

基于多层深度特征的尺度相关粒子滤波算法

马媛媛, 杨小军

(长安大学 信息工程学院, 陕西 西安 710064)

摘要:针对常见的视频目标跟踪算法难以适应形变以及遮挡等一些干扰场景, 该文提出了一种基于多层深度特征的尺度相关粒子滤波跟踪方法。与现有的基于相关滤波器和粒子滤波器的跟踪方法相比, 该算法具有许多优点。首先, 利用区分尺度空间的跟踪器, 考虑了目标对象的尺度变化, 在对目标尺度估计方面表现良好。其次, 通过卷积神经网络提取目标特征, 能够处理目标较大变化和防止跟踪器漂移。最后, 通过尺度相关滤波器将采样的粒子引导至目标状态分布模式, 与传统的粒子滤波算法相比, 该算法能够用较少的粒子有效地覆盖目标状态, 具有鲁棒跟踪性能和较低的计算成本。通过在具有挑战性的基准数据集上的大量实验表明, 该跟踪算法与现有的一些跟踪算法相比具有良好的性能。

关键词: 目标跟踪; 粒子滤波; 相关滤波; 尺度变化; 卷积神经网络

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2021)06-0169-06

doi: 10.3969/j.issn.1673-629X.2021.06.030

A Scale Dependent Particle Filter Algorithm Based on Multi-layer Depth Characteristics

MA Yuan-yuan, YANG Xiao-jun

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: In view of the difficulty of common video target tracking algorithms to adapt to some disturbing scenes such as deformation and occlusion, we propose a scale-dependent particle filter tracking method based on multi-layer depth characteristics. Compared with the existing tracking methods based on correlation filter and particle filter, the proposed algorithm has many advantages. Firstly, by using the scale-space discriminating tracker, the scale variation of the target object is taken into account, which performs well in the estimation of the target scale. Secondly, by using the convolutional neural network to extract the target features, the large change of the target can be processed and the tracker drift can be prevented. Finally, the sampled particles are guided to the target state distribution mode by the scale correlation filter. Compared with the traditional particle filter algorithm, the proposed algorithm can effectively cover the target state with fewer particles, and has robust tracking performance and lower computational cost. A large number of experiments on challenging benchmark data sets show that the proposed algorithm has better performance than some existing tracking algorithms.

Key words: target tracking; particle filtering; correlation filtering; scale change; convolution neural network;

0 引言

计算机视觉^[1-3]已经研究了几十年, 目标跟踪问题通常是在目标运动时估计目标在图像平面上的轨迹问题。虽然在文献中已经对建立鲁棒跟踪框架做了大量的努力, 但当目标出现外观突变或遮挡时, 跟踪器往往会受到影响从而导致跟踪目标失败。

近年来, 基于判别式的相关滤波器被证明能够在跟踪速度上有显著提高且能精准跟踪目标, 因此得到了广泛应用。为了跟踪, 一个相关过滤器通过计算每个学习模板(或过滤器)相对于测试图像样本的点积

来评估相似度。利用卷积定理可以加快相关滤波器的计算速度, 它表明空间域的卷积可以计算为频域傅里叶变换的元素乘。虽然 CSK^[4] 和 KCF^[5] 方法在准确性和鲁棒性方面都达到了最优, 但这些基于相关滤波器的跟踪器并不处理尺度变化。为了处理目标对象的尺度变化, Danelljan 等人^[6]提出了一种新的 DSST 跟踪器, 通过分别学习平移和尺度估计的相关滤波器, 提出一种新的尺度自适应方法来精确估计目标的大小。尽管 DSST 跟踪器在学习基于判别相关滤波器的鲁棒尺度估计方面表现良好, 但在金字塔尺度表示中, 它不

收稿日期: 2020-06-20

修回日期: 2020-10-23

基金项目: 国家自然科学基金项目(61473047)

作者简介: 马媛媛(1999-), 女, 硕士研究生, 通信作者, 研究方向为计算机视觉单目标跟踪; 杨小军, 博士/博士后, 教授, 研究方向为多源交通信息融合、统计信号处理等。

能很好地处理局部遮挡和完全遮挡。

另一方面,粒子滤波^[7-8]可用于处理大尺度变化和局部遮挡。通过增加粒子数来提高跟踪性能。但是,由于计算量的增加,跟踪效率会降低。将外观自适应模型^[9]集成到粒子滤波框架中,其中粒子数取决于噪声方差。与之前的方法不同,粒子滤波框架的主要区别在于观测模型,在观测模型中采用相关滤波器来减少粒子的数量。

然而,由于粒子滤波采样过程计算量大,其实时应用受到限制。针对以上问题,该文提出了一种新的目标跟踪算法,用尺度相关滤波器来训练卷积神经网络特征,通过对所获得的置信图进行加权融合,以得出目标的位置。该算法有效克服了上述问题,具有较强的鲁棒性。

1 粒子滤波器

提出的粒子滤波框架是基于贝叶斯顺序的重要性抽样技术,该方法利用有限的加权样本集递归逼近后验分布,以估计动态系统的状态变量的后验分布。对象在时刻 t 状态变量的参数为 s^t, y^t 。 s^t 表示目标状态变量, y^t 表示观测变量。因此,视频跟踪可以建模为这样一个问题:

$$s^t = p(s^t | y^{1:t-1}) = \int p(s^t | s^{t-1}) p(s^{t-1} | y^{1:t-1}) ds^{t-1} \quad (1)$$

当一个新的观测变量产生时,用贝叶斯定理更新目标状态:

$$s^t = p(s^t | y^{1:t}) = \frac{p(y^t | s^t) p(s^t | y^{1:t-1})}{p(y^t | y^{1:t-1})} \quad (2)$$

其中, $p(y^t | s^t)$ 表示观测的可能性。 $p(s^t | y^{1:t})$ 由带有归一化权值 $\{\omega_i^t\}_{i=1}^n$ 的粒子集 $\{s_i^t\}_{i=1}^n$ 近似:

$$p(s^t | y^{1:t}) \approx \sum_{i=1}^n \omega_i^t \delta(s^t - s_i^t) \quad (3)$$

ω_i^t 是与粒子相关的重要权重,所以定义了一个重要分布 $q(s^t | s^{1:t-1}, y^{1:t})$ 来近似 $p(s^t | y^{1:t})$ 。这里,权值更新为:

$$\omega_i^t \propto \omega_i^{t-1} \cdot \frac{p(y^t | s_i^t) p(s_i^t | s_i^{t-1})}{q(s^t | s^{1:t-1}, y^{1:t})} \quad (4)$$

在粒子滤波器中,一阶马尔可夫过程 $q(s^t | s^{t-1})$ 经常被用来描述重要性采样 $q(s^t | s^{1:t-1}, y^{1:t})$, 因此 $\omega_i^t = \omega_i^{t-1} p(y^t | s_i^t)$ 。由于粒子退化问题的存在,在更新阶段引用重采样技术,重采样步骤派生出的粒子依赖上一步的权值,然后所有的新粒子会接收一个初始权值,粒子权值被设置为: $\omega_i^t = \frac{1}{n}$, 将由下一帧似然函数更新。

2 尺度相关粒子滤波器

2.1 尺度相关滤波器

KCF 跟踪^[5]的基本思想是在探索循环矩阵结构的同时,利用大量的负样本来增强检测器跟踪的鉴别能力,以获得高效率。KCF 跟踪器使用在 $P \times Q$ 像素的图像 x 上训练的相关滤波器 w 来建模目标对象的外观,所有的循环移位 $X_{p,q}$, 生成高斯函数标签 $r_{p,q}$ 的训练样本。目的是找到最优权重 w 。

$$w = \underset{w}{\operatorname{argmin}} \sum_{p,q} |\langle f(x_{p,q}), w \rangle - r_{p,q}|^2 + \lambda \|w\|^2 \quad (5)$$

其中, f 是到内核空间的映射, λ 是一个正则化参数。利用快速傅里叶变换 (FFT) 计算相关性,将目标函数最小化为 $w = \sum_{p,q} \alpha(p,q) f(x_{p,q})$, 求出的系数为:

$$\alpha = F^{-1} \left(\frac{F(r)}{F(\langle f(x), f(x) \rangle) + \lambda} \right) \quad (6)$$

F, F^{-1} 分别表示傅里叶变换及其反变换, 其中 $r = \{r(p,q)\}$ 。给定学习的 α 和目标外观模型 x , 通过计算响应映射, 在新的一帧中进行任务跟踪。

通过使用基于 KCF 跟踪提出的区分尺度空间的跟踪器, 计算图像位置尺度维度的分数。设 S 为尺度滤波器的大小, 提取一个以 $P \times R$ 像素目标中心的图像补丁 I_n , 将训练样本 $f_{t,\text{scale}}$ 在规模水平 n 下的值 $f_{t,\text{scale}}(n)$ 设为 I_n 的 d 维特征描述符, 然后构建尺度样本 $f_{t,\text{scale}}$, 去训练一个尺度滤波器 $h_{t,\text{scale}}$ 。

$$h_{t,\text{scale}} = \frac{\bar{G} F^t}{\sum_{k=1}^d \bar{F}^k F^k + \lambda} \quad (7)$$

式中, G, F 是 g, f 的频域, g 是训练样本期望, f 是图像特征; \bar{G}, \bar{F} 为 G, F 的共轭复数。

为了估计目标的平移量, 采用标准平移滤波器。通常情况下, 两帧之间的目标尺度差相对于平移的差异较小。因此对给定的新的一帧, 首先应用平移过滤器 $h_{t,\text{trans}}$, 然后用尺度滤波器 $h_{t,\text{scale}}$ 估计目标新位置, 用提取训练样本的方法在该位置提取一个尺度训练样本 $z_{t,\text{scale}}$, 通过在傅里叶域计算最大尺度相关性得分, 得到最终的尺度估计结果来估计当前目标状态。

$$Y_t = \frac{\sum_{l=1}^d \overline{A_{t-1}^l} Z_t^l}{B_{t-1} + \lambda} \quad (8)$$

其中, A, B 是前一帧中更新的滤波器的分子和分母。

最后, 利用下式对尺度滤波器进行更新, 用新样本进行尺度变换。在这种情况下, 使用一维高斯函数作为期望的相关输出。

$$A_t^l = (1 - \eta) A_{t-1}^l + \eta \overline{G} F_t^l \quad (9)$$

$$B_t = (1 - \eta) B_{t-1} + \eta \sum_{k=1}^d \overline{F_t^k} F_t^k \quad (10)$$

其中, η 为尺度滤波器模型更新的学习率, t 为第 t 个图像块。

2.2 尺度相关粒子滤波器

通过尺度相关滤波器,利用粒子的循环位移信息,将粒子引导到目标物体的局部模式。为了简单起见,定义尺度相关滤波算子,对于状态: $S_{cf} R^d \rightarrow R^d$ 。

在这里, d 是状态空间的维数, S_{cf} 是一个 d 维的算子,每一个 S_{cf} 在子空间上独立地引导生成的粒子,是新的状态变量。定义尺度相关滤波器对粒子的最大响应为 $R_{cf}(s_i^t)$ 。

$$R_{cf}(s_i^t) = \sum f^{-1}(f(\alpha_k) \cdot f(\langle \Phi(y_i^t), \Phi(\bar{X}) \rangle)) \quad (11)$$

其中, y_i^t 是粒子的观测值,基于上述响应,定义 $p(y^t | s_i^t) = R_{cf}(s_i^t)$,粒子的权值与尺度相关滤波器的响应成正比:

$$w_i^t \propto R_{cf}(s_i^t) \quad (12)$$

使用加权权重,在时刻 t 目标对象的最佳状态估计为:

$$E[s^t | y^{1:t}] \approx \sum_{i=1}^n w_i^t S_{cf}(s_i^t) \quad (13)$$

3 基于多层深度特征的尺度相关粒子滤波

3.1 目标特征的提取

当出现严重的遮挡或物体外观变异时,基于像素的浅层特征不能很好地处理。故采用已经训练好的卷积神经网络特征(deep convolutional neural network feature, CNN deep feature)^[10-11]对目标的外观进行编码,这样特征既能有很好的层表达,又能保留准确的空间信息。由 CNN 层数与语义信息和空间细节的关系可以看出,在靠前的层中,更容易确定图像的一些定位信息,但是图像的语义信息并不丰富;而在靠后的层中,图像的语义信息更容易确定,空间信息则不易于获取。因此可以得出低层特征具有更多的空间信息,对目标能够进行更为精准的定位,而高层特征则具有较多的语义信息,能够更好地处理目标剧烈变化以及防止跟踪器漂移,可以对目标进行范围定位。

采用 CNN 提取目标特征,不仅保留其最后一层输出结果,同时中间层的输出也保留下来,具体保留的是 3、4、5 三个层。

在以目标为中心的帧中裁剪出搜索窗口,用 CNN 获取搜索的窗口特征。但在 CNN 中,由于池操作的存在,使得随着层深度的增加,特征的空间分辨率会逐渐减小。需要用余弦窗对每一层提取的特征通道进行加权,对特征进行升采样之后才能够提取特征。

$$x_i = \sum_k \alpha_{ik} h_k \quad (14)$$

其中, x_i 表示升采样后的特征图, h_k 表示升采样前的特征图, i 表示位置, k 表示 k 领域的特征向量, α 表示插值的权值。

在此过程用到 CNN 的 3、4、5 层,每一层的输出特征为 x (大小为 $M \times N \times D$),将 M 维和 N 维上的所有循环移位作为训练样本,每一个移位样本 $x(m, n)$,都有一个高斯函数标签,然后通过公式(8)学习到跟 x 同样尺寸的尺度相关滤波器。

空间域的卷积对应于傅里叶域中的逐元素乘法,所以卷积层的滤波器频域表示可以写作公式(15), Y 是高斯标签的频域表示。这样使得目标函数最小化。

$$w^d = \frac{Y \circ \bar{X}^d}{\sum_{i=1}^d X^i \circ \bar{X}^i + \lambda} \quad (15)$$

其中,运算符 \circ 是 Hadamard 乘积,给定下一帧图像补丁,第 l 层特征向量为 z ,大小为 $M \times N \times D$,计算 l 层相关滤波响应图 $f_l = F^{-1}(\sum_{d=1}^D W^d \circ \bar{Z}^d)$ 。

然后通过搜索大小为 $M \times N$ 的相关响应图 f_l 的最大值位置来估计目标在 l -th 卷积层上的位置。给定相关响应映射集合 $\{f_l\}$,分层推断每一层的目标转换,即利用最后一层最大值的位置作为正则化,寻找前一层的最大值。使 $(\hat{m}, \hat{n}) = \arg\max_{m,n} f_l(m, n)$ 表示第 l 层的最大响应图位置,那么目标在第 $(l-1)$ 层的最优位置公式为:

$$\arg\max_{m,n} f_{l-1}(m, n) + \gamma f_l(m, n) \quad (16)$$

$$\text{s. t. } |m - \hat{m}| + |n - \hat{n}| \leq \gamma$$

约束表明,在第 l 个相关响应图中,只搜索 (\hat{m}, \hat{n}) 的 $r \times r$ 邻域。将最后一层的响应值用正则化项 γ 加权,然后向后传播到早期层的响应图。最后在空间分辨率最优的层上通过最大化式(16)得出估计目标位置的响应图。

3.2 跟踪以及模板更新

文中算法是在粒子滤波和相关滤波的基础上提出的。首先,使用粒子滤波器的概率系统转移模型绘制样本并重新采样,然后使尺度相关滤波器应用于每个粒子,计算每个样本三层 CNN 特征的响应图进行融合,利用尺度相关滤波器的响应更新权重。最后,根据这些响应图计算加权平均得到目标的位置。

通过学习速率参数的取值,分别自适应地对第 k 个滤波器 α_k 和第 k 个学习目标特征 \bar{x}_k 进行更新:

$$F(\alpha_k)^t = (1 - \eta) F(\alpha_k)^{t-1} + \eta F(\alpha_k) \quad (17)$$

$$F(\bar{x}_k)^t = (1 - \eta) F(\bar{x}_k)^{t-1} + \eta F(\bar{x}_k) \quad (18)$$

其中, η 是学习速率参数。

4 实验结果与分析

4.1 实验环境及参数

实验运行平台为 MATLAB R2016b, 所有实验均在 Inter Core i7 CPU, 主频 2.00 GHz, 8 GB 内存的电脑上完成。实验对所有视频序列采用一致参数: 正则化参数设置为 $\lambda = 0.01$, 尺度金字塔取 33 层。另外, 通过增加粒子数可以提高性能, 但运行时性能会受到影响。设置合理的粒子数为 6, 这样可以有效地权衡跟踪器的准确性和效率。

4.2 数据和评估方法

选取标准测试数据集 OTB100^[12] 中的视频序列进行实验, 视频序列为 Biker、Bird1、Bird2、Blurbody、BlurCar1、BlurCar2、BlurCar3、Box、Car1、ClifBar, 这些视频序列包含目标变化中的各类颜色以及运动挑战, 如运动模糊、尺度变化、快速运动和旋转、遮挡等。

为了评估这些跟踪器, 在测试序列中从第一帧的基础真值位置初始化它们, 并采用距离精度 (distance precision, DP)、成功率 (success rate, SR) 作为评价指

标^[13]。其中距离精度 (DP) 表示中心位置误差 (CLE) 小于或者等于阈值像素 T_0 的帧数 (一般选取阈值为 20 像素) 占视频总帧数的百分比。中心位置误差 (CLE) 如下, 表示预估目标与实际目标中心的欧氏距离。

$$DP = l_{DP} (CLE \leq T_0) / l \quad (19)$$

其中, l_{DP} 表示中心位置误差小于或者等于阈值的帧数, l 是当前视频的总帧数。而成功率 (SR) 表示跟踪所得的目标区域 M_T 和目标真实区域 M_G 所得的重叠率大于等于阈值 T_1 (一般设为 0.5), 表示跟踪成功。通常用成功率图曲线下方的面积 (AUC) 去评估是否跟踪成功。

$$SR = L_{SR} \left\{ \frac{|M_T \cap M_G|}{|M_T \cup M_G|} \geq T_1 \right\} / l \quad (20)$$

4.3 定量分析

将该文提出的算法与其他较为优越的算法进行比较。比较的算法有 CSK^[4]、KCF^[5]、DCF^[14]、fDSST^[6] 和 SRDCF^[15]。提出的算法由于使用多层深度特征的尺度相关粒子滤波器, 因此在面对各种运动挑战时, 更具鲁棒性。在不同的视频帧中, 提出的算法都取得了较好的结果。跟踪结果精确度和成功率如图 1 所示。

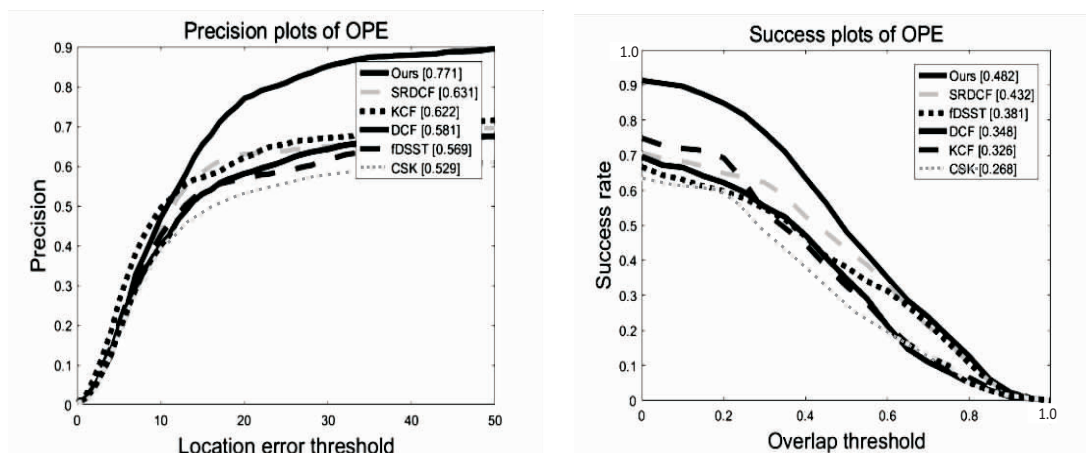
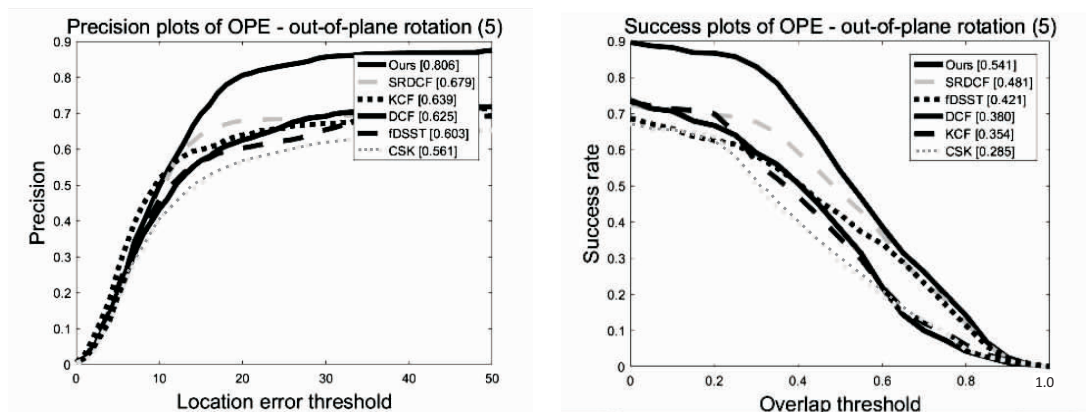


图 1 算法精确度、成功率

为了进一步验证提出的算法在跟踪过程中的鲁棒性, 图 2 给出了该算法与其他算法在旋转、尺度变化、

遮挡等各种干扰因素下的算法精确度与成功率的对比。



(a) 平面外旋转

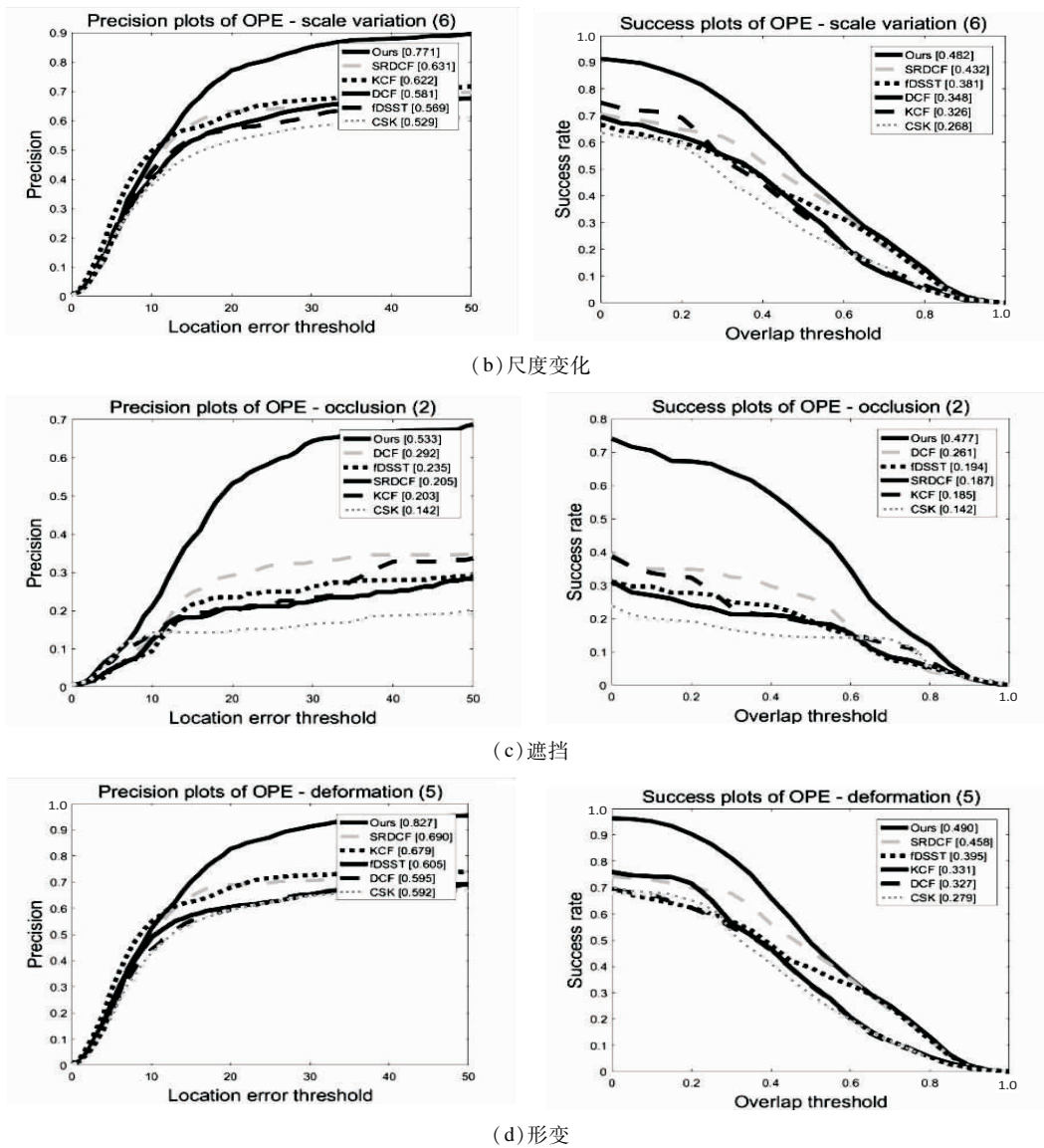


图 2 各种干扰因素下两种图的对比

从图 2 得出,提出的算法基于粒子滤波框架,对每一个模拟目标的状态分布的粒子,提取 CNN 特征,并用尺度相关滤波器来训练特征,对所得置信图进行加权计算,最后得出目标的位置。因此该算法能够在大多数干扰因素下取得较好的跟踪效果。

4.4 定性分析

为了体现提出的算法的跟踪性能,在标准测试数据集上选取了一些视频序列进行测试,对不同算法的跟踪结果进行对比,如图 3 所示。



(a) Bird2_1



(b) Biker_1

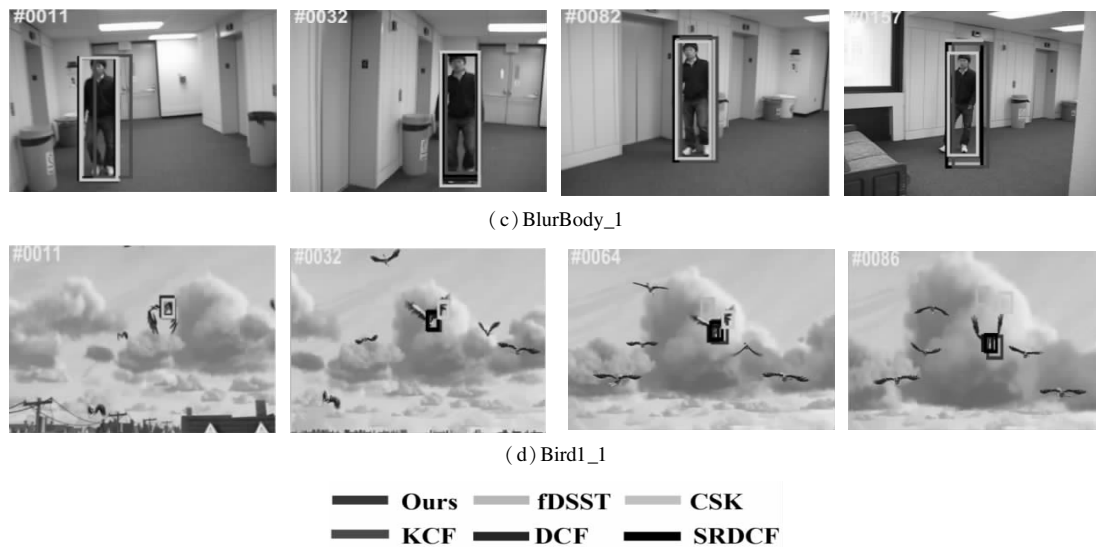


图 3 算法在不同视频下的跟踪结果

综合上述实验的对比结果,可知文中算法对目标在快速运动以及尺度变化和遮挡旋转方面更为优越。从图 3 中可以看出,在 Bird2_1 视频序列中,由于目标存在快速运动以及尺度的变化,对算法跟踪产生干扰,由 99 帧可以看出,文中算法较其他算法能够准确地跟踪目标。在 Birker_1 序列中,文中算法由于使用基于贝叶斯顺序重要性抽样技术的粒子滤波框架,能够更好地跟踪目标。在 BlurBody_1 视频序列中,存在运动模糊的干扰,但文中算法依然精确地跟踪到目标。在 Bird1_1 视频序列中第 64 帧目标在快速运动时被跟丢,但文中算法却由于跟踪精度较高,仍能有效地跟踪到目标,具有较好的鲁棒性,同时也说明,文中算法在处理快速运动时更为优越。

5 结束语

该文提出了一种基于多层深度特征的尺度相关粒子滤波方法。基于粒子滤波框架,对每一个模拟目标的状态分布的粒子,提取 CNN 特征,用尺度相关滤波器来训练特征,对每一层作置信度评分之后由粗到精地对所获得的三个置信图融合得到训练特征之后的响应图,最后根据这些响应图计算加权平均,确定目标的位置。该算法可以解决部分或者全部遮挡的问题,在准确性以及鲁棒性方面有较好的提高。在选定的视频序列上的对比结果表明,该算法得益于新的尺度相关粒子滤波器模式,优于相关的一些算法。

参考文献:

- [1] 文春武,宋杰,姚家振.基于 RSSI 校正的无线传感器网络定位算法[J].传感器与微系统,2014,33(12):134-136.
- [2] AUB-SHABAN Z,ZHOU X,ABHAYAPALA T D. A novel TOA-based mobile localization technique under mixed LOS/

- NLOS conditions for cellular networks[J]. IEEE Transactions on Vehicular Technology,2016,65(11):8841-8853.
- [3] 翟彦蓉,黄欢,张申,等.改进粒子群优化算法在 TDOA 定位中的应用[J].传感器与微系统,2013,32(4):145-148.
- [4] HENRIQUES J F,CASEIRO R,MARTINS P,et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//Computer vision-ECCV 2012,Florence,Italy:Springer,2012:702-715.
- [5] HENRIQUES J F,CASEIRO R,MARTINS P,et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(3):583-596.
- [6] DANELLJAN M,HÄGER G,KHAN F S,et al. Accurate scale estimation for robust visual tracking[C]//Proceedings of British machine vision conference. Nottingham,UK:BMVA Press,2014:1-11.
- [7] ARULAMPALAM M S,MASKELL S,GORDON N,et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking[J]. IEEE Transactions on Signal Processing,2002,50(2):174-188.
- [8] ISARD M,BLAKE A. CONDENSATION—conditional density propagation for visual tracking[J]. International Journal of Computer Vision,1998,29(1):5-28.
- [9] ZHOU S K,CHELLAPPA R,MOGHADDAM B. Intra-personal kernel space for face recognition[C]//IEEE international conference on automatic face and gesture recognition. Seoul:IEEE,2004:235-240.
- [10] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus, Ohio:IEEE,2014:580-587.
- [11] KRIZHEVSKY A,SUTSKEVER I,HINTON G. ImageNet

(下转第 180 页)