

低秩重检测的多特征时空上下文的视觉跟踪*

郭文^{1,2}, 游思思^{1,2}, 张天柱³, 徐常胜³



¹(山东工商学院 信息与电子工程学院, 山东 烟台 264009)

²(山东省高校感知技术与控制重点实验室, 山东 烟台 264009)

³(模式识别国家重点实验室(中国科学院 自动化研究所), 北京 100190)

通讯作者: 郭文, E-mail: grewen@126.com

摘要: 时空上下文跟踪算法充分地利用空间上下文中包含的结构信息能够有效地对目标进行跟踪, 实时性优良。但该算法仅利用单一的灰度信息, 使得目标的表现表达缺乏判别性, 而且该方法在由于遮挡等问题造成的跟踪漂移后无法进行初始化。针对时空上下文算法存在的弱点, 提出了一种基于低秩重检测的多特征时空上下文跟踪方法。首先, 利用多特征对时空上下文进行多方面的提取, 构建复合时空上下文信息, 充分利用目标周围的特征信息, 提高目标表现表达的有效性。其次, 利用简单、有效的矩阵分解方式将跟踪到的历史跟踪信息进行低秩表达, 将其引入有效的在线重检测器中来保持跟踪结构的一致稳定性, 解决了跟踪方法在跟踪失败后的重定位问题。在一系列跟踪数据集上的实验结果表明, 该算法与原始算法及当前的主流算法相比有更好的跟踪精度与鲁棒性, 且满足实时性要求。

关键词: 低秩近似矩阵分解; 时空上下文; 多特征融合; 目标跟踪

中图分类号: TP391

中文引用格式: 郭文, 游思思, 张天柱, 徐常胜. 低秩重检测的多特征时空上下文的视觉跟踪. 软件学报, 2018, 29(4): 1017-1028. <http://www.jos.org.cn/1000-9825/5402.htm>

英文引用格式: Guo W, You SS, Zhang TZ, Xu CS. Object tracking via low-rank redetection based multiple feature fusion spatio-temporal context learning. Ruan Jian Xue Bao/Journal of Software, 2018, 29(4): 1017-1028 (in Chinese). <http://www.jos.org.cn/1000-9825/5402.htm>

Object Tracking via Low-Rank Redetection Based Multiple Feature Fusion Spatio-Temporal Context Learning

GUO Wen^{1,2}, YOU Si-Si^{1,2}, ZHANG Tian-Zhu³, XU Chang-Sheng³

¹(School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264009, China)

²(Key Laboratory of Sensing Technology and Control in Universities of Shandong, Yantai 264009, China)

³(National Laboratory of Pattern Recognition (Institute of Automation, The Chinese Academy of Sciences), Beijing 100190, China)

Abstract: The spatio-temporal tracking (STC) algorithm can effectively track object using the structural information contained in the context around the object in real time. However the algorithm only exploits single gray object feature information in order to make the object representation discriminative. Moreover, it fails to initialize when tracking drift due to occlusion problems. Aiming at the existing

* 基金项目: 国家自然科学基金(61572296, 61472227, 61303086, 61328205), 山东省自然科学基金(ZR2015FL020); 模式识别国家重点实验室开放课题(201600024)

Foundation item: National Natural Science Foundation of China (61572296, 61472227, 61303086, 61328205); Natural Science Foundation of Shandong Province, China (ZR2015FL020), Open Project Program of the National Laboratory of Pattern Recognition (201600024)

本文由“多媒体大数据处理与分析”专题特约编辑赵耀教授、李波教授、华先胜研究员、文继荣教授、蒋刚毅教授、常冬霞副教授推荐。

收稿时间: 2017-04-26; 修改时间: 2017-06-26; 采用时间: 2017-10-13; jos 在线出版时间: 2017-12-01

CNKI 网络优先出版: 2017-12-04 06:46:45, <http://kns.cnki.net/kcms/detail/11.2560.TP.20171204.0646.003.html>

weaknesses of the spatio-temporal context algorithm, a novel low-rank redetection based multiple feature fusion STC tracking algorithm is proposed in this paper. Firstly, multiple feature fusion based spatio-temporal context is extracted to construct complicated spatio-temporal context information, which improves the effectiveness of object representation by taking full advantage of the feature information around the object. Then, a simple and effective matrix decomposition method is used to give a low rank expression of the history tracking information, which can be embedded into the online detector. As a result, the uniform structure stability of the tracking algorithm is maintained to solve the relocation problem after the tracking failure. Experimental results on a series of tracking benchmark show the proposed algorithm has a better tracking precision and robustness than several state-of-the-art methods, and it also have a good real-time performance.

Key words: low-rank approximate matrix decomposition; spatio-temporal context (STC); multiple feature fusion; object tracking

目标的视觉跟踪是计算机视觉领域非常有挑战意义的课题之一,被广泛地应用到视觉监控、人机交互、军事导航、赛事直播等多个领域^[1].在过去的十几年里,视觉跟踪方法层出不穷,涌现了大量的优秀算法,但是视觉跟踪目标易于受到光照变化、尺度变化、嘈杂背景、遮挡和运动模糊等因素的影响,如何设计一个在复杂场景下能够对目标快速准确跟踪的方法依然是一个开放性问题^[2].

视觉目标跟踪一般分成产生式跟踪方法和判别式跟踪方法,基于产生式跟踪方法是直接对目标表观提取显著特征进行建模,并且在后续序列中找到与之匹配的最优区域.Comaniciu 等人^[3]提出的均值漂移跟踪算法,曾经在跟踪领域引起很长时间的关注,作为一种基于核函数密度梯度的无参估计方法在目标跟踪领域具有运算简单、跟踪快速等优点,但其简单的颜色直方图的目标表观表达也限制了其性能的提高.Wen^[4]等人则通过对均值漂移算法的收敛性问题进行研究,提高了该算法在目标跟踪的应用基础.Yun 等人^[5]结合了粒子滤波与均值漂移的方法,通过采样粒子迁移和聚类动态建立目标的状态模型和量测模型,用均值漂移对所有的状态特征进行可靠性聚类,这样可以在不采用相似度计算的情况下就能对目标进行有效的跟踪.增量学习跟踪算法^[6]则是学习一个低维的子空间并且利用增量学习方法在线地更新子空间,这种方法能够很好地适应目标的表现变化,但是缺乏空间结构信息,容易造成跟踪漂移.文献[7]提出了一种多任务学习的稀疏跟踪方法,通过多任务学习来刻画粒子之间的相似度,以便提高跟踪效果,Mei 等人^[8]采用 L_1 范数的最优化进行候选目标表观特征的稀疏性构建,以提高跟踪算法对遮挡的效果,但是,这种基于稀疏表达的方式由于优化需要的时间过长,不利于实时跟踪.Wang 等人^[9]采用了一种基于部件库的特征补集的在线表观模型更新算法,来减弱形变等复杂因素对表观模型的影响,以更好地保持目标特征.总体来讲,产生式的跟踪方法能够建立强有力的表观表达并且在小范围内进行搜索时能够得到更为精确的结果,而且可以增量学习,但其对背景的相应干扰较为敏感.

而判别式跟踪方法则更加注重目标与背景的区别性,将跟踪看成一个目标/背景二值化分类问题,去寻找一个最优分类面来区分目标与背景.这种方法起源于 Avidan^[10]提出的支持向量机跟踪方法,通过分类器来区分跟踪的目标与背景区域,建立了跟踪器与分类器之间的联系,但是作为基于检测的跟踪方法的雏形,其简单的更新策略使得跟踪方法对噪声极为敏感.Grabner 等人^[11]提出的在线提升算法(Boosting)将多个弱分类器提升为强特征学习以进行特征选择,从而保持特征始终具有强有力的判别性.多示例学习(multiple instance learning)的跟踪算法^[12]则将跟踪问题定义为一个多示例的学习问题,算法将包(bag)的概念采用了一定的概率表达,包内存在正样本就是正包,提高了跟踪效果,但是这种假设也导致该算法对噪声非常敏感.Kalal 等人^[13]提出的 TLD 跟踪算法将长期的跟踪分解为跟踪、学习和检测(tracking-learning-detection,简称 TLD)3 个模块,在跟踪过程中采用了一种所谓的“P-N”学习的调节机制来约束正负样本,并且抑制在线检测器的预测误差.Hare 等人^[14]提出的结构跟踪方法是一种有代表意义的方法,它从表观模型的设计角度,对输出学习进行结构化描述,并且采用了核操作技术,取得较好的跟踪效果.Ma 等人^[15]提出了基于加权熵的多特征融合的跟踪方法,充分利用了多特征对特征描述子判别性的贡献.总体来说,如果有足够多的训练样本在线更新分类器的话,判别式跟踪方法要比产生式的跟踪方法好很多^[16].本文没有涉及深度学习的跟踪方法^[17],是因为基于深度学习的方法需要大量的数据进行训练,其精度有了很大的提高,但其实时性受到很大的限制.

最近,基于相关滤波的判别式跟踪方法引起了跟踪领域的极大兴趣,相关滤波是通过学习一个模板或者滤波器与测试图像的卷积运算来计算两者的相似度,相似度的峰值会出现在目标的区域,那么跟踪问题就变成找

到一个合适的滤波器使其与目标的输出响应最大的问题,如图 1 所示。



Fig.1 Correlation filter tracker

图 1 相关滤波跟踪

Bolme 等人^[18]提出误差最小平方和滤波器(minimum output sum of squared error,简称 MOSSE),开创了相关滤波在目标跟踪领域的先河.此后,许多优秀的相关滤波跟踪改进算法^[19-22]涌现出来,其中,Henriques 等人^[21]采用循环结构和核技巧操作使得多通道特征集成更加灵活,非线性回归的函数更加鲁棒,而且实时性达到了高速效果.而 Zhang 等人^[22]提出的时空上下文(spatio-temporal context,简称 STC)跟踪是从概率密度分布角度去解释的一种新颖的相关滤波方法,跟踪目标周围关键点的一致性跟踪算法中极为有效的信息,该算法充分利用了目标周围稠密信息,使得跟踪的性能有了较大的提高。

虽然,STC 算法在利用上下文信息上取得了不错的成绩,文献[23]从考虑上下文区域贡献的角度对其进行了权重分配,但是我们认为仍然存在一些重要的不足限制了其性能的进一步提高.首先,STC 采用单一的灰度信息特征,没有合理地利用所有颜色通道里视觉关注的不同;其次,算法简单的模板更新策略以及缺乏重检测机制,使得算法在遮挡等导致的跟踪失败发生时无法重新初始化.针对上述不足,从解决上述问题出发提出一种低秩重检测的多特征时空上下文跟踪算法,本文采用多特征融合的上下文信息提高表达的判别性,利用低秩的滤波重检测模块提高算法的鲁棒性,实验结果表明,本文的算法与原始的 STC 以及一些当前主流算法相比,性能有了很大的提高。

1 时空上下文跟踪简介^[22]

如果在跟踪过程中只考虑目标本身的特征,目标的表现观由于受到姿态变换、遮挡等因素的影响导致跟踪困难,假如能够同时考虑目标周围的信息,则可一定程度上可以降低这种风险,STC 算法^[22]就是从这个角度考虑的视觉跟踪方法,它把跟踪过程设计成一个学习似然度函数的过程,而这个过程采用的也是寻找滤波器在跟踪区域的相关响应的方法,其差别在于算法把目标位置周围的信息定义成一个先验概率密度函数.如果在当前帧,我们令 x_c 为目标的位置坐标, $\Omega_c(x)$ 表示目标中心周围的上下文信息,那么上下文特征集合可以定义为

$$P^c = \{c(x') = (I(x'), x' | x' \in \Omega_c(x_o))\},$$

$I(x)$ 表示图像 x' 位置处的密度,这样,通过全概率公式可以把位置的似然分布函数 $\ell(x)$ 表示为

$$\ell(x) = P(x|o) = \sum_{c(x') \in P^c} P(x, c(x')|o) = \sum_{c(x') \in P^c} P(x|c(x'), o)P(c(x')|o) \quad (1)$$

这里, x, x' 是位置坐标, $P(c(x')|o)$ 表示目标的表现先验模型, $P(x|c(x'), o)$ 是一个描述空间纹理信息与目标位置关系的空间上下文,STC 算法的核心就在于如何得到空间上下文模型。

1.1 上下文先验表达

$P(c(x')|o)$ 表示目标的表现先验模型,在 STC 算法^[22]里,将其定义为

$$P(c(x')|o) = I(x')\omega_\sigma(x' - x_o) \quad (2)$$

这里, ω_σ 表示一个加权高斯函数,其定义为

$$\omega_\sigma(x' - x_o) = \alpha \cdot \exp(-\|x - x'\|^2 / \sigma^2) \quad (3)$$

这里, α 表示归一化参数, σ 则是一个调节高斯函数平滑的参数,通过高斯权重函数我们可以得知,靠近中心

的目标被赋予了更高的权重,因此,跟踪器对中心区域显示了更高的关注度.

1.2 空间上下文表达

现实中没有直接的空间上下文表达函数,STC^[22]用衡量不同位置差别的运算符 $h(x-x')$ 定义这个函数:

$$P(x|c(x'),o) = h(x-x') \quad (4)$$

设定该函数是非径向对称函数可以分析目标邻域里相似区域造成的不确定,对目标跟踪这个实际问题,通过初始化可知目标处于跟踪区域的中心,因此可以假设此时的似然分布函数服从高斯分布:

$$\ell(x) = P(x|o) = b \cdot \exp\left(-\left\|\frac{x-x'}{\alpha}\right\|^\beta\right) \quad (5)$$

这里, b 是归一化参数, α 是高斯函数的平滑参数, β 是密度函数形状控制参数. 有了这个初始的似然分布函数,我们就可以将上述算式带入公式(1),得到:

$$\ell(x) = P(x|o) = b \cdot \exp\left(-\left\|\frac{x-x'}{\alpha}\right\|^\beta\right) = \sum_{c(x') \in P^c} h(x-x') I(x') \omega_\sigma(x'-x_o) = h(x) \otimes (I(x) \omega_\sigma(x-x_o)) \quad (6)$$

这里, \otimes 表示卷积运算,通过卷积理论,令 F 表示傅里叶变换, F^{-1} 表示傅里叶逆变换,利用快速傅里叶变换可以得到空间上下文的推导:

$$h(x) = F^{-1} \left(\frac{F \left(b \cdot \exp\left(-\left\|\frac{x-x_o}{\alpha}\right\|^\beta\right)\right)}{F(I(x) \omega_\sigma(x-x_o))} \right) \quad (7)$$

1.3 基于时空上下文的目标定位

在确定好时空上下文的表达推导后,跟踪就变成了如何找到似然分布函数最大值的问题,如果假设第 n 帧的空间上下文信息已经得到,那么第 $n+1$ 帧的目标中心位置 x_o^{n+1} 就转化为求取第 $n+1$ 帧似然分布函数最大值的坐标,可以表示为

$$x_o^{n+1} = \arg \max_{x \in \Omega_o(x_o^n)} \ell^{n+1}(x) \quad (8)$$

$\ell^{n+1}(x)$ 表示第 $n+1$ 帧的似然分布函数,可以通过如下公式计算得到:

$$\ell^{n+1}(x) = F^{-1}(F(H^{n+1}(x)) \odot F(I^{n+1}(x) \omega_\sigma(x-x_o^n))) \quad (9)$$

这里, $H^{n+1}(x)$ 表示第 $n+1$ 帧的时空上下文表达, \odot 表示点乘,根据算法可以得知能够使得似然分布函数取得最大值的坐标位置就是新的目标位置,更多的细节请详读文献[22].

2 低秩重检测的多特征时空上下文跟踪算法

为了进一步提高时空上下文跟踪方法的性能,本文提出多特征融合的方法来更加精确地确定位置的似然分布函数 $\ell(x)$,同时,采用低秩 CUR 滤波的方法构建一个在跟踪失败发生时的重检测模块,嵌入到跟踪器中实现基于低秩重检测的多特征时空上下文跟踪方法,跟踪方法流程图如后文图 2 所示.

2.1 多特征融合的时空上下文目标定位

在目标跟踪领域,对目标进行表现建模的常用特征有梯度直方图(HOG),颜色、局部二值模式(LBP)、Haar 特征等,文献[24]表明,多特征融合技术可以提高跟踪性能,而 STC 算法仅使用灰度这单一特征,会造成表现模型的判别性随着跟踪的进行而产生波动.在本文中,我们从多个特征的角度出发,将表示空间信息与位置关系的似然分布函数 $\ell(x)$ 重新定义为

$$\ell(x) = P(x_k|o) = \sum_{c(x'_k) \in P^c} P(x_k, c(x'_k)|o) = \sum_{c(x'_k) \in P^c} P(x_k|c(x'_k), o) P(c(x'_k)|o) = h(x_k) \otimes (I(x_k) \omega_\sigma(x_k - x_{ok})) \quad (10)$$

其中, k 表示特征的个数,对公式(10)进行相关的傅里叶变换,并经过运算可以得到此时的多特征融合的空间上

下文表达的计算公式:

$$h(x_k) = F^{-1} \left(\frac{F \left(b \cdot \exp \left(-\frac{\|x - x_{ok}\|^\beta}{\alpha} \right) \right)}{F(I(x_k)\omega_\sigma(x - x_{ok}))} \right) \quad (11)$$

如果假设第 t 帧的多特征空间上下文信息已经得到,那么第 $t+1$ 帧的目标中心位置 x_o^{t+1} 就转化为求取第 $t+1$ 帧多特征融合的总体似然分布函数最大值的坐标,可以表示为

$$x_o^{t+1} = \arg \max_{x \in \Omega_k(x_o^t)} \sum_k \varphi_k \ell^{t+1}(x_k) \quad (12)$$

φ_k 代表各个特征在形成的总体似然分布函数中的权重,这个值的大小显示了特征贡献的大小,它的选取可以选为固定的先验,例如我们的特征选择颜色三通道 R、G、B,那么我们可以选 φ_k 为 $\frac{1}{3}$, $\ell^{t+1}(x_k)$ 表示第 $t+1$ 时刻的第 k 个特征产生的似然分布函数,其计算采用如下方式:

$$\ell^{t+1}(x_k) = F^{-1}(F(H^{t+1}(x_k)) \odot F(I^{t+1}(x_k)\omega_\sigma(x - x_o^t))) \quad (13)$$

这里, $H^{t+1}(x_k)$ 表示第 $t+1$ 时刻的第 k 个特征的时空上下文信息表达,由历史信息的空间信息上下文的线性插值在线更新:

$$H^{t+1}(x_k) = (1 - \rho)H^t(x_k) + \rho h^t(x_k) \quad (14)$$

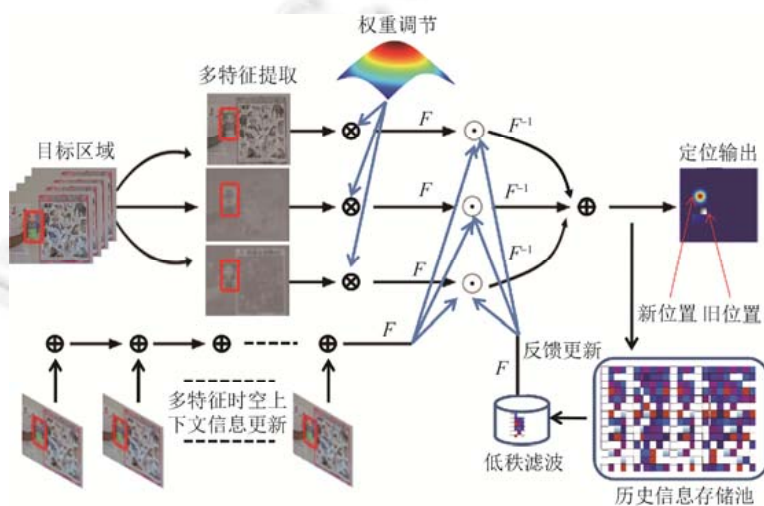


Fig.2 The flowchart of the proposed algorithm

图2 本文算法流程图

2.2 基于在线低秩滤波器的重定位

STC 跟踪方法在发生遮挡或者长时间的视角外跟踪会导致跟踪失败,为了使 STC 跟踪实现可靠长时间的有效跟踪,一个重检测模块是必须的,这种策略在很多已有的跟踪方法中得到了运用,如文献[13]利用到了光流法重检测,文献[25]则利用到了随机蕨丛(fern)的在线检测方法.我们知道,由于检测器需要足够多的训练样本才能完成,因而非常有效的检测器的设计是非常困难的.如果在线的有标签的样本信息无法得到保障的话,则这个工作就更无法开展,但是我们也知道,如果能够合理地挖掘历史信息,充分运用得到的正样本的时空结构信息,就可以得到一个鲁棒的重检测器.与上述工作采用大量的信息进行训练且只是保证每一帧的重定位不同,我们想保证已经跟踪到的正确样本的整体结构信息的稳定性,于是利用 CUR 矩阵低秩表达^[26]的方式来拟合已经得到历史信息,并且简单、快速地学习一个重检测器,以进一步地提高多特征 STC 跟踪的效果.

2.2.1 CUR 矩阵分解

在目标跟踪领域,需要处理大量的视频序列数据矩阵,而对于处理大尺度的矩阵运算,由于计算资源所限,精确地进行矩阵逆变换或者分解几乎是不可能的,如何找到一种快速次优的解决上述问题的矩阵分解算法成为关键.不同于主成分分析(PCA)、奇异值分解(SVD)等方法计算复杂度大,这里采用一种称为 CUR 矩阵分解的矩阵低秩近似方法,它的优点在于:在有足够误差精度保证的前提下,能够快速、简单地进行计算.

CUR 矩阵分解^[27]是寻找数据 K 矩阵, $K \in R^{m \times n}$ 的所有列阵的 c 个子列构成一个 C 矩阵, $C \in R^{m \times c}$, 所有行阵的 r 个子行构成一个 R 矩阵以及一个交互 U 矩阵 $U \in R^{c \times r}$, 使其满足 $\|K - CUR\|$ 的某种范数误差最小, $K \approx CUR$ 被称为矩阵的 CUR 分解.在本节我们想利用少量的采样数就能够合理地拟合历史信息,确保跟踪过程的重检测的精确性,因此,本文仅采用了随机采样的方式选择列矩阵,这种方法能够实现我们的要求.下面给出能够确保随机采样算法有效性的误差上限的定理^[28].

定理 1. 给定一个矩阵秩 ρ 的矩阵 K , $K \in R^{m \times n}$, 以相同次优行、列选择的方式:

$$c = \frac{2k}{\varepsilon} (1 + o(1)), r = \frac{2k}{\varepsilon} (1 + o(1))$$

选择数据矩阵 C , $C \in R^{m \times c}$ 和数据矩阵 R , $R \in R^{r \times n}$, k 为期望的秩, $2 \leq k \leq \rho$, ε 为误差控制参数, $\varepsilon \in (0, 1)$, 那么下面的不等式可以满足:

$$E \|K - CC^+KR^+R\|_F^2 \leq (1 + \varepsilon) \|K - K_k\|_F^2 \quad (15)$$

2.2.2 基于在线的 CUR 低秩滤波器的重定位

有了定理 1 的理论保证,我们假设在 t 帧得到的跟踪目标的外观表达矩阵为 K_t , 将该矩阵按照列堆砌成列向量 k_t , 作为目标历史表观表达信息存储到历史信息记录池矩阵 K 中, 作为历史信息记录池矩阵的一列. 然后, 按照公式 (16) 对矩阵 K 进行 CUR 分解以进行低秩表达. 随后, 随机地采样产生 CUR 分解的 C 矩阵, 在得到矩阵 C 后, 将其进行按照列的求和平均加以处理:

$$o_t = \frac{1}{c} \sum_{i=1, \dots, c} C(:, i) \quad (16)$$

然后, 将得到的列向量 o_t 进行矩阵化操作, 也就是按照表观表达矩阵 K_t 堆砌成列向量的逆操作, 得到类似于目标模板空间表达的矩阵 O_t , 此时的矩阵可以看成历史信息记录池矩阵 K 的一个低秩表达, 我们称其为低秩 CUR 滤波器. 随后, 我们使用低秩的 CUR 滤波器对跟踪图像区域进行稠密的卷积采样运算, 与 STC 的计算过程一样, 这个过程也可以视为利用可靠的历史信息记录池的信息对跟踪的过程进行反馈调节. 由于该算法仅利用到了参数 c 、 k 、 ε 这几个常规的数字设置参数, 令 $k=2$, $\varepsilon=0.2$, 则根据公式 (16) 可以得到 $c=20$, 因而非常快速. 虽然这里也进行了相关运算的步骤, 但是我们清楚地看出 STC 的相关运算在于保证上下文信息在时空上的一致性, 而 CUR 滤波器在于考虑整个跟踪过程目标表观的内在稳定性.

在多特征的 STC 跟踪器得到可能的最佳跟踪位置后, 记为粗定位状态, 其表观表达矩阵为 S_t , 我们调用在线的 CUR 低秩滤波对当前跟踪框进行相关检测, 得到有排序的前 N 个最大可能的区域矩阵, 并将其归一化为与目标模板尺度一致的矩阵 D_t . 如果此时粗定位状态的表观表达矩阵 S_t 与检测到 N 个可能的区域矩阵的相似度大于阈值 θ , 则此时跟踪到的区域是正确的区域, 否则, 我们将 CUR 滤波器的表观表达反馈给多特征融合的 STC 跟踪器, 作为此时的初始模板进行重新定位. 需要指出的是, 历史表观的 CUR 矩阵的更新采用增量学习跟踪算法的更新方法^[6], 对交互矩阵 U 进行更新. 但是, 本文从实时性的角度出发, 采用的是随机均匀采样. 文献[29]指出, 随机投影在有历史信息的大尺度矩阵处理中有非常重要的意义, 我们把本文的这个采样过程看成随机投影矩阵 R .

3 实验结果分析

为了验证本文提出的低秩重检测的多特征时空上下文跟踪算法的有效性, 我们在跟踪算法的基准库上进行了相关的实验对比. 所有的视频都来自于文献[2]中提供的基准库, 由于本文采用的是颜色三通道 RGB 的像素

值作为特征,我们选取了里面的彩色图像序列,视频里所有图像的原始标注(ground truth)均来自文献[2]原始的数据,而且所有参与评测的算法也保持原有的初始条件及参数不变.本文选取了 9 个公开代码的跟踪算法进行比较,有时空上下文跟踪算法(STC)、核相关滤波跟踪算法(KCF)^[21]、加速近端梯度 L1 算法(L1APG)^[30]、多任务稀疏学习算法(MTT)^[7]、增量式学习算法(IVT)^[6]、自适应结构局部稀疏算法(ASLA)^[31]、迭代凸优化算法(ORIA)^[32]、分布场跟踪算法(DFT)^[33]、压缩感知跟踪算法(CT)^[34].

3.1 实验细节

本文的算法是在原始的时空上下文的基础上,进行了多特征融合以及低秩 CUR 重检测机制的改进后完成.程序的执行程序为 Matlab,实验的硬件为 Intel I5,3.2GHz 处理器,8GB 内存,实验算法的平均速度约为 25 帧每秒(FPS),时空上下文(STC)、多特征时空上下文(MFSTC)和本文算法(Ours)在 David 视频上平均处理时间为 120 FPS、50FPS 和 25FPS.

上下文信息的选择区域为目标周围的两倍面积,公式(5)中先验似然函数参数 $\alpha=2.25$, $\beta=1$,公式(15)中更新率参数 $\rho=0.075$,公式(16)中采样参数 $c=20$,CUR 重检测相似度参数 $\theta=0.05$.需要指出的是,本文仅使用了颜色特征,没有测试灰度视频序列,只要选择合适的其他特征,例如灰度值、LBP、HOG 等进行融合,本文的算法依然有效.为了保持原样,其余所有跟踪算法都使用的是原始参数,而且参数在所有的视频跟踪过程中保持不变.本文选取的 12 个视频序列包含了复杂情况下跟踪的各种难度属性,这也是跟踪问题的难点所在,各评测视频的难点属性为:Cardark (1,4),Carscale(3,5,6,7,8),Coke(4,5,6,7,8),Crossing(3),David(2,3,4,5,6,7,8),David3(5,8),Deer(6,7),Girl(3,5,6,7,8),Jogging1(5,8),Jogging2(5,8),Bike(5,6,7,8),Shaking(3,4,6,7,8),其中,难点属性(1~8)分别是“1”表示复杂背景(background clutter)、“2”表示运动模糊(motion blur)、“3”表示尺度变化(scale variation)、“4”表示光照变化(illumination variation)、“5”表示遮挡(occlusion)、“6”表示快速运动(fast motion)、“7”表示平面内旋转(in plane)、“8”表示平面外旋转(out off plane).

3.2 评价指标

为了定量地进行本文提出的低秩重检测多特征时空上下文跟踪方法的性能,我们采用两个评价指标:精确度和成功率.精确度常用中心位置误差:跟踪轨迹的中心位置与人工标记的基准结果中心位置的欧式距离.成功率通常采用重合率^[35]:即得到的跟踪的目标框与基准标注目标框的交集面积与其并集面积的比值.我们使用平均中心位置误差和平均重合率度量算法在一个视频上的跟踪效果,使用一次性评估方式(OPE),根据成功率曲线和精确度曲线^[2]评价算法在所有视频上的整体效果.其中,成功率曲线定义如下:将重合率大于给定阈值的视频帧作为成功跟踪的帧,当阈值从 0 变化到 1 时,记录成功跟踪的帧数占视频总帧数的比率,得到成功率曲线.在成功率曲线下的面积(AUC)越大,说明该目标跟踪算法在该视频上整体跟踪的性能越高.

首先,我们验证一下本文算法在多特征融合和低秩重检测上对原始 STC 在跟踪失败后重检测的改进效果,选取 Coke、Jogging 视频对其进行了评测.图 3 是 Jogging 1 图像的跟踪中心位置误差曲线,该图像主要存在遮挡,从跟踪的误差曲线可见,在图像 70 帧左右由于遮挡的发生使得时空上下文跟踪方法发生了漂移,而单独加入多特征融合的时空上下文则能在遮挡发生时根据周围的颜色信息进行跟踪.可以看出,加入历史性信息低秩表达的重检测机制后,我们的跟踪算法有效地排除了遮挡带来的跟踪困难,跟踪效果良好.

图 4 是 Coke 图像的跟踪中心位置误差曲线,该图像序列主要存在遮挡、光照变化和旋转,这个视频需要算法有很好的处理遮挡光照和旋转的能力,从跟踪的误差曲线我们可以看出,在 60 帧左右存在光照的变化,由于采用的是颜色特征,多特征的 STC 对光照变化的干扰非常敏感,在 250 帧左右发生的全遮挡使得 STC 失去了跟踪目标,而我们的算法加入了低秩重检测机制,使得算法能够顺利地重新定位目标,良好地处理了这种问题.

3.2.1 全面定量实验对比

几种算法在所有视频上的总体成功率曲线和精确度曲线如图 5 和图 6 所示,图 5 为所有视频上的总体成功率曲线,图 6 为所有视频上的总体精确度曲线,本文提出的跟踪器的成功率曲线和精确度曲线在所有的 10 个参与测评跟踪器的上方,具有最大曲线下面积,数值分别达到了 0.593 和 0.786,时空上下文跟踪算法(STC)的数值

则分别为 0.428 和 0.593;核相关滤波跟踪算法(KCF)的跟踪结果则为 0.562 和 0.684,这些结果表明,我们的跟踪方法在测试视频集上具有较好的整体性能,这是由于,多特征信息和低秩的重检测机制使得跟踪效果更加鲁棒.为了使结果有一定的可视性,图 7 给出了部分跟踪结果的示例序列.

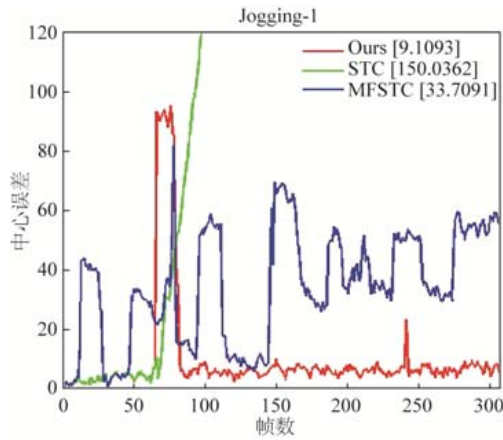


Fig.3 Center location error of Jogging
图 3 Jogging 视频中心位置误差曲线

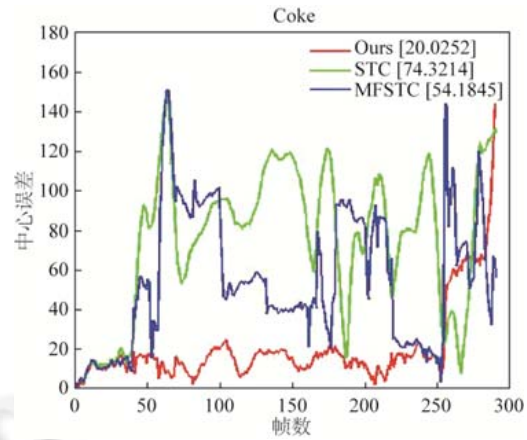


Fig.4 Center location error of Coke
图 4 Coke 视频中心位置误差曲线

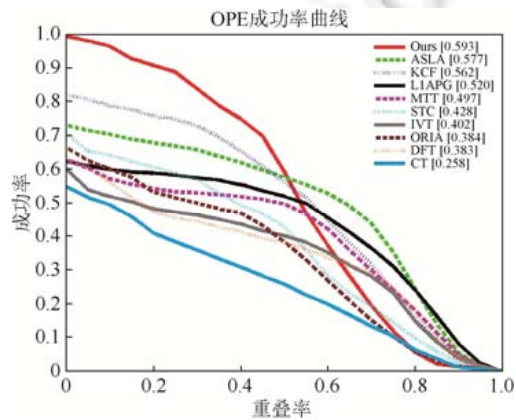


Fig.5 Overall success rate curves
图 5 总体成功率曲线

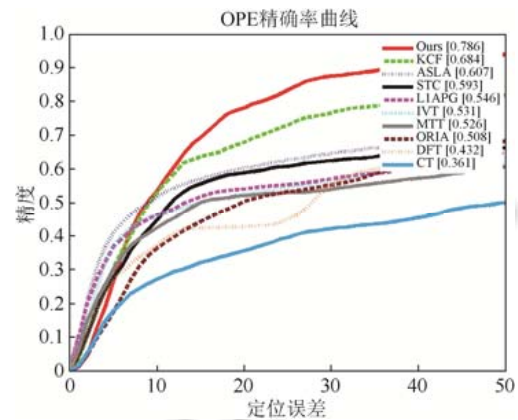


Fig.6 Overall precision rate curves
图 6 总体精确度曲线

为了全面地评估跟踪算法在不同难点上的性能,图 8 和图 9 给出各个跟踪器在各个难点属性上成功率和精确度的分析曲线.通过分析曲线我们可以看出,在 8 个难点属性里,本文提出的算法在以下 3 个属性上取得了跟踪精度第一:复杂背景、遮挡、平面外旋转,分别为 0.899、0.704、0.732,在其余除运动模糊和尺度变换两个属性外都排名前三.由于本文没有考虑尺度上的更新,使得在跟踪尺度变化上还有改进空间.综上,由所有的实验结果我们可以看出,本文算法在整个对比实验中有最好的整体跟踪性能,在单个的难度属性上也大部分取得了很好的效果,虽然在部分属性上未取得与相关算法相比的全面超越,但本文算法无论从精度、成功率和算法复杂度上都有不错的平衡,整体性能超过了参与比较的主流算法.

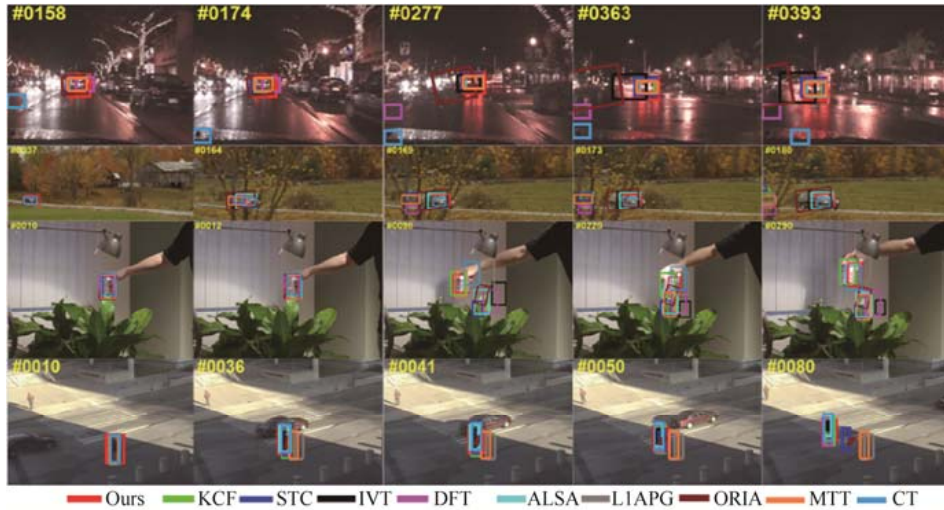


Fig.7 The tracking results illustration

图7 部分跟踪结果显示

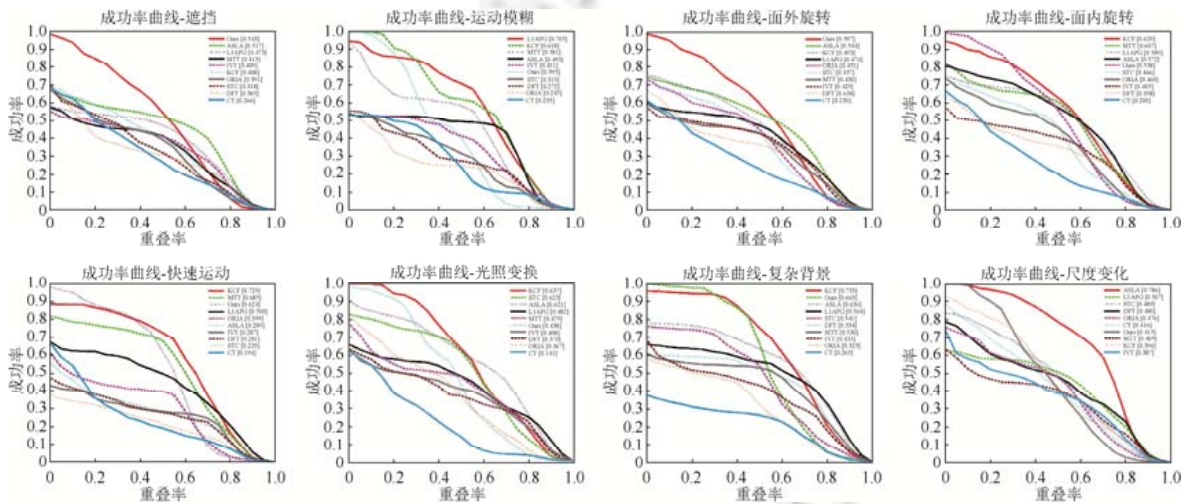


Fig.8 Tracking success curves on different attributes

图8 在不同跟踪视频属性上的成功率曲线

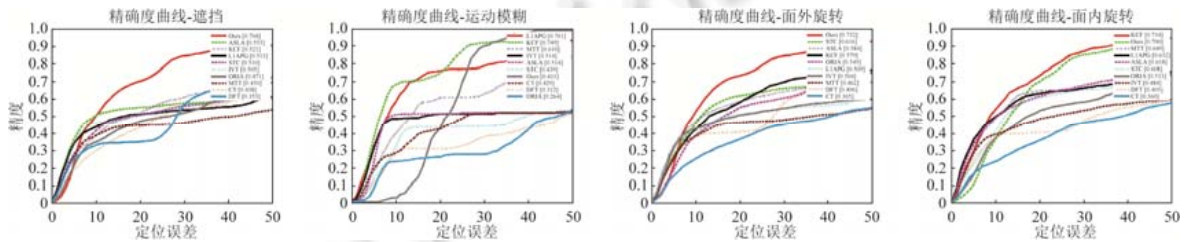


Fig.9 Tracking precision curves on different attributes

图9 在不同跟踪视频属性上的精确率

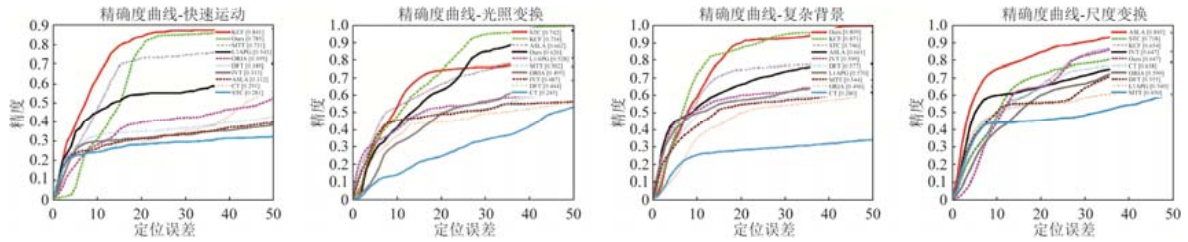


Fig.9 Tracking precision curves on different attributes (Continued)

图9 在不同跟踪视频属性上的精确率(续)

4 结论

本文提出了一种低秩重检测的多特征时空上下文的跟踪方法,通过利用多特征融合的技术改善了上下文的信息表达,完善了目标周围的结构信息,通过有效的低秩矩阵近似分解的方法构建了鲁棒的在线重检测器,该重检测器可以保持历史跟踪信息结构的一致性,在跟踪器跟踪失败后对目标周围进行目标的重定位,实现跟踪方法的长时有效跟踪.通过大量的实验评测可以得出,本文算法能够有效地处理遮挡等引起的跟踪失败的重定位问题,实验结果也表明,本文算法在跟踪精度、成功率和复杂度上取得了不错的性能,在总体性能上超过了当前一些先进的跟踪方法.需要指出的是,由于本文的多特征仅使用简单的颜色三通道,如果利用更加有判别性的特征提取算法,本文的跟踪算法将有进一步的提高.另外,目前基于深度学习的跟踪方法也日渐流行,利用卷积网络的特征也是提高跟踪效果的一个改进方向.

致谢 我们向本文的编辑及审稿人表示感谢,感谢他们提出的深入而有建设性的修改意见.

References:

- [1] Hou ZQ, Han CZ. A survey of visual tracking. *Acta Automatica Sinica*, 2006,32(4):603–617 (in Chinese with English abstract). [doi: 10.16383/j.aas.2006.04.016]
- [2] Wu Y, Lim J, Yang MH. Online object tracking: A benchmark. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2013. 2411–2418. [doi: 10.1109/CVPR.2013.312]
- [3] Comaniciu D, Ramesh V, Meer P. Kernel-Based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003,25(5):564–577. [doi: 10.1109/TPAMI.2003.1195991]
- [4] Wen ZQ, Cai ZX. Convergence analysis of Mean Shift algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2007,18(2):205–212 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/205.htm>
- [5] Yun TJ, Guo YC, Gao C. Human tracking in infrared images based on particle mean shift migration algorithm. *Chinese Journal of Computers*, 2009,32(6):1222–1228 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2009.01222]
- [6] Ross DA, Lim J, Lin RS, Yang MH. Incremental learning for robust visual tracking. *Int'l Journal of Computer Vision*, 2008,77(13):125–141. [doi: 10.1007/s11263-007-0075-7]
- [7] Zhang T, Ghanem B, Liu S, Ahuja N. Robust visual tracking via structured multi-task sparse learning. *Int'l Journal of Computer Vision*, 2013,101(2):367–383. [doi: 10.1109/CVPR.2012.6247908]
- [8] Mei X, Ling H. Robust visual tracking using L1 minimization. In: *Proc. of the Int'l Conf. on Computer Vision*. Kyoto: IEEE, 2009. 1436–1443.
- [9] Wang MH, Liang Y, Liu FM, Luo XN. Object tracking based on component-level appearance model. *Ruan Jian Xue Bao/Journal of Software*, 2015,26(10):2733–2747 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4737.htm> [doi: 10.13328/j.cnki.jos.004737]

- [10] Avidan S. Support vector tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2004,26(8):1064–72. [doi: 10.1109/TPAMI.2004.53]
- [11] Grabner H, Grabner M, Bischof H. Real-Time tracking via on-line boosting. In: *Proc. of the British Machine Vision Conf.* 2006. 47–56. [doi: 10.5244/C.20.6]
- [12] Babenko B, Yang MH, Belongie S. Robust object tracking with online multiple instance learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2011,33(8):1619–1632. [doi: 10.1109/TPAMI.2010.226]
- [13] Kalal Z, Matas J, Mikolajczyk K. P-N learning: Bootstrapping binary classifiers by structural constraints. In: *Proc. of the Int'l Conf. on Computer Vision*. Piscataway: IEEE Press, 2010,238(6):49–56. [doi: 10.1109/CVPR.2010.5540231]
- [14] Hare S, Saffari A, Torr PHS. Struck: Structured output tracking with kernels. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. Piscataway: IEEE Press, 2011. 263–270. [doi: 10.1109/ICCV.2011.6126251]
- [15] Ma L, Lu JW, Feng JJ, Zhou J. Multiple feature fusion via weighted entropy for visual tracking. In: *Proc. of the Int'l Conf. on Computer Vision*. Piscataway: IEEE Press, 2015. 3128–3136. [doi: 10.1109/ICCV.2015.358]
- [16] Bolme DS, Beveridge JR, Draper BA, Lui YM. Visual object tracking using adaptive correlation filters. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2010. 2544–2550. [doi: 10.1109/CVPR.2010.5539960]
- [17] Hu JL, Lu JW, Tan YP. Deep metric learning for visual tracking. *IEEE Trans. on Circuits and Systems for Video Technology*, 2016,26(11):2056–2068. [doi: 10.1109/TCSVT.2015.2477936]
- [18] Lasserre J A, Bishop CM, Minka TP. Principled hybrids of generative and discriminative models. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2006. 87–94. [doi: 10.1109/CVPR.2006.227]
- [19] Danelljan M, Khan FS, Felsberg M, Weijer J. Adaptive color attributes for real-time visual tracking. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2014. 1090–1097. [doi: 10.1109/CVPR.2014.143]
- [20] Danelljan M, Hager G, Khan FS, Felsberg M. Accurate scale estimation for robust visual tracking. In: *Proc. of the British Machine Vision Conf.* 2015. [doi: 10.5244/C.28.65]
- [21] Henriques JF, Caseiro R, Martins P, Batista J. High-Speed tracking with kernelized correlation filters. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2015,37(3):583–596. [doi: 10.1109/TPAMI.2014.2345390]
- [22] Zhang K, Zhang L, Liu Q, Zhang D, Yang M-H. Fast visual tracking via dense spatio-temporal context learning. In: *Proc. of the European Conf. on Computer Vision*. Marseille: Springer-Verlag, 2014,8693:127–141. [doi: 10.1007/978-3-319-10602-1_9]
- [23] XU JQ, LU Y. Robust visual tracking via weighted spatio-temporal context learning. *Acta Automatica Sinica*, 2015,41(11): 1901–1912 (in Chinese with English abstract). [doi: 10.16383/j.aas.2015.c150073]
- [24] Yang M, Lü F, Xu W, Gong Y. Detection driven adaptive multi-cue integration for multiple human tracking. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2009. 1554–1561. [doi: 10.1109/ICCV.2009.5459252]
- [25] Ma C, Yang X, Zhang C, Yang MH. Long-Term correlation tracking. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2015. 5388–5396. [doi: 10.1109/CVPR.2015.7299177]
- [26] Zhu G, Wang J, Wu Y, Lu H. Collaborative correlation tracking. In: *Proc. of the British Machine Vision Conf.* 2015. 1–12. [doi: 10.5244/C.29.184]
- [27] Mahone MWy. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, 2011,3(2):123–224.
- [28] Wang S, Zhang Z. Improving CUR matrix decomposition and the nystrom approximation via adaptive sampling. *Journal of Machine Learning Research*, 2013,14(9):2729–2769.
- [29] Drineas P, Mahoney MW. RandNLA: Randomized numerical linear algebra. *Communications of the ACM*, 2016,59(6):80–90. [doi: 10.1145/2842602]
- [30] Bao C, Wu Y, Ling H, Ji H. Real time robust L1 tracker using accelerated proximal gradient approach. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2012. 1830–1837. [doi: 10.1109/CVPR.2012.6247881]
- [31] Jia X, Lu H, Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2012. 1822–1829. [doi: 10.1109/CVPR.2012.6247880]

- [32] Wu Y, Shen B, Ling H. Online robust image alignment via iterative convex optimization. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012. 1808–1814. [doi: 10.1109/CVPR.2012.6247878]
- [33] SevillaLar L, Learned-Miller E. Distribution fields for tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012. 1910–1917. [doi: 10.1109/CVPR.2012.6247891]
- [34] Zhang K, Zhang L, Yang MH. Real-Time compressive tracking. In: Proc. of the European Conf. on Computer Vision. Marseille: Springer-Verlag, 2012,7574(1):864–877. [doi: 10.1007/978-3-642-33712-3_62]
- [35] Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. Int'l Journal of Computer Vision, 2010, 88(2):303–338. [doi: 10.1007/s11263-009-0275-4]

附中文参考文献:

- [1] 侯志强,韩崇昭.视觉跟踪技术综述.自动化学报,2006,32(4):603–617. [doi: 10.16383/j.aas.2006.04.016]
- [4] 文志强,蔡自兴.Mean Shift 算法的收敛性分析.软件学报,2007,18(2):205–212. <http://www.jos.org.cn/1000-9825/205.htm>
- [5] 云廷进,郭永彩,高潮.基于粒子 mean shift 迁移的红外人体目标跟踪算法.计算机学报,2009,32(6):1222–1228. [doi: 10.3724/SP.J.1016.2009.01222]
- [9] 王美华,梁云,刘福明,罗笑南.部件级表观模型的目标跟踪方法.软件学报,2015,26(10):2733–2747. <http://www.jos.org.cn/1000-9825/4737.htm> [doi: 10.13328/j.cnki.jos.004737]
- [23] 徐建强,陆耀.一种基于加权时空上下文的鲁棒视觉跟踪算法.自动化学报,2015,41(11):1901–1912. [doi: 10.16383/j.aas.2015.c150073]



郭文(1978—),男,山东烟台人,博士,副教授,CCF 专业会员,主要研究领域为计算机视觉,多媒体计算.



张天柱(1982—),男,博士,副研究员,CCF 专业会员,主要研究领域为计算机视觉,多媒体计算.



游思思(1994—),女,硕士生,CCF 学生会员,主要研究领域为计算机视觉.



徐常胜(1969—),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为多媒体内容分析/检索,模式识别,计算机视觉.