doi:10.6041/j.issn.1000-1298.2017.11.005

利用候选区域的多模型跟踪算法

毕笃彦　张园强　查宇飞　库　涛　吴　敏　唐书娟
(空军工程大学航空航天工程学院,西安 710038)

摘要:跟踪过程中发生的尺度变化、形变、遮挡是导致模型漂移的重要原因。为了克服模型漂移对鲁棒跟踪的影响,本文提出了一种利用多判别式模型和候选区域的跟踪算法。首先,该算法采用候选区域替代传统的滑动采样,适应跟踪过程中目标的位移和尺度变化。接下来,为了提高目标的表征能力,先用预训练网络提取整幅图片的深度特征,再通过感兴趣区域采样层(ROI pooling layer)快速提取每一个候选区域的深度特征,进一步提高跟踪算法的鲁棒性。最后,运用多模型选择机制进行回撤过去错误的模型更新,并通过调整搜索区域实现对目标的重检测,有效抑制了模型漂移对鲁棒跟踪的影响。实验中,本文算法与相关算法在 OTB 2013 数据库和 UAV 20L 数据库上进行了对比。结果表明,本文算法在精确度与成功率上均取得了最优性能,并能有效抑制模型漂移对鲁棒跟踪的影响。

关键词:目标跟踪;候选区域;重检测
中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2017)11-0035-08

Multiple Model Tracking Algorithm Using Object Proposals

BI Duyan　ZHANG Yuanqiang　ZHA Yufei　KU Tao　WU Min　TANG Shujuan
(*Institute of Aeronautics and Astronautics Engineering*, *Air Force Engineering University*, *Xi'an* 710038, *China*)

Abstract: The scale variation, deformation and occlusion are the important reasons for model drift. In order to overcome the effect of model drift on robust tracking, a multiple model tracking algorithm based on object proposals was proposed. Firstly, as object proposals can reflect the general object material properties, the proposed tracker replaced traditional sliding sampling with object proposals to adapt the displacement and scale variation in the tracking process. And then, in order to enhance the object representation ability, the deep convolutional feature was used to characterize the target. During this process, although the previous size of object proposals may be different, the deep convolutional feature of each object proposal can be extracted quickly by a ROI pooling layer, and each object proposals feature had the same length, which can help to model training and further improve the robustness of the tracker. Lastly, the multi-models selection mechanism was used to undo past bad model updates by selecting the best tracking model, and adjusting the searching area can achieve object re-detection. These measures can inhibit the effect of model drift on robust tracking. In order to verify the superiority of the algorithm, the OTB 2013 benchmark and UAV 20L benchmark, and some classic contrast algorithms recently were used to evaluate the proposed tracker. The results showed that the proposed tracker achieved the best performance on precision and success rate, and the effect of model drift on robust tracking can be effectively suppressed.

Key words: object tracking; object proposals; re-detection

引言

目标跟踪被广泛应用于人机交互、军事导航以及机器人等众多领域[1-2]。但是在跟踪过程中出现的尺度变化、形变以及遮挡是导致模型漂移的一个重要原因。近几年,基于相关滤波的跟踪算法如 KCF[3]、SRDCF[4],利用循环矩阵的性质提高了跟踪的性能与效率。但其适应目标的尺度变换是通过在

收稿日期:2017-03-14　修回日期:2017-04-01
基金项目:国家自然科学基金项目(61472442)和航空科学基金项目(20155596024)
作者简介:毕笃彦(1962—),男,教授,博士生导师,主要从事计算机视觉与图像处理研究,E-mail:463431261@qq.com

初始帧长宽比不变的情况下进行同比例缩放,不能很好地反映跟踪目标的尺度变化。其他的一些算法如 GOTURN[5]、MDNet[6],利用深度学习的强大表征能力提高了跟踪性能,但其尺度变化亦不能反映目标的物质属性。候选区域(Object proposals)[7]是一种最开始被应用于图像检测的方法。它克服了传统的滑动采样的劣势,能够大量减少候选样本的数目,并反映目标的物质属性和尺度属性,适应检测中目标的长宽比变化。EBT[8]、sPST[9]是近两年将候选区域与传统算法相结合的跟踪算法,在 2015 年的视觉目标跟踪竞赛[10](Visual object tracking 2015, VOT2015)中取得了不错的成绩。但是,这些算法一方面没有利用深度学习的强大表征能力,另一方面,它们缺少对模型的纠错机制。所以一旦目标发生遮挡或者形变很容易发生模型漂移,进而导致跟踪失败。因而,本文提出一种利用候选区域的多模型跟踪算法,该算法利用候选区域适应目标的尺度变化,建立多模型选择机制以纠正模型。

1 算法流程

在本文算法中,首先候选区域被用来适应目标的尺度变化,在当前帧的深度特征与候选区域的位置信息被输入感兴趣区域采样层[11](Region of interest pooling layer, ROI pooling layer)后,每一个候选区域的深度特征可以被快速获取;接下来,通过每隔 φ 帧保存跟踪模型而建立的多模型系统对这些样本进行评估。在这过程中,每一个模型的判决结果都会被熵决策机制评估,并且当感知到模型发生漂移时,该文算法可以通过扩大搜索区域对目标实现重检测。最后,最好的跟踪模型确定当前帧目标的位置,并对其进行模型更新。整个跟踪算法的流程如图 1 所示。

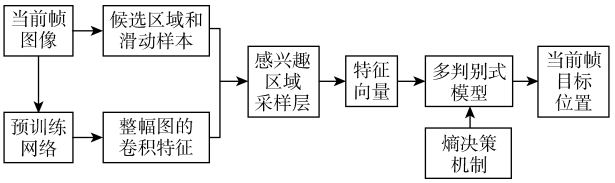


图 1 整个跟踪算法的流程图

Fig. 1 Flow chart of whole tracking algorithm

2 初始化

2.1 特征提取

如图 1 所示,为获取目标的特征表征,本文算法首先将该帧图像输入一个预训练网络获取该帧的特征。接下来,通过把整幅图像的特征与每一个候选区域的位置信息输入感兴趣区域采样层(ROI pooling layer)[11],最终快速获取每个候选区域的深度卷积特征。

在这里需要注意的是,感兴趣区域采样层的输出是相同长度的特征向量,也就是说,感兴趣区域采样层能够将不同尺寸的候选区域变成相同长度的特征向量。

在本文中,基于区域块的快速卷积神经网络(Fast region-based convolutional network, FR-CNN)[11]被作为预训练网络,并提取其第 1 层全连接输出作为特征。但是其他的网络结构也可以被用来提取样本的特征。从整个网络结构来看,FR-CNN 中的感兴趣区域采样层能够一次性地快速提取所有候选区域的特征,因而本文选取 FR-CNN 作为预训练网络来提取样本特征。

2.2 样本选取与模型训练

在获取每个候选区域的深度特征之后,需进行正负样本的选取。正负样本的选取被定义为

$$y(i)=\begin{cases}1 & \left(\dfrac{s(\boldsymbol{x}^*)\cap s(\boldsymbol{x}_i)}{s(\boldsymbol{x}^*)\cup s(\boldsymbol{x}_i)}>0.9\right)\\ -1 & \left(\dfrac{s(\boldsymbol{x}^*)\cap s(\boldsymbol{x}_i)}{s(\boldsymbol{x}^*)\cup s(\boldsymbol{x}_i)}<0.5\right)\end{cases}\tag{1}$$

式中 $s(\boldsymbol{x}_i)$——第 i 个候选样本的面积

$\boldsymbol{x}_i$——第 i 个候选样本

$\boldsymbol{x}^*$——当前帧算法预测的目标区域

$s(\boldsymbol{x}^*)$——算法在当前帧预测的目标区域面积

$y(i)$——第 i 个样本的标签

从式(1)可以看出比率在 0.5 ~ 0.9 之间的样本没有参与正负样本的选择,这是为了避免模型漂移的发生,提高跟踪算法的鲁棒性。

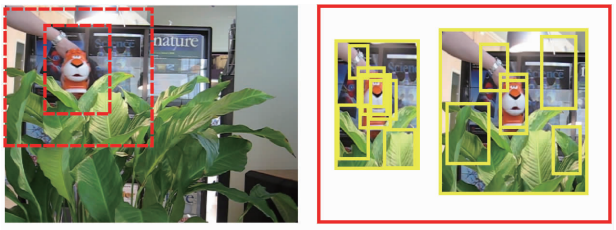
得到正负样本之后,接下来是对支持向量机(Support vector machine, SVM)的训练。在本文中,采用来源于 MEEM[12]算法中的 SVM 模型,与传统 SVM 模型不同的是,它使用 1 个标准样本集来概括先前的支持向量与当前的正负样本,并用这个样本集来更新跟踪模型。当支持向量的数目超过某个阈值的时候,通过计算样本之间的距离来使最相近的支持向量进行融合,最终使得支持向量的数目不超过某个阈值。

总的来说,本文算法所提出的跟踪器具有很强的鲁棒性,主要体现在:来自于 edgebox 算法[7]的候选区域不仅能够适应跟踪过程中目标长宽比变化,而且候选区域所具有的物质属性为模型的训练更新提供了高质量的样本;本文算法使用深度特征来表征每一个候选区域,它能够提高目标的表征能力,进而提高跟踪算法的鲁棒性。

3　利用候选区域的多模型跟踪

3.1　熵决策机制

因为图像检测的目的是检测出图像中所有目标，而跟踪却只要求跟踪某个特定的目标。目标框中轮廓线条的多少与目标框包含物体的概率有着密切的关系，DOLLAR 基于这样的观察提出了 edgebox 算法[7]，它能够快速精准地产生包含目标的候选区域，但是在跟踪过程中，生成的候选区域与搜索区域的大小有关。因为一个大的搜索框通常含有更多的线条，而当目标相对于该帧图像尺寸很小时，此时就很难产生针对该目标的候选区域了。针对此问题，本文比较了不同搜索区域大小对生成候选区域的影响。其结果如图 2 所示，从图 2 可以看出，一个比目标尺寸稍大的搜索区域能产生许多高质量的候选区域。因而，比目标尺寸稍大的搜索区域更适合用于目标跟踪任务。然而，当目标发生模型漂移时，一个小的搜索区域很容易导致目标跟踪失败。



(a) 搜索区域　　(b) 候选区域

图 2　不同搜索区域大小对生成候选区域的影响

Fig. 2　Effects of different searching area sizes on generated object proposals

针对这个问题，本文一方面利用熵决策机制自动决定搜索区域的大小。另一方面，多模型选择机制被用来撤销由模型漂移导致的错误模型更新。具体来说，当模型发生漂移并被熵决策机制感知时，搜索区域将会被放大以实现对目标的重检测。同时，通过选择先前的跟踪模型，可以撤销先前错误的模型更新。

熵决策机制最早出现于文献[13]，而后被适当修改后应用于 MEEM[12] 跟踪算法中，本文算法类似于 MEEM 算法，多模型选择机制被用以减少模型漂移的影响，但是基于熵决策机制，一种新的目标重检测策略被用来进一步提高目标跟踪的鲁棒性，这是 MEEM 算法所忽略的。除此之外，本文算法将候选区域融入跟踪框架，并用深度特征来表征它，这使得本文算法能够适应目标的尺度变化，因而本文算法能够更好地应对模型漂移对鲁棒跟踪的影响。

对于熵决策机制，本文首先定义候选样本 $\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$ 为可能的目标集。$\boldsymbol{y}_i=(\omega_i,l_i)$ 是某个候选样本 $\boldsymbol{x}_i$ 对应的标签。式中 $\omega_i\in(0,1)$ 表示目标的前景与背景的标签，l_i 表示样本 x_i 的位置。$z=((\omega_1,l_1),\cdots,(\omega_n,l_n))$ 代表可能的标签集。损失函数被定义为

$$J_m(\boldsymbol{x},z)=-L(\boldsymbol{\theta}_m;\boldsymbol{x},z)+\lambda H(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m) \tag{2}$$

其中

$$L(\boldsymbol{\theta}_m;\boldsymbol{x},z)=\max_{\boldsymbol{y}\in z}\lg P(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}_m) \tag{3}$$

$$P(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}_m)=\prod_i P(\omega_i,\boldsymbol{l}_i|\boldsymbol{x}_i;\boldsymbol{\theta}_m)=\prod_i P(\boldsymbol{l}_i|\omega_i,\boldsymbol{x}_i)P(\omega_i|\boldsymbol{x}_i;\boldsymbol{\theta}_m)=\prod_i P(\boldsymbol{l}_i|\omega_i)P(\omega_i|\boldsymbol{x}_i;\boldsymbol{\theta}_m) \tag{4}$$

式中　$L(\boldsymbol{\theta}_m;\boldsymbol{x},z)$——模型的对数似然函数

$H(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)$——经验条件熵

λ——对数似然函数和经验条件熵之间的权衡因子

$\boldsymbol{\theta}_m$——模型的参数

由式(3)和式(4)可得知，标签 $\boldsymbol{y}_i=(\omega_i,l_i)$ 仅仅取决于样本 $\boldsymbol{x}_i$ 的特征和位置信息。详细来说，样本 $\boldsymbol{x}_i$ 是通过空间的先验概率 $P(\boldsymbol{l}_i|\omega_i)$ 和后验概率 $P(\omega_i|\boldsymbol{x}_i;\boldsymbol{\theta}_m)$ 来确定目标位置的。其中，$P(l_i|\omega_i)$ 由一个高斯函数来表征，它的中心是上一帧目标的中心位置。从式(3)和式(4)也可看出，$P(\omega_i|\boldsymbol{x}_i;\boldsymbol{\theta}_m)$ 的值越大，似然函数 $L(\boldsymbol{\theta}_m;\boldsymbol{x},z)$ 的值也将越大。

经验条件熵被定义为

$$H(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)=-\sum_{\boldsymbol{y}\in z}P(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)\lg P(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m) \tag{5}$$

其中

$$P(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)=\frac{\delta_z(\boldsymbol{y})P(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}_m)}{\sum_{\boldsymbol{y}'\in z}\delta_z(\boldsymbol{y}')P(\boldsymbol{y}'|\boldsymbol{x};\boldsymbol{\theta}_m)} \tag{6}$$

式中　$\delta_z(\boldsymbol{y})$、$\delta_z(\boldsymbol{y}')$——脉冲函数

$\boldsymbol{y}'$——候选样本 $\boldsymbol{x}$ 的标签

唯有 $\boldsymbol{y}\in z$ 时，$\delta_z(\boldsymbol{y})$ 的取值为 1，否则为 0。从式(5)和式(6)可以看出，当 N 个模型在目标位置上没有分歧时，$P(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)$ 的值为 1，熵正则化项 $H(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)$ 的值为 0。但是，当 N 个模型在目标位置上存在分歧时，熵正则化项将不会为 0，因为 $P(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta}_m)$ 的值将不再等于 1。总的来说，在目标位置上的分歧越大，不确定性与熵将越大。

3.2　模型选择与候选区域的生成

一般来说，最好的模型具有最强的判别力，因而它在几个可能的目标位置上的得分差异更大。基于这样的共识，假设长时间的遮挡已经发生，通过每隔 ϕ 帧保存 1 个模型，最新的 N 个模型将会被用来评估当前的样本集，不同的模型对目标的位置预测有着不同的答案。

在这里，m_c 代表当前最好的模型，m_t 代表之前的某个模型。相比于当前的模型 m_c，之前的模型在

可能的目标位置上分歧将会更大。原因在于当前最好的模型由于学习到了遮挡时错误的背景信息,而之前的模型 m_t 由于没有学到错误的背景信息,因而它的判别能力更强,它在几个可能的目标位置上的得分差异性将更大。

而熵是不确定性的度量,不确定性越大,熵越大。所以通过式(2)对每个跟踪模型进行评估,最好的模型将会具有最小的损失值。不失一般性,本文通过计算最近若干帧的损失值之和来决定最好的跟踪模型。对最好模型的选择基于

$$M^* = \arg\min_{m\in M}\sum_{k=t-\Delta}^{t} J_{mk} \tag{7}$$

其中 $$M = (m_{t_1}, m_{t_2}, \cdots)$$

式中 M^*——最好的模型

M——模型集合

J_{mk}——某个模型在第 k 帧的损失值

除了通过熵决策机制选择最好的跟踪模型之外,本文利用最好的模型在最近几帧的损失值来确定搜索区域的大小。由于在大部分情形中,多个模型在目标的位置上是一致的,也就是说对当前帧的目标位置不确定性为零,所以式(2)中熵正则化项的值为零,模型的损失值很小,但是当模型漂移发生时,此时多个模型存在对目标位置的分歧,不确定性不为零,从而式(2)中的熵正则化不为零,所以模型的损失值将会增大。基于这样的发现,本文提出了一种通过扩大搜索区域来实现目标重检测的方法,其中搜索区域的改变基于

$$\eta = \frac{\sum_{k=t-2}^{t} J_{mk}(\boldsymbol{x},z)}{\sum_{k=t-5}^{t-3} J_{mk}(\boldsymbol{x},z)} \tag{8}$$

式中 t——当前帧 η——损失值比率

当比率 η 超出某个阈值时,模型漂移很有可能发生,此时通过扩大搜索区域对目标进行重检测,模型漂移的影响将会被减小。

3.3 候选样本的生成

在大部分情况下,搜索区域被设置成上一帧预测目标大小的 2 倍。当模型漂移发生时,搜索区域被设置成目标大小的 4 倍。在本文中,被用来预测目标位置的样本一部分来自于 edgebox 算法产生的候选区域,记为 $\boldsymbol{x}_E$。另外一部分样本来自于以目标位置为中心的滑动采样,记为 $\boldsymbol{x}_R$。$\boldsymbol{x}_E$ 和 $\boldsymbol{x}_R$ 之和记为 $\boldsymbol{x}_{E\cup R}$。对于模型更新,除了样本 $\boldsymbol{x}_E$、$\boldsymbol{x}_R$ 之外,本文在距目标中心较远的位置随机采样若干个样本对目标模型进行更新,这些样本被记为 $\boldsymbol{x}_{\tilde{R}}$。

对于候选样本 $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$ 的选择是基于

$$S_{mi} = P(\boldsymbol{l}_i \mid +)P(+ \mid \boldsymbol{x}_{E\cup Ri}; \boldsymbol{\theta}_m) \tag{9}$$

其中 $$P(+ \mid \boldsymbol{x}_{E\cup Ri}; \boldsymbol{\theta}_m) = \int_{-\infty}^{f(\boldsymbol{x}_{E\cup Ri})} \mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u \tag{10}$$

式中 $\boldsymbol{l}_i$——某个样本的中心位置

$f(\boldsymbol{x}_{E\cup Ri})$——样本 $\boldsymbol{x}_{E\cup Ri}$ 被某个模型计算后的得分

一个标准差为 15 的二维高斯函数被用来代表空域优先概率 $P(\boldsymbol{l}_i \mid +)$。在所有的样本 $\boldsymbol{x}_{E\cup R}$ 经过式(9)和式(10)计算之后,得分值被归一化为 0 ~ 1,并经过非极大值抑制处理后,只有得分值大于 0.9 的样本将会被作为候选样本。

4 仿真

在实验中,样本 $\boldsymbol{x}_R$ 和候选区域 $\boldsymbol{x}_E$ 被分别设置为 256 和 200;在 3.3 节中,分别分析了样本 $\boldsymbol{x}_R$ 和候选区域 $\boldsymbol{x}_E$ 的结合方式和数量对跟踪性能的影响以及算法的复杂度,样本 $\boldsymbol{x}_{\tilde{R}}$ 的个数和 η 的值通过在一个小的样本集上遍历寻优被分别设置为 300 和 3。剩余的参数与文献[12]中的参数设置一致,如式(2)的 λ 为 10,保存一个模型间隔 ϕ 为 50 帧,跟踪模型的数量 N 为 4。所有的实验都是基于 CPU 为 Intel Core i3 -2120,主频 3.30 GHz,8 G 内存的台式机上进行的,软件平台为 Matlab2016a。OTB 2013 数据集(Online Object Tracking 2013 Benchmark)[14] 和 UAV 20L 数据集(Unmanned aerial vehicles long-term tracking benchmark)[15] 被用来验证本文算法。

4.1 OTB 2013 数据集

在 OTB 2013 数据集中,本文算法与 SRDCF[4]、MEEM[12]、CNN-SVM[16] 等 30 种较新的算法进行了对比分析。这些算法都是近几年排名靠前的算法。在本次实验中,精确度与成功率这 2 个指标被用来评估本文算法与对比算法的性能。

其中精确度是一项基于中心位置误差的指标。中心位置误差是指预测的目标位置中心与实际的目标位置中心之间的欧氏距离。而精确度是指距离小于某个门限值的帧数占整个视频帧数的百分比。

成功率是另外一项基于重叠率的性能指标。重叠率表示预测区域与实际区域的交集与其并集之比,成功率表示重叠率大于某个阈值的帧数占视频总帧数的比重。

图 3 为各种算法在 OTB 2013 数据库中的总体效果。图中只列出排名靠前的 10 种算法。由图 3 可以看出,在精确度与成功率这 2 个性能指标上,本文算法均取得了最好的效果。

在精确度上,CNN-SVM 获得了一个得分为

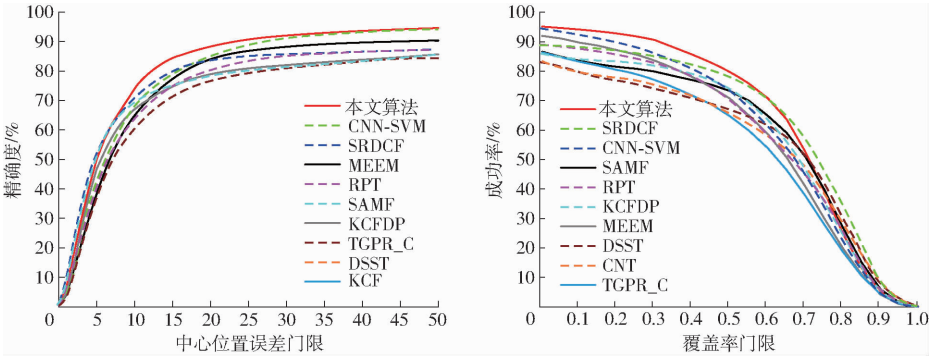


图 3　各算法在 OTB 2013 数据库上的性能对比

Fig. 3　Performance comparisons of algorithms in OTB 2013 database

77.7% 的次优结果。它是利用卷积神经网络获取指定目标的显著性映射的一种方法。SRDCF 比 CNN-SVM 得分低 1.7 个百分点，获得了第 3 名的结果。由于本文算法的候选样本融入了候选区域，而这些候选区域能够自然地适应目标的长宽比变化，除此之外，当模型发生漂移时，自适应的调整搜索区域实现对目标的重检测，能够很好的减轻模型漂移对跟踪模型的影响。所以本文算法获得了最高得分 80.2%，达到了最好的跟踪性能。

在成功率这个评价指标上，由于 MEEM 不能适应目标的尺度变化，所以它在成功率上的指标比较靠后。SRDCF 是一种通过压制边界效应的相关滤波算法，它获得了一个 62.6% 的得分。本文算法得分为 63.6%，获得了最好的性能结果。

为进一步分析本文算法在解决模型漂移上的优越性，本文对比了不同算法在不同属性指标上的性能，部分实验结果如图 4 所示。

在尺度变化中，一些视频序列比如滑雪，目标涉及到突然的长宽比变化，由图 4 可以看出，本文算法效果明显比 CNN-SVM[16]、DSST[17] 等算法要好。长宽比变化自适应是本文算法取得最优效果的一个重要原因。

在形变、平面旋转以及遮挡等属性中，模型漂移很容易发生。尤其是在发生长时间的遮挡后，由于错误的模型更新，背景信息不可避免的被跟踪模型学到，所以跟踪失败很容易发生在接下来的跟踪过程中。但是对于本文跟踪算法，一方面，候选区域具有物质属性，这在一定程度上能够减轻形变或平面

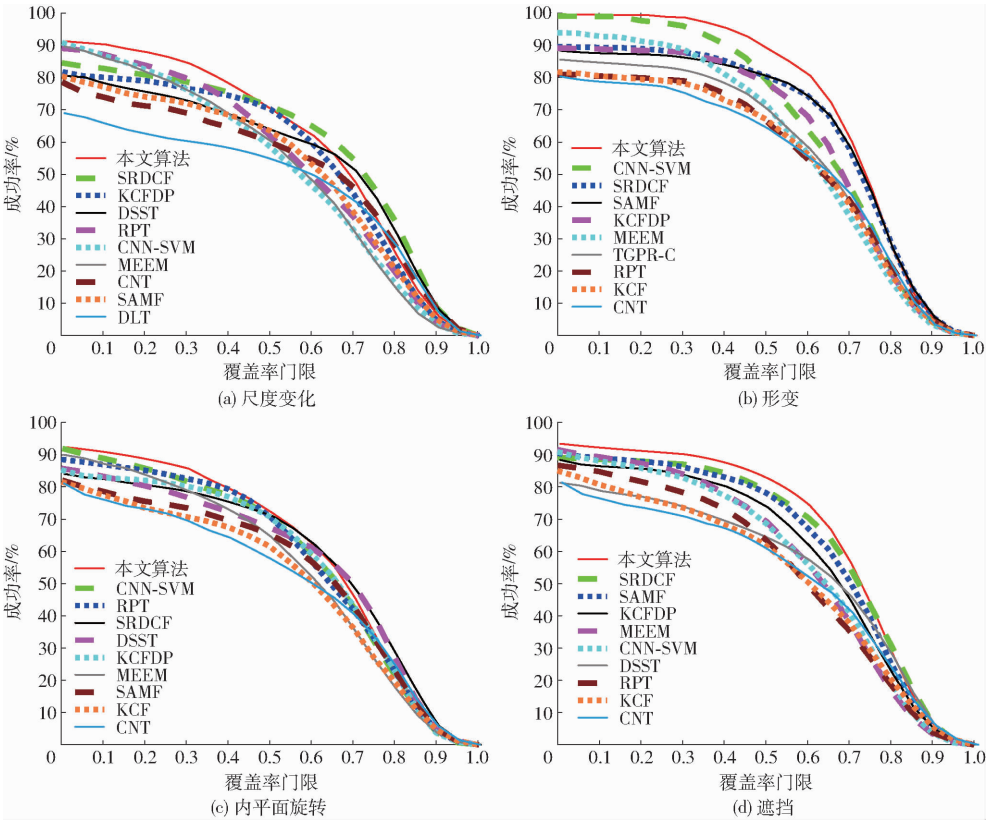


图 4　各算法在不同属性上的性能对比

Fig. 4　Performance comparisons of different attributes of algorithm

旋转对跟踪性能的影响。另一方面,通过保存之前跟踪模型建立的多模型选择机制,在多个模型中选择最具有判别力的跟踪模型,这样就能减轻由于遮挡造成的模型漂移对跟踪性能的影响。所以在这3个属性指标上,本文算法均实现了最好性能。

4.2 UAV 20L 数据库

最近发布的 UAV 20L 数据库包含了 20 个长视频序列,共有 58 670 帧图像。基于当前无人机跟踪在搜索与营救、障碍规避等计算机视觉领域的广泛应用,UAV 20L 数据库被用来评估本文算法在无人机长期跟踪上的鲁棒性。精确度与成功率被用来评估本文算法在此数据库上的有效性。

在本次实验中,15 个包含有 MEEM[12]、DSST[17]、SAMF[18]、MUSTER[19]、Struck[20]等先进跟踪算法被用来验证本文算法在 UAV 20L 数据库上的优越性,图 5 是它们在整个数据库上的整体性能比较,图 6 展示了本文算法与部分对比算法的跟踪结果。

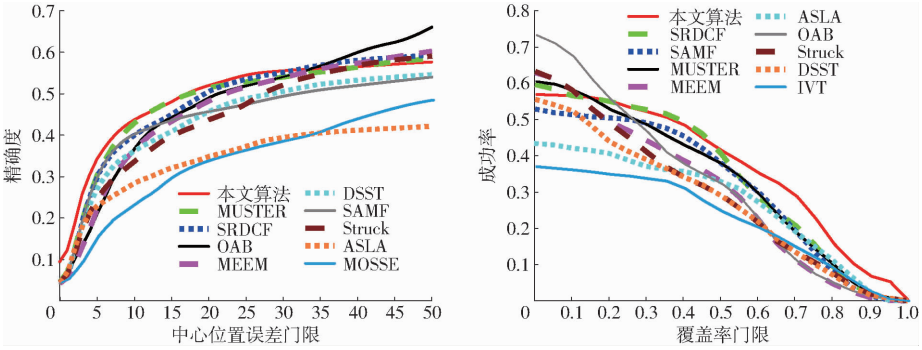


图 5 各算法在 UAV 20L 数据库上的性能对比

Fig. 5 Performance comparisons of algorithms in UAV 20L database



图 6 跟踪效果示意图

Fig. 6 Sketch images of tracking effect

与各个算法在 OTB 2013 数据库上的得分相比,在 UAV 20L 数据库上的得分普遍偏低,这说明 UAV 20L 相比于 OTB 2013 数据库更具挑战性。因为在无人机视频跟踪过程中,一方面无人机可能因为前方建筑等障碍物需实时转换角度或高度;另一方面,也因为目标所处地理环境复杂,无人机必须得实时

转换角度和位置才能适应目标的变化。这些因素最终导致的尺度变化和形变是 UAV 20L 数据库上的主要属性，而且在尺度变化中，长宽比变化相比于 OTB 2013 数据库更为常见。

由图 5 可以得知，SRDCF 在成功率上取得了次优结果，它的成功率为 0.413。由于本文算法融入了候选区域，这些候选区域能够自适应目标的长宽比变化，因而本文算法在尺度变化上更为灵活，这是一些经典算法如 SRDCF、SAMF 所不能实现的。在模型更新上，本文引入了多模型选择机制来回撤错误模型更新，并通过改变搜索区域实现目标的重检测以应对跟踪过程中遮挡的影响，因而本文算法在成功率上取得了最优性能。

4.3　参数分析与算法复杂度

为进一步分析本文算法各个环节对跟踪性能的影响，本文算法在 OTB 2013 数据库上分别分析了样本 $\boldsymbol{x}_R$ 和候选区域 $\boldsymbol{x}_E$ 的加入分别对跟踪性能的影响，成功率被用来对比它们之间的性能差异。结果如表 1 所示。

表 1　$\boldsymbol{x}_R$ 和 $\boldsymbol{x}_E$ 不同结合方式对跟踪性能的影响

Tab. 1　Effects of different combinations of $\boldsymbol{x}_R$ and $\boldsymbol{x}_E$ on tracking performance

结合方式	$\boldsymbol{x}_R$	$\boldsymbol{x}_E$	$\boldsymbol{x}_R+\boldsymbol{x}_E$
成功率	0.583	0.608	0.636

从表 1 中可以看出，当只把滑动样本作为候选样本时，其得分比只把候选区域作为候选样本的得分值低，滑动样本没有尺度变化是一个重要原因。但它比 MEEM 的得分值要高，说明深度特征比传统的特征要好，$\boldsymbol{x}_R$ 和 $\boldsymbol{x}_E$ 结合的方式取得了最高分，说明在一些复杂场景如部分遮挡时，候选区域并不能很好的框住目标，而滑动采样弥补了这一缺陷，使其取得了最好的跟踪效果。除此之外，本文还进一步分析了候选区域的数量对跟踪性能的影响，其结果如表 2 所示。

表 2　不同数量的候选区域对跟踪性能的影响

Tab. 2　Effects of different numbers of object proposals on tracking performance

数量	0	50	100	150	200	250	300
成功率	0.583	0.607	0.617	0.628	0.636	0.634	0.636

从表 2 可以看出，当候选区域的数量达到 200 个时，跟踪算法的性能基本保持稳定，此时增加候选区域的样本只能增加计算的冗余度，因而本文算法选取 200 个候选样本用于跟踪。

在该文算法中，深度特征的提取是本文算法比较耗时的一个部分，为此，本文使用 GPU 对特征提取部分加速，使得特征提取部分速度得到提高，最终使得算法跟踪速度达到了 4 帧/s。

5　结束语

本文提出了一种利用候选区域的多判别式模型跟踪算法。在本文算法中，深度特征表征的候选区域被用来适应目标的尺度变化，基于熵决策机制的多判别式模型被用来修正错误的模型更新，对目标的重检测有效避免了遮挡对鲁棒跟踪的影响。在实验中，OTB 2013 数据库与 UAV 20L 数据库被用来验证本文算法的优越性，实验结果表明本文算法取得了最好的性能结果，并能有效克服模型漂移对鲁棒跟踪的影响。

参　考　文　献

1　李盛辉，田光兆，姬长英，等. 自主导航农业车辆的全景视觉多运动目标识别跟踪[J/OL]. 农业机械学报，2015，46(1)：1－7. http：// www. j-csam. org/jcsam/ch/reader/view_abstract. aspx? flag = 1&file_no = 20150101&journal_id = jcsam. DOI：10.6041/j. issn. 1000-1298. 2015. 01. 001.
LI Shenghui, TIAN Guangzhao, JI Changying, et al. Multiple moving objects tracking based on panoramic vision for autonomous navigation of agricultural vehicle[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 1－7. (in Chinese)

2　张园强，毕笃彦，查宇飞，等. 一种利用最大间隔相关滤波的鲁棒目标跟踪方法[J]. 空军工程大学学报：自然科学版，2017，18(2)：63－68.
ZHANG Yuanqiang, BI Duyan, ZHA Yufei, et al. A robust object tracking algorithm using maximum margin correlation filter[J]. Journal of Air Force Engineering University: Natural Science Edition, 2017, 18(2): 63－68. (in Chinese)

3　HENRIQUES J, 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.

4　DANELLJAN M, HAGER G, SHAHBAZ K F, et al. Learning spatially regularized correlation filters for visual tracking[C]// IEEE International Conference on Computer Vision, 2015: 4310－4318.

5　HELD D, THRUN S, SAVARESE S. Learning to track at 100 fps with deep regression networks[EB/OL]. [2016]. https：// arxiv. org/abs/1604. 01802.

6　NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[EB/OL]. [2015]. https：// arxiv. org/

abs/1510.07945.

7 ZITNICK C L, DOLLAR P. Edge boxes: locating object proposals from edges[C]// European Conference on Computer Vision, 2014:391 - 405.

8 ZHU G, PORIKLI F, LI H. Beyond local search: tracking objects everywhere with instance-specific proposals[EB/OL]. [2016]. https://arxiv.org/abs/arXiv:1605.01839.

9 HUA Y, ALAHARI K, SCHMID C. Online object tracking with proposal selection[C]// IEEE International Conference on Computer Vision, 2015:3092 - 3100.

10 KRISTAN M, MATAS J, LEONARDIS A, et al. The visual object tracking VOT2015 challenge results[C]// IEEE International Conference on Computer Vision Workshop, 2015:564 - 586.

11 GIRSHICK R. Fast r-cnn[C]// IEEE International Conference on Computer Vision, 2015:1440 - 1448.

12 ZHANG J, MA S, SCLAROFF S. MEEM: robust tracking via multiple experts using entropy minimization[C]// European Conference on Computer Vision, 2014:188 - 203.

13 GRANDVALET Y, BENGIO Y. Semi-supervised learning by entropy minimization[C]// Neural Information Processing Systems, 2005:529 - 536.

14 WU Y, LIM J, YANG M H. Online object tracking: a benchmark[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411 - 2418.

15 MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for uav tracking[C]// European Conference on Computer Vision, 2016: 445 - 461.

16 HONG S, YOU T, KWAK S, et al. Online tracking by learning discriminative saliency map with convolutional neural network [EB/OL]. [2015]. https://arxiv.org/abs/1502.06796.

17 DANELLJAN M, HAGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking[C]// British Machine Vision Conference, 2014:65.1 - 65.11.

18 LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]// European Conference on Computer Vision, 2014:254 - 265.

19 HONG Z, CHEN Z, WANG C, et al. Multi-store tracker (muster): a cognitive psychology inspired approach to object tracking [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015:749 - 758.

20 HARE S, SAFFARI A, TORR P H S. Struck: structured output tracking with kernels[C]// IEEE International Conference on Computer Vision,2011:263 - 270.