第 58 卷 第 16 期/2021 年 8 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于卷积的稀疏跟踪算法

许奇[1]，韩俊波[1,2*]，黎海霞[2,3]

[1]巢湖学院信息工程学院，安徽 合肥 230031；

[2]南京航天航空大学计算机科学与技术学院/人工智能学院，江苏 南京 211106；

[3]浙江警官职业学院，浙江 杭州 310018

摘要 传统的稀疏表示旨在通过字典的线性结合构建跟踪目标的表观模型，忽视了目标的分层结构特征，因此难以处理复杂的跟踪环境。针对该问题，提出一种新颖的基于卷积的稀疏跟踪算法(CSTA)。在目标区域中提取局部图像块作为局部描述子，依据稀疏表示从中选取一组图像块作为固定卷积核与输入的图像进行卷积运算，能够有效保留跟踪目标的层次化结构特征；同时提出一种新的选择性在线模型更新机制，有效避免错误模型更新导致跟踪结果漂移的问题。所提 CSTA 在公开数据集中与现有稀疏表示算法进行定量、定性的分析比较，结果表明，CSTA 的准确度、鲁棒性均优于现有的稀疏跟踪算法。

关键词 机器视觉；目标跟踪；稀疏表示；表观模型；模型更新

中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP202158.1615004

Convolution-Based Sparse Tracking Algorithm

Xu Qi[1], Han Junbo[1,2*], Li Haixia[2,3]

[1] *College of Information Engineering, Chaohu University, Hefei, Anhui* 230031, *China*;

[2] *College of Computer Science and Technology/College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu* 211106, *China*;

[3] *Zhejiang Police Vocational Academy, Hangzhou, Zhejiang* 310018, *China*

Abstract Traditional sparse representation algorithms attempt to build a robust appearance model to track targets according to the linear combination of sparse dictionaries. However, such algorithms ignore the hierarchical structure features of the tracking object; thus, handling complex tracking scenery is difficult. In this paper, an innovative convolution-based sparse tracking algorithm (CSTA) is proposed to address this limitation. Local image patches extracted within the object region serve as local descriptors. According to the sparse representation theory, a group of sparse image blocks is selected as the fixed convolution kernel, and the results obtained by convoluting the convolution kernel with the input image demonstrate that the hierarchical structure of tracking objects has been preserved. In addition, a selective online updating mechanism is presented to avoid the drift problem caused by erroneous model updating. Quantitative and qualitative analyses are conducted, and the proposed CSTA and advanced sparse representation algorithms are compared using open datasets. The experimental results demonstrate that the proposed CSTA outperforms state-of-the-art sparse tracking algorithms in terms of accuracy and robustness.

Key words machine vision; object tracking; sparse representation; appearance model; model update

OCIS codes 150.0155; 150.1135; 100.2960

收稿日期：2020-10-14；**修回日期**：2020-11-29；**录用日期**：2020-12-08

基金项目：国家自然科学基金(61370075)、江苏省自然科学基金面上项目(BK20191274)、浙江省教育厅一般科研项目(Y202043143)、巢湖学院重点学科招标项目(ZDXK-201816)

通信作者：*243805091@qq.com

1 引 言

目标跟踪是计算机视觉中的一个重要的研究方向，在交通监管、人机交互、自动化控制等诸多领域有着广泛的应用前景。近年来，相关研究取得了突破性的进展，但是仍然存在许多的挑战性问题尚未解决。这些问题可以概括为跟踪目标自身的形状变化问题（如目标形变、平面旋转等）和外在干扰问题（如尺度变化、遮挡等）。

虽然跟踪算法的设计需要根据实际需求而定，但大多数跟踪任务都遵循一次成功原则（OPE）[1]，即系统只会标注跟踪目标在第一帧的位置状态，隐藏之后的视频序列信息，要求跟踪算法输出之后每一帧的中心位置和相应的跟踪框大小。影响跟踪结果的因素有很多，其中最为重要的是选取合适的特征和方法对跟踪目标的表观建模，即构建表观模型。近年来，基于稀疏表示的相关算法被广泛应用于目标跟踪中，核心思想是利用稀疏表示理论构建鲁棒的稀疏表观模型，在粒子滤波[2]跟踪框架下，寻求与目标模板最为相似的图像区域。现有的稀疏表观模型主要分为两种：全局表观模型[3-5]和局部表观模型[6-9]。Mei 等[3]提出的全局稀疏模型将目标所在的图像区域作为一个整体用以学习稀疏字典，并通过线性结合的方式构建表观模型，之后的跟踪过程可被视为求解 L1 最小化问题。实验结果表明：利用稀疏模板描述跟踪目标，可以有效地处理表观变化问题；但由于缺乏局部特征，对遮挡等外在干扰问题处理乏力。与全局稀疏表观模型不同，局部表观模型通过局部图像块构建稀疏表观模型。例如 Jia 等[6]提出一种基于局部稀疏的跟踪算法，基本思想是目标的整体结构可以由局部图像块的集合构成；在稀疏约束下，利用局部图像块构建稀疏字典，再利用线性结合还原目标的整体结构，最后通过重构误差找到最佳候选粒子。相比于全局稀疏，局部稀疏模型更加注重对目标局部特征的提取，因此能够更好地处理一些外在干扰问题。即便如此，局部稀疏模型仍然依赖于字典的线性结合。文献[10]指出基于线性结合的稀疏表观模型难以开发出跟踪目标的分层结构特征，限制了稀疏跟踪算法的性能，尤其是难以应对尺度变化、重度遮挡等具有挑战性的难题。

为了解决传统稀疏表示难以提取跟踪目标分层结构特征的问题，本文提出一种新颖的基于卷积的稀疏表观模型。根据稀疏表示理论，提取局部图像块作为卷积核，使表观模型在保留了局部结构特征的基础上，能够分层地显示出目标的层次结构。概括地说，本文主要贡献如下：

1）提出一种全新的方法构建稀疏表观模型，与传统的线性结合不同，所提表观模型建立在基于卷积运算的基础上；

2）创新性地将稀疏表示理论与卷积运算结合，有效地提取了跟踪目标的分层结构特征，并扩展了稀疏表示理论在目标跟踪中的应用前景；

3）提出一种新颖的在线选择更新机制，避免了错误的模型更新问题；

4）结合粒子滤波框架，提出一种基于卷积的稀疏跟踪算法（CSTA），在 100 个视频序列中的实验结果表明，基于卷积的稀疏模型可以显著地提高稀疏跟踪算法的准确度和鲁棒性。

2 基于卷积的稀疏跟踪算法

2.1 跟踪框架

与传统的稀疏跟踪算法类似，所提算法依然采用粒子滤波作为跟踪框架。在贝叶斯滤波的基础上，粒子滤波通过估计最大后验概率预测跟踪目标在第 t 帧的位置状态 $\boldsymbol{x}_t=[x_t,y_t,s_t]$，其中 (x_t,y_t) 为跟踪框的中心位置，s_t 为尺度大小。

$$p(\boldsymbol{x}_t|\boldsymbol{z}_t)=p(\boldsymbol{z}_t|\boldsymbol{x}_t)\int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\times p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{t-1})\mathrm{d}\boldsymbol{x}_{t-1},\tag{1}$$

式中：$\boldsymbol{z}_t$ 为模型的观测值；$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ 为运动模型，用于提取一组候选图像区域，或称为“粒子”；$p(\boldsymbol{z}_t|\boldsymbol{x}_t)$ 为观测模型，用于选取最合适的粒子。为了方便计算，采用简易的布朗运动模拟粒子的运动方式，具体表现为高斯分布。

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})=N(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{V}),\tag{2}$$

式中：对角协方差矩阵 $\boldsymbol{V}=\mathrm{diag}(v_x,v_y,v_s)$，$v_x$，$v_y$，$v_s$ 分别对应目标状态 $\boldsymbol{x}_t=[x_t,y_t,s_t]$ 的标准方差。通过蒙特卡罗模拟和重要性采样[11]，粒子滤波选取一组候选粒子 $\{\boldsymbol{x}_t^i\}_{i=1}^N$ 估计目标的后验状态。简而言之，根据(2)式，在前一帧跟踪结果的基础上通过高斯分布随机选取 N 块候选图像区域 $\{\boldsymbol{x}_t^i\}_{i=1}^N$，之后最大后验概率可近似求解为最佳粒子 $\hat{\boldsymbol{x}}_t$ 所对应的图像区域。$\hat{\boldsymbol{x}}_t$ 的计算公式为

$$\hat{\boldsymbol{x}}_t=\arg\max_{\{\boldsymbol{x}^i\}_{i=1}^N} p(\boldsymbol{z}_t|\boldsymbol{x}_t^i)p(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1})。\tag{3}$$

观测模型 $p(\boldsymbol{z}_t|\boldsymbol{x}_t)$ 为一个存储的目标模板 $\boldsymbol{T}$ 与任一候选粒子 $\boldsymbol{x}_t^i$ 之间的相似度，表示为

$$p(\boldsymbol{z}_t \mid \boldsymbol{x}_t^i)=\exp(-\|\mathrm{Vec}(\boldsymbol{T})-\mathrm{Vec}(\boldsymbol{C}_t^i)\|_2), \quad (4)$$

式中：$\boldsymbol{C}_t^i$ 为在第 t 帧，从第 i 个粒子 $\boldsymbol{x}_t^i$ 所对应的图像区域中构建的基于卷积的稀疏表观模型。使得(4)式值最大化的候选粒子所对应的图像区域为算法输出的跟踪结果。

2.2 基于卷积的稀疏表观模型

2.2.1 学习稀疏字典

给定目标在第一帧的图像区域，先通过线性插值的方式将其尺寸固定为 $m\times m$ 像素，之后采用 $n\times n$ 大小的滑动窗口提取一组局部图像块作为局部特征集，表示为 $\boldsymbol{Y}=[\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_K]\in\mathbb{R}^{n\times n\times K}$，其中 K 为局部图像块的个数，共计 $K=(m-n+1)^2$。所有图像块都通过 l_2 正则化处理，以应对光照变化问题。

滑动窗口所提取的局部图像块可以尽可能地表示目标的局部结构，有利于提取结构特征。然而，过多的图像块存在大量冗余的部分，一方面增加了算法的运算量，另一方面也会携带一些包含噪声在内的无用特征。因此，从局部特征集 $\boldsymbol{Y}=[\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_K]$ 中提取一组过完备稀疏字典 $\boldsymbol{D}=[\boldsymbol{D}_1,\cdots,\boldsymbol{D}_W]$ 作为局部描述子。$\boldsymbol{D}$ 与 $\boldsymbol{Y}$ 的关系为

$$\boldsymbol{Y}_{L\times K}=\boldsymbol{D}_{L\times W}\boldsymbol{X}_{W\times K}, \quad (5)$$

式中：$L=n\times n$；$\boldsymbol{X}$ 为系数矩阵。而字典学习的目的是求解字典 $\boldsymbol{D}$ 使得 $\boldsymbol{X}$ 尽可能稀疏，即

$$\min_{\boldsymbol{D},\boldsymbol{X}}\|\boldsymbol{x}_u\|_0^2 \text{ s.t. } \|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\|_F^2\leqslant e, \quad (6)$$

式中：e 为预设的稀疏误差。(6)式是一个 Non-deterministic Polynomial(NP)问题。可以先假定字典 $\boldsymbol{D}$ 是固定的，再利用正交匹配追踪(OMP)算法求解次优解：

$$\min_{\boldsymbol{X}}\|\boldsymbol{y}_u-\boldsymbol{D}\boldsymbol{x}_u\|_0^2 \text{ s.t. } \|\boldsymbol{x}_u\|_0\leqslant T_0, \quad (7)$$

式中：T_0 为预设的稀疏度。当更新字典 $\boldsymbol{D}$ 时，$\boldsymbol{D}\boldsymbol{X}$ 可被写成

$$\boldsymbol{D}\boldsymbol{X}=\sum_{u=1}^{W}\boldsymbol{d}_u\boldsymbol{x}_u^{\mathrm{T}}, \quad (8)$$

式中：$\boldsymbol{d}_u$ 为矩阵 $\boldsymbol{D}$ 的第 u 列；$\boldsymbol{x}_u^{\mathrm{T}}$ 为矩阵 $\boldsymbol{X}$ 的第 u 行。通过 W 次迭代，可以依次更新字典的每一列。当更新字典 $\boldsymbol{D}$ 的第 k 列时，其他列的值保持不变，因此目标函数可被重写为

$$\min_{\boldsymbol{D}}\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\|_F^2=\min_{\boldsymbol{d}_k}\left\|\boldsymbol{Y}-\sum_{u=1}^{W}\boldsymbol{d}_u\boldsymbol{x}_u^{\mathrm{T}}\right\|_F^2=\min_{\boldsymbol{d}_k}\left\|\left(\boldsymbol{Y}-\sum_{u\neq k}\boldsymbol{d}_u\boldsymbol{x}_u^{\mathrm{T}}\right)-\boldsymbol{d}_k\boldsymbol{x}_k^{\mathrm{T}}\right\|_F^2=\min_{\boldsymbol{d}_k}\|\boldsymbol{E}_k-\boldsymbol{d}_k\boldsymbol{x}_k^{\mathrm{T}}\|_F^2, \quad (9)$$

式中：误差矩阵 $\boldsymbol{E}_k=\boldsymbol{Y}-\sum_{u\neq k}\boldsymbol{d}_u\boldsymbol{x}_u^{\mathrm{T}}$。由于 $\boldsymbol{E}_k$ 值是不变的，因而可以依据 $\boldsymbol{E}_k$ 值更新 $\boldsymbol{d}_k$ 和 $\boldsymbol{x}_k^{\mathrm{T}}$。首先对 $\boldsymbol{E}_k$ 进行奇异值分解：

$$\boldsymbol{E}_k=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}, \quad (10)$$

式中：$\boldsymbol{U}$ 和 $\boldsymbol{V}^{\mathrm{T}}$ 为一组正交基；$\boldsymbol{\Lambda}$ 为对角矩阵。理论上，可以直接选取与对角矩阵 $\boldsymbol{\Lambda}$ 中最大值对应的 $\boldsymbol{U}$ 和 $\boldsymbol{V}^{\mathrm{T}}$ 中的向量分别更新 $\boldsymbol{d}_k$ 和 $\boldsymbol{x}_k^{\mathrm{T}}$。然而，此举可能会改变未更新前 $\boldsymbol{x}_k^{\mathrm{T}}$ 中非零元素的值和位置，进而降低 $\boldsymbol{X}$ 的稀疏性。为了处理这种情况，可以先提取 $\boldsymbol{x}_k^{\mathrm{T}}$ 中所有非零元素，用以构建一个新的稀疏矩阵 $\boldsymbol{E}_0$，使之满足

$$\hat{\boldsymbol{E}}_k=\boldsymbol{E}_k\boldsymbol{E}_0, \quad (11)$$

再对 $\hat{\boldsymbol{E}}_k$ 进行奇异值分解，即可选取合适的向量更新字典。

2.2.2 基于卷积的表观模型

得到稀疏字典 $\boldsymbol{D}=[\boldsymbol{D}_1,\cdots,\boldsymbol{D}_W]$ 后，对于一个给定的候选图像区域，通过线性插值的方式将其尺寸固定为 $m\times m$ 像素，表示为 $\boldsymbol{I}\in\mathbb{R}^{m\times m}$。基于卷积的稀疏表观模型定义为 $\boldsymbol{C}=[\boldsymbol{C}_1,\cdots,\boldsymbol{C}_W]$，其中

$$\boldsymbol{C}_j=\boldsymbol{D}_j\otimes\boldsymbol{I},\ j=1,\cdots,W。 \quad (12)$$

与传统的稀疏表示所采用的线性结合不同，利用稀疏滤波器和原图像的卷积构建稀疏表观模型，其理论思想主要受到最新的生物视觉感知理论的启发[12]。该理论指出：大脑皮层中视觉感知的前向路径是从初级视觉皮层到前额叶皮层的腹侧流处理路径，该腹侧流处理可以被建模为一个越来越稀疏的感知层次。

(12)式中的卷积核由目标前景区域内提取的稀疏图像块构成，在尽可能保留局部特征的基础上，剔除了冗余的噪声。而目标前景区域与输入的候选图像区域都被固定为同一尺寸，使得一些局部特征的位置并未发生较大的变化。由于卷积运算是基于像素的层次化运算的，因此(12)式中的卷积结果能够分层地响应相似的局部特征的位置，进而构建出分层结构特征。从图 1 可以看出，当跟踪目标的尺度发生变化或者被严重遮挡时，这些分层结构特征所在的位置并未发生较大变化，因而能够处理这些复杂的跟踪环境。此外，卷积运算可以通过快速傅里叶变换(FFT)转换成频域内的点乘运算，加速求解过程。

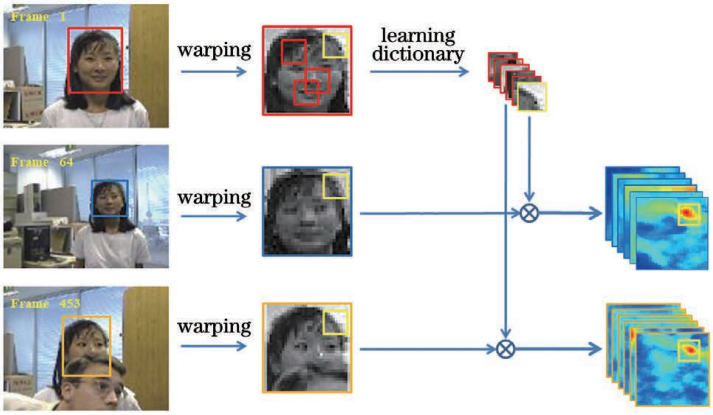

图 1　基于卷积的稀疏表观模型构建分层结构特征示意图

Fig. 1　Feature diagram of layered structure based on convolution sparse surface model

2.3　模型在线更新

由于跟踪是一个持续的过程，目标的表观会随时可能发生改变。为了适应这些变化，可以结合在线学习[13]的方法更新目标模板。然而，现有的稀疏跟踪算法[2-9]会在每一帧都无差别地更新模板，当目标被严重遮挡甚至完全消失时，跟踪的结果是不准确的。此时若错误地更新模型，可能会在之后的视频序列中产生漂移问题[14]。为了应对该问题，(4)式的计算结果既能用于估计目标模板和候选图像区域的相似度，同时对于最佳候选图像区域又能评价其准确度。因此，先通过(4)式计算出最佳候选粒子的相似度 $s_t=p(\boldsymbol{z}_t|\boldsymbol{x}_t^i)$，再定义一种新颖的在线度量准则，计算公式为

$$M_{\text{measure}}=\frac{(s_t-s_{\min})^2}{\text{mean}\left[\sum_{t=t-5}^{t}(s_{\max}-s_{\min})^2\right]},\tag{13}$$

式中：$s_{\min}$ 和 $s_{\max}$ 分别为在过去 5 帧中所有最佳候选粒子的相似度中的最小值和最大值。

M_{measure} 值注重于跟踪准确度的波动性，因而能够反映跟踪结果的可靠性。从图 2 可以看出，当目标被严重遮挡时，跟踪准确度波动性较大，此时 $s_t\approx s_{\min}$，M_{measure} 值较小；当遮挡逐渐结束时，跟踪准确度快速提升，此时 $s_{\max}\approx s_{\min}$，M_{measure} 值迅速增大。基于以上观察，所提在线模板更新方法为

$$\boldsymbol{T}=\rho\hat{\boldsymbol{C}}_t+(1-\rho)\boldsymbol{T}\ \text{ s.t. }\ M_{\text{measure}}>0.6,\tag{14}$$

式中：ρ 为固定的学习参数；$\hat{\boldsymbol{C}}_t$ 为最佳候选粒子对应的表观模型。

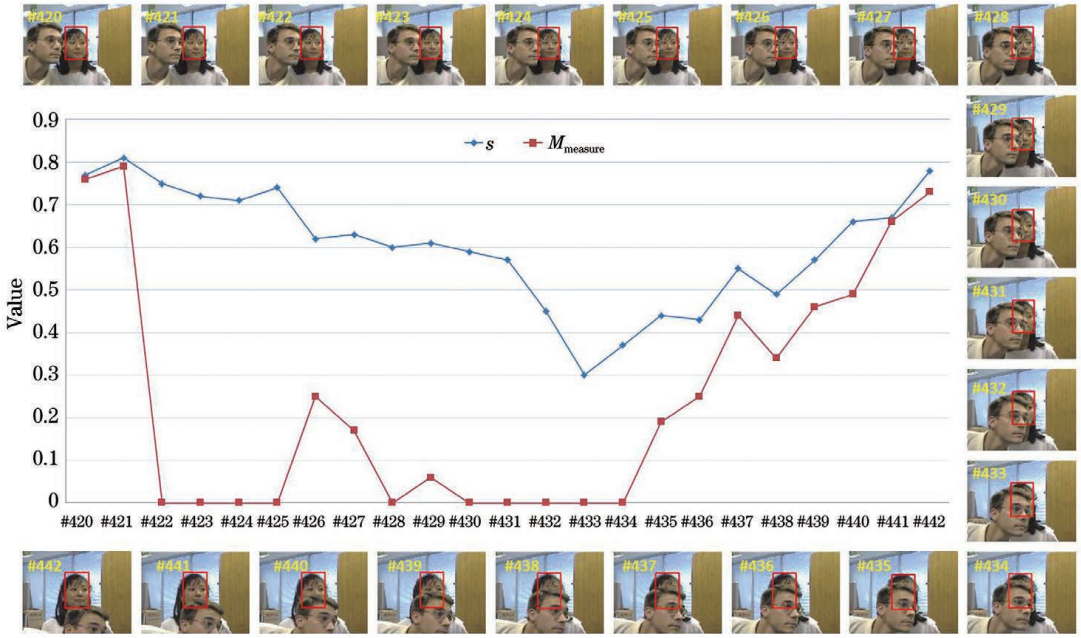

图 2　M_{measure} 与目标被遮挡的关系

Fig. 2　Relationship between the M_{measure} value and corresponding occluded target

CSTA 的详细步骤如下。

输入：视频序列（共计 T 帧），目标在第一帧的位置状态 $\boldsymbol{x}_1$。

输出：跟踪目标在第 t（$1<t\leqslant T$）帧的位置状态 $\boldsymbol{x}_t$。

For $t=1$ to T

If $t=1$　/* 初始化 */

固定目标区域 $\boldsymbol{x}_1$ 为 $m\times m$ 像素，表示为 $\boldsymbol{I}_1$，并通过(9)～(11)式习得稀疏字典 $\boldsymbol{D}=[\boldsymbol{D}_1,\cdots,\boldsymbol{D}_W]$。

For $j=1$ to W

$\boldsymbol{C}_j=\boldsymbol{D}_j\otimes\boldsymbol{I}_1$;

End For

$\boldsymbol{T}=[\boldsymbol{C}_1,\cdots,\boldsymbol{C}_W]$;

Else　/* 跟踪开始 */

通过(2)式采集 N 个粒子 $\{\boldsymbol{x}_t^i\}_{i=1}^N$，并将其固定为 $m\times m$ 像素，表示为 $\{\boldsymbol{I}_t^i\}_{i=1}^N$。

For $i=1$ to N

For $j=1$ to W

$\boldsymbol{C}_j=\boldsymbol{D}_j\otimes\boldsymbol{I}_t^i$;

End For

$\boldsymbol{C}_t^i=[\boldsymbol{C}_1,\cdots,\boldsymbol{C}_W]$;

End For

$\boldsymbol{x}_t=\hat{\boldsymbol{C}}_t=\underset{\boldsymbol{c}_t^j}{\operatorname{argmax}}\exp(-\|\mathrm{Vec}(\boldsymbol{T})-\mathrm{Vec}(\boldsymbol{C}_t^j)\|_2)$; /* 输出最佳粒子对应的目标区域 */

$s_t=\exp(-\|\mathrm{Vec}(\boldsymbol{T})-\mathrm{Vec}(\hat{\boldsymbol{C}}_t)\|_2)$;

$$M_{\mathrm{measure}}=\frac{(s_t-s_{\min})^2}{\mathrm{mean}\left[\sum_{t=t-5}^{t}(s_{\max}-s_{\min})^2\right]};$$

If $M_{\mathrm{measure}}>0.6$　/* 模板更新 */

$\boldsymbol{T}=\rho\hat{\boldsymbol{C}}_t+(1-\rho)\boldsymbol{T}$;

End If

End If

End For

3　实验与分析

给出 CSTA 与先进的稀疏跟踪算法对比的实验结果，并通过定量比较、定性比较两个方面分析 CSTA 的优劣性。

3.1　实验细节

对比实验采用目标跟踪中使用最为广泛的两个标准数据集：OTB50[15] 和 OTB100[14]。由于前者是后者的一个子集，因此主要分析在 OTB100 上的实验结果。

为了充分分析所提稀疏跟踪算法的优劣，对其与最近几年先进的稀疏跟踪算法进行了对比，包括 ASLA[6]、SCM[3]、SST[16]、MTT[17]、CST[8]、L1APG[5]、MSRT[7]、MJDL[7] 和 RSST[10]。上述 9 个算法都是基于粒子滤波框架的稀疏跟踪算法，其表观模型都是基于线性结合的全局稀疏表观模型或局部稀疏表观模型。

相关参数设定：预设的稀疏误差 $e=0.015$；学习参数 $\rho=0.05$；线性插值后图像尺寸 $m=32$；滑动窗口尺寸 $n=6$；字典数量 $W=30$；采样的粒子数量 $N=600$；运动模型参数 $\boldsymbol{V}=\mathrm{diag}(0.01,0.01,0.05)$。

3.2　定量分析

在 OTB100 数据集中，采用两种准则定量地比较算法的跟踪准确度：基于中心误差的精度图（precision plot）和基于重叠率的成功图（success plot）。

定义中心误差为

$$e_{\mathrm{error}}=\sqrt{(x_{\mathrm{T}}-x_{\mathrm{G}})^2+(y_{\mathrm{T}}-y_{\mathrm{G}})^2},\tag{15}$$

式中：$(x_{\mathrm{T}},y_{\mathrm{T}})$ 和 $(x_{\mathrm{G}},y_{\mathrm{G}})$ 分别为跟踪算法输出的中心位置和数据集给定的参考值。精度图用于反映中心误差小于给定阈值的视频帧数占所有帧数的比例，同时定义精度值为 $e_{\mathrm{error}}\leqslant 20$ 的帧数占总帧数的比值，用于算法的排名。

定义重叠率为

$$R=\frac{B_{\mathrm{T}}\cap B_{\mathrm{G}}}{B_{\mathrm{T}}\cup B_{\mathrm{G}}},\tag{16}$$

式中：B_{T} 和 B_{G} 分别为算法输出的跟踪框所占图像区域和数据集给定的目标区域。成功图显示出在阈值 $t_0\in[0,1]$ 时满足 $R\geqslant t_0$ 的帧数占所有帧数的比例，同时采用曲线下面积（AUC）衡量算法的名次。

图 3 展示了 10 种稀疏跟踪算法在 OTB100 数据集上的总体表现效果。可以看出，所提 CSTA 在精度图和成功图上均排名第一。在精度图上，CSTA 的精度值为 0.690，略优于第二的 RSST，比第三的 MSRT 高 0.062；在成功图上，CSTA 的 AUC 值为 0.522，比第二的 RSST 高 0.025，比第三的 SST 高 0.054。实验结果表明，所提 CSTA 在跟踪精度上比现有的先进的稀疏跟踪算法更加准确。

图 4 和图 5 分别显示了不同挑战因素下的精度图和成功图。可以清晰地看出：CSTA 在大部分挑战因素下排名前二；此外，在光照变化（illumination variations）、平面外旋转（out-of-plane rotation）、遮

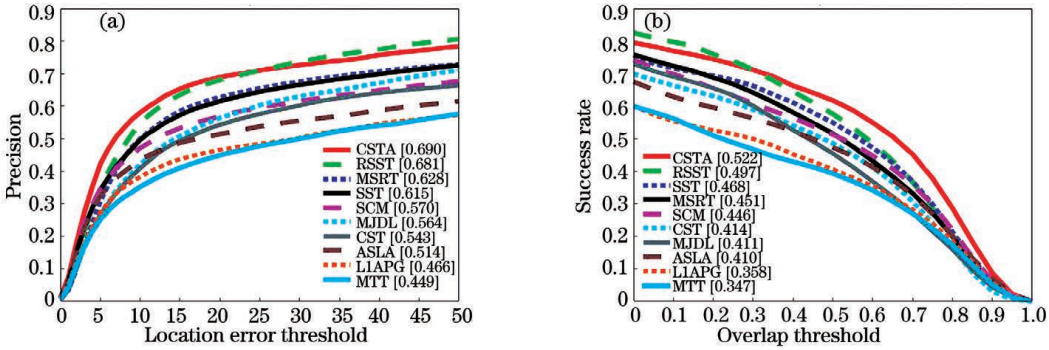

图 3　10 种稀疏跟踪算法在 OTB100 数据集上的精度图和成功图。(a)精度图;(b)成功图

Fig. 3　Precision and success plots of 10 sparse tracking algorithms on OTB100. (a) Precision plot; (b) success plot

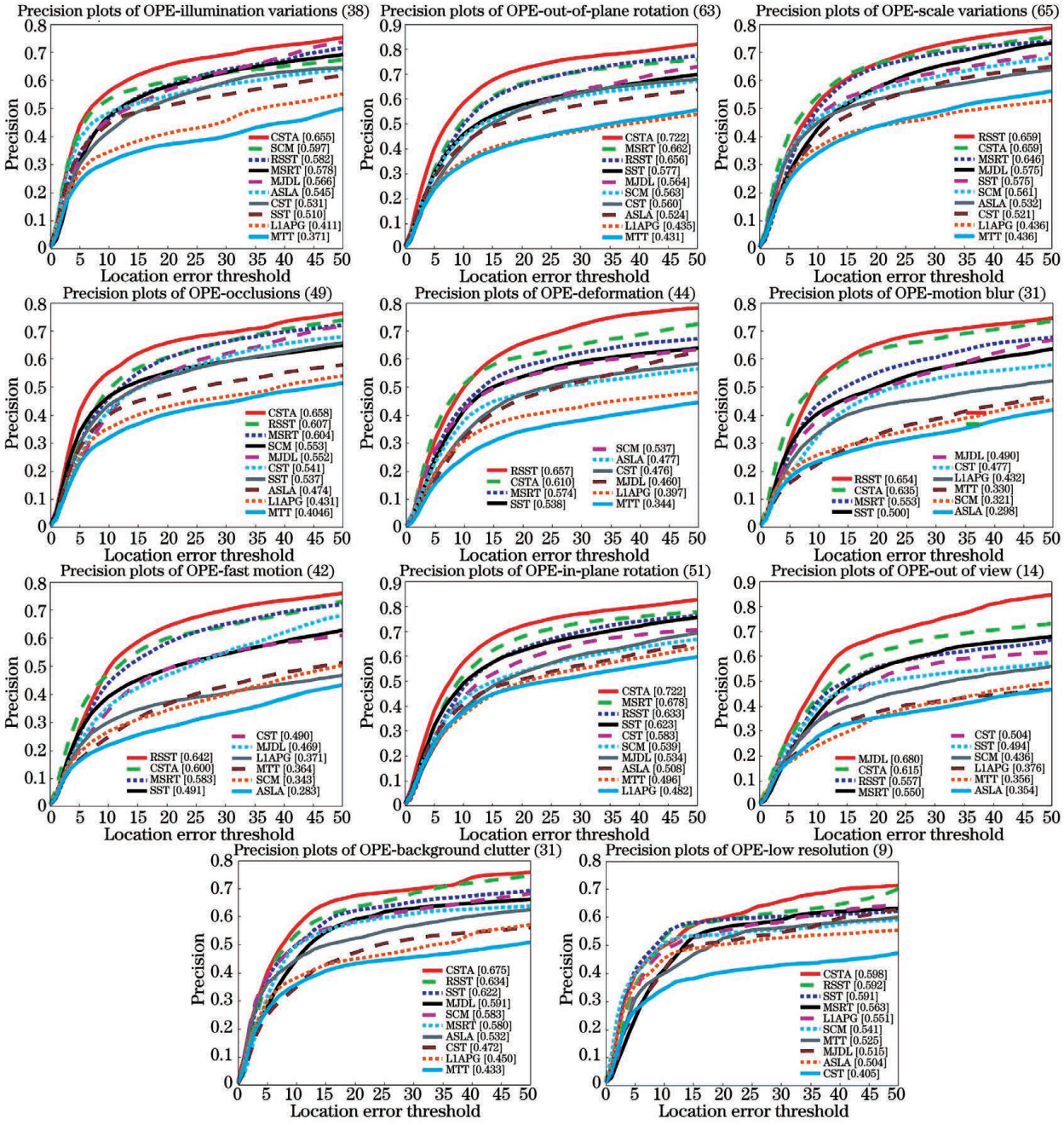

图 4　10 种稀疏跟踪算法在不同挑战因素下的精度图

Fig. 4　Precision plots of 10 sparse tracking algorithms in different challenge attributes

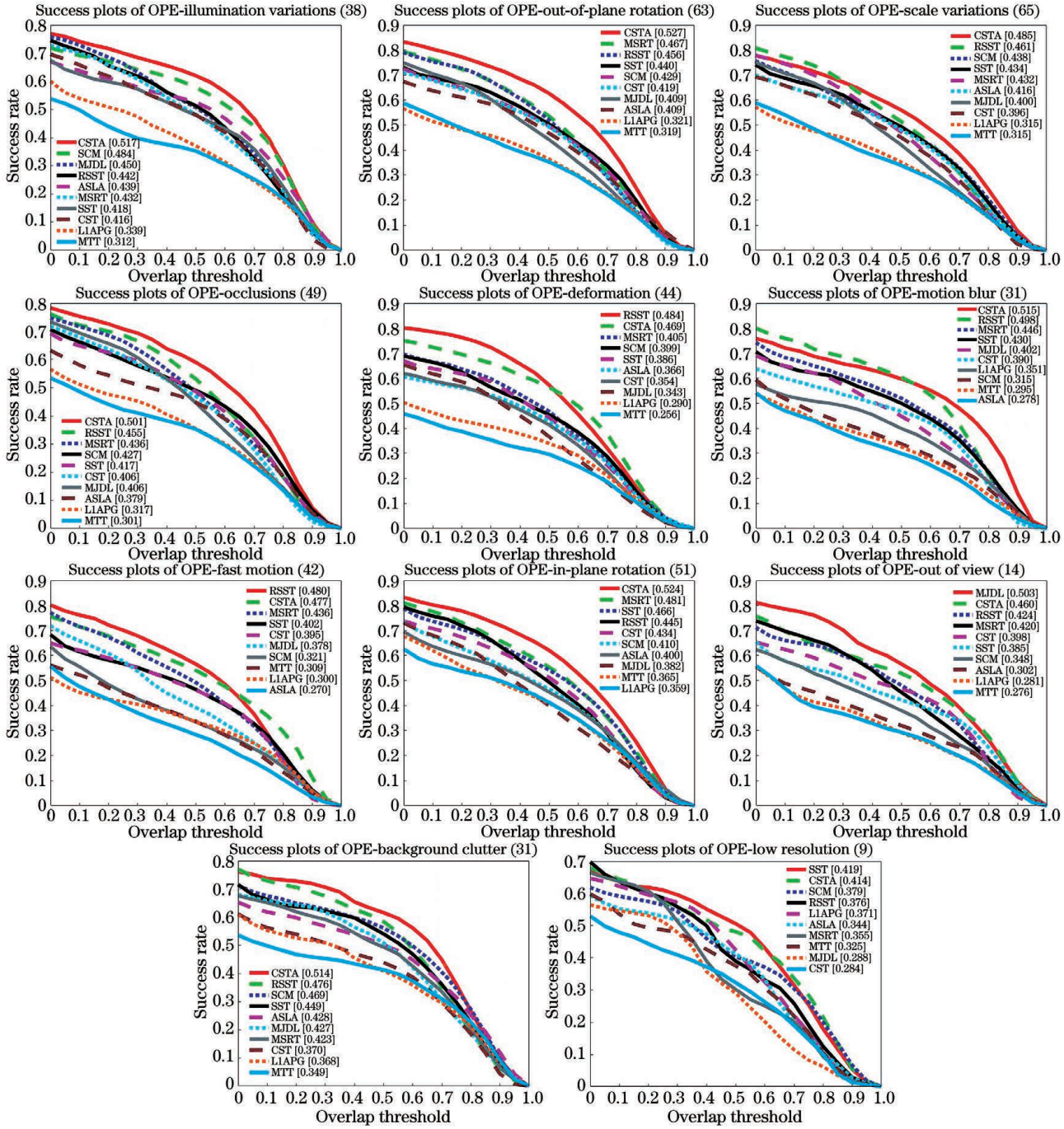

图 5　10 种稀疏跟踪算法在不同挑战因素下的成功图

Fig. 5　Success plots of 10 sparse tracking algorithms in different challenge attributes

挡(occlusions)、平面内旋转(in-plane rotation)、背景干扰(background clutter)等复杂的跟踪环境下，CSTA 均表现最优。这些实验结果表明，所提基于卷积的稀疏表观模型比传统的基于线性结合的稀疏表观模型更加鲁棒。

3.3　定性分析

为了更加直观地展现 CSTA 处理复杂跟踪环境的能力，在 OTB100 数据集中选取 6 个极具挑战性的视频序列并作定性分析，包括 coke、dragonbaby、human7、human9、liquor 和 singer2。图 6 展示了每个视频序列中部分帧实验结果的截图。

在 coke 视频序列中，跟踪目标首先被其他物体遮挡，SST、ASLA 和 CST 算法错误地更新模板，导致在之后的跟踪过程中错误地跟踪了其他物体；在 #140 帧之后，跟踪目标发生平面内旋转，MJDL 产生了漂移的问题，进而丢失了目标；之后在 #170 帧上下，跟踪目标发生尺度变化和快速运动，SCM、MTT 和 L1APG 算法跟踪失败。纵观整个视频序

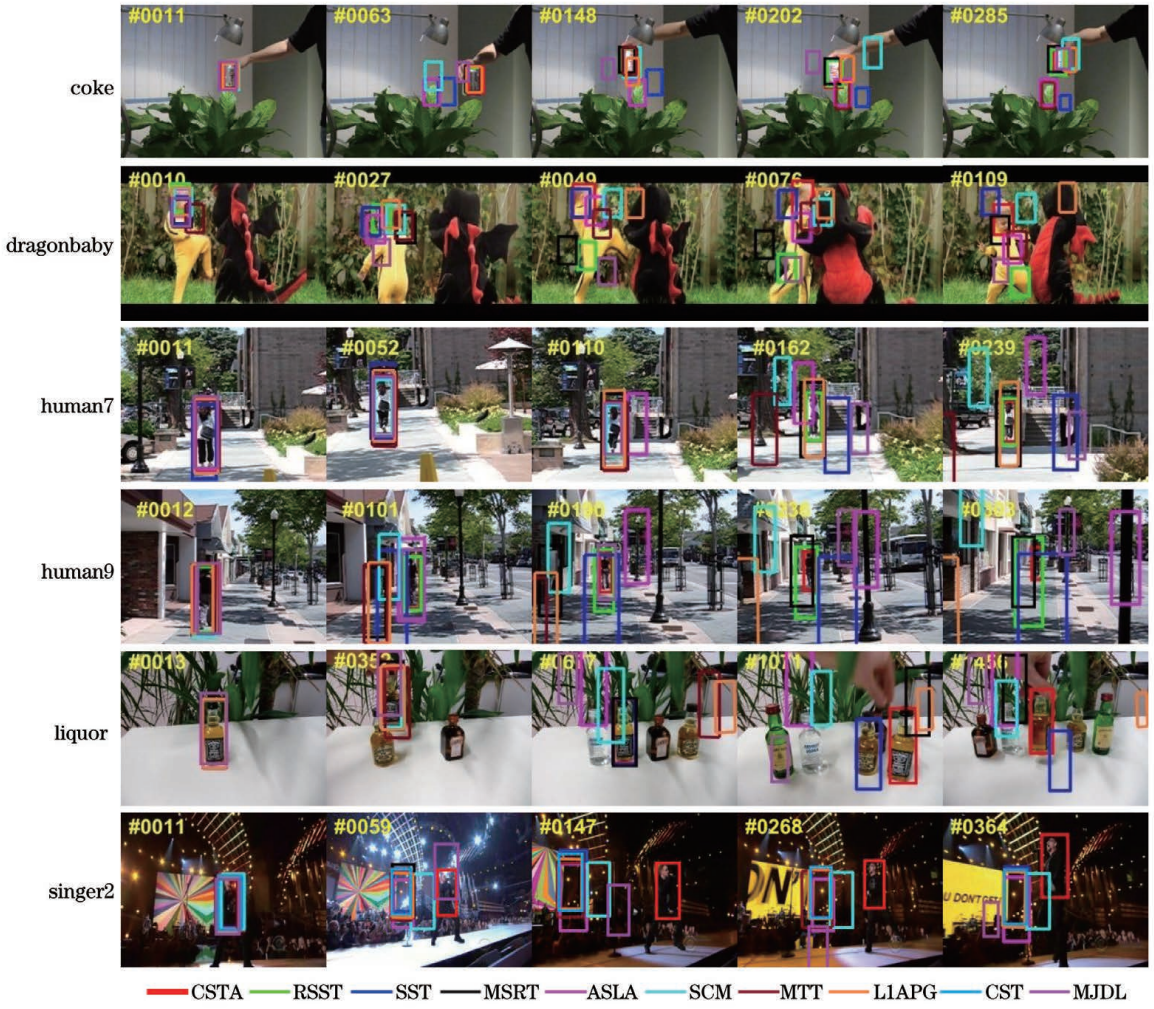

图 6　10 种稀疏跟踪算法在 OTB100 视频集中的部分视频序列中的实验结果

Fig. 6　Experimental results of 10 sparse tracking algorithms in some video sequences at OTB100 video set

列，仅有 CSTA、RSST 和 MSRT 表现优异。

在 dragonbaby 视频序列中，跟踪目标在 # 6 帧至 # 9 帧中发生平面内旋转，MTT 算法跟踪失败；在 # 13 帧至 # 43 帧中，跟踪目标发生平面外旋转，同时伴随着自身形状的变化，大多数算法皆丢失目标，只有 CSTA 和 SST 跟踪准确；之后在 # 60 帧至 # 72 帧时，跟踪目标被遮挡甚至完全消失，SST 算法产生漂移问题。在整个视频序列中，只有 CSTA 跟踪准确且稳定。

在 human7 和 human9 两个视频序列中，跟踪目标的尺度在不停发生变化，跟踪算法输出的跟踪框的大小随之改变。大多数稀疏跟踪算法皆无法适应这些变化，从而发生漂移或完全丢失目标。在 human7 视频序列中，仅有 CSTA 和 RSST 跟踪良好。在 Human9 视频序列中，只有 CSTA 始终准确地估计目标的尺度大小。

在 liquor 视频序列中，跟踪目标首先发生快速移动和尺度变化，ASLA、SCM、MTT、L1APG 和 MJDL 相继丢失目标；在之后的跟踪过程中，跟踪目标被完全遮挡，除了 CSTA，其他所有的稀疏跟踪算法都错误地跟踪了其他的目标；在之后的跟踪过程中，也仅有 CSTA 能够准确地跟踪目标。

在 singer2 视频序列中，跟踪目标发生了光照变化、尺度变化、平面旋转和形状变化；同时目标所处的背景杂乱，存在一定的干扰性。除了 CSTA，其余稀疏跟踪算法相继地发生漂移问题而跟踪失败。

在 6 个视频序列中，CSTA 之所以能够表现优异，原因是采用了基于卷积的稀疏表观模型，开发出跟踪目标内部的分层结构特征；同时通过选择性的模板更新策略，避免了错误的模型更新问题，从而有效地应对了这些复杂的跟踪环境。

4　结　　论

所提新颖的基于卷积的稀疏跟踪算法首先在跟

踪目标的前景区域内提取一组稀疏图像块作为固定的卷积核，然后对其与输入的图像块进行卷积运算。由于相似的图像块位置并未发生显著变化，因此卷积结果可以分层地保留跟踪目标的局部结构特征，从而提升算法处理复杂跟踪环境的能力。实验结果表明，所提基于卷积的稀疏跟踪算法优于现有的稀疏跟踪算法。

所提算法仍有较大的提升空间，未来的工作可以从以下几个方面考虑。首先，可以设计不同的稀疏字典作为卷积核，尤其结合判别式的稀疏字典，在理论上可以提升算法对背景干扰的能力；其次，所提字典学习是依赖于目标在第一帧的图像区域内提取的图像块，如何充分利用之后帧的图像块，在保持字典稀疏性的基础上，进一步地提升算法的性能是值得研究的问题；最后，多任务学习是稀疏表示在目标跟踪中的研究热点之一，如何利用多个粒子之间的相关性，提升字典的表示能力，亦是可行的方向。

参 考 文 献

[1] Qian Q S, Hu Y H, Zhao N X, et al. Object tracking algorithm based on global feature matching processing of laser point cloud[J]. Laser & Optoelectronics Progress, 2020, 57(6): 061012.
钱其姝, 胡以华, 赵楠翔, 等. 基于激光点云全局特征匹配处理的目标跟踪算法[J]. 激光与光电子学进展, 2020, 57(6): 061012.

[2] Yang G W, Yan S M, Wang Y Z. V-shaped seam tracking based on particle filter with histogram of oriented gradient[J]. Chinese Journal of Lasers, 2020, 47(7): 0702002.
杨国威, 闫树明, 王以忠. 基于方向梯度直方图粒子滤波的 V 型焊缝跟踪[J]. 中国激光, 2020, 47(7): 0702002.

[3] Mei X, Ling H B. Robust visual tracking using ℓ_1 minimization[C]//2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE Press, 2009: 1436-1443.

[4] Zhong W, Lu H C, Yang M H. Robust object tracking via sparse collaborative appearance model[J]. IEEE Transactions on Image Processing, 2014, 23(5): 2356-2368.

[5] Bao C L, Wu Y, Ling H B, et al. Real time robust L1 tracker using accelerated proximal gradient approach[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 1830-1837.

[6] Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 1822-1829.

[7] Fan H, Xiang J H. Robust visual tracking with multitask joint dictionary learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(5): 1018-1030.

[8] Zhang T Z, Bibi A, Ghanem B. In defense of sparse tracking: circulant sparse tracker[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3880-3888.

[9] Li J X, Zong Q. Object tracking based on multi-feature and local joint sparse representation[J]. Laser & Optoelectronics Progress, 2017, 54(10): 101502.
李敬轩, 宗群. 基于多特征和局部联合稀疏表示的目标跟踪[J]. 激光与光电子学进展, 2017, 54(10): 101502.

[10] Zhang T Z, Xu C S, Yang M H. Robust structural sparse tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(2): 473-486.

[11] Xu Q, Wang H B, Zhou J, et al. Swarm intelligence filtering for robust object tracking[J]. CAAI Transactions on Intelligent Systems, 2019, 14(4): 697-707.
许奇, 王华彬, 周健, 等. 用于目标跟踪的智能群体优化滤波算法[J]. 智能系统学报, 2019, 14(4): 697-707.

[12] Cai B L. Object tracking based on biologically inspired model[D]. Guangzhou: South China University of Technology, 2016.
蔡博仑. 基于生物启发模型的视觉跟踪[D]. 广州: 华南理工大学, 2016.

[13] Shen Y L, Wu Z D, Zhao R J, et al. Long-term object tracking based on model updating and fast re-detection[J]. Acta Optica Sinica, 2020, 40(3): 0315002.
沈玉玲, 伍忠东, 赵汝进, 等. 基于模型更新与快速重检测的长时目标跟踪[J]. 光学学报, 2020, 40(3): 0315002.

[14] Wu Y, Lim J, Yang M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

[15] Wu Y, Lim J, Yang M H. Online object tracking: a

benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 2411-2418.

[16] Zhang T Z, Liu S, Xu C S, et al. Structural sparse tracking[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 150-158.

[17] Zhang T Z, Ghanem B, Liu S, et al. Robust visual tracking via multi-task sparse learning[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2042-2049.