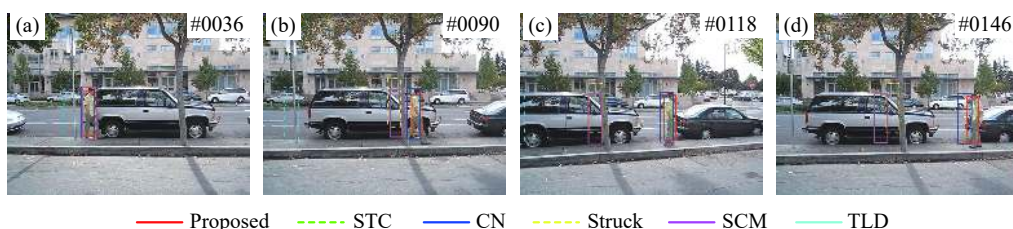


图 10 文中算法与其他 5 种对比算法在 David3 序列中的跟踪结果

Fig.10 Tracking results of the proposed algorithm and other five comparison algorithms on David3 sequence

战是复杂背景、遮挡和目标非平面旋转。如图 10(a), TLD 算法的跟踪框在第 36 帧停留在遮挡物体上, 图 10(b) 中 Struck 算法和 SCM 算法在同时出现复杂背景和目標被遮挡时, 也都出现了跟踪失败, 第 118 帧当目标出现非平面旋转时, STC 算法只跟踪到目标的局部, CN 算法的跟踪框包含了较多非目标信息, 在第 146 帧时, 它们的跟踪框都包含了大量背景信息, 已不能有效地跟踪目标, 而文中算法则一直能较好地跟踪目标。

通过将文中算法与其他 5 种对比算法在 Coke、Jogging1、Human4、Shaking、Bolt 和 David3 序列中的实际跟踪结果进行对比, 可以看出文中算法具有较强的抗遮挡能力, 同时能有效克服复杂环境下的光照变化、背景干扰、目标形变以及非平面旋转等多种挑战。

4 结 论

文中提出了一种采用时空上下文的抗遮挡实时目标跟踪算法。通过提取目标及其上下文区域的 CN 特征并进行自适应降维来构建目标的外观模型, 提高了算法在复杂环境下对目标的辨别能力; 利用置信图响应的峰值和 PSR 值对目标跟踪的置信度进行评估, 判断是否出现了外界干扰导致目标跟踪波动; 然后根据跟踪过程中目标模板的相关系数的变化进一步判别目标的被遮挡程度; 并在目标跟踪置信度较高时正常跟踪, 当目标跟踪受到外界干扰导致其置信度降低时, 通过 Kalman 滤波修正目标位置, 并降低目标模型更新速度, 而当目标被严重遮挡时则利用 Kalman 滤波得到的目标状态预测目标位置, 并停止更新目标模型, 以便在目标脱离遮挡后可以再次捕获目标并进行跟踪。通过 OTB 数据集中的 36 组具有多种挑战因素的彩色视频序列测试了文中算法的目标跟踪性能, 并与其他性能优异的算法进行了对比分析。实验结果表明, 文中算法具有较强的抗遮挡能力, 并且在出现光照变化、目标形变及复杂背景等干扰时仍具有较好的跟踪稳定性, 同时能够满足目标跟踪的实时性要求。另外, 实验中也发现文中算法适应目标尺度变化的能力还有待提升, 后续可改进尺度估计策略进一步提高算法的目标跟踪性能。

参考文献:

- [1] Luo Haibo, Xu Lingyun, Hui Bin, et al. Status and prospect of target tracking based on deep learning [J]. *Infrared and Laser Engineering*, 2017, 46(5): 0502002. (in Chinese)
罗海波, 许凌云, 惠斌, 等. 基于深度学习的目标跟踪方法研究现状与展望[J]. 红外与激光工程, 2017, 46(5): 0502002.
- [2] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.
- [3] Chen Hongyu, Luo Haibo, Chen Faling, et al. Scale adaptive part-based tracking method using multiple correlation filters [J]. *Optical Engineering*, 2019, 58(5): 053106.
- [4] Fang Shengnan, Gu Xiaojing, Gu Xingsheng. Infrared target tracking with correlation filter based on adaptive fusion of responses [J]. *Infrared and Laser Engineering*, 2019, 48(6): 0626003. (in Chinese)
房胜男, 谷小婧, 顾幸生. 基于自适应响应融合的相关滤波红外目标跟踪[J]. 红外与激光工程, 2019, 48(6): 0626003.
- [5] Song Yibing, Ma Chao, Wu Xiaohu, et al. VITAL: Visual tracking via adversarial learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8990-8999.
- [6] Bhat G, Johnander J, Danelljan M, et al. Unveiling the power of deep tracking[C]//Proceedings of the European Conference on Computer Vision, 2018: 493-509.
- [7] Gao Peng, Ma Yipeng, Yuan Ruyue, et al. Learning cascaded siamese networks for high performance visual tracking[C]//Proceeding of the IEEE International Conference on Image Processing, 2019: 3078-3082.
- [8] Zhang Kaihua, Zhang Lei, Yang Minghsuan, et al. Fast visual tracking via dense spatio-temporal context learning[C]//Proceedings of the European Conference on Computer Vision, 2014: 127-141.
- [9] Wei X G, Zhang S, Chan S C. A novel visual object tracking algorithm using multiple spatial context models and Bayesian Kalman filter[C]//Proceeding of the IEEE International Symposium on Circuits and Systems, 2015: 1034-1037.
- [10] Zhai Jingmei, Liu Kun. Object tracking and location with spatio-temporal context assisted by key points [J]. *Chinese Journal of Scientific Instruments*, 2017, 38(11): 2839-2848. (in Chinese)
翟敬梅, 刘坤. 特征点辅助的时空上下文目标跟踪与定位[J]. 仪器仪表学报, 2017, 38(11): 2839-2848.
- [11] Guo Wen, You Sisi, Zhang Tianzhu, et al. Object tracking via

- low-rank redetection based multiple feature fusion spatio-temporal context learning [J]. *Journal of Software*, 2018, 29(4): 1017-1028. (in Chinese)
- 郭文, 游思思, 张天柱, 等. 低秩重检测的多特征时空上下文的视觉跟踪[J]. 软件学报, 2018, 29(4): 1017-1028.
- [12] Wang Xiangjun, Guo Zhiyi. Long-term target tracking algorithms in force-constrained platform [J]. *Infrared and Laser Engineering*, 2020, 49(1): 0126003. (in Chinese)
- 王向军, 郭志翼. 算力限制平台下的长时目标跟踪[J]. 红外与激光工程, 2020, 49(1): 0126003.
- [13] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1090-1097.
- [14] Su Yun, Wang Ting, Yao Chen, et al. A target tracking method of UAV based on cooperative target [J]. *Robot*, 2019, 41(4): 425-432. (in Chinese)
- 苏贇, 王挺, 姚辰, 等. 基于合作目标的无人机目标跟踪方法[J]. 机器人, 2019, 41(4): 425-432.
- [15] Wu Yi, Lim Jongwoo, Yang Minghsuan. Object tracking benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834-1848.



第一作者简介: 陈法领 (1984-), 男, 博士生。2006年毕业于郑州大学, 获得工学学士学位; 2009年毕业于中国科学院沈阳自动化研究所, 获得工学硕士学位; 现在中国科学院沈阳自动化研究所攻读博士学位; 研究方向包括实时图像处理和目标跟踪等。Email: chfling@sia.cn



导师简介: 丁庆海 (1969-), 男, 博士, 研究员。1998年毕业于南京理工大学, 获得工学博士学位; 目前为航天恒星科技有限公司研究员; 研究方向包括数据链、自动目标识别和目标跟踪等。Email: l3693689880@163.com



导师简介: 罗海波 (1967-), 男, 博士, 研究员。1990年毕业于哈尔滨工业大学, 获得工学学士学位; 2009年毕业于中国科学院沈阳自动化研究所, 获得工学博士学位; 目前为中国科学院沈阳自动化研究所研究员; 研究方向包括光电成像、实时图像处理、自动目标识别和目标跟踪等。Email: luohb@sia.cn