

基于残差矩阵估计的稀疏表示目标跟踪算法

陈典兵 朱明 高文 王慧利 杨航

Visual tracking based on the estimation of representation residual matrix

Chen Dian-Bing Zhu Ming Gao Wen Wang Hui-Li Yang Hang

引用信息 Citation: *Acta Physica Sinica*, 65, 194201 (2016) DOI: 10.7498/aps.65.194201

在线阅读 View online: <http://dx.doi.org/10.7498/aps.65.194201>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2016/V65/I19>

您可能感兴趣的其他文章

Articles you may be interested in

基于激光散斑成像的零件表面粗糙度建模

Surface roughness modeling based on laser speckle imaging

物理学报.2015, 64(23): 234203 <http://dx.doi.org/10.7498/aps.64.234203>

基于指导滤波的图像盲复原算法

Guided filter-based blind image restoration method

物理学报.2015, 64(13): 134202 <http://dx.doi.org/10.7498/aps.64.134202>

基于尺度不变特征变换和区域互信息优化的多源遥感图像配准

Multi-source remote sensing image registration based on scale-invariant feature transform and optimization of regional mutual information

物理学报.2015, 64(12): 124204 <http://dx.doi.org/10.7498/aps.64.124204>

基于多相组重建的航空图像超分辨率算法

Super resolution of aerial image by means of polyphase components reconstruction

物理学报.2015, 64(11): 114208 <http://dx.doi.org/10.7498/aps.64.114208>

编码孔径光谱成像仪光学简化彗差对图谱反演误差分析

Analysis on the simplified optic coma effect on spectral image inversion of coded aperture spectral imager

物理学报.2015, 64(5): 054205 <http://dx.doi.org/10.7498/aps.64.054205>

基于残差矩阵估计的稀疏表示目标跟踪算法*

陈典兵¹⁾²⁾ 朱明^{1)†} 高文¹⁾ 王慧利¹⁾²⁾ 杨航¹⁾

1) (中国科学院长春光学精密机械与物理研究所, 长春 130033)

2) (中国科学院大学, 北京 100039)

(2016年4月5日收到; 2016年7月14日收到修改稿)

基于稀疏表示的目标跟踪算法多数利用稀疏系数计算目标位置信息, 而忽略了稀疏表示过程中的残差所包含的信息. 因此, 本文设计了一种基于残差矩阵估计的跟踪模型. 该模型在粒子滤波的框架下利用 L_1 范数分别约束稀疏表示系数与残差矩阵, 并且利用 L_2 范数建立残差矩阵与观测模型之间的联系. 本文给出了相应求解模型的表示系数与残差矩阵的迭代算法, 并利用残差矩阵更新模板字典. 相比应用稀疏系数的跟踪算法, 本文算法考虑了残差矩阵对跟踪结果的影响, 使得算法对于候选目标的评估更加精确, 同时在模板更新部分引入残差矩阵, 使得字典能够更好地描述目标的变化. 实验数据表明, 本文算法优于现今主流算法.

关键词: 稀疏表示, 正则化, 残差矩阵, 目标跟踪

PACS: 42.30.Va, 42.30.Tz, 07.05.Pj, 07.05.Rm

DOI: 10.7498/aps.65.194201

1 引言

目标跟踪技术一直是计算机视觉领域的研究热点之一^[1], 其研究的主要目的是设计可应用于现实场景, 有效克服各种跟踪障碍的实时跟踪算法. 目前, 目标跟踪技术已经取得了长足的进步^[2-16], 且目标跟踪技术已在多个领域被广泛地应用. 但仍然很难有效处理跟踪过程中遇到的问题, 如: 目标快速运动和运动模糊, 目标遮挡, 背景光照变化, 背景混杂等干扰因素.

近年来, 稀疏表示理论已成功应用于图像处理的多个领域, 如图像分类^[2]、图像显著性检测^[3]、图像融合^[4]等领域. Mei和Ling^[5]首次提出将稀疏表示应用在目标跟踪领域, 他们提出一种 L_1 范数最小化优化的目标跟踪方法, 该算法在粒子滤波的框架下, 利用包含目标模板和遮挡模板的字典表示每个粒子区域, 通过求解具有非负约束的 L_1 范数优化问题得到粒子的稀疏系数. Liu等^[6]求得局部图像块的稀疏表示, 然后在均值漂移 (mean-shift)

框架下求解优化公式 Jia等^[7]提出了一种结构化的局部稀疏表示模型, 该模型利用池化技术综合了目标的局部和全局信息. Liu和Sun^[8]提出了一种基于反向稀疏表示模型的跟踪方法, 利用粒子和遮挡模板组成字典, 反向表示一个静态的目标模板, 进而求解 L_1 范数最小化问题. 王保宪等^[9]设计了一种结合正向稀疏表示与反向稀疏表示的跟踪算法. Liu等^[10]设计了同时具有稀疏性和判别性的特征来表示跟踪目标. Wang等^[11]通过求解局部图像块的稀疏系数, 并利用线性分类器区分前景和背景, 可以有效克服背景混杂问题. Bao等^[12]在文献^[5]的基础上利用加速逼近梯度 (accelerated proximal gradient, APG) 算法求解 L_1 范数优化公式, 进一步提高了跟踪精度.

本文提出了一种基于残差矩阵估计稀疏表示模型, 并在粒子滤波框架下实现目标跟踪. 本文与其他研究的不同点在于: 读取图像后, 本文利用部分奇异值重构跟踪图像, 有效减少跟踪图像的混杂干扰信息; 基于残差矩阵估计的稀疏表示模型可有效估计每个粒子相应的残差矩阵; 更新部分采用残

* 国家自然科学基金 (批准号: 61401425) 和吉林省科技发展计划青年科研基金 (批准号: 20150520057JH) 资助的课题.

† 通信作者. E-mail: zhu_mingca@163.com

差矩阵更新字典, 并保留目标首帧真值, 可以有效保留目标初始真值信息. 本文的主要内容如下: 第2部分简述了粒子滤波与稀疏表示的基本原理; 第3部分详细描述了本文设计的稀疏表示模型并展示了相应的求解方法; 第4部分展示了本文算法与5种现今主流算法在8个不同序列上实验结果, 比较了平均覆盖率与平均中心误差. 本文算法的跟踪精度优于5种现今主流的目标跟踪算法.

2 基本原理

2.1 粒子滤波

粒子滤波^[13]是通过估计变量的后验概率分布来描述一个动态随机系统. 定义第 t 帧的目标观测集 $\mathbf{Z}_t = [z_1, z_2, \dots, z_t]$, 目标状态 $\mathbf{x}_t$. 那么第 t 帧的目标, 可以通过最大后验估计 (maximum a posteriori, MAP) 计算得出:

$$\hat{\mathbf{x}}_t = \underset{\mathbf{x}_t^i}{\operatorname{argmax}} p(\mathbf{x}_t^i | z_t), \quad (1)$$

式中 $\mathbf{x}_t^i$ 表示第 i 个粒子的状态.

后验概率 $p(\mathbf{x}_t | z_t)$ 可以通过贝叶斯定理递归推出:

$$p(\mathbf{x}_t | z_t) \propto p(z_t | \mathbf{x}_t) \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) \times p(\mathbf{x}_{t-1} | z_{t-1}) d\mathbf{x}_{t-1}, \quad (2)$$

式中 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ 是两个连续帧之间目标状态的转移模型. 本文采用6个独立的仿射参数描述目标状态, 即 $\mathbf{x}_t = \{r, c, \theta, s, \alpha, \phi\}$, 分别对应着目标的水平位移、垂直位移、角度、尺度、宽高比和倾斜角. 由于转移模型可以用高斯状态分布描述, 因此 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ 可表示为

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{x}_t; \mathbf{x}_{t-1}, \mathbf{\Omega}), \quad (3)$$

其中 $\mathbf{\Omega}$ 表示对角矩阵, 它的元素为状态变量 $\mathbf{x}_t$, 其六个仿射参数的标准差为

$$\mathbf{\Omega} = \operatorname{diag}(\sigma_r^2, \sigma_c^2, \sigma_\theta^2, \sigma_s^2, \sigma_\alpha^2, \sigma_\phi^2). \quad (4)$$

利用如上方法, 根据不同粒子状态 $\mathbf{x}_t^i$, 可得到第 t 帧的 n 个候选粒子 $\mathbf{Y}$, 即

$$\mathbf{Y} = [y_1, y_2, \dots, y_n] \in \mathbf{R}^{d \times n}, \quad (5)$$

其中 d 表示候选粒子样本的特征维度, n 是粒子的数量. (2)式中, $p(z_t | \mathbf{x}_t)$ 反映了目标的状态 $\mathbf{x}_t$ 与观测状态 z_t 之间的相似度. 本文中 $p(z_t | \mathbf{x}_t)$ 反比于粒子重构误差的大小.

2.2 稀疏表示理论^[17]

若给定模板字典 $\mathbf{D} = [d_1, d_2, \dots, d_m] \in \mathbf{R}^{d \times m} (m \gg d)$, 字典中包含 m 个模板图像, 每个图像为 $d_q \in \mathbf{R}^{d \times 1}$, $q \in (1, 2, \dots, m)$. 通过粒子滤波得到的粒子经过向量化后可表示为 $y_i \in \mathbf{R}^{d \times 1}$, $i \in (1, 2, \dots, n)$. y_i 可以用模板字典线性表示为

$$\mathbf{y}_i \approx \mathbf{D} \cdot \mathbf{c}_i = c_{i1}d_1 + c_{i2}d_2 + \dots + c_{im}d_m$$

$$i \in (1, 2, \dots, n), \quad (6)$$

其中, $c_i \in \mathbf{R}^{m \times 1}$ 是用模板字典 $\mathbf{D}$ 线性表示粒子 y_i 时的系数. 由于 $(m \gg d)$, (6)式是欠定方程, 有无数个解. 为了在这无数个解中找到最稀疏的解, 需要为系数限定稀疏性质. 优化(7)式可以达到求解稀疏表示系数 c 目的:

$$\min_{c_i} \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}c_i\|_2^2 + \lambda \|\mathbf{c}_i\|_0, \quad (7)$$

(7)式中符号 λ 是一个标量. 由于(7)式中 L_0 范数的非连续性, 无法直接求解, Donoho和Tsaig^[18]提出 L_1 范数的解可以逼近 L_0 范数问题的解. 因此, (7)式可重写为

$$\min_{c_i} \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}c_i\|_2^2 + \lambda \|\mathbf{c}_i\|_1. \quad (8)$$

本文中, 字典 $\mathbf{D}$ 的构建是由首帧以高斯分布选取的 m 个样本图像组成, $\mathbf{D}$ 中的每列元素 d_q , $q \in (1, 2, \dots, m)$, 对应相应的样本图像的列向量化后的结果.

3 本文算法

3.1 模型建立

自从2006年文献^[17, 18]完成了较为严密的工作之后, 压缩感知与稀疏表示算法被广泛应用于计算机视觉. 在很多计算机视觉的应用中, 测试图像很可能被遮挡或者损毁, 所以(6)式应该更加准确地被描述为

$$\mathbf{y}_i = \mathbf{D} \cdot \mathbf{c}_i + \mathbf{e}_i, \quad (9)$$

式中是 $\mathbf{e}_i \in \mathbf{R}^{d \times 1}$ 是残差向量. 当模板图像集合不能准确描述测试图像时, $\mathbf{e}_i$ 可以描述这个残差. 由于目标在运动过程, 只有图像中一小部分被遮挡或损毁, 相应的残差向量具有稀疏的性质^[19]. 因此, 考虑残差向量可提高粒子重构误差的计算精度, 本文设计了一种包含残差矩阵估计的稀疏表示模型. 该模型利用 L_1 范数约束残差向量的稀疏性, 并且

利用 L_2 范数构建残差向量与观测模型之间的关系.

$$\min_{c_i, e_i} \frac{1}{2} \|y_i - Dc_i - e_i\|_2^2 + \lambda_1 \|c_i\|_1 + \lambda_2 \|e_i\|_1, \quad (10)$$

其中 i 表示粒子的标号 $i \in (1, 2, \dots, n)$, y_i 表示每个粒子 i 的向量化图像, 字典 D 由 2.2 节描述的方式构建, c_i 是 y_i 利用字典 D 表示的稀疏系数, e_i 为相应粒子 i 的残差向量, λ_1 是调节 c_i 稀疏度的标量, λ_2 是调节残差向量稀疏度的标量.

写成矩阵形式:

$$\min_{C, E} \frac{1}{2} \|Y - DC - E\|_2^2 + \lambda_1 \|C\|_1 + \lambda_2 \|E\|_1, \quad (11)$$

其中 $C = [c_1, c_2, \dots, c_n] \in \mathbf{R}^{m \times n}$ 是系数矩阵, $E = [e_1, e_2, \dots, e_n] \in \mathbf{R}^{d \times n}$ 是残差向量组成的残差矩阵, $Y = [y_1, y_2, \dots, y_n] \in \mathbf{R}^{d \times n}$ 是粒子群的观测矩阵, 每一列是一个粒子的观测向量.

求解优化问题 (10) 可以得到粒子群的系数矩阵与残差矩阵. 由于 (11) 式是一个非确定多项式难题 (non-deterministic polynomial hard, NP-hard), 所以无法直接一步求解, 因此需要利用迭代的方法逐个求解变量.

3.2 模型求解 [20]

由于优化问题 (10) 是一个 NP-hard 的问题, 无法直接求得系数矩阵 C 和残差矩阵 E , 但是若 C 和 E 中, 固定一个变量, 另一个就可以通过现有方法求解, 因此我们可以通过迭代的方法求解优化问题 (10) [21]

3.2.1 固定 E , 更新 C

固定 E , 优化问题 (10) 就化为 (12) 式的形式:

$$\min_C \frac{1}{2} \|(Y - E) - DC\|_F^2 + \lambda_1 \|C\|_1, \quad (12)$$

优化问题就变为经典的最小绝对收缩与选择 (least absolute shrinkage and selectionator operator, LASSO) 算法模型, LASSO 模型可以通过多种算法求解, 本文采用的求解方法为最小角回归 (least angle regression, LARS) 算法 [22].

3.2.2 固定 C 更新 E

固定 C , 优化问题 (10) 就变为 (13) 式的形式:

$$\min_E \frac{1}{2} \|(Y - DC) - E\|_F^2 + \lambda_2 \|E\|_1. \quad (13)$$

这是一个凸优化问题, 全局最小值可以通过收缩算子 [23]:

$$S_\tau(x) = \text{sgn}(x) \cdot (|x| - \tau). \quad (14)$$

求得残差矩阵 E :

$$E = S_{\lambda_2}(Y - DC). \quad (15)$$

综上所述, 算法通过分别迭代 C 和 E , 就可求得优化问题 (10) 中 C 和 E 最优值.

3.3 相似度测量

获得粒子的系数 C 与残差矩阵 E 后, 我们通过 (16) 式

$$Y' = D \cdot C \quad (16)$$

重构每个粒子组成矩阵 $Y' = [y'_1, y'_2, \dots, y'_n] \in \mathbf{R}^{d \times n}$, 每一列对应相应粒子的重构图像. 根据残差矩阵的每一个元素大小构建相应位置元素的权重矩阵 W , 残差大的元素权重小, 残差小的权重重大

$$\begin{aligned} W_{j,i} &\propto \exp(|E_{j,i}|) \\ i &\in (1, 2, \dots, n), \\ j &\in (1, 2, \dots, d). \end{aligned} \quad (17)$$

W 中每个元素对应的是 E 中相同位置的元素绝对值的自然指数值, 即是图像中相应位置的元素的权重. 将粒子群的观测矩阵 Y 、残差矩阵 E 以及权重矩阵 W 代入 (18) 式, 就可计算每个粒子 i 的重构误差累加和 $Residual(i)$.

$$\begin{aligned} Residual(i) &= \sum_{j=1}^d |(Y_{j,i} - Y'_{j,i} - E_{j,i}) \odot W_{j,i}| \\ j &\in (1, 2, \dots, d), \\ i &\in (1, 2, \dots, n), \end{aligned} \quad (18)$$

其中 $\odot$ 是 **Hadamard** 乘积, 即矩阵对应元素乘积.

那么, 在粒子滤波框架下, 每个 $p(z_t|x_t^i)$ 与 $Residual$ 的值成反比, 即 $Residual$ 中的值越大, 相应粒子的 $p(z_t|x_t^i)$ 值越小. 因此, $p(z_t|x_t^i)$ 与 $Residual$ 的关系可由 (19) 式描述:

$$\begin{aligned} p(z_t|x_t) &= \min_i (Residual) \\ &= \max_i (-Residual). \end{aligned} \quad (19)$$

3.4 字典更新

给定 t 时刻的字典 $\mathbf{D}_t = [d_1, d_2, \dots, d_m] \in \mathbf{R}^{d \times m}$ ($m \gg d$)、当前时刻的定目标 $\mathbf{Target}_t(\hat{i})$ 以及相应的残差向量 $\mathbf{E}(\hat{i})$, $\hat{i}$ 表示被选定为目标候选粒子标号. 利用(20)式测量选定目标与模板向量的欧氏距离, 并依距离值降序排列, 获得顺序索引 $\mathbf{Dictindex} \in \mathbf{R}^{m \times 1}$:

$$\mathbf{Dictindex}(q) = \text{sort}(\|\mathbf{Target}_t(\hat{i}) - d_q\|_2) \\ q \in (1, 2, \dots, m). \quad (20)$$

依据(21)式构建新的模板向量 d_p :

$$d'_p = \alpha_1 \cdot d'_p + \alpha_2 \cdot \mathbf{E}(\hat{i}) + \alpha_3 \cdot \mathbf{Y}'(\hat{i}), \quad (21)$$

其中 p 为 $\mathbf{Dictindex}$ 首元素值, p' 为 $\mathbf{Dictindex}$ 末元素值, $\mathbf{E}(\hat{i})$ 和 $\mathbf{Y}'(\hat{i})$ 分别为残差矩阵 $\mathbf{E}$ 和重构矩阵 $\mathbf{Y}'$ 的第 $\hat{i}$ 列. 其中 $\alpha_1 \in (0, 1)$ 和 $\alpha_3 = 1 - \alpha_1$ 控制字典模板向量描述的准确性与时效性, $\alpha_2 \in (0, 1)$ 控制残差向量中目标细节信息的更新比重, 因此, 本文的参数设置为 $\alpha_1 = 0.5$, $\alpha_2 = 1$, $\alpha_3 = 0.5$. 求得新的模板向量 d'_p 后, 根据 p 和 p' 的取值, 分情况对字典 $\mathbf{D}_t$ 进行更新:

a) 当 $p' = 1$, 代表当前时刻 t 选定的目标与字典的首模板欧氏距离最小(即最相似), 采用 $\mathbf{D}_t(\mathbf{Dictindex}(2)) = d'_p$, $\mathbf{D}_t(\mathbf{Dictindex}(3)) = \mathbf{D}_t(1) = d_1$;

b) 当 $p' \neq 1$, $p = 1$, 代表时刻 t 的目标与字典首模板欧式距离最大(即最不相似), 采用 $\mathbf{D}_t(\mathbf{Dictindex}(2)) = d'_p$;

c) 当 $p' \neq 1$, $p \neq 1$, 代表时刻 t 的目标与字典首模板的欧式距离非最大也非最小, 采用替换方式更新字典 $\mathbf{D}_t(\mathbf{Dictindex}(1)) = d'_p$.

综合以上a), b), c)三种情况, 本文的字典更新策略能保留首帧真值模板, 并适时利用首帧真值模板清理字典中的累积误差, 有效减少目标跟踪过程中的误差累积, 提高了跟踪精度, 实现长时间连续的稳定跟踪. 算法1中给出了本文跟踪算法的简要概述.

算法1 本文跟踪算法流程

输入: 图像 $\mathbf{I}$, 帧号 $Frame$

输出: 目标信息

初始化: $E = 0$, λ_1 , λ_2 , 迭代次数 $iteratenum$, t , d , m , n , 奇异值阈值 L

1) 对图像进行奇异值分解, 取部分奇异值重构图像 $\mathbf{I}'$.

$$[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{SVD}(\mathbf{I}), \quad (22)$$

$\mathbf{\Sigma}'$ 的前 L 个奇异值取 $\mathbf{\Sigma}$ 的前 L 个奇异值, 其余取0; 利用 $\mathbf{\Sigma}'$ 重构跟踪图像:

$$\mathbf{I}' = \mathbf{U} \cdot \mathbf{\Sigma}' \cdot \mathbf{V}. \quad (23)$$

2) 如果 $Frame = 1$: 依照2.2节构建模板字典 $\mathbf{D} = [d_1, d_2, \dots, d_m] \in \mathbf{R}^{d \times m}$.

如果 $Frame \neq 1$: 依照2.2节抽取粒子群 $\mathbf{Y} = [y_1, y_2, \dots, y_n] \in \mathbf{R}^{d \times n}$.

3) 通过(11), (12), (14)式迭代求解 $\mathbf{C}$ 和 $\mathbf{E}$.

4) 依据(14), (15), (16), (17)式确定当前目标.

5) 依据3.4部分更新字典模板.

6) $t = t + 1$.

4 实验结果

本文提出的算法运行在CPU为Intel core i3-3220, 运行内存3.39G, 编译环境为MATLAB2011b的PC机上. 对于每一个实验序列, 在第一帧手动标记目标位置. 算法参数设置为: 图像的模板大小为 32×32 (即 $d = 1024$)像素, 字典 $\mathbf{D}$ 模板数为20个(即 $m = 20$). 为了平衡计算效率与算法的有效性, 我们采用粒子数为600个(即 $n = 600$), 字典的更新方式是帧帧更新. 在所有实验中正则化常数 λ_1 为0.5, λ_2 为0.1, 优化公式求解迭代次数为5次, 我们在8个图像序列上测试算法, 实验序列采用文献[24]中提供的序列以及真值采用随序列提供的真值信息. 表1列出了所有实验序列, 并给出了各

表1 8个图像序列
Table 1. Eight image sequence.

序列名称	序列特点
David	非共面旋转, 尺度变化, 光照变化, 遮挡
Car4	尺度变化, 光照变化
Singer1	尺度变化, 光照变化, 非共面旋转, 遮挡
CarDark	背景混杂, 尺度变化
Dudek	遮挡, 尺度变化, 快速运动, 背景混杂, 姿态变化
BlurCar2	快速运动, 尺度变化, 运动模糊
Walking	尺度变化, 遮挡, 姿态变化
BlurFace	运动模糊, 快速运动, 共面旋转

序列的特点, 本实验只采用图像的灰度信息. 本实验将提出算法与5种现今主流算法做对比, 5种算法分别为: MTT^[14], APG L1^[12], SCM^[16], IVT^[25], TLD^[26]. 为了公平起见, 实验所对比的算法的源码均为原论文中提供的代码, 且在第一帧以相同初始位置标记目标.

4.1 定性分析

我们在八个视频序列的实验中对比了6个算法, 实验部分结果见图4. 首先, BlurFace与BlurCar2序列中的主要的跟踪问题为运动模糊, 在全部跟踪过程中, 仅有本文算法与TLD算法实现了稳定的跟踪, 其他算法均在跟踪过程中丢失目标. 表明本文算法可以很好地处理目标的运动模糊与快速运动的问题. 其次, 在Singer1和Car4序列中, 目标分别出现了较强的光照变化, 在Singer1(102, 142帧)与Car4(192, 232, 315帧)出现明暗变化时,

所有实验算法均能准确预测目标位置, 但当光照变化消失后, 只有本文算法在两个序列上准确预测了目标位置, SCM算法虽在Singer1中实现了稳定跟踪, 但在Car4序列中从第232帧开始丢失目标. 而IVT算法在Singer1中错误抓取了目标的左半身作为目标. 可以看出, 本文算法对于光照的变化具有较强的处理能力. 接下来, 在Dudek, David, Singer1和Walking序列均出现了目标的姿态变化、遮挡以及尺度变化, David(414, 460, 600, 715帧)目标分别出现了姿态变化、遮挡、尺度变化, 本文算法能够较好地处理这些跟踪问题, 实现较准确的稳定跟踪. 在Dudek中, 本文算法能够准确跟踪目标, 始终能够正确标记出目标的位置与大小, 可以看出, 本文算法具有处理目标旋转、尺度变化以及遮挡的能力, 但若目标同时出现多种综合变化, 本文会产生细微的偏差与漂移. 最后, 在CarDark序



图1 (网刊彩色) 八个序列部分实验结果图

Fig. 1. (color online) partial results of eight sequence: (a) David; (b) Car4; (c) Singer1; (d) CarDark; (e) Walking; (f) BlurFace; (g) BlurCar; (h) Dudek.

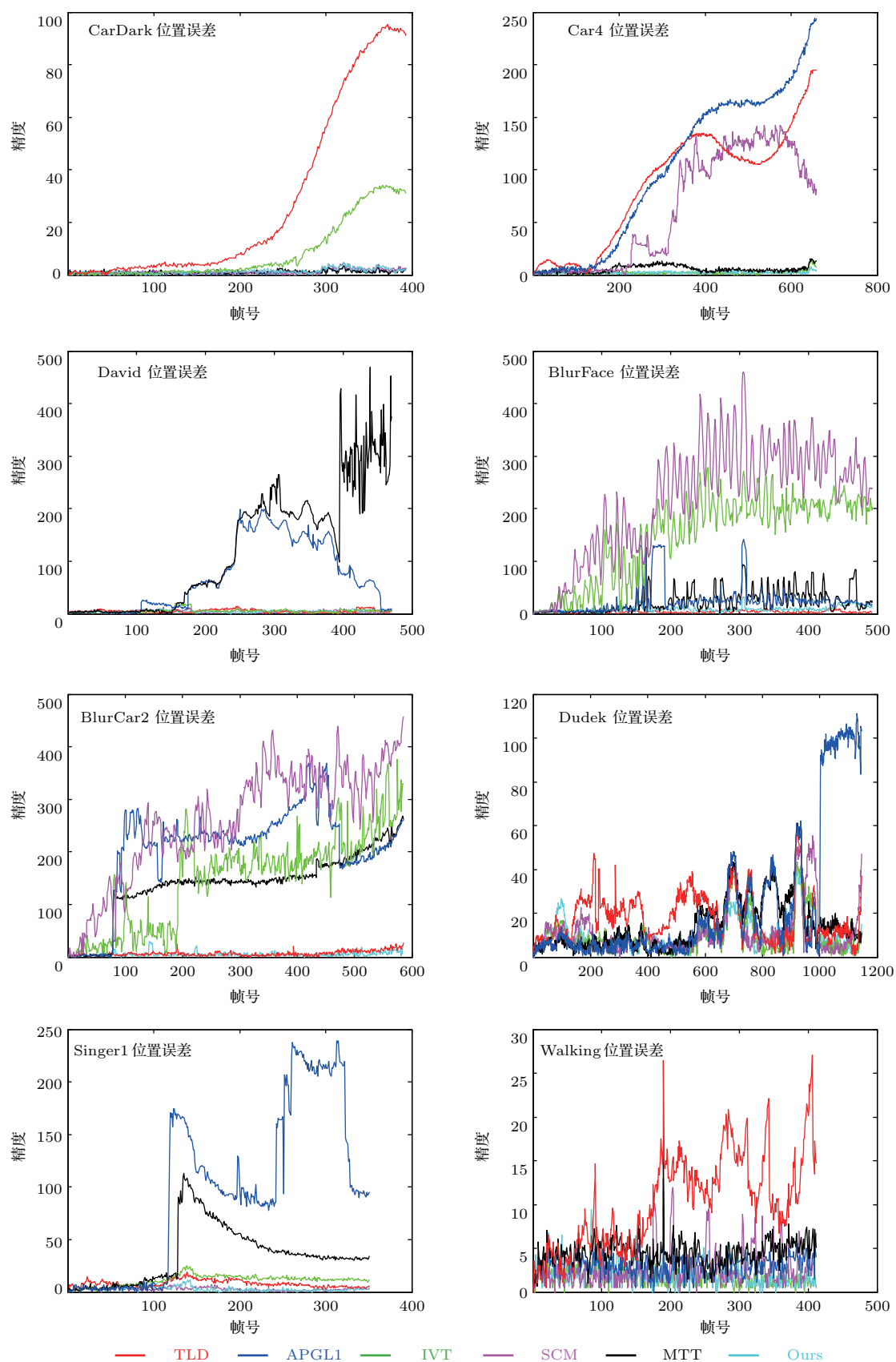


图2 (网刊彩色) 八个序列逐帧像素误差

Fig. 2. (color online) location error at each frame of eight sequence.

列的测试中,除了TLD算法丢失目标,IVT算法产生漂移现象,其他算法均能稳定且准确地跟踪目标.综合以上八个序列的实验结果分析,本文提出的算法可以有效处理光照变化、短时遮挡、目标的共面旋转、非共面旋转以及运动模糊与快速运动等主要跟踪问题.相比于其他5种对比算法,对不同场景中的跟踪问题具有更好的鲁棒性与适应性.

4.2 定量分析

为了评估提出的跟踪器的性能,我们采用两种标准对实验结果进行评估,分别为平均中心误差和平均覆盖率.需要说明的是,较小的平均误差或较高的覆盖率表明实验结果更加精确.在获得每一帧的跟踪结果 $\mathbf{R}_T$ 以及相应的每帧的真值 $\mathbf{R}_G$,我们可通过PASCAL VOC^[27]标准获得覆盖率值:

$$\text{score} = \frac{\text{area}(\mathbf{R}_T \cap \mathbf{R}_G)}{\text{area}(\mathbf{R}_T \cup \mathbf{R}_G)}. \quad (24)$$

表2和表3展示了对比结果的平均中心误差与平均覆盖率,图2展示了6个算法在8个实验序列上的实验结果与目标真值之间的逐帧误差.

表2 平均覆盖率(%)

Table 2. Average overlap rate (%).

视频	APG L1	MTT	SCM	TLD	IVT	Ours
David	24.7	23.7	74.2	71.8	65.2	60.7
Singer1	24.4	34.8	86.4	72.5	57.3	84.7
Dudek	69.2	72.4	75.9	64.7	77.4	78.6
Car4	24.2	53.1	34.2	20.5	86.	87.6
Cardark	88.5	86.6	84.3	44.8	66.3	87.7
BlurFace	62.5	62.6	8.3	88.	14.7	79.6
BlurCar2	12.2	12.2	5.5	73.	17.4	85.8
Walking	74.1	65.7	70.1	44.6	76.6	77.7
Average	47.5	51.4	54.9	60.	57.6	80.3

从表2和表3中可以看出,本文算法虽仅在序列Dudek和Blurcar2实验中为最优算法(平均覆盖率与平均中心误差均为最优),但根据全部序列的平均值统计,本文算法的在8个序列上的实验的平均覆盖率达80.3%,优于第二位TLD算法的60%,而中心误差平均值统计为4.6,优于第二位算法TLD的20.6%.因此,依据结果统计,本文算法相对于对比算法具有更好的跟踪稳定性,可以很好地

适应多种跟踪环境.在图2中,根据各序列的逐帧中心误差曲线可以看出,本文算法在各个序列的误差曲线相对更加平坦,表明本文对于目标的位置信息能够准确地预测.综上所述,本文算法优于其他5种算法.

表3 平均中心位置误差(单位:像素)

Table 3. Average center location error (unit: pixel).

视频	APG L1	MTT	SCM	TLD	IVT	Ours
David	70.6	122.	4.	5.1	5.1	4.6
Singer1	97.5	34.6	2.7	7.9	11.3	2.9
Dudek	23.4	14.5	11.6	18.	9.7	9.6
Car4	101.3	5.5	62.8	86.1	2.1	2.3
Cardark	0.8	1.1	1.3	27.4	8.4	1.4
BlurFace	25.2	23.5	217.4	3.7	148.1	8.7
BlurCar2	203.	138	258.	6.5	147.9	5.2
Walking	3.	4.5	2.6	10.2	1.6	2.1
Average	65.6	43.	70.1	20.6	41.8	4.6

5 结 论

本文设计了一种基于残差矩阵估计的稀疏跟踪模型.该模型在粒子滤波的框架下利用 L_1 范数与 L_2 范数构建包含残差矩阵估计的稀疏表示模型,并且给出了求解模型的迭代算法,最后利用残差矩阵更新模板字典.首先,该算法利用奇异值重构图像滤除混杂信息,减少了图像的干扰信息,使得图像对于跟踪器更加友好;其次,算法采用包含重构稀疏重构误差约束的稀疏模型求解表示系数与重构稀疏重构误差,使得候选目标的评估结果更加准确;最后,字典更新保留了首帧真值模板,使得在跟踪中减少了误差的累积,防止了模板漂移的问题.我们在八种不同场景的实验序列上与5种现今主流算法对比,分别从两种角度分析所提出算法的性能和准确度.根据实验结果可以得出结论:本文提出的算法可在不同的环境下克服多种跟踪障碍,实现稳定的跟踪.相比5种实验对比算法,本文算法具有更好的精度与跟踪鲁棒性.

参考文献

- [1] Gao W, Tang Y, Zhu M 2015 *Acta Phys. Sin.* **64** 014205 (in Chinese) [高文, 汤洋, 朱明 2015 物理学报 **64** 014205]

- [2] Xu Y, Zhang B, Zhong Z F 2015 *Pattern Recogn. Lett.* **68** 9
- [3] Fan Q, Qi C 2016 *Neurocomputing* **175** 81
- [4] Kim M, Han D K, Ko H 2016 *Information Fusion* **27** 198
- [5] Mei X, Ling H B 2009 *Proceedings of IEEE International Conference on Computer Vision* Kyoto, Japan, September 27–October 4, 2009 p1436
- [6] Liu B Y, Huang J Z, Yang L, Kulikowsk C 2011 *Proceedings of IEEE Computer Vision, Pattern Recognition* Colorado, Springs, June 21–25 2011 p1313
- [7] Jia X, Lu H C, Yang M H 2012 *Proceedings of IEEE Computer Vision, Pattern Recognition* Providence, Rhode Island, June 16–21, 2012 p1822
- [8] Liu H P, Sun F C 2010 *Proceedings of International Conference on Pattern Recognition* Istanbul, Turkey, August 23–26, 2010 p1702
- [9] Wang B X, Zhao B J, Tang L B, Wang S G, Wu J H 2014 *Acta Phys. Sin.* **63** 234201 (in Chinese) [王保宪, 赵保军, 唐林波, 王水根, 吴京辉 2014 物理学报 **63** 234201]
- [10] Liu B Y, Yang L, Huang J Z, Meer P, Gong L G, Kulikowski C 2010 *Proceedings of the 11th European Conference on Computer Vision* Crete, Greece, September 5–11, 2010 p624
- [11] Wang Q, Chen F, Xu W L, Yang M H 2012 *Proceedings of IEEE Workshop on Applications of Computer Vision* Breckenridge, CO, January 9–11, 2012 p425
- [12] Bao C L, Wu Y, Ling H B, Ji H 2012 *Proceedings of IEEE Computer Vision, Pattern Recognition* Providence, Rhode Island, June 16–21, 2012 p1830
- [13] Pérez P, Hue C, Vermaak J, Gangnet M 2002 *European Conference on Computer Vision* Copenhagen, Denmark, May 28–31, 2002 p661
- [14] Zhang T Z, Ghanem B, Liu S, Ahuja N 2013 *Int. J. Comput. Vision* **101** 367
- [15] Zhuang B H, Lu H C, Xiao Z Y, Wang D 2014 *IEEE Trans. Image Proces.* **23** 1872
- [16] Zhong W, Lu H C, Yang M H 2012 *Proceedings of IEEE Computer Vision, Pattern Recognition* Providence, Rhode Island, June 16–21, 2012 p1838
- [17] Donoho D L 2006 *IEEE Trans. Inform. Theory* **52** 1289
- [18] Donoho D L, T SA IG Y 2006 *Signal Proces.* **86** 533
- [19] Rao S R, Tron R, Vidal R, Ma Y 2009 *IEEE Trans. PAMI.* **32** 1832
- [20] Wang D, Lu H C 2012 *IEEE Signal Proces. Lett.* **19** 711
- [21] Yan H, Yang J 2016 *Neurocomputing* **173** 1936
- [22] Efron B, Hastie T, Johnstone I, Tibshirani R 2004 *Ann. Statist.* **32** 407
- [23] Hale E T, Yin W, Zhang Y 2008 *SIAM J. Opt.* **19** 1107
- [24] Wu Y, Lim J, Yang M H 2013 *Proceedings of IEEE Computer Vision, Pattern Recognition* Portland, Oregon, June 23–28, 2013 p2411
- [25] Ross D, Lim J, Lin R, Yang M H 2008 *Int. J. Comput. Vision* **77** 125
- [26] Kalal Z, Mikolajczyk K, Matas J 2012 *IEEE Trans. on PAMI* **34** 1409
- [27] Everingham M, Gool L V, Williams C K I, Winn J M, Zisserman A 2010 *Int. J. Comput. Vision* **88** 303

Visual tracking based on the estimation of representation residual matrix^{*}

Chen Dian-Bing¹⁾²⁾ Zhu Ming^{†1)} Gao Wen¹⁾ Wang Hui-Li¹⁾²⁾ Yang Hang¹⁾

1) (*Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*)

2) (*University of Chinese Academy of Science, Beijing 100039, China*)

(Received 5 April 2016; revised manuscript received 14 July 2016)

Abstract

In recent years, sparse representation theory has acquired considerable progress and has extensively been used in visual tracking. Most trackers used the sparse coefficients to merely calculate the position of the target according to the reconstruction error relative to sparse coefficients, and often neglected the information contained by representation residual matrix in representing step. Consequently, we present a novel sparse representation based tracker which takes representation residual matrix into consideration. First of all, at initialization of a new frame, we reconstruct the frame by singular value decomposition (SVD) to eliminate noise and useless information, which contributes a friendly frame for the following representation step. To obtain the compact representation of the target, we build L_2 -norm regularization according to the distance between the candidates wrapped in particle framework and the reconstruction calculated by dictionary templates and residual matrix. Additionally, we use the L_1 -norm constraint to restrict the sparse coefficients and the residual matrix of each candidate. Secondly, as the built optimization problem does not have closed-form solution, we design a method to compute the coefficients and the residual matrix iteratively. During each iteration, the coefficients are obtained by solving classical least absolute shrinkage and selectionator operator (LASSO) model, and the residual matrix is achieved by shrinkage operation. After solving the optimization problem, we compute the score of each candidate for evaluating the truth target with considering coefficients and residual matrix. The score is formulated as weighted reconstruction error which consists of dictionary templates, candidates, coefficients and residual matrix. The weight is the exponential value of absolute value of elements in residual matrix. Finally, for capturing the varying appearance of target in series, we update the dictionary template with assembled template, which is composed of residual matrix, selected candidate and dictionary template. In this paper, the template to be replaced is determined according to the score which is inversely proportional to the distance between the selected candidate and each dictionary template. Then we update the dictionary frame by frame during tracking process. Contributions of this work are threefold: 1) the representation model captures holistic and local features of target and makes the tracker robust to varying illumination, shape transformation, and background clutter, profiting from preprocessing of SVD reconstruction, the model exhibits a more compact representation of target without disturbance of noisy variance; 2) we employ a weight matrix to adjust reconstruction error in candidate evaluation step, as described above, the weight matrix strengthens the effect of error in residual matrix for evaluating candidates from which target is selected, it is noted that weights are all greater than one, which leads to reconstruction error expanding according to the error value of residual matrix, and keeps pixels where there is small error value believable for evaluation; and 3) we adopt an assembled template to update dictionary

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 61401425), and the Science and Technology Development Plan for Youth Foundation of Jilin Province, China (Grant No. 20150520057JH).

[†] Corresponding author. E-mail: zhu_mingca@163.com

template rather than selected candidate. The template updated includes information about both high-score dictionary template and reconstruction of coefficients of selected candidate, which alleviates dictionary degradation and tracking drift problems and provides an accurate description of new appearance of target. In order to illustrate the performance of the proposed tracker, we enforce the algorithm on several challenging sequences and compare the proposed algorithm with five state-of-art methods, whose codes are all supplied by the authors. For complete illustration, both qualitative evaluation and quantitative evaluation are presented in experiment section. Through the experimental results, we could conclude that the proposed algorithm has a more favorable and robust performance than other state-of-art algorithms when dealing with kinds of situations during tracking.

Keywords: sparse representation, regularization, representation residual matrix, visual tracking

PACS: 42.30.Va, 42.30.Tz, 07.05.Pj, 07.05.Rm

DOI: [10.7498/aps.65.194201](https://doi.org/10.7498/aps.65.194201)