基于判别性局部联合稀疏模型的多任务跟踪

黄丹丹 [1] 孙 怡 [1]

摘 要 目标表观建模是基于稀疏表示的跟踪方法的研究重点, 针对这一问题, 提出一种基于判别性局部联合稀疏表示的目标表观模型, 并在粒子滤波框架下提出一种基于该模型的多任务跟踪方法 (Discriminative local joint sparse appearance model based multitask tracking method, DLJSM). 该模型为目标区域内的局部图像分别构建具有判别性的字典, 从而将判别信息引入到局部稀疏模型中, 并对所有局部图像进行联合稀疏编码以增强结构性. 在跟踪过程中, 首先对目标表观建立上述模型; 其次根据目标表观变化的连续性对采样粒子进行初始筛选以提高算法的效率; 然后求解剩余候选目标状态的联合稀疏编码, 并定义相似性函数衡量候选状态与目标模型之间的相似性; 最后根据最大后验概率估计目标当前的状态. 此外, 为了避免模型频繁更新而引入累积误差, 本文采用每 5 帧判断一次的方法, 并在更新时保留首帧信息以减少模型漂移. 实验测试结果表明 DLJSM 方法在目标表观发生巨大变化的情况下仍然能够稳定准确地跟踪目标, 与当前最流行的 13 种跟踪方法的对比结果验证了 DLJSM 方法的高效性.

关键词 目标跟踪, 表观建模, 稀疏表示, 多任务跟踪, 粒子滤波

引用格式 黄丹丹, 孙怡. 基于判别性局部联合稀疏模型的多任务跟踪. 自动化学报, 2016, **42**(3): 402−415

DOI 10.16383/j.aas.2016.c150416

Tracking via Multitask Discriminative Local Joint Sparse Appearance Model

HUANG Dan-Dan[1] SUN Yi[1]

Abstract Appearance modeling is the research focus in tracking method based on sparse representation. In this paper, a discriminative local joint sparse appearance model based multitask tracking method (DLJSM) is proposed within particle filter framework. The proposed model builds a discriminative dictionary for each image patch within the object-region in order to introduce the discriminative information into the local sparse model, and enhances the structure feature via joint sparse representation. During tracking, the target appearance is modeled firstly. Then the sampling particles are pre-selected according to the target appearance′s consecutive changes characteristic to improve efficiency of the algorithm. Next, joint sparse representations of all the candidates are solved jointly. Furthermore, a function is defined to measure the similarities between candidates and the target model. Lastly, the target state is estimated by the maximum posterior probability. Besides, update is judged every five frames to avoid the accumulative error caused by frequent update and the target information in the first frame is reserved to alleviate drifting. Test results show that the proposed DLJSM tracker can maintain a stable and accurate tracking when the target appearance undergoes huge variations. Comparison results on challenging benchmark image sequences show that the DLJSM method out performs 13 other state-of-the-art algorithms.

Key words Object tracking, appearance modeling, sparse representation, multitask tracking, particle filter

Citation Huang Dan-Dan, Sun Yi. Tracking via multitask discriminative local joint sparse appearance model. *Acta Automatica Sinica*, 2016, **42**(3): 402−415

目标跟踪是计算机视觉领域一个重要的研究课题, 也是许多实际应用系统不可缺少的部分, 例如视频监控、智能交通、增强现实以及运动分析等. 由于其重要性, 研究者们提出了大量的目标跟踪算法[1−3], 并取得了显著的进展. 然而如何在复杂环境中对外观时刻变化的目标进行长时间稳定地跟踪仍是当前的研究热点.

收稿日期 2015-06-29 录用日期 2015-10-23
Manuscript received June 29, 2015; accepted October 23, 2015
本文责任编委 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 大连理工大学信息与通信工程学院 大连 116024
1. School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024

目前常见的跟踪方法可分为两种: 生成式方法和判别式方法. 生成式跟踪方法首先对目标表观建模, 然后在后续图像中搜索与该模型最相似的区域作为跟踪结果, 这类方法最重要的部分是构建有效的表观模型. 如文献 [4] 通过对目标区域中具有重要特征的子区域计算直方图来建模目标; 视觉跟踪分解 (Visual tracking decomposition, VTD) 算法[5] 则将目标的表观模型分解为多个基本的小模型, 每个小模型描述目标表观的一种变化. 以上两种方法都利用了局部的概念并将目标建模为局部模型. 此外, 增量视觉跟踪 (Incremental visual tracking, IVT)[6] 采用目标区域的低维子空间对目标表观进行

全局建模, 并且通过增量的学习子空间来自动更新目标模型. 以上方法虽然对目标遮挡和形变具有一定的鲁棒性, 但是由于没有考虑背景信息, 因此当目标处于复杂背景时很难维持长时间的跟踪. 判别式跟踪方法综合考虑目标与背景信息, 并且将目标跟踪建模为二类的分类问题, 通过正负样本训练一个二类的分类器将目标区域从背景中区分出来. 由于判别式的方法很大程度上依赖于分类器的训练, 因此一个重要问题是如何减少跟踪中由累积误差引起的模型漂移. 典型的判别式跟踪方法有多样例学习 (Multiple instance learning, MIL) 和跟踪学习检测 (Tracking-learning detection, TLD). MIL[7] 方法采用目标位置附近的多个样本构成正样本集, 从而包容正样本中引入的背景信息, 缓解了分类器更新中的漂移问题. TLD[8] 则通过不断学习正负样本的结构信息而提高分类器的准确率, 从而保证跟踪的精度.

近几年, 稀疏编码理论在计算机视觉领域取得了巨大成功, 因此也被引入到目标跟踪中. 在基于稀疏表示的跟踪方法中, 字典的构造方法是表观建模以及相似性计算的基础, 根据字典构造方式的不同可将算法分为两类: 基于整体模板和基于部分模板的方法. 前者直接将整个目标区域作为基底来构造字典, 最典型的是 l_1 跟踪[9] 中的字典构造方法. 该方法将目标区域和平移该区域所得到的目标模板直接列向量化作为字典的基底, 并在字典中加入噪声模板来处理部分遮挡问题. 其他基于整体模板的字典构造方法都是文献 [9−10] 中方法的变形. 例如, 文献 [11] 构造的字典中不仅包含目标模板, 同时还加入了背景模板, 使得字典具有判别性. 文献 [12] 则用目标模板的独立分量分析矢量代替目标模板构成字典. 文献 [13] 将目标模板的稀疏表示进一步扩展为核稀疏表示. 基于整体模板构造字典的方法, 通常将目标表观建模为整体稀疏模型, 这类模型对目标的全局特征描述的较好, 但是对局部遮挡或剧烈形变的描述则较差. 与基于整体模板的方法不同, 基于部分模板的字典构造方法提取目标区域中的小图像块作为字典的基底. 局部稀疏选择模型 (Local sparse k-selection, LSK)[14] 选择目标区域中最具有表示能力的小图像块构成字典, 以此建立基于局部稀疏的表观模型. 自适应结构局部稀疏表观模型 (Adaptive structural local sparse appearance, ALSA)[15] 则利用多个模板的局部空间结构信息构建字典, 并建立结构性的局部稀疏模型. 文献 [16] 在构建字典时加入了判别信息, 并训练线性支持向量机跟踪目标. 总体来说, 加入结构信息或判别信息的局部稀疏模型能更好地描述目标的特征. 除了单独使用整体模板和部分模板的方法, 文献 [17] 将二者联合使用, 该联合模型结合了目标的整体特征和局部特征, 因此能在跟踪中有效地适应目标外观的多种变化.

上述介绍的方法虽然在跟踪中各有优势, 但是它们在求解稀疏编码时, 采用的都是分别计算的方法, 即对每个候选区域的稀疏编码进行独立求解, 这种计算方法不但计算速度慢而且忽略了候选区域的结构信息. 这里的结构信息不仅是指单个候选区域内部的局部图像块之间的结构, 还包括多个候选区域中处于相同位置的图像块之间的结构关系, 而现有的基于局部稀疏表示的模型只利用了单个区域内部的结构信息, 如文献 [14−15]. 此外, 文献 [18−21] 在多任务框架下对所有候选区域进行联合稀疏编码, 在提高计算速度的同时还共享了联合稀疏编码的结构性. 但是它们仅考虑候选区域间整体的结构信息, 却忽略了区域内部的结构关系. 同时, 上述方法均将目标表观建模为整体稀疏模型, 因此仍然不能很好地处理目标的部分遮挡和视角旋转等问题. 综合以上分析, 本文提出一种基于判别性局部联合稀疏模型的多任务跟踪方法 (Discriminative local joint sparse appearance model based multitask tracking method, DLJSM). 它将目标建模为目标区域内所有局部图像块的联合稀疏编码矩阵, 并在多任务框架下联合求解所有候选区域内图像块的稀疏表示. 此外, 该方法还定义一个联合相似性函数来综合衡量候选区域与目标模型间的相似性, 相似性最大的候选区域即为跟踪结果, 最后通过在线更新来调整模型以适应目标的表观变化. 与现有的基于多任务的跟踪方法不同, DLJSM 采用局部模型描述目标, 因此具有局部稀疏模型在处理遮挡、形变等问题上的优势; 同时, 它为目标区域内的每个局部块分别构造字典, 字典中不仅包含目标模板还加入了背景模板, 因此在跟踪中具有更强的判别能力; 此外, 对所有局部图像块进行联合稀疏编码则使得 DLJSM 能更充分地利用多个候选区域的结构性, 从而更准确地跟踪目标. 多个测试视频的跟踪结果验证了 DLJSM 的高效性.

1 基于多任务的目标跟踪

在文献 [9] 提出的 l_1 跟踪框架中, 首先由目标模板 D 和噪声模板 I 构成稀疏字典 B, 即 $B = [D\ I]$; 然后利用字典 B 对每个候选区域进行稀疏编码, 得到稀疏系数矢量后, 根据候选区域的重建误估计目标的当前状态. l_1 跟踪对每个候选区域独立地计算稀疏编码, 因此运算速度较慢. 在此基础上, MTT (Multi-task tracking)[18] 将每个候选区域的稀疏编码视为一个独立的任务, 在多任务学习框架下对所有候选区域的稀疏编码进行联合求解. MTT 方法

首先构造观测矩阵 X, X 的每个列向量为一个候选区域, 则 X 可线性表示为 $X = [D\ I]\begin{bmatrix} Z \\ E \end{bmatrix} = BC$. 其中, Z 和 E 分别是稀疏系数矩阵和误差系数矩阵, 二者的第 i 列分别为第 i 个候选区域的稀疏表示系数和误差系数. 系数矩阵 C 根据 APG (Accelerated proximal gradient) 算法进行联合求解. 对于单个的候选区域而言, 上述求得稀疏编码的过程即称为该候选区域被联合稀疏表示. MTT 算法同样根据各候选区域的重建误差对目标状态进行估计.

类似地, 文献 [20] 也在多任务学习框架下进行目标跟踪. 不同于 MTT 算法将每个候选区域的稀疏编码作为独立任务, 该方法对每个候选区域提取多种特征, 将每种特征的稀疏编码作为一个独立任务, 最后对单个候选区域的多种特征进行联合稀疏编码, 所有特征的重建误差之和最小的候选区域即为跟踪结果. MTMV (Multi-task multi-view)[19] 跟踪综合了文献 [18] 与文献 [20] 的方法, 将候选区域的一种特征的稀疏编码当作独立任务, 对所有候选区域的所有特征进行联合稀疏表示. 记 $X^k = [\boldsymbol{x}_1^k, \cdots, \boldsymbol{x}_n^k]$ 为 n 个候选区域的第 k 种特征构成的观测矩阵, $B^k = [D^k\ I^k]$ 为目标模板的第 k 种特征构成的字典, 其中, $k = 1, \cdots, K$, K 为特征总数. 观测矩阵 $X = [X^1, \cdots, X^K]$ 在字典 $B = [B^1, \cdots, B^K]$ 上的稀疏编码 $C = [C^1, \cdots, C^K]$ 可通过以下问题求解:

$$C = \arg\min_{C} \sum_{k=1}^{K} \left\| B^k C^k - X^k \right\|_F^2 + \lambda \left\| C^k \right\|_{1,2} \quad (1)$$

其中, λ 为平衡系数. 最后利用候选区域的重建误差构造似然函数, 根据最大后验概率估计跟踪结果, 即

$$i = \arg \max_{i=1,\cdots,n} \exp\left(-\alpha \sum_{k=1}^{K} \left\| D^k Z^k - X^k \right\|^2 \right) \quad (2)$$

其中, α 为常数, Z^k 为编码矩阵, C^k 中对应于字典 D^k 的部分. 除上述三种方法外, 文献 [21] 也利用联合稀疏表示的方法建模, 但是其采用反向稀疏的方法, 即利用候选区域来构造字典, 对目标和背景模板进行稀疏编码, 并构造相似性图来跟踪目标. 本文只考虑基于正向稀疏表示的方法, 即利用目标模板构造字典, 从而对候选区域进行稀疏编码的建模方法[18−20]. 此外, 文献 [22] 提出一种基于结构稀疏模型的粒子滤波跟踪方法, 该方法利用多层金字塔结构的全局和局部窗口对目标进行信息提取和表观建模, 由于同时考虑了全局和局部特征, 因此对目标的描述更加全面和准确. 文献 [23] 针对跟踪中出现的目标遮挡问题, 提出一种复合约束的稀疏多任务学习方法, 该方法综合考虑粒子间相关性, 对遮挡区域和非遮挡区域进行分别约束, 因此能够更准确地重构遮挡区域, 降低遮挡对跟踪的影响.

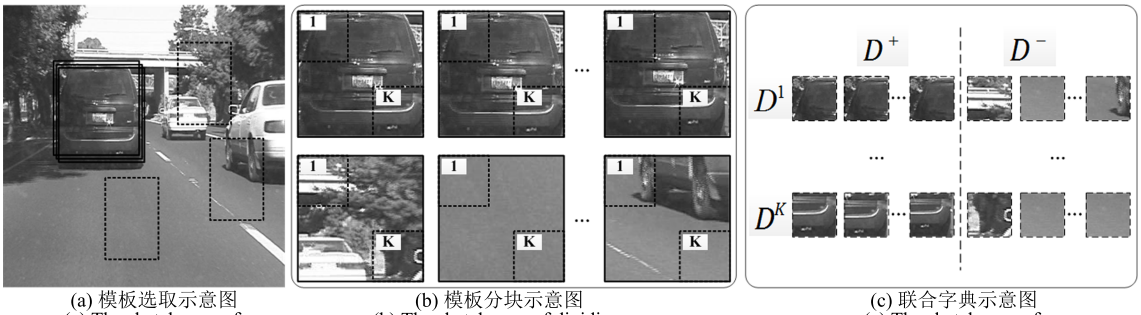
上述基于多任务联合稀疏表示的跟踪方法, 对候选区域的编码矩阵进行两个约束, 一是每个列向量都具有少量的非零元素, 以保证每个候选区域或每个特征由有限的几个目标模板或特征线性表示; 二是所有列向量中非零元素的位置分布相同, 使得所有的候选区域或特征由共同的目标模板或特征线性表示. 这两个约束合称为结构稀疏性, 这种特性使得算法对候选区域的稀疏编码更具有针对性, 从而使目标跟踪的结果更稳定. 然而这类方法大多采用基于整体模板的方法构造字典, 因此将目标表观建模为整体稀疏模型. 尽管文献 [19−20] 使用了多种特征描述目标, 但是这类方法仍然存在整体模型固有的缺点. 例如, 很难处理目标大面积遮挡和大角度旋转等. 针对这一问题, 本文提出一种基于判别性局部联合稀疏模型的目标跟踪方法. 它将目标外观建模为联合稀疏编码矩阵, 矩阵的每一列均为目标区域中一个小图像块的联合稀疏表示矢量. 该方法对所有候选区域的小图像块进行联合稀疏编码, 并且构造似然函数计算目标模型和候选区域间的相似性, 最后利用最大后验概率估计目标状态. 此外, 目标模型通过在线更新实时地调整以适应由于光照、形变、视角变化等原因引起的表观变化.

2 基于判别性局部联合稀疏表示的目标表观模型

本节着重介绍用于描述目标表观的判别性局部联合稀疏模型. 首先给出联合字典的学习方法; 然后介绍如何根据联合字典对目标的表观建模; 最后给出联合字典以及目标模型的在线更新方法.

2.1 联合字典的学习

如前文所述, 加入结构信息和判别信息的局部稀疏模型在跟踪中表现的更加鲁棒, 因此本文采用基于部分模板的字典构造方法. 为了增强判别性, 构成字典的部分模板由局部目标模板和局部背景模板构成, 这些局部模板取自于 10 个目标模板和 10 个背景模板. 其中目标模板的选取与文献 [9] 中的方法相同, 即由指定的目标区域以及该区域向各个方向平移得到的图像构成, 以此保证目标模板能够准确的表达目标并减轻漂移, 如图 1 (a) 中的目标车辆外围的矩形所示. 背景模板则从目标区域外的背景图像中采样得到, 如图 1 (a) 中远离目标车辆的矩形所示. 目标模板和背景模板合称为整体模板, 将这些整体模板进行归一化, 归一化后的结果如图 1 (b) 所示.



(a) 模板选取示意图
(a) The sketch map of templates choice

(b) 模板分块示意图
(b) The sketch map of dividing templates into patches

(c) 联合字典示意图
(c) The sketch map of collaborative dictionary

图 1　联合字典的学习过程

Fig. 1　The flowchart of dictionary learning

为使图像清晰明了, 图 1 中仅给出部分整体模板. 为了得到部分模板, 将每个整体模板划分为 K 个尺寸为 $m \times m$ 的图像块, 如图 1 (b) 中虚线框所示, 并按照从左到右, 从上到下的顺序对整体模板中的图像块进行标记, 每个图像块即为一个部分模板. 为了更好的保留整体模板的结构信息, 本文将整体模板中相邻的两个图像块的重合率设为 0.5, 即相邻的两个图像块有一半的区域是重合的. 取每个模板内相同位置的图像块 (图 1 (b) 中具有相同序号的图像块) 构成一个字典, 共得到 K 个字典, 如图 1 (c) 所示. 将初始字典表示为 $D = \left\{D^1, \cdots, D^k, \cdots, D^K\right\}$, 其中 $D^i = [\boldsymbol{d}_1^i, \cdots, \boldsymbol{d}_{20}^i] \in \mathbf{R}^{m^2 \times 20}$, $\boldsymbol{d}_1^i \in \mathbf{R}^{m^2 \times 1}$ 为第一个模板的第 i 个图像块列向量化构成的基向量. 联合字典的学习过程如图 1 所示, 图 1 (c) 中 D^+ 表示字典中由目标模板构成的部分, D^- 表示字典中由背景模板构成的部分. 该字典具有两个优点: 1) 由部分模板构造, 因此能更好地处理目标的部分遮挡以及局部形变; 2) 加入背景模板, 因此具有更强的判别性, 更适合目标跟踪.

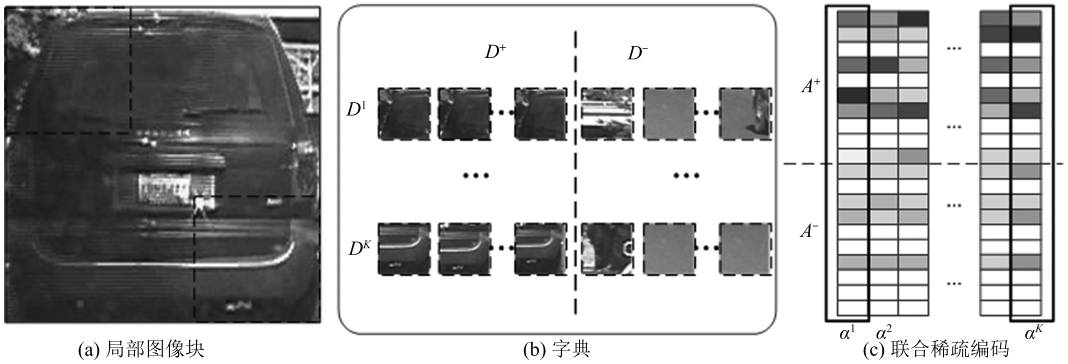
2.2　目标表观模型

从联合字典的构造方法可知, 目标区域经过归一化后, 其内的每个图像块 $\boldsymbol{x}^k$ 均对应于一个字典, 那么 $\boldsymbol{x}^k$ 可由对应的字典 D^k 进行稀疏编码. 考虑到联合稀疏表示在描述目标表观上的优势, 本文对所有图像块的稀疏编码进行联合求解. 令 $X = [\boldsymbol{x}^1, \cdots, \boldsymbol{x}^k, \cdots, \boldsymbol{x}^K] \in \mathbf{R}^{m^2 \times K}$ 表示由 K 个图像块构成的观测矩阵, $A = [\boldsymbol{\alpha}^1, \cdots, \boldsymbol{\alpha}^k, \cdots, \boldsymbol{\alpha}^K] \in \mathbf{R}^{20 \times K}$ 表示观测矩阵 X 的联合稀疏编码矩阵, 那么 A 可通过式 (1) 求解, 即

$$A = \arg\min_A \sum_{k=1}^{K} \left\|\boldsymbol{x}^k - D^k \boldsymbol{\alpha}^k\right\|^2 + \lambda \|A\|_{1,2} \quad (3)$$

其中, $\|A\|_{1,2} = \left(\sum_{k=1}^{K} \left(\|\boldsymbol{\alpha}^k\|_1\right)^2\right)^{\frac{1}{2}}$, 本文利用文献 [24] 中提出的方法对式 (3) 进行优化求解. 由于在求解时对编码矩阵 A 进行了结构稀疏性的约束, 因此, A 中每一列均包含有限个数的非零元素, 并且所有的非零元素分布在相同的行上. 令 $A = [A^+, A^-]^{\mathrm{T}}$, $A^+ \in \mathbf{R}^{10 \times K}$ 为编码矩阵中对应于 D^+ 的部分, $A^- \in \mathbf{R}^{10 \times K}$ 为编码矩阵中对应于 D^- 的部分. 将目标表观模型定义为 A^+, 即联合稀疏编码中对应于目标模板的部分. 建模的过程如图 2 所示, 首先将目标区域分块, 如图 2 (a) 中左上角和右下角的区域所示, 然后对所有的局部图像根据对应的字典 (图 2 (b) 中的 D^k) 进行联合稀疏表示, 解得的稀疏编码中对应于 D^+ 的部分即为目标的表观模型, 如图 2 (c) 中 A^+ 所示. 其中左边和右边矩形框内的矢量分别对应于图 2 (a) 中左上角和右下角的图像块. 经过这个过程, 目标区域内的所有图像块均被联合稀疏表示. 图 2 (c) 中 A^+ 和 A^- 分别与字典 D^+ 和 D^- 相对应, 稀疏编码中的非零元素由灰色方块表示, 值越大则颜色越深. 从图中可见, 尽管两个图像块的稀疏表示系数各不相同, 但是具有相同的分布, 即建立的目标模型具有结构稀疏性.

从现有文献的跟踪效果来看, 局部稀疏模型已经被证明是一种比较有效的目标表观描述模型[14−15], 加入结构性信息后, 目标特征的位置分布被编码到目标模型中, 能更准确地描述目标区域, 同时也避免在跟踪中引入过多的背景信息而导致模型出现漂移. 本文使用局部联合稀疏表示对目标表观建模, 是基于这样一个事实, 即线性描述局部图像块的基底, 应该来自于相同的目标模板. 这种结构稀疏性约束, 在限制每个局部块被稀疏表示的同时, 还对表示局部图像块的模板进行整体结构上的约束. 因此本文建立的联合稀疏模型, 除了具有局部模型在处理目标遮挡、形变等问题时的优点, 还具备结构性; 此外字典中包含的背景模板使得该模型同时具有更强的判别能力. 该模型与第 2.3 节的更新方法相结合, 能及时准确地捕捉目标的表观变化, 更有利于长时间稳定地跟踪.



(a) 局部图像块 (a) The local image patches　(b) 字典 (b) The dictionary　(c) 联合稀疏编码 (c) The joint sparse code

图 2 目标表观建模示意图

Fig. 2 The sketch map of modeling the target appearance

2.3 联合字典与表观模型的更新方法

跟踪过程中, 目标的表观会由于光照、本身形变、摄像角度和遮挡等原因产生部分或整体的变化. 这些变化随着时间逐渐累加, 固定的目标模型会因为不能适应这些表观变化而失去效果, 从而导致跟踪失败. 因此目标模型的更新是构成稳定跟踪系统必不可少的部分. 本文建立的目标模型是基于联合稀疏表示的, 联合稀疏字典是构建模型的基础, 所以目标模型的更新过程实际上就是联合字典的更新过程. 而本文中的字典由目标模板和背景模板构成, 因此联合字典更新主要集中于模板的更新.

算法 1. 联合字典与目标模型的更新方法

1) 计算权重最小的目标模板与当前跟踪结果之间的相似性, 初始状态下, 权重为均值分布, 目标模板与跟踪结果之间的相似性定义为二者的反余弦函数值.

2) **If** 计算得到的相似性处于阈值 (θ_1, θ_2) 内
   a) 用当前的跟踪结果代替权重最小的目标模板;
   b) 重新计算每个目标模板的权重, 计算方法与文献 [19] 相同;
   c) 按照第 2.1 节的方法重新学习联合字典;
   d) 将目标模型更新为当前跟踪结果的联合稀疏编码矩阵;

Else

目标模板、模板权重、联合字典与目标模型都保持不变.

End

如第 2.1 节所述, 背景模板是从当前帧的背景区域采样得到的, 在更新时, 只需对背景区域重新采样即可. 对于目标模板的更新, 本文采用类似于文献 [19] 的方法. 在跟踪中, 目标的表观变化具有连续性, 因此在一段时间之内, 可以认为目标的表观是不变的. 在此假设下, 对模型进行每帧一次的更新不仅浪费时间而且可能由于频繁的更新而引入大量的累积误差, 最终导致模型漂移. 为避免上述情况发生, 本文每 5 帧判断一次, 根据判断结果来决定联合字典和目标模型是否需要更新, 具体的更新方法如表 1 所示. 此外, 为了保证在跟踪过程中目标模板不会发生重大的漂移, 在字典更新时, 保留首帧中由指定的目标区域构成的目标模板. 第 4 节的实验证明该更新算法与本文构建的目标模型相结合能更准确地描述目标表观, 进而获得更稳定的跟踪结果.

3 DLJSM 算法实现

本文提出的 DLJSM 方法将判别性局部联合稀疏模型嵌入到粒子滤波框架下, 并根据最大后验概率对目标状态进行估计. 粒子滤波方法包括预测和更新两步[25]. 在预测阶段, 目标在第 t 帧的状态矢量 s^t 可由第一帧到第 $(t-1)$ 帧的观测 $z_{1:t-1}$ 根据式 (4) 得到

$$P\left(s_t|z_{1:t-1}\right) = P\left(s_t|s_{t-1}\right) P(s_{t-1}|z_{1:t-1}) d_{s_{t-1}} \quad (4)$$

其中, $s_t = (c_t^x, c_t^y, w_t, h_t, r_t, v_t)$ 为第 t 帧的目标状态. c_t^x 和 c_t^y 分别表示目标的中心位置坐标; w_t 和 h_t 分别表示目标在横纵坐标轴上的尺度; r_t 和 v_t 分别表示目标在横纵坐标轴上的角度. $P\left(s_t|s_{t-1}\right)$ 是自动模型, 本文采用仿射变换对其建模, 即

$$P\left(s_t|s_{t-1}\right) = N\left(s_t; s_{t-1}, \Sigma\right) \quad (5)$$

其中, Σ 为方差矩阵, 其形式为对角矩阵并且对角元素为仿射变换参数. 根据式 (5) 可对第 t 帧的目标状态进行采样, 每个采样状态对应一个候选区域. 那么在已知观测 $z_{1:t}$ 的情况下, 第 t 帧的目标状态更新为

$$P\left(s_t|z_{1:t}\right) = P\left(z_t|s_t\right) P\left(s_t|z_{1:t-1}\right) \quad (6)$$

其中, $P\left(z_t|s_t\right)$ 为观测模型, 反映了候选区域与目标模型之间的相似性, 本文将观测模型定义如下:

$$P\left(z_t|s_t^i\right) \propto f\left(C_t^i, B_t\right) \tag{7}$$

其中, s_t^i 为采样得到的第 i 个目标状态; C_t^i 为状态 s_t^i 对应的候选区域的表观模型; B_t 为当前的目标表观模型; f 为计算候选区域与目标模型间相似性的函数, 具体定义在后文中给出. 为避免退化, 粒子滤波根据粒子的重要性权重进行重采样, 第 i 个粒子的重要性权重定义为: $g_t^i = g_t^{i-1} P\left(z_t|s_t^i\right)$. 目标当前的状态 s_t^* 通过最大后验概率估计

$$s_t^* = \arg\max_{s_t^i} P\left(s_t^i|z_{1:t}\right) \tag{8}$$

在上述跟踪框架下, DLJSM 方法从视频序列的第 2 帧开始, 循环地对每帧图像进行候选区域采样、目标状态估计和重采样的过程. 为了减少多余计算, 本文采用文献 [26] 中的方法对每次采样得到的候选区域进行筛选, 并根据目标表观变化的连续性去除偏移实际位置较大的粒子, 如图 3 (a) 中远离目标的矩形框所示. 令 $D_t^k = [D_t^{k+}\ D_t^{k-}]$ 表示第 t 帧的第 k 个字典, 其中 D_t^{k+} 与 D_t^{k-} 分别表示字典中由目标模板和背景模板构成的部分, X_t^i 为第 t 帧中第 i 个粒子中所有图像块构成的观测矩阵. 根据表观变化连续性, 目标当前状态附近的粒子与前一帧的目标状态接近, 转化为数学描述则为 $e^i = \sum_{k=1}^{K} \|D_t^{k+} B_{t-1}^k - \boldsymbol{x}_t^{i,k}\|^2$ 的值比较小, 其中 $\boldsymbol{x}_t^{i,k}$ 为 X_t^i 的第 k 列, B_{t-1}^k 为第 $t-1$ 帧的目标稀疏编码矩阵的第 k 列. DLJSM 算法通过滤除具有较大 e^i 值的粒子, 减少了候选区域的个数, 提高了计算速度. 实际跟踪中, 在合适的采样参数下, 大部分的粒子分布在目标周围, 根据表观变化的连续性, 这部分粒子都具有较小的 e^i 值. 同时将去除多余粒子的阈值 ε 设置为最大 e^i 值的 1/2, 这使得阈值与粒子的位置分布相关, 并保证了总有一部分具有较小 e^i 值的粒子会被保留下来, 从而避免了所有粒子都被滤除而引起的粒子耗尽问题. 在第 4 节的实验设置下, 平均每帧图片中大约有 1/3 的粒子被滤除. 多余粒子的去除过程如图 3 所示, 图 3 (a) 中的矩形框表示粒子滤波采样得到的候选区域. 直观上看, 最外面的矩形框偏离目标实际的位置较远, 这样的粒子参与联合稀疏编码不仅浪费时间而且有可能对编码结果产生影响, 因此通过图 3 (b) 所示的滤波器后, 保留的候选区域如图 3 (c) 所示, 可见距离目标较远的矩形框均被去除, 而其他的粒子则并不受影响.

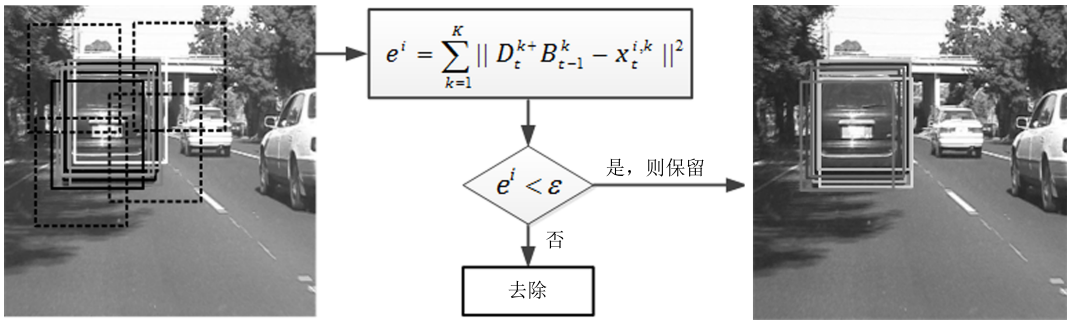
经过粒子去除后, DLJSM 方法对剩下的候选区域进行联合稀疏表示. 首先将候选区域分块, 图像块的大小与目标区域分块的大小相同, 为避免混淆, 仍令 $\boldsymbol{x}_t^{i,k}$ 表示第 t 帧中第 i 个候选区域的第 k 个图像块, 其中 $i = 1, \cdots, L$; $k = 1, \cdots, K$. L 为剩余候选区域的个数, K 为每个候选区域获得的图像块个数. 则 $\boldsymbol{x}_t^{i,k}$ 可由字典 D_t^k 线性表示为

$$\boldsymbol{x}_t^{i,k} = D_t^k \boldsymbol{\alpha}_t^{i,k} + \boldsymbol{o} \tag{9}$$

其中, $\boldsymbol{\alpha}_t^{i,k}$ 为 $\boldsymbol{x}_t^{i,k}$ 在字典 D_t^k 上的稀疏编码, 令 $A_t = [\boldsymbol{\alpha}_t^{1,1}, \cdots, \boldsymbol{\alpha}_t^{L,1}, \cdots, \boldsymbol{\alpha}_t^{i,k}, \cdots, \boldsymbol{\alpha}_t^{L,K}] \in \mathbf{R}^{20\times(L\times K)}$ 表示联合稀疏编码矩阵, 则 A_t 可由式 (10) 联合求解

$$\min_{A_t} \sum_{k=1}^{K} \sum_{i=1}^{L} \left\|\boldsymbol{x}_t^{i,k} - D_t^k \boldsymbol{\alpha}_t^{i,k}\right\|^2 + \lambda \sum_{i=1}^{L} \left\|A_t^i\right\|_F \tag{10}$$

其中, $A_t^i = \left(\boldsymbol{\alpha}_t^{i,1}, \cdots, \boldsymbol{\alpha}_t^{i,k}, \cdots, \boldsymbol{\alpha}_t^{i,K}\right)$, 本文利用文献 [24] 中提出的方法对上式进行优化求解. 通过求解式 (10), 每个候选区域均被稀疏编码为一个矩阵, 编码矩阵中对应于目标模板的部分即为候选区域的表观模型, 用 C_t^i 表示第 i 个候选区域的表观模型, 即 $C_t^i = \left[\boldsymbol{c}_t^{i,1}, \cdots, \boldsymbol{c}_t^{i,k}, \cdots, \boldsymbol{c}_t^{i,K}\right] \in \mathbf{R}^{10\times K}$, 其中, $\boldsymbol{c}_t^{i,k}$ 为 $\boldsymbol{\alpha}_t^{i,k}$ 中对应于 D_t^{k+} 的部分.



(a) 初始粒子
(a) The initial particles

(b) 滤波器
(b) The filter

(c) 剩余粒子
(c) The residual particles

图 3　多余粒子的去除

Fig. 3　The elimination of extra particles

为了估计当前的目标状态, 需要计算候选区域与目标模型之间的相似性, 本文从整体和局部两个方面综合衡量二者之间的相似性, 相似性函数定义如下:

$$f\left(C_t^i, B_t\right) = f_h \times f_l = \exp\left(-\beta \times E_t^i\right) \sum_{k=1}^{K} \frac{\min\left(\boldsymbol{c}_t^{i,k}, B_t^k\right)}{B_t^k} \tag{11}$$

其中, $f_h = \exp\left(-\beta \times E_t^i\right)$ 为整体相似性函数, $E_t^i = \sum_{k=1}^{K} \left\| \boldsymbol{x}_t^{i,k} - D_t^{k+} \boldsymbol{c}_t^{i,k} \right\|^2$ 表示第 i 个候选区域的整体重建误差, β 为归一化参数. $f_l = \sum_{k=1}^{K} \min(\boldsymbol{c}_t^{i,k}, B_t^k)/B_t^k$ 为部分相似性函数, 若将目标与候选区域模型视为直方图, 那么 f_l 是利用直方图相交来计算二者之间的相似性. 由于同时考虑整体和局部的相似性, 式 (11) 能更全面地衡量目标模型与候选区域之间的相似性, 为目标状态估计提供更准确的观测模型, 因此更有助于长时间的目标跟踪.

为了更清楚地说明 DLJSM 算法, 本节将该算法的各个步骤进行总结, 并给出流程图 (图 4) 以助于直观理解 DLJSM 算法.

算法 2. DLJSM 算法流程

初始化阶段:

1) 指定目标区域, 在目标区域周围提取目标模板, 并在背景中采样得到背景模板;

2) 将目标模板和背景模板分块, 构造初始字典;

3) 根据式 (3) 构造初始目标模型.

跟踪阶段:

For $i = 1$: FrameNum

a) 根据粒子滤波方法对目标状态进行采样;

b) 根据目标表观变化的连续性去除多余粒子;

c) 根据式 (10) 对剩余粒子进行联合稀疏编码;

d) 根据式 (11) 计算候选区域与目标模型间的相似性;

e) 根据式 (7) 和式 (8) 估计当前目标状态;

f) 更新粒子的重要性权重, 并进行重采样;

If i 是 5 的整数倍

根据第 2.3 节方法更新字典以及目标模型;

End

End

4 实验结果及分析

本节对 DLJSM 算法的有效性进行评估, 为了全面地说明本文算法的跟踪性能, 从定性、定量以及算法复杂度三个角度对 DLJSM 算法进行分析, 并选取 13 种当前目标跟踪领域比较流行的跟踪算法在 12 个测试视频上进行跟踪效果的对比. 实验在 Intel Core2 Duo 2.93 GHz 处理器, 内存为 2.96 GB

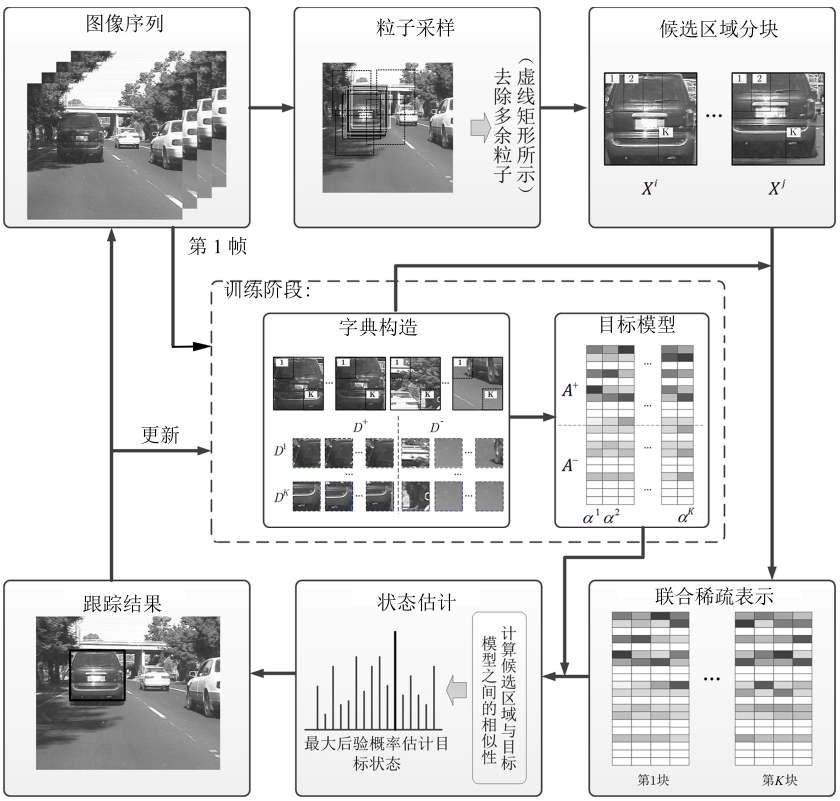


图 4 DLJSM 跟踪算法流程图

Fig. 4 The flowchart of DLJSM tracking algorithm

的计算机上由 Matlab 2011a 执行. 选取的测试视频中包含目标跟踪中的大部分难点, 例如: 遮挡、位姿变化、运动模糊、光流变化、尺度变化以及复杂背景等. 用于对比的跟踪方法可分为三类: 1) 非稀疏的方法: Frag (Fragments-based tracking)[4], VTD[5], IVT[6], MIL[7] 与 TLD[8]; 2) 基于单个稀疏的方法: l_1[9], APG-l_1[10], LSK[14], ALSA[15] 以及 SCM (Sparsity-based collaborative model)[17]; 3) 基于联合稀疏的方法: MTT[18], MTMV[19] 和 DSSM (Discriminative sparse similarity map)[21]. 这些对比方法均为当前跟踪领域内效果较好的方法, 为了公平起见, 本节中所有的跟踪算法均使用作者提供的程序代码, 并赋值相同的初始状态. 除此之外, 对基于贝叶斯推论的方法设置相同的采样个数. DLJSM 算法的实验参数的设置如下: 如果目标区域的初始宽高比小于 1/2, 则将归一化尺寸定义为 24 像素 (宽) × 48 像素 (高); 如果初始宽高比大于 2, 则将归一化尺寸定义为 48 像素 (宽) × 24 像素 (高); 否则将归一化尺寸定义为 32 像素 × 32 像素. 构成字典的小图像块的尺寸为 8 像素 × 8 像素, 相邻图像块间的重叠率为 0.5. 图像块的个数 K 与归一化的尺寸有关, 当归一化的尺寸为 24 像素 × 48 像素或 48 像素 × 24 像素时, $K = 55$; 否则 $K = 49$. 粒子个数为 300, 字典更新的阈值 θ_1 和 θ_2 分别为 50 和 70, 去除多余粒子的阈值 $\varepsilon = 0.5 \times \max(e^i)$, 相似性函数中的归一化参数 $\beta = 0.5$. 本节的实验结果均在上述参数设置下获得.

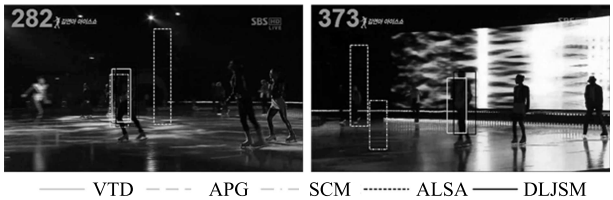
4.1　定性分析

本节对 DLJSM 算法进行定性的分析, 并给出它在各个测试视频上的跟踪结果. 为了在图像中更清晰地显示 DLJSM 算法的跟踪结果, 每个测试视频只给出 14 种方法中效果最好的 5 个结果.

测试 1. 当目标的光流与尺度发生剧烈变化时, DLJSM 算法的跟踪效果. 在图 5 (a) 所示的 Singer1 序列中, 目标的表观由于光流和尺度的剧烈改变而快速变化, 其中 MTMV 在第 9 帧就因为光流忽然变化而失去目标真正的尺度, 并且在后续的跟踪中不能再恢复, 而 VTD, SCM 和 APG-l_1 在跟踪中均有不同程度的偏离. 尽管这几种方法都能成功跟踪目标, 但是在准确性上 DLJSM 算法仍然领先于其他方法, 如第 4.2 节中表 1 和表 2 所示. 图 5 (b) 所示的 Car4 序列中, 目标经历了光线的明暗以及尺度不断变化的过程, 并且伴随由抖动产生的模糊. 由于该视频序列中的目标为刚性物体, 在整个过程中并没有发生形变, 因此模型的更新效果是成功跟踪的一个重要因素. 如果在更新中没有引入过多的累积误差, 那么模型不会产生重大的漂移, 对目标的描述就更加准确. 在 185 帧, 目标由光线明亮的区域驶入阴影区域, SCM, ALSA 和 DSSM 不能正确地捕捉到目标的尺度变化. 从 428 帧可见, DSSM 已经严重偏离了目标的实际位置, 而 SCM 和 ALSA 的跟踪结果区域仍然不准确. DLJSM 算法和 IVT 方法在该视频序列上的良好的跟踪结果得益于二者的模型更新方法. DLJSM 算法规定的每 5 帧判断是否需要更新减少了频繁更新带来的累积误差, 而在更新中保留首帧的目标区域则进一步减少了漂移的可能. 在图 5 (c) 的 Skating1 序列中目标不仅经历了光流与尺度的变化, 同时还存在严重的形变以及遮挡. 从结果图中可见, APG-l_1 方法采用整体稀疏模型建模目标表观, 因此不能很好地处理目标的非刚性形变, 在跟踪开始后不久就失去了目标. 而 SCM 和 ALSA 也在中途失去了目标, 从而跟踪失败. 在 373 帧, 虽然 VTD 方法还能定位目标, 但是偏差较大. 而 DLJSM 方法由于采用局部联合稀疏模型, 同时加入了结构和区分信息, 因此能在整个序列中稳定地跟踪目标.

(a) Singer 1 视频序列
(a) Sequence Singer 1

(b) Car 4 视频序列
(b) Sequence Car 4



(c) Skating 1 视频序列
(c) Sequence Skating 1

图 5　目标的光流与尺度发生剧烈变化时的跟踪结果

Fig. 5　Tracking results when targets undergo drastic changes of illumination and scale

测试 2. 当目标发生巨大形变时, DLJSM 算法的跟踪效果. 图 6 (a) 中目标的位姿不断变化, 并且经历尺度以及光流变化. 在 154 帧, 目标的表观发生

(a) David 视频序列
(a) Sequence David

(b) Girl 视频序列
(b) Sequence Girl

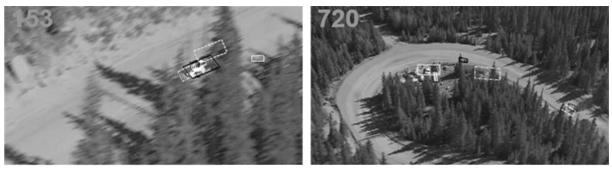
(c) Sylv 视频序列
(c) Sequence Sylv

图 6 目标发生较大形变时的跟踪结果

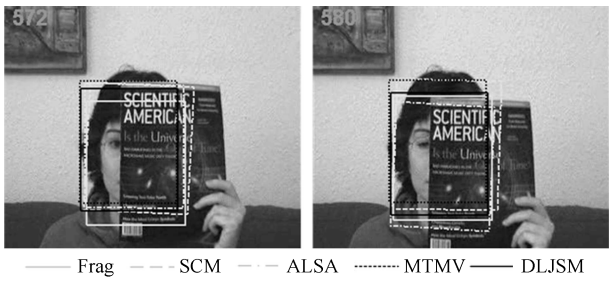
Fig. 6 Tracking results when targets′ appearance deform

很大变化, SCM 和 ALSA 跟踪结果不能准确覆盖目标区域, IVT 和 MTMV 的定位偏离实际位置较大. 从 307 帧可见, SCM 和 ALSA 没有恢复对目标表观的正确描述, 仍然不能准确跟踪到目标区域. 而 DLJSM 算法采用判别性的局部联合稀疏模型建模目标, 使得该模型能够正确捕捉到目标的局部形变; 提出的相似性计算函数则对相邻两帧间的表观相似性衡量的更准确. 因此 DLJSM 算法在整个跟踪过程中都能很好地处理由目标形变带来的表观变化, 准确的定位目标位置以及目标区域. 在图 6 (b) 的 Girl 序列中目标由于旋转而使表观彻底改变, LSK 和 DSSM 对目标表观建立的模型在 210 帧已经产生漂移, 并且随着跟踪进行不能恢复. 而在 330 帧, 除了 LSK 和 DSSM 不能准确跟踪目标外, SCM 的跟踪区域明显小于实际的目标区域, 而 MTMV 方法的跟踪区域则明显大于实际区域. DLJSM 算法则在整个序列中保持稳定准确的跟踪. 对于图 6 (c) 中的视频序列, 目标由于位姿和光流变化而导致表观改变, 尽管这种改变是刚性的, 但在 606 帧, SCM 和 ALSA 偏离了目标的正确位置, 并且在目标的表观恢复后仍然不能准确地跟踪, 如 627 帧所示. 此外, DSSM 由于加入了辨别信息而使得模型能够适应目标早期的表观变化, 但是随着跟踪的进行, 在 627 帧, DSSM 仍然产生了漂移. 而 DLJSM 算法除了加入辨别信息外还将结构信息整合到目标模型中, 因此虽然 DSSM 和 DLJSM 算法都具有最小的位置误差, 但是 DLJSM 算法的跟踪准确性最高.

测试 3. 当目标在跟踪中发生部分或严重遮挡时, DLJSM 算法的跟踪效果. 图 7 (a) 中所示的 Race 视频序列是一个极具挑战性的视频, 目标不仅经历多次的严重遮挡, 还存在光流变化, 巨大的形变以及剧烈的尺度变化. 因此, 该序列对跟踪算法中目标的建模, 模型的更新以及相似性的判定都是极大的考验. 从图 7 (a) 的 153 帧可见, 目标经历严重遮挡, 除了 DLJSM 算法和 DSSM 仍能定位目标外, 其余的方法均失去了目标, 但是 DSSM 方法的定位并不准确. 在目标经历过多次部分或完全遮挡, 以及尺度和位姿变化后, 只有 DLJSM 算法仍然能跟踪上目标, 如 720 帧所示. 对该视频的跟踪结果充分证明了判别性局部联合稀疏模型在描述目标表观上的准确性以及 DLJSM 算法在处理各种跟踪难点问题上的有效性. 图 7 (b) 所示的 Faceocc 视频序列中目标多次发生长时间的部分遮挡, 虽然图中的 5 种方法都能大致跟踪上目标, 但是当发生遮挡时, Frag, SCM, ALSA 和 MTMV 均有不同程度的偏移, 如第 572 帧和 580 帧所示. 相反地, DLJSM 算法在跟踪中一直保持着稳定的结果.



(a) Race 视频序列
(a) Sequence Race



(b) Faceocc 视频序列
(b) Sequence Faceocc

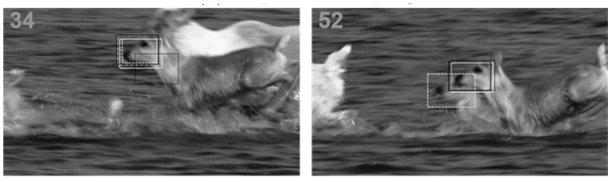
图 7 目标发生遮挡时的跟踪结果

Fig. 7 Tracking results when targets are occluded

测试 4. 当目标处于快速运动并产生运动模糊时, DLJSM 算法的跟踪效果. 图 8 (a) 中目标区

域由于快速跳跃而产生严重的运动模糊. 可见在 43 帧时, DSSM 和 MTMV 严重偏离目标的真实位置, 而在 95 帧仍然没有恢复. 虽然 SCM, ALSA 和 DLJSM 算法都能成功地跟踪目标, 但是 DLJSM 算法的误差更小, 跟踪更准确. 图 8 (b) 的 Animal 序列中, 目标区域不仅存在运动模糊, 而且在背景中有相似区域出现. 在 34 帧, MTT 方法出现漂移, 而在 52 帧, VTD 方法将背景中的相似区域错误地判别为目标. 而 DLJSM 算法在整个过程中都能准确而稳定地跟踪目标, 原因是用来描述目标的判别性局部联合稀疏模型能够有效从背景中区分出目标, 而对所有的候选区域进行联合稀疏编码又进一步的共享了多个任务间的有效信息, 从而使跟踪更加鲁棒.

(a) Jumping 视频序列
(a) Sequence Jumping



(b) Animal 视频序列
(b) Sequence Animal

图 8　目标快速运动时的跟踪结果
Fig. 8　Tracking results when targets undergo rapid movement

测试 5. 当目标处于复杂背景时, DLJSM 算法的跟踪效果. 图 9 (a) 的 Car11 视频序列中, 目标车辆行驶于低照度环境下, 因此背景与目标的分界并不清晰. 图中的 5 种跟踪方法都能成功地跟踪目标, 只有 ALSA 方法在 287 帧附近产生漂移. 对于利用增量子空间方法的 IVT 算法而言, 由于对图像数据进行了归一化以及子空间投影, 因此能够解决目标与背景间的低对比度问题. 而从 SCM, DSSM 和 DLJSM 的跟踪结果可见, 对于利用稀疏表示的方法, 加入判别信息的局部模型能更好地处理这类低对比度问题. 图 9 (b) 的 Stone 视频序列中目标处于杂乱的背景环境下, 并且在背景中存在多个与目标表观相似的区域. 尽管图中的方法都能对目标进行定位, 但是在 521 帧, SCM 和 ALSA 已经出现漂移并在后续跟踪中没有恢复. 而 MTT 方法和 DLJSM 方法的成功跟踪则说明, 联合稀疏编码能够有效地处理由于相似区域而引起的误跟踪.

(a) Car11 视频序列
(a) Sequence Car11

(b) Stone 视频序列
(b) Sequence Stone

图 9　目标处于复杂背景时的跟踪结果
Fig. 9　Tracking results when targets are in complex background

4.2　定量分析

本节采用中心位置误差和重合面积参数对各种跟踪方法的性能进行定量分析. 中心位置误差 PosErr 定义为

$$\text{PosErr} = \sqrt{(c_x^g - c_x^t)^2 + (c_y^g - c_y^t)^2} \tag{12}$$

其中, c_x^t, c_y^t 表示当前跟踪结果的中心位置坐标, c_x^g, c_y^g 表示中心位置坐标的参考值. 重合面积 F_{score} 定义为

$$F_{\text{score}} = \frac{2R_g \cap R_t}{R_g \cap R_t + R_g \cup R_t} \tag{13}$$

其中, R_g 和 R_t 分别表示目标区域的参考值和实际跟踪值.

1) 与基于非稀疏的跟踪方法比较

表 1 给出了 DLJSM 算法与 5 种非稀疏的跟踪方法在测试视频上的平均中心位置误差值和平均重合面积参数, 其中最好的两个结果分别用粗体和斜体标示, "–" 表示跟踪不稳定, 没有连续的跟踪结果. 稀疏表示利用多个模板对目标进行线性表示, 能够对目标特征进行多方面的提取, 相较于非稀疏的方法能够更及时地捕捉目标外观的变化. 而在跟踪过程中, 对稀疏编码的联合求解使得本文的联合稀

疏模型能够更加准确地建模目标表观. 同时, 综合考虑了整体与局部相似的相似性函数为当前的观测模型提供了更全面的衡量结果, 因此 DLJSM 算法的跟踪效果远优于表 1 中的非稀疏方法, 尤其在处理运动模糊 (Jumping) 和剧烈形变 (Race, Girl) 方面.

2) 与基于单个稀疏表示的跟踪方法比较

表 2 给出了 DLJSM 算法与 5 种基于单个稀疏表示方法的对比结果. 由于大量文献已经叙述了基于稀疏的方法在目标跟踪中取得的良好结果, 本节仅从目标表观建模的角度对以上方法进行对比与分析. 在这 5 种方法中, l_1 和 APG-l_1 方法基于整体稀疏模型; ALSA 和 LSK 方法基于局部稀疏模型; SCM 方法基于稀疏判别模型; 本文提出的 DLJSM 方法基于局部联合稀疏模型, 并且加入结构和判别信息. 由于局部模型在处理部分遮挡以及剧烈形变方面更具有优势, 因此基于局部模型的方法优于基于整体模型的方法; 而基于判别模型的方法则同时利用目标和背景的信息建模, 所以跟踪效果更好. DLJSM 方法综合局部模型和判别模型的优点, 并采用联合稀疏表示建模, 增强了结构性, 因此更适合在跟踪中描述目标表观. 表 2 中的结果证明, 相较于基于单个稀疏的跟踪方法, DLJSM 获得了更好的跟踪结果.

3) 与基于联合稀疏表示的跟踪方法比较

表 3 给出了 DLJSM 算法与基于联合稀疏表示跟踪方法的对比结果. 尽管都是基于多任务联合稀疏表示的方法跟踪目标, DSSM, MTT 与 MTMV 方法使用整体模板构造字典, 因此在跟踪性能上不如本文所使用的局部模型. 从表 3 可见, DLJSM 算法在部分遮挡 (Race, Animal)、剧烈形变以及复杂背景 (Skating1) 等情况下的跟踪结果远优于其他方法.

4) 整体跟踪性能分析

为了更直观地表现包括 DLJSM 算法在内的 14 种方法的跟踪结果, 图 10 (a) 和图 10 (b) 分别给出这

表 1 DLJSM 算法与非稀疏跟踪方法的结果对比

Table 1 Comparison of the results between DLJSM algorithm and the methods not based on sparse representation

	中心误差 (pixel)						*F*-参数					
	IVT	VTD	Frag	MIL	TLD	DLJSM	IVT	VTD	Frag	MIL	TLD	DLJSM
Girl	29.6	*23.8*	81.6	31.3	–	**14.4**	0.703	*0.740*	0.134	0.681	–	**0.836**
Singer1	9.1	*3.7*	42.1	241.0	27.5	**3.2**	0.642	*0.898*	0.394	0.021	0.444	**0.904**
Faceocc	11.2	*9.5*	89.5	18.6	16.0	**6.3**	0.891	0.903	**0.940**	0.838	0.786	*0.938*
Car4	**4.0**	144.8	180.5	142.1	–	*4.5*	*0.937*	0.341	0.263	0.262	–	**0.939**
Sylv	5.9	21.5	45.1	6.9	*5.6*	**5.1**	*0.837*	0.672	0.809	*0.837*	0.835	**0.867**
Race	176.4	*82.2*	221.4	310.6	–	**2.7**	0.025	*0.372*	0.053	0.013	–	**0.721**
Jumping	34.8	111.9	*21.2*	41.8	–	**5.2**	0.273	0.175	*0.429*	0.255	–	**0.787**
Animal	*10.5*	11.8	45.7	252.6	–	**9.7**	0.736	**0.765**	0.120	0.014	–	*0.748*

表 2 DLJSM 算法与基于单个稀疏跟踪方法的结果对比

Table 2 Comparison of the results between DLJSM algorithm and the methods based on single sparse representation

	中心误差 (pixel)						*F*-参数					
	l_1	APG-l_1	SCM	ALSA	LSK	DLJSM	l_1	APG-l_1	SCM	ALSA	LSK	DLJSM
Animal	23.1	23.9	20.2	289.5	*10.2*	**9.7**	0.583	0.619	0.652	0.046	*0.732*	**0.748**
David	20.1	13.7	**9.8**	11.4	11.8	**9.3**	0.605	0.652	*0.759*	0.707	0.713	**0.772**
Car11	33.7	2.9	*2.1*	2.3	73.3	**2.0**	0.501	0.857	*0.895*	**0.897**	0.09	**0.897**
Singer1	5.6	3.8	*3.7*	5.1	7.7	**3.2**	0.780	0.870	**0.910**	0.887	0.742	*0.904*
Race	214.7	203.9	*28.7*	245.5	217.2	**2.7**	0.049	0.059	*0.628*	0.062	0.017	**0.721**
Jumping	38.0	16.4	*6.1*	12.3	63.5	**5.2**	0.256	0.582	*0.767*	0.748	0.214	**0.787**
Skating1	137.5	60.5	*37.0*	64.5	106.4	**8.1**	0.221	0.475	*0.628*	0.580	0.335	**0.789**

表 3　DLJSM 算法与基于联合稀疏表示跟踪方法的结果对比

Table 3　Comparison of the results between DLJSM algorithm and the methods based on joint sparse representation

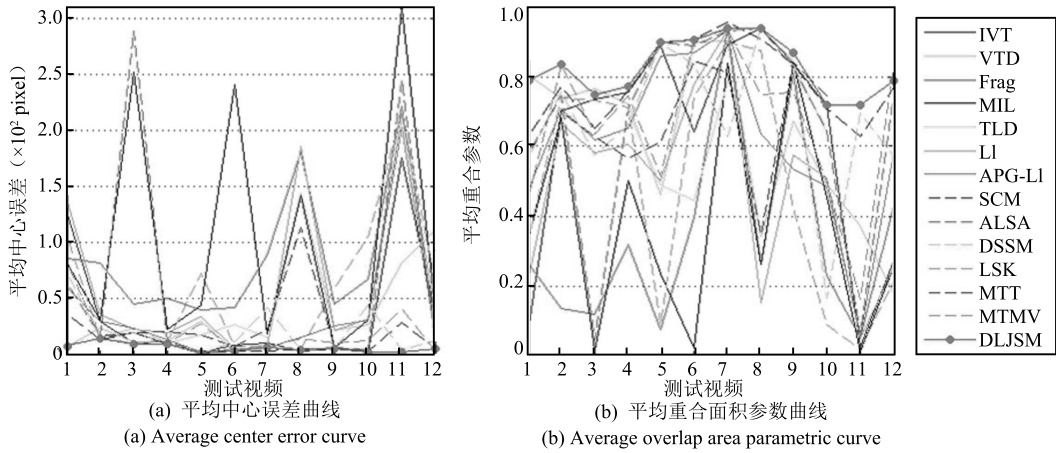
	中心误差 (pixel)				F-参数			
	MTT	MTMV	DSSM	DLJSM	MTT	MTMV	DSSM	DLJSM
Car11	17.4	27.7	**2.0**	**2.0**	0.612	0.514	0.896	**0.897**
David	21.4	10.2	10.4	**9.3**	0.565	0.745	0.663	**0.772**
Race	–	41.2	4.3	**2.7**	–	0.163	0.695	**0.721**
Skating1	–	81.9	73.8	**8.1**	–	0.451	0.569	**0.789**
Animal	19.4	19.5	23.7	**9.7**	0.630	0.635	0.574	**0.748**
Stone	3.3	12.5	43.9	**2.8**	**0.746**	0.50	0.166	0.720

些算法在全部测试视频上的平均中心位置误差和平均重合面积参数. 平均中心误差越小表示跟踪结果越准确, 而平均重合面积参数越大则表示跟踪效果越好. 横轴序号代表的视频序列依次为: 1 代表 Skating1, 2 代表 Girl, 3 代表 Animal, 4 代表 David, 5 代表 Car11, 6 代表 Singer1, 7 代表 Faceocc, 8 代表 Car4, 9 代表 Sylv, 10 代表 Stone, 11 代表 Race, 12 代表 Jumping. 图 10 中带有标记符的曲线表示 DLJSM 方法的跟踪结果, 可见其在图 10 (a) 中几乎全部处于最下方, 而在图 10 (b) 中则大部分都处于最上方, 图 10 证明 DLJSM 方法的跟踪效果优于其他方法.

除了上述平均跟踪结果, 图 11 还给出了 benchmark[2] 中用于衡量整体跟踪性能的曲线: OPE (One-pass evaluation) 的曲线下面积 (Area under curve, AUC). 其中图 11 (a) 表示精确度曲线, 横轴为误差阈值, 纵轴为误差小于阈值的帧数占总帧数的百分率. 图 11 (b) 表示成功率曲线, 横轴为跟踪成功阈值, 若跟踪面积与参考面积的比值大于阈值, 则认为跟踪成功, 纵轴表示成功跟踪的帧数占总帧数的百分率, 成功率中对重合参数的定义与文献 [2] 相同, 即 $(R_g \cap R_t)/(R_g \cup R_t)$. 从图 11 可知, 当误差阈值设为 20 时, DLJSM 算法的跟踪精确度达到 97 %; 当成功跟踪阈值设为 0.6 时, 它的成功跟踪率达到 77 %. 这说明 DLJSM 算法在整个跟踪过程中能够准确稳定的跟踪目标.

4.3　算法的复杂度分析

从算法复杂度的角度分析, DLJSM 算法的耗时主要集中在联合稀疏编码的求解上. 由于本文采用文献 [24] 中提出的算法计算联合稀疏编码, 因此 DLJSM 算法的算法复杂度与其相同, 均为 $\mathrm{O}(KLNm^2 + 2TKLNm^2)$, 其中 K 为字典的个数, L 为滤除多余粒子后保留下来的粒子个数, N 为整体模板的个数, m 为局部图像块的尺寸, T 为算法的迭代次数. 实际的运行中, 为了提高算法的运算速度, 在编写 DLJSM 算法的代码时采用 Matlab 自带的并行计算功能, 同时将计算联合稀疏编码的迭代次数设为 20, 文中所有实验结果均在此设置下得到. 在本台计算机上, 针对所有测试视频, 跟踪每帧需要

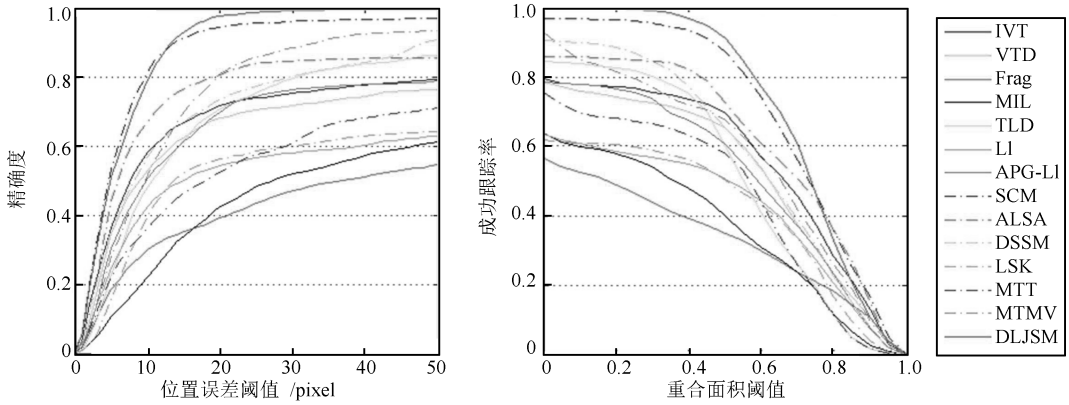


(a) 平均中心误差曲线　(a) Average center error curve

(b) 平均重合面积参数曲线　(b) Average overlap area parametric curve

图 10　所有跟踪方法在全部测试视频上的跟踪性能

Fig. 10　Performance of all the tracking methods in test sequences



(a) 跟踪精度的统计曲线
(a) Precision plots of OPE

(b) 跟踪成功率的统计曲线
(b) Success plots of OPE

图 11 OPE 曲线

Fig. 11 One-pass evaluation curves

的处理时间平均为 0.38 s.

5 总结

本文提出一种基于判别性局部联合稀疏表示模型的多任务目标跟踪方法 (DLJSM). 该方法在字典中加入背景模板而使模型具有判别性, 并采用联合稀疏编码增强模型的结构性, 因此相较于其他基于稀疏表示的目标模型, 本文提出的模型具有更好的判别能力和描述能力. 为了提高跟踪效率, DLJSM 对采样得到的候选区域根据目标表观变化的连续性进行预处理并对剩余的候选区域进行联合稀疏编码. 文中定义的相似性函数综合考虑候选区域与目标模型在整体结构与部分特征上的相似性, 因此能够为状态估计提供更准确的观测模型. 此外, DLJSM 采用的更新方法能够有效地捕捉目标的表观变化却不会引入过多的累积误差而导致模型漂移, 因此更适合跟踪任务. 与 13 种跟踪算法的对比结果证明 DLJSM 跟踪方法的高效性.

References

1 Yilmaz A, Javed O, Shah M. Object tracking: a survey. *ACM Computing Surveys (CSUR)*, 2006, **38**(4): Article No. 13

2 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. Portland, USA: IEEE, 2013. 2411−2418

3 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

4 Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 798−805

5 Kwon J, Lee K M. Visual tracking decomposition. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 1269−1276

6 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

7 Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami FL, USA: IEEE, 2009. 983−990

8 Kalal Z, Matas J, Mikolajczyk K. P-N learning: bootstrapping binary classifiers by structural constraints. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 49−56

9 Mei X, Ling H B. Robust visual tracking using L_1 minimization. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 1436−1443

10 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. Providence, USA: IEEE, 2012. 1830−1837

11 Zhang S P, Yao H X, Zhou H Y, Sun X, Liu S H. Robust visual tracking based on online learning sparse representation. *Neurocomputing*, 2013, **100**: 31−40

12 Wang D, Lu H C, Yang M H. Online object tracking with sparse prototypes. *IEEE Transactions on Image Processing*, 2013, **22**(1): 314−325

13 Wang L F, Yan H P, Lv K, Pan C H. Visual tracking via kernel sparse representation with multikernel fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014, **24**(7): 1132−1141

14 Liu B Y, Huang J Z, Yang L, Kulikowsk C. Robust tracking using local sparse appearance model and k-selection. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2011. 1313−1320

15 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. Providence, USA: IEEE, 2012. 1822−1829

16 Xie Y, Zhang W S, Li C H, Lin S Y, Qu Y Y, Zhang Y H. Discriminative object tracking via sparse representation and online dictionary learning. *IEEE Transactions on Cybernetics*, 2014, **44**(4): 539−553

17 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. Providence, USA: IEEE, 2012. 1838−1845

18 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. Providence, USA: IEEE, 2012. 2042−2049

19 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. Sydney, NSW: IEEE, 2013. 649−656

20 Dong W H, Chang F L, Zhao Z J. Visual tracking with multifeature joint sparse representation. *Journal of Electronic Imaging*, 2015, **24**(1): 013006

21 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

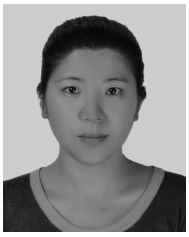
22 Zhang T Z, Liu S, Xu C S, Yan S C, Ghanem B, Ahuja N, Yang M H. Structural sparse tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 150−158

23 Wang Meng. Multi-Task Visual Tracking Using Composite Sparse Model [Master dissertation], Shanghai Jiao Tong University, China, 2014.
(王梦. 基于复合稀疏模型的多任务视频跟踪算法研究 [硕士学位论文], 上海交通大学, 中国, 2014.)

24 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

25 Doucet A, de Freitas N, Gordon N. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

26 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



黄丹丹 大连理工大学信息与通信工程学院博士研究生. 2007 年获得长春理工大学学士学位. 主要研究方向为视频序列中的目标检测与目标跟踪.
E-mail: dlut_huang@163.com
(**HUANG Dan-Dan** Ph. D. candidate at the School of Information and Communication Engineering, Dalian University of Technology. She received her bachelor degree from Changchun University of Science and Technology in 2007. Her research interest covers object detection and object tracking in video sequences.)

孙 怡 大连理工大学信息与通信工程学院教授. 1986 年获得大连理工大学学士学位. 主要研究方向为图像处理, 模式识别与无线通信. 本文通信作者.
E-mail: lslwf@dlut.edu.cn
(**SUN Yi** Professor at the School of Information and Communication Engineering, Dalian University of Technology. She received her bachelor degree from Dalian University of Technology in 1986. Her research interest covers image processing, pattern recognition, and wireless communication. Corresponding author of this paper.)