

基于稀疏表示的视频目标跟踪研究综述

黄宏图^{1,2} 毕笃彦¹ 侯志强^{3,4} 胡长城² 高山¹ 查宇飞¹ 库涛¹

摘要 视频目标跟踪在计算机视觉领域有着广泛应用, 由于目标自身和外界环境变化的复杂性和难以预知性, 使得复杂场景下鲁棒实时目标跟踪成为一项亟待解决的关键问题. 由于视觉信息可以用少量神经元进行稀疏表示, 因此稀疏表示已经广泛应用于人脸识别、目标检测和目标跟踪等计算机视觉领域. 本文旨在对基于稀疏表示的视频目标跟踪算法进行综述. 首先, 介绍了基于稀疏表示的视频目标跟踪算法中的字典组成; 其次, 介绍了稀疏模型的构建及求解算法和模型更新, 并对算法复杂度进行了简要分析; 然后, 对现有公开代码的稀疏表示跟踪算法在测试数据上进行了实验分析, 结合算法模型和实验结果对其进行了分析; 最后, 对基于稀疏表示的视频跟踪算法存在问题进行了讨论, 并对未来的研究趋势进行了展望.

关键词 视频跟踪, 稀疏表示, 算法评估, 实验分析

引用格式 黄宏图, 毕笃彦, 侯志强, 胡长城, 高山, 查宇飞, 库涛. 基于稀疏表示的视频目标跟踪研究综述. 自动化学报, 2018, 44(10): 1747–1763

DOI 10.16383/j.aas.2018.c170209

Research of Sparse Representation-based Visual Object Tracking: A Survey

HUANG Hong-Tu^{1,2} BI Du-Yan¹ HOU Zhi-Qiang^{3,4} HU Chang-Cheng² GAO Shan¹ ZHA Yu-Fei¹ KU Tao¹

Abstract Visual object tracking has been widely used in computer vision. Due to the complexity and unpredictability of the object itself and surroundings' changes, robust and real-time tracking is a key issue in urgent need of settlement in complex scenes. Since vision information can be expressed by few neurons, sparse representation has already been used in face recognition, object detection, visual tracking and so on. This paper aims to review the state-of-the-art of sparse representation-based visual tracking algorithms. Firstly, we introduce the codebook employed in the sparse representation-based trackers. Secondly, sparse model construction, corresponding solution and model update are described. And the algorithm complexity is briefly analyzed. Thirdly, the open sparse representation-based trackers' codes are conducted on the benchmark datasets, and the experimental results are fully analyzed in combination with models. Finally, we discuss the existing problems of sparse representation-based trackers and prospect the future.

Key words Visual tracking, sparse representation, algorithm evaluation, experiment analysis

Citation Huang Hong-Tu, Bi Du-Yan, Hou Zhi-Qiang, Hu Chang-Cheng, Gao Shan, Zha Yu-Fei, Ku Tao. Research of sparse representation-based visual object tracking: a survey. *Acta Automatica Sinica*, 2018, 44(10): 1747–1763

视频目标跟踪是计算机视觉领域的关键问题和研究热点, 是任何一个以视频为输入的实际应用系统的关键技术, 是后续目标识别、行为分析、视频压缩编码和视频理解等高级视频处理任务的基础^[1]. 视频目标跟踪的目的是在连续的图像序列中标记出

特定目标, 估计目标的位置、尺度或区域, 确定目标的速度以及轨迹等运动信息. 实际应用中完整的视频目标跟踪系统通常包括目标的检测、提取、识别和跟踪 4 个环节. 一般为了研究方便, 将其进行一定简化, 例如检测、提取和识别通过人工方式完成, 从而突出跟踪环节算法的研究^[2].

在民用领域, 视频目标跟踪在智能视频监控^[3]、智能交通系统、人机交互、视觉导航、无人驾驶汽车^[4]、医学图像分析、虚拟现实、运动分析、行为识别和视频检索等方面有较多应用. 在军事领域, 视频目标跟踪已经广泛应用于战场侦察与监视、景象匹配制导和无人机对地目标跟踪等领域^[5].

近 30 多年来得益于计算机技术的发展, 视频目标跟踪研究已经取得较大进展^[6]. 但是目前复杂环境下持续鲁棒的视频目标跟踪仍然比较困难, 没有哪个跟踪算法能够成功地应用于所有的视频跟踪任务, 大多数算法在跟踪一段时间之后, 通常就会因

收稿日期 2017-04-19 录用日期 2017-09-23
Manuscript received April 19, 2017; accepted September 23, 2017

国家自然科学基金 (61472442, 61773397, 61473309, 61773397) 资助

Supported by National Natural Science Foundation of China (61472442, 61773397, 61473309, 61773397)

本文责任编辑 桑农

Recommended by Associate Editor SANG Nong

1. 空军工程大学航空工程学院 西安 710038 2. 中国人民解放军 95972 部队 酒泉 735018 3. 空军工程大学信息与导航学院 西安 710077 4. 西安邮电大学计算机学院 西安 710112

1. Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038 2. 95972 Troops of PLA, Jiuquan 735018 3. Information and Navigation Institute, Air Force Engineering University, Xi'an 710077 4. School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710112

为某些原因丢失目标^[7]. 鲁棒视频目标跟踪的难点在于随着时间和空间的推移, 目标自身变化和外界场景变化的复杂性和难以预知性^[8], 例如从 3D 空间投影到 2D 平面时的信息损失、光照变化、尺度变化、遮挡、形变、运动模糊、快速运动、旋转、复杂场景、低分辨率图像、目标移出视角、摄像机视角改变和噪声等. 从控制论观点看, 视频目标跟踪是一个从输入到输出的单向开环系统, 由于没有反馈输入, 很难确认跟踪的正确与否, 随着时间的推移, 模型更新中误差累积会产生漂移, 最终因系统发散导致跟踪失败^[9]. 此外某些算法的时间复杂度较高难以实现实时跟踪^[7]. 因此鲁棒实时的视频目标跟踪仍然是一项亟待解决的关键问题, 具有重要的理论意义和实用价值.

在信号与信息处理领域, 一直高度重视信号描述的“简单性”^[10]. 从信息论角度看, 如果信号是稀疏的, 或具有某种结构, 或可用某个确定的模型表示, 这样的信号称之为“简单”信号^[11]. “简单性”是简单信号的固有特性, 通常表现为稀疏性、低秩性和低熵性等. 稀疏表示作为信号“简单性”的描述, 是近几年研究热点. 生物学家研究发现^[12-13] 哺乳类动物在长期进化过程中, 视神经形成了快速、准确、低能耗地表示自然图像的能力, 关键在于哺乳动物在感知视觉信息时, 大脑视皮层 V1 区中只有少量神经元被激活, 即视觉信息可以用少量神经元进行稀疏表示.

稀疏表示广泛应用于人脸识别^[14]、图像超分辨率重建、图像去噪和恢复^[15]、图像分割、特征提取和融合^[16]、图像显著性检测^[17]、背景建模和图像分类^[18] 等计算机视觉领域^[19]. 稀疏表示有助于获取描述符的显著模式, 能够在有限的样本容量下, 使所得模型参数呈现某种稀疏性, 提高模型的可靠性和可解释性, 有利于采用模型对实际问题进行解释和指导, 并且重建性能好. 受到基于稀疏表示的人脸识别^[14] 的影响, 以及视频连续性产生的图像帧与帧之间的冗余性, 2009 年国际计算机视觉大会上, Mei 等首次将稀疏表示应用到视频目标跟踪中^[20], 构建由目标模板和单位矩阵组成的冗余字典, 核心思想是将候选目标图像表示为目标图像的稀疏线性组合, 从稀疏角度得到目标最紧致的表示. 后续出现了大量基于稀疏表示的视频目标跟踪算法^[21], 并取得了较好的跟踪性能^[22]. 其中基于灰度特征字典的稀疏表示跟踪算法对遮挡和噪声等具有一定的鲁棒性, 对于判别式跟踪算法而言在稀疏表示下目标和背景更加线性可分. 但算法时间复杂度一般较高, 难以实现实时跟踪和算法快速优化. 由于算法时间复杂度与字典维数相关, 难以使用高维的鲁棒特征, 导致算法鲁棒性低于某些算法.

文献 [21] 按照稀疏表示的用途和阶段不同将稀疏表示跟踪算法分为基于稀疏表示的表观建模和基于稀疏表示的目标搜索两类, 其中表观建模中将稀疏表示看成是对目标建模表示的过程, 目标搜索中将稀疏表示的过程看成是目标搜索的过程. 本文在粒子滤波框架下将基于稀疏表示的视频目标跟踪算法分为 4 个组成部分: 字典构建、稀疏模型的构建及求解、观测模型的构建和模型更新. 其中字典构建是基础, 因为后续的一切处理都是在字典中原子张成的子空间内进行. 稀疏模型的构建及求解是核心, 稀疏模型构建是在重构误差和稀疏性先验之间寻求某种微妙的平衡, 快速有效的模型求解算法是关键, 因为这直接涉及跟踪算法的处理速度. 观测模型构建是根本, 最终决定了目标匹配的相似性度量函数. 模型更新是重点, 由于跟踪过程目标是不断变化的, 在跟踪结果的基础上适时地对模型进行在线更新是鲁棒跟踪所必需的. 下面分别对上述 4 个组成部分进行分析.

1 字典构建

基于稀疏表示的视频目标跟踪算法本质上是冗余字典下的稀疏逼近问题^[23], 在冗余字典下能更有效地找出隐含在输入数据内部的结构与模式. 目前信号在冗余字典下的稀疏表示研究集中在以下两个方面: 1) 构建适合某一类信号的冗余字典; 2) 设计快速有效的稀疏分解算法. 如图 1 所示, 字典构建包括特征选择和字典组成两个步骤.

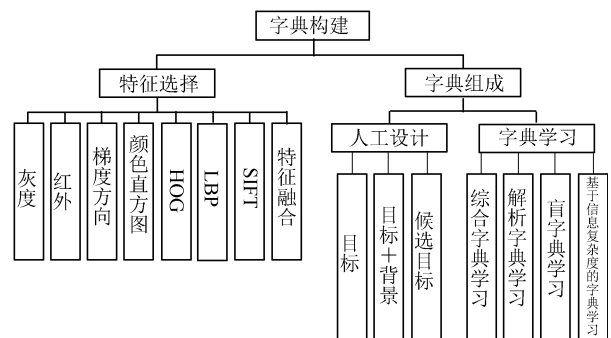


图 1 字典构建方法

Fig. 1 The codebook construction method

1.1 特征选择

受到基于稀疏表示的人脸识别的影响, 以及灰度特征的简单性和有效性, 大多数算法利用目标的全局模板或局部图像块的灰度特征^[24]. 但是灰度特征涉及像素点的对齐问题, 一般是通过以下两种方法实现像素点的对齐^[25]: 1) 在目标周围进行稠密采样, 使得字典中能够尽可能地包括图像的转换形式; 2) 候选样本的每列通过几何转换与字典中

的目标模板对齐. 后续跟踪算法引入了其他特征例如: 红外特征^[26]、颜色直方图、HOG (Histograms of oriented gradients)^[27]、像素点梯度方向的正余弦^[28]、SIFT (Scale-invariant feature transform)^[29]和几何模糊^[30]等. 文献 [31] 直接将颜色直方图、灰度、HOG 和 LBP (Local binary pattern) 堆叠实现简单融合. 文献 [32] 将像素点的坐标、灰度、梯度幅值和方向使用协方差矩阵进行融合, 在不同图像上的区域协方差描述符获得了一定的尺度和旋转不变性. 协方差矩阵将不同类型的特征进行有效融合, 利用了特征之间的空间特性、统计特性和特征之间的相关性且维数较小. 虽然协方差矩阵位于黎曼流形上, 但将其进行对数转换后可以在欧氏空间上进行度量. 多特征融合^[33] 的目的在于利用特征之间的互补性弥补单一特征的不足. 由于特征维数直接决定了后续稀疏求解算法的复杂度, 因此在特征选择过程中必须考虑特征维数对算法求解速度的影响.

1.2 字典组成

字典组成的方法分为人工设计方法和机器学习方法. 由于人工设计方法的简单性和有效性, 大多数字典构建方法直接使用预先指定的字典. 基于机器学习的字典构建方法大多需要大量的训练数据.

如图 1 所示, 人工设计的方法按照字典的构成可以分为: 基于目标特征的字典构建方法、基于目标特征和背景特征的字典构建方法和基于候选目标特征的字典构建方法.

基于目标特征的字典构建方法大多利用目标模板或者主成分分析 (Principle component analysis, PCA) 来构建字典, 为了处理遮挡和噪声等异常, 字典中一般还包括微模板^[20, 34-37] (与字典中原子维数同样大小的单位矩阵, 以下统称微模板). 由于稀疏表示反映了候选目标和字典中原子的线性相关性, 为了保证稀疏表示的非负性引入了负微模板. 后续算法为了减少字典中原子的个数提高算法效率, 将负微模板和非负性约束去掉, 利用目标特征和正微模板构建字典^[38]. 文献 [39-42] 使用局部图像块代替全局模板使得字典本身对于部分遮挡具有一定的鲁棒性.

为了提高模型处理复杂背景的能力, 文献 [43] 同时利用目标特征和背景特征构建字典, 使得字典本身具有判别性, 进而使得稀疏表示中包含有判别信息, 从而使算法具有较强的区分目标和背景的能力.

利用候选目标特征构建字典的方法^[44-45] 是将跟踪问题看作识别问题, 求解目标在候选目标组成字典下的稀疏表示, 根据稀疏表示即可从候选目标中确定目标的位置, 理论上只需要求解几次稀疏表

示, 相比传统方法大大减少了稀疏表示的求解次数.

目前稀疏表示跟踪算法中基于学习的字典构建方法还比较少. 字典学习^[10] 通过优化相应的目标函数, 获得能够对信号进行稀疏表示的字典, 或从分析的角度看, 通过优化目标函数使变换系数最稀疏. 字典学习根据数据或信号本身来学习冗余字典, 这类字典中的原子与训练集中的数据本身相适应. 与基于解析方法的字典相比, 通过学习获得的字典原子数量更多, 形态更丰富, 能更好地与信号或图像本身的数据结构匹配, 具有更稀疏的表示, 比解析方法构造的字典有更出色的性能.

综合字典学习都采用系数更新和字典更新交替优化的方式. 字典学习算法的区别主要在于字典更新方式, 而系数更新没有本质区别. 固定字典更新稀疏表示是标准的稀疏编码问题, 理论上任何一种稀疏编码方法都可以用于系数更新^[46-47]. 固定稀疏表示更新字典则是字典学习算法最为关注的环节. 综合字典学习试图找到一组能够反映信号本征空间的基, 解析字典学习则是从对偶分析的角度考虑稀疏表示问题, 试图找到信号正交空间的基. 上述两种模型的训练样本为特定的信号库, 盲字典学习的训练样本是待重构信号的测量值. 上述三种模型均以 ℓ_p 范数描述信号的稀疏性, 而基于信息复杂度的字典学习则是利用信息的复杂度描述信号的“简单性”. 信号越稀疏, 其复杂度越低, 因此它仍然属于广义的稀疏表示模型.

2 稀疏模型的构建及求解

稀疏模型的构建是在重构误差一定的前提下, 在稀疏性约束中加入一些目标的先验信息, 得到目标更加紧凑的表示^[48], 从而提取出数据潜在的内部结构. 由于大多数基于稀疏表示的视频目标跟踪算法是在粒子滤波框架下进行, 因此按照观测模型将其分为生成式模型、判别式模型和混合式模型.

2.1 生成式模型

目前基于稀疏表示的视频跟踪算法大多为生成式算法, 如图 2 所示, 生成式模型中根据目标匹配相似性度量函数的不同, 分为基于重构误差的生成式算法、基于稀疏表示系数的生成式算法和基于稀疏编码直方图的生成式算法.

基于重构误差的生成式算法一般是首先求解候选目标在稀疏模型下的表示系数, 而后求解候选目标基于目标特征的重构误差, 最后选择重构误差最小的候选目标作为跟踪位置. 如图 2 所示, 按照稀疏模型的不同分为 L_1 跟踪算法、加权稀疏编码跟踪算法、基于结构稀疏表示的跟踪算法、基于非局部自相似正则化的稀疏表示跟踪算法、多任务稀疏表示

跟踪算法和多任务多视角稀疏表示跟踪算法.

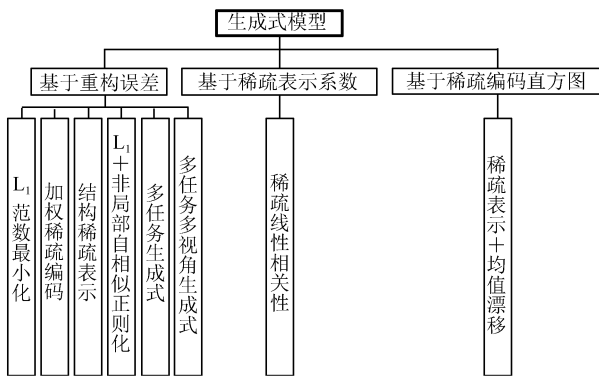


图2 基于稀疏表示的生成式模型

Fig. 2 The sparse representation-based generative model

基于目标特征重构误差的生成式算法的典型代表是 L_1 跟踪算法^[20, 24]，其算法框架如图3所示。

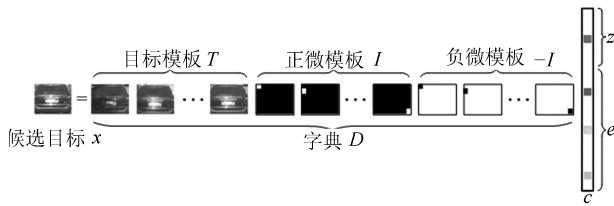


图3 L_1 跟踪算法框架

Fig. 3 The L_1 tracker framework

L_1 跟踪算法的稀疏表示模型为

$$\mathbf{c} = \arg \min_{\mathbf{c}} \|\mathbf{x} - D\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 \quad (1)$$

s. t. $\mathbf{c} \geq 0$

其中, $\mathbf{x} \in \mathbf{R}^d$ 为候选目标. $D = [T, I, -I] \in \mathbf{R}^{d \times (m+2d)}$ 为字典, $T \in \mathbf{R}^{d \times m}$ 为 m 个目标模板, $I \in \mathbf{R}^{d \times d}$ 为单位矩阵构成的微模板, 目标模板用于描述目标外观变化, 微模板用来显式地建模异常噪声点. $\mathbf{c} = [\mathbf{z}^T, \mathbf{e}^T]^T \in \mathbf{R}^{(m+2d)}$ 为候选目标 \mathbf{x} 在当前字典下的稀疏表示, 其中, $\mathbf{z} \in \mathbf{R}^m$ 为对应目标模板的系数, $\mathbf{e} \in \mathbf{R}^{2d}$ 为对应微模板的系数, λ 为稀疏性约束参数. 式 (1) 的求解算法为内点法^[49], 算法时间复杂度为 $O(d^2 + dm)$. 实验结果表明 L_1 跟踪算法对于光照变化、部分遮挡和噪声具有一定的鲁棒性^[24]. 缺点是模型更新中根据相似度将跟踪结果用于字典原子替换, 一旦引入背景信息会导致后续跟踪失败, 且为了提高算法速度采用低分辨率的图像灰度特征.

为了提高 L_1 跟踪算法的速度, Bao 等将加速最近梯度 (Accelerated proximal gradient, APG) 算法引入到稀疏模型的求解中, 并对微模板系数附加

l_2 范数约束^[35]:

$$\mathbf{c} = \arg \min_{\mathbf{c}} \frac{1}{2} \|\mathbf{x} - D\mathbf{c}\|_2^2 + \lambda_1 \|\mathbf{c}\|_1 + \frac{\lambda_2}{2} \|\mathbf{e}\|_2^2$$

s. t. $\mathbf{z} \geq 0$

(2)

其中, $D = [T, I] \in \mathbf{R}^{d \times (m+d)}$ 为字典, $\mathbf{c} = [\mathbf{z}^T, \mathbf{e}^T]^T \in \mathbf{R}^{(m+d)}$ 为候选目标 \mathbf{x} 在当前字典下的稀疏表示, 其中, $\mathbf{z} \in \mathbf{R}^m$ 为对应目标模板的系数, $\mathbf{e} \in \mathbf{R}^d$ 为对应微模板的系数, λ_1 为稀疏性约束, λ_2 控制微模板中的能量, 并使用基于重构误差的遮挡检测对 λ_2 进行动态调整, 当检测到遮挡时 $\lambda_2 = 0$, 当没有检测到遮挡时 $\lambda_2 > 0$. 当特征维数较低时, 算法速度提高较为明显, 随着特征维数的升高算法速度剧烈下降.

传统的稀疏表示模型中重构误差通常使用 l_1 范数或者 l_2 范数, 分别对应拉普拉斯误差分布和高斯误差分布. 然而在实际中重构误差并不一定是理想的拉普拉斯分布或高斯分布, 为了解决这个问题 L_1 跟踪算法使用微模板以较高的计算代价来解决遮挡或者噪声等异常情况. Yan 等^[50] 从重构误差的概率分布出发, 通过加权稀疏编码误差最小化来提高跟踪算法的鲁棒性, 旨在解决超出拉普拉斯和高斯噪声的异常, 理论上使算法模型适用于更为一般的噪声和异常. 因此字典中并不需要微模板, 减少了字典中原子个数一定程度上降低了算法复杂度.

基于 PCA 的重构误差生成式算法使用 PCA 代替目标模板, 主要原因有: 1) L_1 跟踪算法字典中只是利用目标模板来处理目标的变化, 而目标模板张成的子空间的表达能力是有限的, 难以处理目标较大的视角和姿态变化. 2) 字典一旦使用背景图像或者严重遮挡的跟踪结果更新后, L_1 跟踪算法很容易失败, 而基于 PCA 的字典能够最大限度地保留目标的类内方差, 从而获得目标模板丰富的冗余结构信息, 利用历史数据使用增量子空间学习得到目标表示的时间相关性, 由其张成的最优子空间不仅对当前目标, 而且对历史目标都有较小的重构误差, 因此本征模板集合相比目标模板能够提供更丰富的表达. 3) 使用 PCA 代替目标模板后, 字典中的微模板能够克服 PCA 对噪声太敏感只能对一些类似高斯分布的数据有效的不足.

Wang 等^[51] 利用 PCA 替换 L_1 跟踪算法中的目标模板, 同时利用了子空间对目标变化的建模能力和稀疏表示对遮挡等异常的鲁棒性. 候选目标 $\mathbf{x} \in \mathbf{R}^d$ 可以使用 PCA 子空间线性表示 $\mathbf{x} = U\mathbf{z} + \mathbf{e}$, 其中字典 $U \in \mathbf{R}^{d \times m}$ 由正交基向量组成, \mathbf{z} 为稀疏表示, \mathbf{e} 为重构误差. 其稀疏表示通过下式求解:

$$[\mathbf{z}, \mathbf{e}] = \arg \min_{\mathbf{z}, \mathbf{e}} \frac{1}{2} \|\mathbf{x} - U\mathbf{z} - \mathbf{e}\|_2^2 + \lambda \|\mathbf{e}\|_1 \quad (3)$$

其中, 由于 PCA 基向量是正交的, 因而对应正交基向量的系数 \mathbf{z} 是稠密, 而对应微模板的系数 \mathbf{e} 是稀疏的, 上述模型可以通过迭代优化求解.

Pan 等^[52] 在文献 [51] 的基础上从理论上证明当字典为正交字典时, 使用 l_0 范数约束表示目标比 l_1 和 l_2 范数约束更优. 将算法模型修正为

$$[\mathbf{z}, \mathbf{e}] = \arg \min_{\mathbf{z}, \mathbf{e}} \frac{1}{2} \|\mathbf{x} - U\mathbf{z} - \mathbf{e}\|_2^2 + \lambda_1 \|\mathbf{z}\|_0 + \lambda_2 \|\mathbf{e}\|_1 \quad (4)$$

其中, $\|\mathbf{e}\|_1$ 用于拒绝异常样本, $\|\mathbf{z}\|_0$ 用来实现特征选择, 目的是在保留子空间内有用成分的同时移除冗余特征. 上述模型能够使用加速最近梯度算法^[53] 有效求解.

Bai 等^[37] 利用遮挡的连续空间分布考虑图像的结构信息提出了基于结构稀疏表示的跟踪算法, 利用结构单元子空间的线性组合表示目标. 字典 $D = [U, I] \in \mathbf{R}^{d \times (m+d)}$ 由本征模板 $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m] \in \mathbf{R}^{d \times m}$ 和部分微模板集合 I 组成. 由于部分遮挡通常具有连续的空间分布, 将字典 D 分成 $(R+1)$ 个长度为 m 的组, $d = Rm$. $D[v] \in \mathbf{R}^{d \times m}$ 为 D 的第 v 组, 显然组内的原子之间是相互正交, 且字典的第 1 个组 $D[1] = U$ 是由目标模板经过奇异值分解得到的主成分组成. $\mathbf{c}[v] \in \mathbf{R}^m$ 为 \mathbf{c} 的第 v 组. 字典 D 和稀疏表示 \mathbf{c} 的分组分别为

$$D = \underbrace{[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]}_{D[1]} \underbrace{[\mathbf{i}_{m+1}, \mathbf{i}_{m+2}, \dots, \mathbf{i}_{2m}]}_{D[2]} \dots \underbrace{[\mathbf{i}_{Rm+1}, \mathbf{i}_{Rm+2}, \dots, \mathbf{i}_{(R+1)m}]}_{D[R+1]} \quad (5)$$

$$\mathbf{c} = \underbrace{[\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m]}_{\mathbf{c}^T[1]} \underbrace{[\mathbf{c}_{m+1}, \mathbf{c}_{m+2}, \dots, \mathbf{c}_{2m}]}_{\mathbf{c}^T[2]} \dots \underbrace{[\mathbf{c}_{Rm+1}, \mathbf{c}_{Rm+2}, \dots, \mathbf{c}_{(R+1)m}]}_{\mathbf{c}^T[R+1]} \quad (6)$$

Lu 等^[54] 考虑候选目标的稀疏表示之间的相关性, 在稀疏表示中引入了非局部自相似正则项 (Non-local self-similarity regularized sparse coding, NLSSSC), 利用最近邻来编码目标中的结构信息, 提高了算法的判别力. $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$ 为 n 个候选目标, $D = [T, I] \in \mathbf{R}^{d \times (m+d)}$ 为字典, 候选目标 \mathbf{x}_i 的稀疏表示为 $\mathbf{c}_i = [\mathbf{z}_i^T, \mathbf{e}_i^T]^T$. 对非局部自相似正则项附加 l_1 范数约束得到 NLSSSC

的目标函数为

$$\mathbf{c}_i = \arg \min_{\mathbf{c}_i} \frac{1}{2} \|\mathbf{x}_i - D\mathbf{c}_i\|_2^2 + \lambda_1 \|\mathbf{c}_i\|_1 + \lambda_2 \left\| \mathbf{c}_i - \sum_{i'=1}^{n_0} w_{i'} \mathbf{c}_{i'} \right\|_1 \quad (7)$$

其中, $i = 1, 2, \dots, n$, $i' = 1, 2, \dots, n_0$. 在候选目标中选择 \mathbf{x}_i 的前 n_0 个最近邻 $\{\mathbf{x}_{i'}\}$, $\mathbf{c}_{i'}$ 的分配权重为 $w_{i'} = \frac{1}{\hat{h}_i} \exp(-\frac{1}{\hat{\lambda}} \|N(\mathbf{x}_i) - N(\mathbf{x}_{i'})\|_2^2)$, $\hat{\lambda}$ 为强加相似性参数, \hat{h}_i 为归一化常数, $N(\mathbf{x}_i)$ 是 \mathbf{x}_i 附近的近邻向量.

上述算法是基于子空间的结构信息, 而文献 [55] 则是利用像素级遮挡的连续空间分布, 提出了基于结构稀疏学习的遮挡检测跟踪算法. 候选目标中所有像素构成一组节点为 V 、边为 E 的图, 边存在于任意相邻的两个像素点之间. 权重 w_{ml} 表示像素 m 和 l 之间的相关性, 与其灰度值的相关性成正比, 与其欧氏距离成反比. 结构稀疏表示的目标函数为

$$[\mathbf{z}, \mathbf{e}] = \arg \min_{\mathbf{z}, \mathbf{e}} \|\mathbf{z}\|_1 + \lambda \|\mathbf{e}\|_1 + \gamma \sum_{(m,l) \in E} w_{ml} \|e_m - e_l\|_1 \quad (8)$$

s. t. $\mathbf{x} = D\mathbf{z} + \mathbf{e}$

其中, γ 越大融合作用越大, w_{ml} 作为融合权重惩罚使得高度相关的像素有较高的 w_{ml} . 上述模型引入松弛变量后采用近似扩展拉格朗日乘数法 (Inexact augmented lagrange multiplier, IALM) 求解.

利用候选目标稀疏表示之间的相关性的多任务稀疏表示跟踪算法^[56] (Multi-task tracking, MTT) 利用粒子空间位置的相关性产生的稀疏表示的相似性, 对粒子的稀疏表示附加联合稀疏性约束, 即在每帧中尽可能使用较少的几个相同原子来表示所有粒子, 联合稀疏性可以看作是全局的结构正则化, 能够同时作用于所有粒子. n 个候选目标 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$ 可由字典 D 线性表示:

$$X = [T, I] \begin{bmatrix} Z \\ E \end{bmatrix} = DC \quad (9)$$

其中, $Z \in \mathbf{R}^{m \times n}$ 为候选目标的稀疏表示中对应目标模板的系数, $E \in \mathbf{R}^{d \times n}$ 为对应微模板的系数, $C = [Z^T, E^T]^T \in \mathbf{R}^{(m+d) \times n}$ 为候选目标在当前字典下的稀疏表示.

Zhang 等^[57] 在多任务跟踪算法的基础上, 考虑粒子之间的结构相关性和表示的空间平滑性, 将候选目标之间的空间位置关系引入到目标函数中, 从而将多任务跟踪扩展到结构多任务跟

踪^[57] (Structured multi-task tracking, S-MTT). 假定稀疏表示 C 是通过成对的相互作用而相关. 利用这些局部结构先验对粒子的表示附加空间平滑性, 即同一帧中空间位置比较近的粒子应该具有相似的稀疏表示. 定义对称权重矩阵 $W \in \mathbf{R}_+^{n \times n}$, $W_{ii'} = \exp(-\|\mathbf{l}_i - \mathbf{l}_{i'}\|_2^2 / (2\varsigma^2))$ 描述粒子 i 和 i' 的稀疏表示 \mathbf{c}_i 和 $\mathbf{c}_{i'}$ 之间的相似性 (\mathbf{c}_i 和 $\mathbf{c}_{i'}$ 分别为矩阵 C 的第 i 列和第 i' 列), 其中, \mathbf{l}_i 和 $\mathbf{l}_{i'}$ 为粒子的中心坐标, $i = 1, 2, \dots, n$, $i' = 1, 2, \dots, n$, ς 为所有粒子之间距离的平均值. 记 $a_i = \sum_{i'=1}^n W_{ii'}$, $A = \text{diag}\{a_1, a_2, \dots, a_n\}$ 为图的度, $L = A - W$ 为图的拉普拉斯, W 表示图中所有边的权重. 规范的图平滑正则项为规范的稀疏表示中粒子间距离的加权和, 每个距离项的权重反映了粒子之间相关性的强弱. 图正则项为 $\frac{1}{2} \sum_{i=1}^n \sum_{i'=1}^n W_{ii'} \|\mathbf{c}_i / \sqrt{a_i} - \mathbf{c}_{i'} / \sqrt{a_{i'}}\|_2^2 = \text{tr}(C\hat{L}C^T)$, 其中, $\hat{L} = A^{-1/2}LA^{-1/2}$ 为规范化的拉普拉斯矩阵. 所以基于结构的粒子稀疏表示可以通过直接附加图正则项获得:

$$C = \underset{C}{\text{argmin}} \frac{1}{2} \|X - DC\|_2^2 + \frac{\lambda_1}{2} \text{tr}(C\hat{L}C^T) + \lambda_2 \|C\|_{p,q} \quad (10)$$

其中, $q = 1, p \in \{1, 2, \infty\}$. λ_1 和 λ_2 分别用于平衡局部结构正则项和全局结构正则项. 当 $\lambda_1 = 0$ 时为多任务跟踪算法, 当 $\lambda_1 = 0$ 且 $p = 1$ 时为 L_1 跟踪算法. S-MTT 的目标函数 (10) 由凸二次项和非平滑正则项组成, 因此一般使用 APG 算法^[58] 求解. 由于 $d \gg m$, S-MTT 和 MTT 算法的时间复杂度为 $O(mnd\epsilon^{-1/2})$, 迭代次数为 $O(\epsilon^{-1/2})$.

Hong 等^[31] 在多任务跟踪的基础上提出了多任务多视角联合稀疏表示跟踪算法. 假定 n 个粒子每个由 O 个不同的特征表示, 对于第 o ($o = 1, 2, \dots, O$) 个特征, $X^{(o)} \in \mathbf{R}^{d_o \times n}$ 为 n 个特征向量组成的特征矩阵. 第 o 个特征字典 $D^{(o)} = [T^{(o)}, I_{d_o}]$ 由 m 个目标特征 $T^{(o)} \in \mathbf{R}^{d_o \times m}$ 和对应微模板 I_{d_o} 组成. 通过多任务学习来获得 n 个粒子的 O 个特征矩阵 $\{X^{(1)}, X^{(2)}, \dots, X^{(O)}\}$ 的稀疏表示 $\{C^{(1)}, C^{(2)}, \dots, C^{(O)}\}$. $C^{(o)}$ 使得粒子在不同特征下具有不同的表示, 每个特征下的稀疏表示的同一列为同一个样本, 因此同一个样本在每个特征下的稀疏表示具有一定的相似性, 所以能够利用每个特征的独立性并且获得不同的统计特性. 因此所有特征的稀疏表示可以水平方向堆叠起来分别构成矩阵 P 和 Q , 每一个由所有特征下的稀疏表示系数组成. 对于 P 的行组使用组 Lasso 约束来获得所有候选目标在所有特征上的共有特征, 同样的组

Lasso 约束作用于 Q 的列组来同时确定异常样本. 所以多任务多视角联合稀疏表示的目标函数为

$$[C, P, Q] = \underset{C, P, Q}{\text{arg min}} \frac{1}{2} \sum_{o=1}^O \|D^{(o)}C^{(o)} - X^{(o)}\|_2^2 + \lambda_1 \|P\|_{1,2} + \lambda_2 \|Q^T\|_{1,2} \quad (11)$$

其中, $C^{(o)} = P^{(o)} + Q^{(o)}$, $P = [P^{(1)}, P^{(2)}, \dots, P^{(O)}]$, $Q = [Q^{(1)}, Q^{(2)}, \dots, Q^{(O)}]$.

多任务跟踪算法利用粒子之间的自相似性一定程度上提高了跟踪性能, 但是当候选目标采样区域较大时粒子之间的差异较大, 通过多任务学习强制所有粒子共享同样的结构会降低跟踪算法性能^[59]. 即 MTT 跟踪算法中利用粒子之间的相关性时没有考虑粒子之间的差异性. 所以鲁棒多任务跟踪算法^[59] 在求解稀疏表示时将稀疏表示系数矩阵分解成两部分, 考虑粒子的相似性对其中的一部分强加联合稀疏正则项, 考虑粒子之间的差异性对另外一部分按照元素附加稀疏正则项. 鲁棒多任务稀疏表示模型为

$$[P, Q] = \underset{P, Q}{\text{arg min}} \frac{1}{2} \|X - D(P + Q)\|_2^2 + \lambda_1 \|P\|_{1,2} + \lambda_2 \|Q\|_{1,1} \quad (12)$$

其中, 候选目标的稀疏表示为 $C = P + Q$, 其中, $\|P\|_{p,q} = \sum_i \|P_i\|_q^p$, P_i 为矩阵 P 的第 i 行. 对 P 附加联合稀疏正则项, 对应共享的结构. 对 Q 按照元素附加稀疏正则项, 对应非共享的特征. 联合稀疏性利用了粒子的相似性而按照元素附加的稀疏约束考虑了粒子之间的差异性. 当选择合适的 λ_1 使得 $P = 0$ 时, 算法为 L_1 跟踪算法, 当选择合适的 λ_2 使得 $Q = 0$, 算法为 MTT 跟踪算法. 对于 P 附加组稀疏正则项, 反映了粒子之间共享的共有结构, Q 按照元素附加稀疏正则项反映了粒子之间的差异性, 使得算法比 L_1 跟踪算法和 MTT 算法更加鲁棒. 上述模型可以基于加速梯度算法进行求解.

为了提高多任务稀疏表示跟踪算法的求解速度, 文献 [60] 提出了基于 l_0 范数的多任务梯度最小化跟踪算法. 实验结果表明随着字典中原子数量的增加, l_1 和 l_0 约束的重构误差基本一致, 但是 l_1 约束下的稀疏表示的求解时间增加很快, l_0 约束下的稀疏表示的求解时间增加较慢. 随着稀疏性正则项约束参数的变大, l_1 约束下的重构误差增加较快, l_0 约束下的重构误差先下降后保持不变, 耗时方面 l_1 约束下的求解时间先下降后增加, l_0 约束下的先增加而后基本保持不变, 但是 l_0 约束下的耗时始终低于 l_1 约束下的耗时.

由于稀疏表示本质上是线性模型, 稀疏表示反

映了候选目标和字典中原子之间的线性相关性, 所以可以根据稀疏表示系数的大小即候选目标和字典中原子相关性的确定目标位置. Jia 等提出了基于结构局部稀疏观模型的跟踪算法^[39-40], 这里的结构实质上指将相同空间位置上的图像块的稀疏表示作为候选目标的特征, 利用稀疏表示构建观测模型. 由于每个图像块的位置固定, 因此所有的局部图像块联合起来能够表示目标完整的结构信息.

文献 [30] 利用跟踪中运动平滑性建立下一帧相同位置图像块与当前帧对应位置图像块的相关性. 候选目标图像块特征与目标对应位置图像块特征之间的相关性使用部分置换矩阵 \mathcal{P}^k 表示^[30]:

$$\mathcal{P}^k = \{P^k | P^k \in \{0, 1\}^{n_k \times n}, 1_{n_k}^T P^k = 1_n^T, P^k 1_n \leq 1_{n_k}, A^k P^k 1_n = 1_n\} \quad (13)$$

其中, n_k 为目标模板的数量, n 为候选目标的数量, $k = 1, 2, \dots, K$ 为单个目标中提取的图像块个数. 式 (13) 三个约束项分别对应每个目标图像块对应候选样本中的一个图像块、每个样本图像块最多对应目标图像中的一个图像块和运动平滑性. $\{P^k\}_{k=1}^K$ 为上式优化后的部分置换矩阵, 基于提取的特征 F , 则有 $D_i = [F^1 \mathbf{p}_i^1, \dots, F^K \mathbf{p}_i^K] \in \mathbf{R}^{d \times K}, i = 1, 2, \dots, n$, 其中, \mathbf{p}_i^k 是矩阵 P^k 的第 i 列, 将第 i 个目标模板和视频中对应图像块的特征堆叠起来构成低秩矩阵 D_i , 理想情况下秩为 1, 考虑到噪声和遮挡优化部分置换矩阵问题可以转换为秩最小化问题:

$$\begin{aligned} & \{P^k, L_i, E_i\} = \\ & \arg \min_{\{P^k \in \mathcal{P}^k\}_{k=1}^K, \{L_i, E_i\}_{i=1}^n} \sum_i \text{rank} \|L_i\| + \lambda \|E_i\|_0 \\ & \text{s. t. } D_i = L_i + E_i \end{aligned} \quad (14)$$

上述模型中将秩使用矩阵的核范数代替, 矩阵的零范数通过一范数代替后可以通过快速一阶交替方向乘数法 (Alternative direction method of multiplier, ADMM) 求解.

Liu 等将基于位置约束的稀疏表示和 Mean-shift 结合提出了基于稀疏编码直方图的生成式算法^[61-62]. 将候选目标中所有的图像块在字典上的稀疏表示按照与目标中心的距离使用核函数加权求和, 归一化后作为候选目标在字典上的稀疏编码直方图. 将稀疏编码直方图与所有图像块的重构误差进行乘性结合作为目标匹配的相似性度量函数, 利用了目标与字典之间的相似性和目标与字典基分布之间的相似性, 最后采用 Mean-shift 迭代得到目标的中心位置和尺度. 相比粒子滤波框架下的稀疏表示跟踪算法, 将稀疏表示和均值漂移结合后经过几次迭代

就可以获得目标的跟踪位置, 大大减少了稀疏表示的求解次数.

2.2 判别式模型

稀疏表示的过程可以看作是二次特征提取的过程, 因此稀疏表示本身可以作为目标的特征, 并且稀疏表示的特征更加线性可分. 判别式模型根据目标匹配相似性度量函数的不同可以分为基于分类器响应的判别式算法^[29, 63]、基于稀疏表示差值的判别式算法^[44] 和多任务稀疏表示差值的判别式算法^[64-65].

Wang 等将稀疏表示和线性分类器结合提出了基于稀疏表示的判别式跟踪算法^[63], 使用 Logistic 回归来学习分类器, 将分类器响应作为目标匹配的相似性度量. \mathbf{y} 为从候选目标中提取的图像块的 SIFT 特征, 基于弹性网^[66] 的稀疏表示模型为

$$\mathbf{b} = \arg \min_{\mathbf{b}} \frac{1}{2} \|\mathbf{y} - D\mathbf{b}\|_2^2 + \lambda_1 \|\mathbf{b}\|_1 + \frac{\lambda_2}{2} \|\mathbf{b}\|_2^2 \quad (15)$$

其中, $\lambda_2 > 0$ 保证了上述优化问题是严格凸的. 弹性网模型适用于字典中原子个数远大于原子维数的情况^[67]. Lasso 模型利用的原子个数最多等于原子维数, 因此当字典中原子个数远大于原子维数时使用 Lasso 模型是不合理的, 而弹性网模型使得每个原子都可能被利用甚至所有的原子都能利用到. Lasso 模型的正则项是凸的, 而弹性网模型的正则项是严格凸的, 能够诱导出组效应^[68].

Zhuang 等^[44] 利用候选目标构建字典, 求解目标模板和背景模板在字典下的稀疏表示, 将样本基于目标的稀疏表示和基于背景的稀疏表示做差值构建观测模型. n 个候选目标为 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$, 模板集合 $T = [T_{\text{pos}}, T_{\text{neg}}] \in \mathbf{R}^{d \times (m+m')}$ 由目标模板 $T_{\text{pos}} \in \mathbf{R}^{d \times m}$ 和背景模板 $T_{\text{neg}} \in \mathbf{R}^{d \times m'}$ 组成. $\mathbf{c}_i = [\mathbf{c}_{i1}, \mathbf{c}_{i2}, \dots, \mathbf{c}_{in}]^T$ 为 T 中第 i 个模板基于 X 的稀疏表示, $i = 1, 2, \dots, (m+m')$. 由模板的重构系数向量构成矩阵 $C = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{m+m'}]$, 表示模板和候选目标之间的相似性. 为了保持相似候选目标之间稀疏表示的相似性, 引入 Laplacian 约束得到 Laplacian 多任务逆稀疏表示:

$$\begin{aligned} C = \arg \min_C & \|T - XC\|_2^2 + \lambda_1 \sum_{i=1}^{m+m'} \|\mathbf{c}_i\|_1 + \\ & \lambda_2 \text{tr}(CLC^T) \\ \text{s. t. } & \mathbf{c}_i \geq 0 \end{aligned} \quad (16)$$

其中, $L = A - B$ 为 Laplacian 矩阵. B 是二值矩阵表示两个候选目标特征之间的相关性, 如果 \mathbf{c}_i 位于 $\mathbf{c}_{i'}$ 的 n_0 个最近邻中 $B_{ii'} = 1$, 否则 $B_{ii'} = 0$. \mathbf{c}_i 的度定义为 $a_i = \sum_{i'=1}^{m+m'} B_{ii'}$,

$$A = \text{diag}\{a_1, a_2, \dots, a_{m+m'}\}.$$

Zhang 等^[65] 基于时间上的连贯性而产生的目标稀疏表示之间的相似性, 将目标跟踪看作是连贯、稀疏和低秩问题. 将候选目标 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ 表示成字典 D 的线性组合 $X = DZ$, 其中, $Z = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n]$ 为对应候选目标在字典上的稀疏表示. 字典 $D = [T_{\text{pos}}, T_{\text{neg}}]$ 由目标模板 T_{pos} 和背景模板 T_{neg} 组成. 由于目标的表现模型在短时间内不会发生较大变化, 因此利用时间连贯性跟踪问题可以表示为

$$\begin{aligned} [Z, E] = \arg \min_{Z, E} & \lambda_1 \|Z\|_* + \lambda_2 \|Z\|_{1,1} + \\ & \lambda_3 \|Z - Z_0\|_{2,1} + \lambda_4 \|E\|_{1,1} \\ \text{s. t. } & X = DZ + E \end{aligned} \quad (17)$$

其中, $\|Z\|_*$ 表示矩阵的核范数, Z_0 的每列均为之前跟踪结果的稀疏表示. $\|Z\|_{1,1}$ 对于遮挡和噪声具有较好的鲁棒性. $\|Z - Z_0\|_{2,1}$ 产生 $(Z - Z_0)$ 列水平上的稀疏性, 使得大多数候选目标的稀疏表示与之前跟踪结果的稀疏表示相似, 同时允许少量候选目标的稀疏表示与之前跟踪结果的稀疏表示不同. $\|E\|_{1,1}$ 确保模型对于稀疏重构误差的鲁棒性, E 的值和列的支撑是包含有信息的, 当 E 的值较大但为稀疏的列支撑时表示候选目标中存在遮挡, 当 E 的值较大但为非稀疏的列支撑时, 表示候选目标中包含较多背景信息.

在式 (17) 中当 $\lambda_1 = \lambda_3 = 0$ 时, 算法为稀疏跟踪算法 (Sparse tracker, ST), 与 L_1 跟踪算法相似. 当 $\lambda_2 = \lambda_3 = 0$ 时, 算法为低秩跟踪算法 (Low rank tracker, LRT), 低秩性利用了候选目标之间的相关性. 当 $\lambda_1 \lambda_2 \neq 0$ 且 $\lambda_3 = 0$ 时, 算法为低秩稀疏跟踪算法^[64] (Low rank sparse tracker, LRST), LRST 算法同时利用了候选目标表示的稀疏性和低秩性. 当 $\lambda_1 \lambda_2 \lambda_3 \neq 0$ 时, 算法为连贯低秩稀疏跟踪算法 (Consistent low rank sparse tracker, CLRST), CLRST 算法将 LRST 算法使用时间连贯性进行了推广. 引入等式约束和松弛变量将问题转换后可以将式 (17) 通过近似扩展拉格朗日乘数法求解.

多任务跟踪算法 MTT 和连贯低秩稀疏跟踪算法 CLRST 都利用了候选目标之间的结构信息, 但是这种结构的假设是不同的. MTT 中通过使用 $\|Z\|_{2,1}$ 约束使得候选目标由几个相同的原子表示, 使得 Z 的所有列彼此相似, 间接地使表示矩阵 Z 的秩为 1. CLRST 算法将跟踪置于低秩学习框架中约束目标的表示位于低维子空间内, 而不要求候选目标使用同样的原子表示, CLRST 算法假定表示矩阵 Z 的秩较低 (为 1 或大于 1). MTT 算法利用的是候选目标之间的空间相关性产生的稀疏表示的相似性,

而 CLRST 算法则是利用时间上的相关性产生的稀疏表示的相似性.

2.3 混合式模型

Zhong 等^[41-42] 将基于重构误差的判别式模型和基于稀疏编码直方图的生成式模型进行乘性结合提出了基于稀疏表示的混合式跟踪算法. 候选目标基于目标模板的稀疏重构误差为 ϵ_{pos} , 基于背景模板的稀疏重构误差为 ϵ_{neg} , 构建的判别式模型的置信度为 $H = \exp(-(\epsilon_{\text{pos}} - \epsilon_{\text{neg}})/\sigma)$. 基于稀疏性的生成式模型为候选目标中所有图像块基于目标字典的稀疏表示按照空间位置依次连接起来构成的稀疏编码直方图, 并利用重构误差阈值将被遮挡图像块的稀疏表示置 0, 去除遮挡后的稀疏编码直方图为 ρ , 使用直方图交叉函数来计算候选目标 ρ 和模板 ρ_0 之间的相似性. 基于稀疏性的混合式模型为

$$p = HF = \exp\left(-\frac{(\epsilon_{\text{pos}} - \epsilon_{\text{neg}})}{\sigma}\right) \times \left(\sum_j \min(\rho^j, \rho_0^j)\right) \quad (18)$$

其中, H 给予正样本 ($\epsilon_{\text{pos}} < \epsilon_{\text{neg}}$) 较高的权重, 因此 H 可以看作是 F 的权重. 对于较难区分的目标 ($\epsilon_{\text{pos}} \approx \epsilon_{\text{neg}}$), $H \approx 1$, 此时 F 起主要作用, 因此在混合式模型中生成式模型起着更为重要的作用. 实验结果表明^[41-42] 大多数情况下混合式模型优于单个的判别式模型和生成式模型. 主要原因是判别式模型主要用来区分目标和背景, 不能够有效处理遮挡, 而基于局部特征的生成式模型能够有效处理遮挡, 而单独的生成式模型不能够有效处理复杂背景. 混合式模型集成了二者的优点使得算法的鲁棒性较好.

3 模型更新

基于稀疏表示的跟踪算法按照算法模型的不同更新的方案也不同. 生成式模型中字典的组成不同, 字典的在线更新方案也不尽相同. 对于由目标模板构建的字典, L_1 跟踪算法是通过计算跟踪结果与字典中原子的相似度通过设定阈值来对原子进行更新^[24]. 由于跟踪结果的正确与否不得而知, 因此一旦将错误的跟踪结果更新到字典中, 会导致模型退化. Zhang 等^[32] 将字典中的模板分为固定模板、稳定模板和变化模板. 其中固定模板为第一帧中在人工标定的基础上提取的目标模板, 固定模板在跟踪过程中保持不变. 固定模板和稳定模板主要用来阻止漂移, 稳定模板用来抓住跟踪过程中目标的稳定特征, 变化模板用来对目标的变化作出响应, 三种模板以不同的方式进行更新. 对于由 PCA 构建的字典, 则可以使用增量子空间学习算法实现 PCA 的

在线更新^[69], 也可将其应用于字典中模板的在线更新^[39-40]. 文献 [70] 将一阶马尔科夫链用于字典中原子的更新.

判别式模型的更新主要是在跟踪结果的基础上提取正负样本对分类器进行在线更新. 由于跟踪结果的正确与否是不可知的, 因此在跟踪结果的基础上采集样本的真实属性也是未知的, 所以分类器的更新过程更多的是半监督训练和非监督训练过程. Wang 等^[29] 将候选目标在初始分类器和新分类器上的响应加权作为候选目标的分类器响应, 由于初始分类器是通过监督训练得到的, 可以一定程度上减轻分类器更新中的漂移.

4 算法复杂度分析

基于稀疏表示的视频目标跟踪的算法复杂度较高^[21], 对于字典 $D \in \mathbf{R}^{m \times n}$, 粒子数量为 N , 基于 Lasso 的稀疏表示求解的算法时间复杂度为 $O(Nm^2n^{3/2})$, 因此一般通过以下三种方法提高算法速度: 1) 减少有效候选目标的数量; 2) 降低字典的维数; 3) 提高每次稀疏表示的求解速度.

减少有效候选目标的数量是为了减少稀疏表示的求解次数, 粒子滤波框架下传统算法的稀疏表示求解次数等于粒子的个数. 为了减少每次跟踪过程中 l_1 范数最小化的求解次数, Mei 等引入最小平方误差界, 在计算稀疏表示之前通过线性最小平方重构误差剔除大量的非重要粒子, 从而减少了稀疏表示的计算次数^[34]. 文献 [44-45] 利用候选目标构建字典, 将跟踪问题看作识别问题, 仅需要求解几个目标模板在由候选目标组成的字典上的稀疏表示, 大大减少了稀疏表示的计算次数.

由于稀疏表示求解的算法复杂度与字典的维数成正比, 因此可以通过降低字典维数来提高算法求解速度. 降低字典维数可以通过以下两种方法实现: 1) 降低原子的维数; 2) 减少字典中原子的个数. 文献 [32] 利用协方差矩阵在实现特征融合的同时降低了特征的维数. 由于目标模板中通常包括一些非线性的背景特征, Liu 等^[38] 通过样本标签分布学习, 利用稀疏模型建立正负样本和标签之间的映射关系, 通过判别式特征选择目标模板中包含的背景特征被剔除, 选择更具有判别力的特征实现特征降维. Li 等^[43] 利用满足约束等距性的测量矩阵同时作用于字典和候选目标的特征实现降维. 基于样本标签学习的特征选择方法降维和基于约束等距性的特征降维虽然形式一样, 都是将高维特征降维至低维空间, 但二者有本质的区别. 基于标签学习的特征选择降维方法是从高维特征空间中更具判别力的特征实现降维. 而基于约束等距性的特征降维是利用满足约束等距性的哈希矩阵将高维特征降维至低维空

间, 在降维过程中保持了原始高维数据的空间结构.

为了在跟踪鲁棒性和算法速度之间做出折中, 很多算法不得不采用低维数的字典, 而低维特征下难以对目标进行精确描述. 对于基于稀疏表示的判别式跟踪算法而言, 低维字典严重限制了稀疏表示的判别力, 降低了目标和背景的可分性. 文献 [71] 提出了基于乘积稀疏表示的支持向量跟踪算法, 将原始稀疏编码问题分解为两个较小字典上的稀疏编码问题, 不仅使得等效字典中原子个数大大增加, 使得目标能够获得更高维的稀疏表示, 而且降低了稀疏性求解过程的计算量, 使得目标和背景在高维的稀疏表示下更加线性可分, 提高了跟踪算法的鲁棒性.

L_1 跟踪算法字典中加入微模板是为了处理遮挡和噪声等异常, Yan 等^[50] 将字典中微模板去除, 从重构误差的概率分布出发提出了加权 Lasso 模型, 使算法模型适用于更为一般的噪声. 另外基于局部特征构建的字典本身对于遮挡具有一定的鲁棒性也不需要微模板. 同时利用目标和背景构建的字典中也不包括微模板.

提高每次 l_1 范数最小化的计算速度, Bao 等将加速最近梯度算法^[35] 引入到稀疏模型求解中, 加速最近梯度算法的有效性在于它的二次收敛性. Li 等^[43] 使用正交匹配追踪算法求解稀疏表示, 显然要比 L_1 跟踪算法中的内点法^[72] 速度要快得多.

5 实验结果与分析

5.1 实验数据分析

如表 1 所示, 自 2013 年以来出现了多个视频目标跟踪算法评估的基准数据库^[7, 22, 73-76]. 其中 VOT2013^[73]、OTB50^[7]、VOT2014^[77] 和 OTB100^[22] 为 GT (Groundtruth) 全标注的基准数据库. PTB 中为附带景深的 RGB-D 图像^[74], 受到景深探测器景深探测范围的限制, 无法获取大景深的视频, 公开的视频数据中只给出 5 个视频的 GT, 并且当目标被全遮挡时 GT 的标注为空, 需要在线提交跟踪结果进行评估. ALOV++ 中按照影响视频目标跟踪的因素将视频分成 13 类, GT 是每隔 5 帧标注一次^[75]. NUS-PRO 中将视频分成 5 类: 人脸视频、行人视频、运动员视频、刚体视频和长视频, 每类视频又分成不同的子类, 所有图像大小均为 1280×720 , 没有公开 GT, 需要在线提交跟踪结果进行评估^[76].

5.2 评估方法讨论

不同于目标检测、图像分类等计算机视觉领域, 目前为止视频目标跟踪还没有统一的评估标准^[78-80].

表 1 视频跟踪评估基准数据

Table 1 Summary of some visual tracking evaluation benchmark datasets

| 数据集 | 时间 | 视频数量 | 地址 |
|----------|------|------|---|
| VOT 2013 | 2013 | 16 | http://www.votchallenge.net/vot2013/ |
| OTB 50 | 2013 | 50 | http://visualtracking.net |
| PTB | 2013 | 100 | http://tracking.cs.prince-ton.edu/submit.php |
| ALOV++ | 2013 | 315 | http://www.alov300.org/ |
| VOT 2014 | 2014 | 25 | http://www.votchallenge.net/vot2014/ |
| OTB 100 | 2015 | 100 | http://pami.visual-tracking.net/ |
| NUS-PRO | 2015 | 365 | http://www.lv-nus.org/pro/nus_pro.html |

中心误差^[78] (Center error in pixel, CE) 定义为算法标定的目标中心和人工标定的目标中心之间的欧氏距离 (像素). 由于中心误差不能反映目标的尺度变化, 因此后续提出了标准化的中心误差, 即将中心误差除以目标的大小. 虽然标准化中心误差一定程度上能够反映目标的尺度变化, 但是中心误差的大小会随着目标大小成比例的变化, 并且跟踪失败后中心误差可能是随机产生的任意值, 并不能反映算法的真实性能.

基于 PASCAL VOC 中目标检测的评估标准^[81], 算法重叠率 (Overlap rate, OR) 能够同时较好地反应算法的跟踪位置和目標尺度, 且重叠率的值有界. 基于重叠率阈值的跟踪成功率 (Success rate, SR) 定义为重叠率大于阈值的比例.

由于不同的跟踪任务对于跟踪精度的要求不同, Wu 等^[7] 提出了精度曲线和成功曲线. 精度曲线^[7] (Precision plot) 是指算法中心误差小于中心误差阈值的比例随中心误差阈值的变化情况, 通常选择中心误差阈值为 20 像素时的值作为算法的跟踪精度. 成功曲线^[7] (Success plot) 是指跟踪成功率随重叠率阈值的变化情况. 通常选择重叠率阈值^[81] 为 0.5, 但是由于跟踪精度的要求不同选择重叠率阈值为 0.5 不具有代表性, Wu 等^[7] 利用成功曲线下的面积 (Success rate area under curve, SR-AUC) 作为算法评估的依据. 当重叠率阈值选择的足够多时, 算法的 SR-AUC 值等于在所有实验视频上重叠率的均值^[78].

传统的评估方法是对算法在整个视频上运行一次的跟踪结果进行评估, 这里称之为单次通过评估 OPE (One-pass evaluation). 但有些算法对于初始化 (初始位置和初始帧) 非常敏感, 不同的初始化导致跟踪结果的差别很大, 并且很多算法跟踪失败后没有重新初始化, 导致失败后的跟踪结果往往是随机的, 没有太大的参考价值. 所以文献 [7] 提出了算法初始化鲁棒性评估标准: 时间鲁棒性评估 (Temporal robustness evaluation, TRE) 和空间鲁棒性评估 (Spatial robustness evaluation, SRE).

时间鲁棒性评估是指从不同的初始帧开始将跟踪算法进行多次评估, 将多次评估的结果进行平均得到 TRE.

空间鲁棒性评估是通过移动初始帧 GT 的位置, 缩放 GT 窗口的大小来评估算法对初始化误差的敏感性. 移动 GT 的位置是将目标的中心位置向上、下、左、右、左上、右上、左下、右下分别移动对应维度的 10%, 其中向上下左右移动时, GT 的尺度不变; 向左上、右上、左下、右下移动时宽度和高度分别增加对应维度的 10%. 缩放 GT 窗口是指 GT 的中心位置不变, 将 GT 的宽度和高度分别变为原来宽度和高度的 80%、90%、110%、120%. 空间鲁棒性评估是上述 12 个评估的均值. 文献 [7] 的实验结果表明同一算法的平均 TRE 一般高于 OPE, 这主要是由于跟踪算法一般在较短的视频上的跟踪效果较好, 这也说明误差累积产生的漂移是导致跟踪失败的重要原因. 平均的 SRE 一般低于 OPE, 这主要是由于 SRE 中不精确的初始化导致后续跟踪漂移较快, 这也说明了目标精确初始化的重要性.

由于跟踪失败之后算法跟踪结果往往是随机产生的, 导致失败后的跟踪结果没有太大的参考价值, 并且没有人为干预时算法一般很难再跟踪上目标. 等价于算法的评估过程中仅使用了视频的一部分, 并没有充分利用整个视频数据. 并且一个视频中可能包含多个挑战因素, 算法可能对于视频中的某个因素是鲁棒的, 而对于另外的因素是不鲁棒的. 因此 VOT2014 中提出了新的鲁棒性评估标准^[77]. 在跟踪过程中当重叠率小于阈值时对目标进行重新初始化, 将重新初始化的次数作为算法鲁棒性的评估标准, 重新初始化的次数越少说明算法的鲁棒性越好, 反之说明算法的鲁棒性越差. 但是由于实际跟踪中导致跟踪失败的原因可能是多个因素共同作用的结果, 上述方法难以对单一挑战因素进行有效剥离.

处理速度是视频目标跟踪算法需要考虑的重要因素, 按照实时性的要求算法处理速度需要达到每秒 20 帧以上, 即单帧处理时间小于 50 ms. 由于视频图像的分辨率不同, 跟踪目标区域的大小不同, 以

及算法实现环境不同, 这些都会影响到算法的处理速度, 因此算法的实时性评估需要综合考虑多方面因素.

由上述分析可知, 尽管目前跟踪算法的评估标准较多, 但是核心的评估标准还是基于 GT 的中心误差和重叠率. 因此本文主要选取中心误差和重叠率的均值和标准差、跟踪成功曲线、跟踪精度曲线、跟踪成功率和算法单帧平均处理时间作为评估标准.

5.3 实验结果与分析

将现有公开代码的基于稀疏表示的跟踪算法总结如表 2 所示. 实验在 Intel (R) Core (TM) i7-3770CPU@3.40 GHz, 内存 16.0 GB 的 64 位计算机上通过 Matlab (R2014a) 软件实现. 测试视频为 OTB50 上的 50 个视频^[7], 共计 29 507 帧. 实验中算法均采用相同的初始位置, 算法中的参数均采用源代码中的默认参数. 算法中如果涉及粒子滤波框架下的仿射变换模型, 则采用相同的仿射变换标准差和粒子个数. 为了获得客观的比较结果, 将算法在测试视频上的 5 次实验结果的平均值作为最终的实

验结果. 需要指出的是由于实验中目标的初始位置、参数设置、粒子个数、实现环境和硬件平台等与相关文献中可能有所不同, 以及算法本身含有的某些随机因素, 某些实验结果和算法速度与相关文献中给出的结果有出入, 但是从大量实验结果的比较中得出的算法总体性能与相关文献一致.

上述基于稀疏表示的跟踪算法在 OTB50 上的跟踪精度^[7] 随中心误差阈值的变化曲线如图 4 所示, 其中右下角为对应算法在中心误差阈值为 20 像素时的跟踪精度, MDNet 为文献 [82] 提出算法, DLSSVM 为文献 [83] 提出算法, CFNet-conv5 为文献 [84] 提出算法.

上述基于稀疏表示的跟踪算法在 OTB50 上的跟踪成功率随重叠率阈值的变化曲线^[7] 如图 5 所示, 其中右上角为对应算法成功率曲线下的面积.

当跟踪成功率的重叠率阈值设为 0.5 时^[81], 上述算法在 OTB50 上的跟踪成功率如表 3 所示.

表 4 给出上述基于稀疏表示的视频目标跟踪算法在 OTB50 上的单帧平均处理时间, 其中实现环境中“M”表示使用 Matlab 编程实现, “MC”表示

表 2 算法简称和论文代码地址

Table 2 The sparse tracker's abbreviation, paper and code's URL

| 算法 | 论文题目和代码地址 |
|-------|---|
| L_1 | Robust visual tracking using l1 minimization. In ICCV, 2009. http://www.dabi.temple.edu/~hbling/publication-selected.htm |
| LSK | Robust tracking using local sparse appearance model and k-selection. In CVPR, 2011. http://www.uky.edu/~lya227/spt.html |
| L1APG | Real time robust l1 tracker using accelerated proximal gradient approach. In CVPR, 2012. http://www.dabi.temple.edu/~hbling/publication-selected.htm |
| ASLA | Visual tracking via adaptive structural local sparse appearance model. In CVPR, 2012. http://ice.dlut.edu.cn/lu/publications.html |
| SCM | Robust object tracking via sparsity-based collaborative model. In CVPR, 2012. http://ice.dlut.edu.cn/lu/publications.html |
| MTT | Robust visual tracking via multi-task sparse learning. In CVPR, 2012. http://faculty.ucmerced.edu/mhyang/pubs.html |
| LRT | Low-rank sparse learning for robust visual tracking. In ECCV, 2012. http://faculty.ucmerced.edu/mhyang/pubs.html |
| CT | Real-time compressive tracking. In ECCV, 2012. http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm |
| DLSR | Online discriminative object tracking with local sparse representation. In WACV, 2012. http://faculty.ucmerced.edu/mhyang/pubs.html |
| SRPCA | Online object tracking with sparse prototypes. In TIP, 2013. http://ice.dlut.edu.cn/lu/publications.html |
| DSSM | Visual tracking via discriminative sparse similarity map. In TIP, 2014. http://ice.dlut.edu.cn/lu/publications.html |
| SST | Structural sparse tracking. In CVPR, 2015. http://nlpr-web.ia.ac.cn/mmc/homepage/tzzhang/index.html |
| CST | In defense of sparse tracking: Circulant sparse tracker. In CVPR, 2016 http://nlpr-web.ia.ac.cn/mmc/homepage/tzzhang/index.html |

表 3 算法跟踪成功率 (%) 比较
Table 3 The tracker's success rate (%) comparison

| 算法 | ASLA | SCM | CST | SST | LRT | LSK | MTT | DSSM | CT | L1APG | L ₁ | DLSR | SRPCA |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------------|-------|-------|
| 成功率 | 70.69 | 68.96 | 68.20 | 59.30 | 59.02 | 56.09 | 54.98 | 45.80 | 42.29 | 41.61 | 35.56 | 34.22 | 25.83 |

表 4 算法单帧平均处理时间比较 (ms)
Table 4 The comparison of tracker's average processing time (ms)

| 算法 | ASLA | SCM | CST | SST | LRT | LSK | MTT | DSSM | CT | L1APG | L ₁ | DLSR | SRPCA |
|------|------|------|-----|-----|------|-----|------|------|----|-------|----------------|-------|-------|
| 实现环境 | MC | MC | M | M | M | ME | M | M | MC | MC | MC | MC | MC |
| 时间 | 241 | 7846 | 454 | 450 | 3152 | 382 | 2279 | 586 | 12 | 79 | 397 | 23030 | 249 |

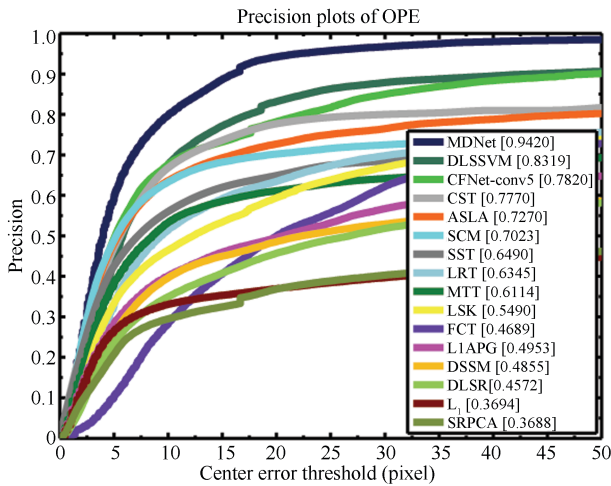


图 4 算法跟踪精度随中心误差阈值的曲线
Fig. 4 The tracker's tracking precision versus center error threshold

使用 Matlab 和 C/C++ 混合编程实现, “E” 表示使用可执行的二进制代码实现. 从表 4 中可以看出, 目前 Matlab 环境下大多数基于稀疏表示的视频目标跟踪算法还难以实现实时跟踪. 这主要是由于此类算法大多是在粒子滤波框架下进行, 粒子数量是直接影响算法速度的主要因素, 另外稀疏表示求解的算法时间复杂度较高也是影响算法速度的重要原因.

表 5 给出上述基于稀疏表示的跟踪算法的模型组成和其在 OTB50 上的重叠率的统计特征比较.

从上述实验结果和算法模型的分析比较可以看出, 特征字典上大多数算法直接使用灰度特征构建字典, 从算法重叠率的比较可以看出基于局部灰度特征构建的字典明显优于基于全局灰度构建的字典, 这主要是由于局部图像块对于遮挡和局部表观变化具有一定的鲁棒性, 所有图像块组合起来又能够表示目标完整的结构信息, 而全局灰度特征显然不具有上述优势.

运动模型上仿射运动能够精确描述目标的尺度变化和旋转, 所以大多数基于稀疏表示的跟踪算法

采用仿射运动模型. 搜索方案上粒子滤波和稠密采样能够有效避免陷入局部最优但计算量较大, 直接表现为算法的单帧处理时间较高, 均值漂移效率较高但是容易陷入局部最优, 导致跟踪性能下降.

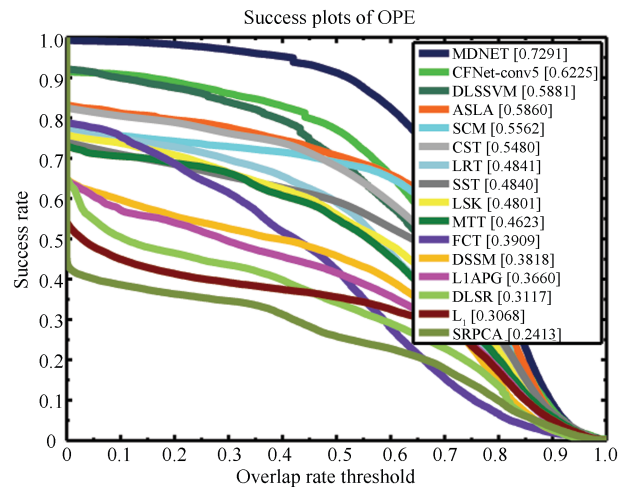


图 5 算法跟踪成功率随重叠率阈值的曲线
Fig. 5 The tracker's success rate versus overlap rate threshold

从重叠率的统计特征比较可以看出 ASLA 算法和 SCM 算法取得了较好的跟踪性能, 远高于其他基于稀疏表示的跟踪算法. 并且文献 [7] 的实验结果表明, 与其他非稀疏表示跟踪算法的横向比较中, ASLA 算法和 SCM 算法的跟踪性能也较为优异. 这主要是由于 ASLA 算法和 SCM 算法均采用了结构化的分块稀疏表示机制, 同时利用了目标的局部表观和空间结构信息. 因此在稀疏表示跟踪算法中分块稀疏表示和空间结构信息对于鲁棒视频目标跟踪是至关重要的.

6 结论与展望

伴随着稀疏表示理论不断发展, 基于稀疏表示的视频目标跟踪研究也取得了较大进展. 但是对

表 5 基于稀疏表示的视频跟踪算法模型和重叠率比较

Table 5 The comparison of the sparse representation-based tracker's model and overlap rate mean and std

| 算法简称 | 特征字典 | 运动模型 | 搜索方案 | 匹配模式 | 模型更新 | 重叠率均值 | 重叠率标准差 |
|----------------|----------|-------|------|------|----------|--------|--------|
| ASLA | 局部灰度 | 仿射运动 | 粒子滤波 | 生成式 | 增量学习 | 0.5860 | 0.3225 |
| SCM | 局部灰度 | 仿射运动 | 粒子滤波 | 混合式 | 模板替换 | 0.5562 | 0.3436 |
| CST | HOG | 仿射运动 | 粒子滤波 | 生成式 | 模板替换 | 0.5480 | 0.3063 |
| LRT | 全局灰度 | 仿射运动 | 粒子滤波 | 判别式 | 模板替换 | 0.4841 | 0.3247 |
| SST | 局部灰度 | 仿射运动 | 粒子滤波 | 生成式 | 模板替换 | 0.4840 | 0.3187 |
| LSK | 局部灰度 | 相似性变换 | 均值漂移 | 生成式 | 加权更新 | 0.4801 | 0.3379 |
| MTT | 全局灰度 | 仿射运动 | 粒子滤波 | 生成式 | 模板替换 | 0.4623 | 0.3411 |
| CT | 扩展类 haar | 平移运动 | 稠密采样 | 判别式 | Bayes 更新 | 0.3909 | 0.2850 |
| DSSM | 全局灰度 | 仿射运动 | 粒子滤波 | 判别式 | 模板替换 | 0.3818 | 0.3520 |
| L1APG | 全局灰度 | 仿射运动 | 粒子滤波 | 生成式 | 模板替换 | 0.3660 | 0.3565 |
| DLSR | 局部灰度 | 仿射运动 | 粒子滤波 | 判别式 | SVM 更新 | 0.3116 | 0.3414 |
| L ₁ | 全局灰度 | 仿射运动 | 粒子滤波 | 生成式 | 模板替换 | 0.3068 | 0.3687 |
| SRPCA | 全局 PCA | 仿射运动 | 粒子滤波 | 生成式 | 增量学习 | 0.2412 | 0.3321 |

于复杂环境下的持续鲁棒跟踪问题, 仍然存在一些亟待解决的问题. 在上述研究的基础上, 作者认为以下几个方面是值得继续研究的方向.

1) 上述稀疏表示模型中的正则项只考虑了信号的稀疏性, 没有考虑图像本身的特性^[85]. 图像作为二维信号, 具有很强的空间相关性, 表现为局部光滑特性. 因此针对图像的重构, 基于图像离散梯度稀疏性的最小全变分模型更适合二维图像重构, 且重构结果精确鲁棒. 所以最小全变分模型框架下的稀疏表示跟踪算法是重要的研究方向.

2) 由于基于稀疏表示的视频跟踪中字典和目标特征是不断变化的, 也就是构成子空间的支撑集和待重构信号是不断变化的, 因此基于稀疏表示的跟踪算法本质上是动态稀疏表示问题^[86], 也就是目标特征是具有动态特性的时变稀疏信号. 而上述模型中为了研究问题的方便均将其看做静态稀疏信号的重构问题, 所以动态稀疏表示下的跟踪算法也是值得研究的重要问题.

3) 由上述研究现状可以看出稀疏表示跟踪算法仍然难以实现实时跟踪, 除了算法优化外, 研究新的域下的稀疏表示的快速求解^[87-88] 是提高算法速度的关键, 也是面向实际应用的重要问题.

4) 目前稀疏表示跟踪算法的字典构建主要是人工设计完成, 基于学习的字典构建方法还比较少. 深度学习在视频目标跟踪中展现出的强大优势^[82, 84], 尤其是在对目标特征提取和表示方面, 因此基于深度学习的字典构建方案也是提高稀疏表示跟踪算法鲁棒性的重要方向.

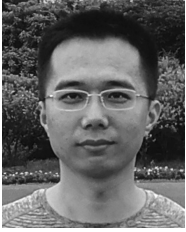
References

- Li X, Hu W M, Shen C H, Zhang Z F, Dick A, Van Den Hengel A. A survey of appearance models in visual object tracking. *ACM Transactions on Intelligent Systems and Technology*, 2013, 4(4): Article No.58
- Wang D. Research of online visual tracking algorithms based on linear representation models [Ph. D. dissertation], Dalian University of Technology, China, 2013.
(王栋. 基于线性表示模型的在线视觉跟踪算法研究 [博士论文]. 大连理工大学, 中国, 2013.)
- Haritaoglu I, Harwood D, Davis L S. W^4 : real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 809-830
- Rios-Cabrera R, Tuytelaars T, Van Gool L. Efficient multi-camera vehicle detection, tracking, and identification in a tunnel surveillance application. *Computer Vision and Image Understanding*, 2012, 116(6): 742-753
- Lu L P. Study on targets tracking algorithm and detection processing technology [Ph. D. dissertation], Nanjing University of Science and Technology, China, 2012.
(卢莉萍. 目标跟踪算法与检测处理技术研究 [博士论文]. 南京理工大学, 2012.)
- Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence. Vancouver, British Columbia, Canada: ACM, 1981. 674-679
- Wu Y, Lim J, Yang M H. Online object tracking: a benchmark. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, Oregon, USA: IEEE, 2013. 2411-2418
- Chu D M, Smeulders A W M. Thirteen hard cases in visual tracking. In: Proceedings of the 7th IEEE International

- Conference on Advanced Video and Signal Based Surveillance (AVSS). Sydney, Australia: IEEE, 2010. 103–110
- 9 Kuang J J. Study on image classification and target tracking based on sparse representation [Ph. D. dissertation], Chongqing University, China, 2013.
(匡金骏. 基于稀疏表示的图像分类与目标跟踪研究 [博士论文]. 重庆大学, 2013.)
 - 10 Lian Qiu-Sheng, Shi Bao-Shun, Chen Shu-Zhen. Research advances on dictionary learning models, algorithms and applications. *Acta Automatica Sinica*, 2015, **41**(2): 240–260
(练秋生, 石保顺, 陈书贞. 字典学习模型、算法及其应用研究进展. 自动化学报, 2015, **41**(2): 240–260)
 - 11 Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289–1306
 - 12 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, **381**(6583): 607–609
 - 13 Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, 1997, **37**(23): 3311–3325
 - 14 Wright J, Yang A Y, Ganesh A, Sastry S S, Ma Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(2): 210–227
 - 15 Oja E, Hyvärinen A, Hoyer P. Image feature extraction and denoising by sparse coding. *Pattern Analysis & Applications*, 1999, **2**(2): 104–110
 - 16 Yu N N, Qiu T S, Bi F, Wang A Q. Image features extraction and fusion based on joint sparse representation. *IEEE Journal of Selected Topics in Signal Processing*, 2011, **5**(5): 1074–1082
 - 17 Li X H, Lu H C, Zhang L H, Ruan X, Yang M H. Saliency detection via dense and sparse reconstruction. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 2976–2983
 - 18 Yuan X T, Liu X B, Yan S C. Visual classification with multitask joint sparse representation. *IEEE Transactions on Image Processing*, 2012, **21**(10): 4349–4360
 - 19 Wright J, Ma Y, Mairal J, Sapiro G, Huang T S, Yan S C. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 2010, **98**(6): 1031–1044
 - 20 Mei X, Ling H B. Robust visual tracking using l_1 minimization. In: Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV). Kyoto, Japan: IEEE, 2009. 1436–1443
 - 21 Zhang S P, Yao H X, Sun X, Lu X S. Sparse coding based visual tracking: review and experimental comparison. *Pattern Recognition*, 2013, **46**(7): 1772–1788
 - 22 Wu Y, Lim J, Yang M H. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1834–1848
 - 23 Shi Guang-Ming, Liu Dan-Hua, Gao Da-Hua, Liu Zhe, Lin Jie, Wang Liang-Jun. Advances in theory and application of compressed sensing. *Acta Electronica Sinica*, 2009, **37**(5): 1070–1081
(石光明, 刘丹华, 高大化, 刘哲, 林杰, 王良君. 压缩感知理论及其研究进展. 电子学报, 2009, **37**(5): 1070–1081)
 - 24 Mei X, Ling H B. Robust visual tracking and vehicle classification via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(11): 2259–2272
 - 25 Peng Y G, Ganesh A, Wright J, Xu W L, Ma Y. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(11): 2233–2246
 - 26 Wu Y, Blasch E, Chen G, Bai L, Ling H B. Multiple source data fusion via sparse representation for robust visual tracking. In: Proceedings of the 14th IEEE International Conference on Information Fusion (FUSION). Barcelona, Spain: IEEE, 2011. 1–8
 - 27 Han Z J, Jiao J B, Zhang B C, Ye Q X, Liu J Z. Visual object tracking via sample-based Adaptive Sparse Representation (AdaSR). *Pattern Recognition*, 2011, **44**(9): 2170–2183
 - 28 Tzimiropoulos G, Zafeiriou S, Pantic M. Sparse representations of image gradient orientations for visual recognition and tracking. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Colorado Springs, CO, USA: IEEE, 2011. 26–33
 - 29 Wang Q, Chen F, Yang J M, Xu W L, Yang M H. Transferring visual prior for online object tracking. *IEEE Transactions on Image Processing*, 2012, **21**(7): 3296–3305
 - 30 Zhang T Z, Jia K, Xu C S, Ma Y, Ahuja N. Partial occlusion handling for visual tracking via robust part matching. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, USA: IEEE, 2014. 1258–1265
 - 31 Hong Z B, Mei X, Prokhorov D, Tao D C. Tracking via robust multi-task multi-view joint sparse representation. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, NSW, Australia: IEEE, 2013. 649–656
 - 32 Zhang X Q, Li W, Hu W M, Ling H B, Maybank S. Block covariance based l_1 tracker with a subtle template dictionary. *Pattern Recognition*, 2013, **46**(7): 1750–1761
 - 33 Hu W M, Li W, Zhang X Q, Maybank S. Single and multiple object tracking using a multi-feature joint sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(4): 816–833
 - 34 Mei X, Ling H B, Wu Y, Blasch E P, Bai L. Efficient minimum error bounded particle resampling l_1 tracker with occlusion detection. *IEEE Transactions on Image Processing*, 2013, **22**(7): 2661–2675
 - 35 Bao C L, Wu Y, Ling H B, Ji H. Real time robust l_1 tracker using accelerated proximal gradient approach. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island, USA: IEEE, 2012. 1830–1837

- 36 Chen F, Wang Q, Wang S, Zhang W D, Xu W L. Object tracking via appearance modeling and sparse representation. *Image and Vision Computing*, 2011, **29**(11): 787–796
- 37 Bai T X, Li Y F. Robust visual tracking with structured sparse representation appearance model. *Pattern Recognition*, 2012, **45**(6): 2390–2404
- 38 Liu B Y, Yang L, Huang J Z, Meer P, Gong L G, Kulikowski C. Robust and fast collaborative tracking with two stage sparse optimization. In: Proceedings of the 11th European Conference on Computer Vision (ECCV). Crete, Greece: Springer, 2010. 624–637
- 39 Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island, USA: IEEE, 2012. 1822–1829
- 40 Jia X, Lu H C, Yang M H. Visual tracking via coarse and fine structural local sparse appearance models. *IEEE Transactions on Image Processing*, 2016, **25**(10): 4555–4564
- 41 Zhong W, Lu H C, Yang M H. Robust object tracking via sparsity-based collaborative model. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island, USA: IEEE, 2012. 1838–1845
- 42 Zhong W, Lu H C, Yang M H. Robust object tracking via sparse collaborative appearance model. *IEEE Transactions on Image Processing*, 2014, **23**(5): 2356–2368
- 43 Li H X, Shen C H, Shi Q F. Real-time visual tracking using compressive sensing. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 1305–1312
- 44 Zhuang B H, Lu H C, Xiao Z Y, Wang D. Visual tracking via discriminative sparse similarity map. *IEEE Transactions on Image Processing*, 2014, **23**(4): 1872–1881
- 45 Wang D, Lu H C, Bo C J. Online visual tracking via two view sparse representation. *IEEE Signal Processing Letters*, 2014, **21**(9): 1031–1034
- 46 Lee H, Battle A, Raina R, Ng A Y. Efficient sparse coding algorithms. In: Proceedings of the 19th International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: ACM, 2006. 801–808
- 47 Yang A Y, Sastry S S, Ganesh A, Ma Y. Fast ℓ_1 -minimization algorithms and an application in robust face recognition: a review. In: Proceedings of the 17th IEEE International Conference on Image Processing (ICIP). Hong Kong, China: IEEE, 2010. 1849–1852
- 48 Hu Zheng-Ping, Li Jing. Face recognition of joint sparse representation based on low-rank subspace recovery. *Acta Electronica Sinica*, 2013, **41**(5): 987–991
(胡正平, 李静. 基于低秩子空间恢复的联合稀疏表示人脸识别算法. 电子学报, 2013, **41**(5): 987–991)
- 49 Kim S J, Koh K, Lustig M, Boyd S, Gorinevsky D. An interior-point method for large-scale ℓ_1 -regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 2007, **1**(4): 606–617
- 50 Yan J C, Tong M L. Weighted sparse coding residual minimization for visual tracking. In: Proceedings of the 2011 IEEE International Conference on Visual Communications and Image Processing (VCIP). Tainan, China: IEEE, 2011. 1–4
- 51 Wang D, Lu H C, Yang M H. Online object tracking with sparse prototypes. *IEEE Transactions on Image Processing*, 2013, **22**(1): 314–325
- 52 Pan J S, Lim J W, Su Z X, Yang M H. L_0 -regularized object representation for visual tracking. In: Proceedings of the 2014 British Machine Vision Conference (BMVC). Nottingham, England: BMVA Press, 2014. 1–12
- 53 Shen Z W, Toh K C, Yun S. An accelerated proximal gradient algorithm for frame-based image restoration via the balanced approach. *SIAM Journal on Imaging Sciences*, 2011, **4**(2): 573–596
- 54 Lu X Q, Yuan Y, Yan P K. Robust visual tracking with discriminative sparse learning. *Pattern Recognition*, 2013, **46**(7): 1762–1771
- 55 Zhang T Z, Ghanem B, Xu C S, Ahuja N. Object tracking by occlusion detection via structured sparse learning. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Portland, Oregon, USA: IEEE, 2013. 1033–1040
- 56 Zhang T Z, Ghanem B, Liu S, Ahuja N. Robust visual tracking via multi-task sparse learning. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island, USA: IEEE, 2012. 2042–2049
- 57 Zhang T Z, Ghanem B, Liu S, Ahuja N. Robust visual tracking via structured multi-task sparse learning. *International Journal of Computer Vision*, 2013, **101**(2): 367–383
- 58 Chen X, Pan W K, Kwok J T, Carbonell J G. Accelerated gradient method for multi-task sparse learning problem. In: Proceedings of the 9th IEEE International Conference on Data Mining. Miami, FL, USA: IEEE, 2009. 746–751
- 59 Bai Y C, Tang M. Object tracking via robust multitask sparse representation. *IEEE Signal Processing Letters*, 2014, **21**(8): 909–913
- 60 Hu H W, Ma B, Jia Y D. Multi-task l_0 gradient minimization for visual tracking. *Neurocomputing*, 2015, **154**: 41–49
- 61 Liu B Y, Huang J Z, Yang L, Kulikowski C. Robust tracking using local sparse appearance model and k-selection. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 1313–1320
- 62 Liu B Y, Huang J Z, Kulikowski C, Yang L. Robust visual tracking using local sparse appearance model and K-selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(12): 2968–2981
- 63 Wang Q, Chen F, Xu W L, Yang M H. Online discriminative object tracking with local sparse representation. In: Proceedings of the 2012 IEEE Workshop on the Applications of Computer Vision (WACV). Breckenridge, Colorado, USA: IEEE, 2012. 425–432

- 64 Zhang T Z, Ghanem B, Liu S, Ahuja N. Low-rank sparse learning for robust visual tracking. In: Proceedings of the 12th European Conference on Computer Vision (ECCV). Firenze, Italy: Springer, 2012. 470–484
- 65 Zhang T Z, Liu S, Ahuja N, Yang M H, Ghanem B. Robust visual tracking via consistent low-rank sparse learning. *International Journal of Computer Vision*, 2015, **111**(2): 171–190
- 66 Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, **67**(2): 301–320
- 67 Xiang Kui, Li Bing-Nan. Sparsity in principal component analysis: a survey. *Acta Electronica Sinica*, 2012, **40**(12): 2525–2532
(向魁, 李炳南. 主元分析中的稀疏性. 电子学报, 2012, **40**(12): 2525–2532)
- 68 Fu G H. Model selection for analyzing high-dimensional, strongly correlated data [Ph.D. dissertation], Central South University, China, 2011.
(付光辉. 高维的强相关数据的模型选择 [博士论文]. 中南大学, 2011.)
- 69 Ross D A, Lim J, Lin R S, Yang M H. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 2008, **77**(1–3): 125–141
- 70 Zarezaade A, Rabiee H R, Soltani-Farani A, Khajenezhad A. Patchwise joint sparse tracking with occlusion detection. *IEEE Transactions on Image Processing*, 2014, **23**(10): 4496–4510
- 71 Huang H T, Bi D Y, Zha Y F, Ma S P, Gao S, Liu C. Robust visual tracking based on product sparse coding. *Pattern Recognition Letters*, 2015, **56**: 52–59
- 72 K Koh, S Kim, S Boyd. l1-ls: Simple Matlab Solver for l1-regularized Least Squares Problems [Online], available: <https://web.stanford.edu/~boyd/software.html>, October 8, 2018.
- 73 Kristan M, Pflugfelder R, Leonardis A, Matas J, Porikli F, Čehovin L, et al. The visual object tracking VOT2013 challenge results. In: Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops (ICCVW). Sydney, Australia: IEEE, 2013. 98–111
- 74 Song S R, Xiao J X. Tracking revisited using RGBD camera: unified benchmark and baselines. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 233–240
- 75 Smeulders A W M, Chu D M, Cucchiara R, Calderara S, Dehghan A, Shah M. Visual tracking: an experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(7): 1442–1468
- 76 Li A N, Lin M, Wu Y, Yang M H, Yan S C. NUS-PRO: a new visual tracking challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(2): 335–349
- 77 Kristan M, Pflugfelder R, Leonardis A, Matas J, Čehovin L, Nebelha G, Vojří T, Fernández G, Lukežič A, Dimitriev A, Petrosino A, Saffari A, Li B, Chan C S, Heng C, Ward D, Kearney D, Monekosso D, Karaimer H C, Rabiee H R, Zhu J K, Gao J, Xiao J J, Zhang J G, Xing J L, Huang K Q, Lebeda K, Cao L J, Maresca M E, Lim M K, El Helw M, Felsberg M, Remagnino P, Bowden R, Goecke R, Stolkin R, Lim S Y, Maher S, Poullot S, Wong S, Satoh S, Chen W H, Hu W M, Zhang X Q, Li Y, Niu Z H. The visual object tracking VOT2014 challenge results. In: Proceedings of the 13th European Conference on Computer Vision Workshops (ECCVW). Zurich, Switzerland: Springer, 2014. 191–217
- 78 Cehovin L, Kristan M, Leonardis A. Is my new tracker really better than yours? In: Proceedings of the 2014 IEEE Winter Conference on Applications of Computer Vision (WACV). Steamboat Springs, Colorado, USA: IEEE, 2014. 540–547
- 79 Pang Y, Ling H B. Finding the best from the second bests-inhibiting subjective bias in evaluation of visual tracking algorithms. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 2784–2791
- 80 SanMiguel J C, Cavallaro A, Martínez J M. Adaptive online performance evaluation of video trackers. *IEEE Transactions on Image Processing*, 2012, **21**(5): 2812–2823
- 81 Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, **88**(2): 303–338
- 82 Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA: IEEE, 2016. 4293–4302
- 83 Ning J F, Yang J M, Jiang S J, Zhang L, Yang M H. Object tracking via dual linear structured SVM and explicit feature map. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA: IEEE, 2016. 4266–4274
- 84 Valmadre J, Bertinetto L, Henriques J F, Vedaldi A, Torr P H S. End-to-end representation learning for correlation filter based tracking. arXiv: 1704.06036, 2017.
- 85 Zhang T Z, Liu S, Xu C S, Yan S C, Ghanem B, Ahuja N, Yang M H. Structural sparse tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, Massachusetts, USA: IEEE, 2015. 150–158
- 86 Jing Nan, Bi Wei-Hong, Hu Zheng-Ping, Wang Lin. A survey on dynamic compressed sensing. *Acta Automatica Sinica*, 2015, **41**(1): 22–37
(荆楠, 毕卫红, 胡正平, 王林. 动态压缩感知综述. 自动化学报, 2015, **41**(1): 22–37)
- 87 Zhang T Z, Bibi A, Ghanem B. In defense of sparse tracking: circulant sparse tracker. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA: IEEE, 2016. 3880–3888
- 88 Bibi A, Itani H, Ghanem B. FFTLasso: large-scale LASSO in the Fourier domain. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, Hawaii, USA: IEEE, 2017. 3820–3829



黄宏图 中国人民解放军 95972 部队工程师. 2016 年获空军工程大学信号与信息处理专业工学博士学位. 主要研究方向为视频目标跟踪. 本文通信作者.

E-mail: huanghongtu@sina.cn

(**HUANG Hong-Tu** Engineer at 95972 Troops of PLA. He received his Ph.D. degree of engineering in signal

and information processing major from Air Force Engineering University in 2016. His main research interest is visual object tracking. Corresponding author of this paper.)



毕笃彦 空军工程大学航空工程学院教授. 主要研究方向为图像处理 and 模式识别. E-mail: biduyan@126.com

(**BI Du-Yan** Professor at the Aeronautics Engineering College, Air Force Engineering University. His research interest covers image processing and pattern recognition.)



侯志强 空军工程大学信息与导航学院教授, 西安邮电大学计算机学院教授. 主要研究方向为模式识别, 计算机视觉, 图像处理, 信息融合.

E-mail: hou-zhq@sohu.com

(**HOU Zhi-Qiang** Professor at the Information and Navigation Institute, Air Force Engineering University and

School of Computer Science and Technology, Xi'an University of Posts and Telecommunications. His research interest covers pattern recognition, computer vision, image processing, and information fusion.)



胡长城 中国人民解放军 95972 部队工程师. 主要研究方向为数据处理.

E-mail: h_ccemail@163.com

(**HU Chang-Cheng** Engineer at 95972 Troops of PLA. His main research interest is data processing.)



高山 空军工程大学航空工程学院讲师. 主要研究方向为图像处理.

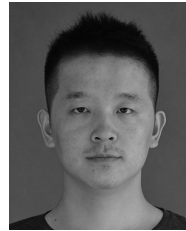
E-mail: gaoshan1114@126.com

(**GAO Shan** Lecturer at the Aeronautics Engineering College, Air Force Engineering University. Her main research interest is image processing.)



查宇飞 空军工程大学航空工程学院副教授. 主要研究方向为计算机视觉和机器学习. E-mail: zhayufei@126.com

(**ZHA Yu-Fei** Associate professor at the Aeronautics Engineering College, Air Force Engineering University. His research interest covers computer vision and machine learning.)



库涛 空军工程大学航空工程学院博士研究生. 主要研究方向为视频目标跟踪. E-mail: keltloja@163.com

(**KU Tao** Ph.D. candidate at the Aeronautics Engineering College, Air Force Engineering University. His main research interest is visual object tracking.)