融合多层卷积特征的相关滤波运动目标跟踪算法

李健鹏 尚振宏 刘 辉

(昆明理工大学信息工程与自动化学院 昆明 650504)

摘 要 在目标跟踪算法中,相关滤波算法近几年来发展迅速,成为了该领域的研究热点。相关滤波跟踪算法具有速度快、效果好等优点,但受限于传统手工特征对目标表达能力不足,仍然难以应对诸如形变、遮挡、模糊等情形。最近,卷积神经网络在诸多领域取得了极大的成功,研究人员将相关滤波与卷积特征相结合,克服了传统手工特征缺少目标语义信息的缺点。为了有效处理目标外观变化,文中提出一种融合多层卷积特征的相关滤波运动目标跟踪算法。该算法将目标跟踪分为预测位置和估计尺度两个步骤:提取多层卷积特征并在每个卷积层上估计目标位置,通过固定权重将所有卷积层的结果融合以确定目标的最终位置;确定位置后通过提取目标多个尺度的方向梯度直方图特征来估计目标的最佳尺度。在公开数据集中选取20段视频来验证所提算法,并将该算法与4种运动目标跟踪算法进行比较。实验数据表明,与次优的基于传统手工特征的DSST算法相比,所提算法的距离精度提高了48.9%,重叠精度提高了51.9%;与同样使用卷积特征的HCFT算法相比,其距离精度提高了19.1%,重叠精度提高了25.2%。文中提出的算法较好地克服了传统手工特征表达能力弱的缺点,其性能优于使用手工特征的传统相关滤波跟踪算法,相比同样使用卷积特征的相关滤波算法也有所提高。在目标发生遮挡、模糊等复杂情况下,该算法仍然能够准确跟踪目标。

关键词 目标跟踪,相关滤波,卷积特征,尺度估计

中图法分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.07.038

Visual Object Tracking Algorithm Based on Correlation Filters with Hierarchical Convolutional Features

LI Jian-peng SHANG Zhen-hong LIU Hui

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China)

Abstract In the visual object tracking, correlation tracking algorithm is a hot topic and has developed rapidly in recent years. Correlation filter tracking algorithms have the advantages of fast speed and good effect. However, the traditional hand-crafted features are insufficient for target discrimination, and fail to handle challenging situations such as deformation, occlusion and blurring. Recently, convolutional neural networks have achieved great success in many fields. Researchers have combined correlation tracking and convolutional features to surmount the shortcomings of hand-crafted features that lack target semantic information. In order to cope well with the above problems, this paper proposed a visual object tracking algorithm based on correlation filter with hierarchical convolutional features. The proposed algorithm divides the object tracking into two steps, including position prediction and scale estimation. Multi-layer convolutional features are trained and the target position on each convolutional layer is predicted with a coarse-to-fine searching approach. The Histogram of Oriented Gradient features is used to estimate the optimal scale of target. Comprehensive experiments on 20 challenging sequences were performed to verify the proposed algorithm, and the proposed algorithm was compared with other four trackers. The results show that the proposed approach significantly improves the performance by 48.9% and 51.9% in distance precision and overlap precision respectively compared to the DSST tracker based on hand-crafted features. Moreover, the proposed method outperforms the HCFT tracker using convolutional features by 19.1% and 25.2% in distance precision and overlap precision, respectively. The proposed algorithm overcomes the shortcomings of poor representation skills of traditional manual features, and its performance is better than the correlation filtering tracking algorithms using manual features. Compared with the same correlation filtering algorithms using convolutional features, the tracking performance has also been improved. The algorithm can accurately track the target in complex situations such as occlusion and blurring.

Keywords Visual object tracking, Convolutional feature, Scale estimation, Correlation filter

到稿日期:2018-08-09 返修日期:2018-10-11 本文受国家自然科学基金项目(61462052,11873027)资助。

李健鹏(1993-),男,硕士生,主要研究方向为目标跟踪、计算机视觉,E-mail:lijianpeng19930712@hotmail.com;**尚振宏**(1975-),男,博士,副教授,CCF会员,主要研究方向为计算机视觉、图像处理,E-mail:shangzhenhong@126.com(通信作者);**刘 辉**(1969-),男,博士,教授,主要研究方向为计算机视觉、模式识别。

1　引言

目标跟踪是计算机视觉的重要研究课题，广泛应用于人机交互、行为识别等领域。近年来，相关滤波发展迅速，自 Bolme 等[1]首次提出误差最小平方和滤波算法（Minimum Output Sum of Squared Error filter，MOSSE）以来，相关滤波算法层出不穷。其中经典算法如 MOSSE，DSST（Discriminative Scale Space Tracker）[2]，KCF（Kernelized Correlation Filters）[3]等具有速度快、效果好的特点。但由于跟踪目标存在诸如形变、光照、尺度变化、遮挡、快速运动等情形，相关滤波跟踪算法面临巨大的挑战[4-5]。

在相关滤波算法中，特征是决定跟踪效果的重要因素之一。传统手工特征，如方向梯度直方图特征（Histogram of Oriented Gradient，HOG）、颜色特征[6]（Color Name，CN），难以捕捉目标的语义信息，对形变、旋转等复杂外观变化没有良好的鲁棒性，容易造成模型漂移，从而导致跟踪失败。最近，深度学习发展迅速，卷积神经网络在图像分类、目标识别领域都获得了极大的成功。研究人员将卷积神经网络应用于目标跟踪领域，取得了良好效果[7-8]。Danelljan 等[9]发现：VGG[10]网络的高层卷积特征具有丰富的语义信息，能够对物体进行类别判别，对遮挡、形变等问题处理良好，但分辨率较低，无法进行精确定位；低层含有丰富的空间细节信息，分辨率较高，能够精确地定位目标。相比于传统手工特征，卷积特征具有易于提取、判别力强等优点，能够有效处理目标遮挡、形变等问题。同时，卷积特征还存在特征维度过高、信息冗余、不同卷积层分辨率不一致等问题[11-12]。针对卷积神经网络不同卷积层分辨率不一致的问题，Ma 等[12]依据卷积神经网络的结构特点，提出了结合多层卷积特征，由粗到细的搜索策略，该策略获得了优异的性能；但是卷积特征数量众多，响应不一，跟踪效果差别较大。将这些卷积特征全部应用于相关滤波框架难免会影响跟踪效果。HCFT（Hierarchical Correlation Features based Tracker）[12]，C-COT（Continuous Convolution Operators for Visual Tracking）[9]等算法并未针对卷积特征数量众多、信息冗余的缺点进行改进，这使得算法的计算量大，跟踪效果会受到影响。

Danelljan 等于 2014 年提出了 DSST 相关滤波跟踪算法，该算法以 MOSEE 为基础，采用了 HOG 和灰度特征，添加了尺度处理，该算法获得了 2014 VOT 比赛的第一名。但是由于传统手工特征对遮挡和目标形变、快速运动等效果不佳，因此该算法仍然具有一定局限性。受到不同层次的卷积特征对目标有良好的特征表示的启发[13-15]，本文提出一种融合多层卷积特征的相关滤波算法，在 DSST 相关滤波跟踪算法框架下，融合多层卷积特征来克服传统手工特征对目标表达能力不足的缺点。我们采用 VGG 网络的 3 个卷积层来提取不同层次的特征并加以筛选，在每层分别训练滤波器，并将响应归一化到[0，1]区间。最后将每层的响应加权融合，响应最大的位置即为预测目标位置，同时提取不同尺度下的 HOG 特征来更新目标尺度以适应目标尺度变化。最后通过预测位置和尺度提取样本来更新滤波器以适应目标的外观变化。

2　DSST 相关滤波跟踪算法

DSST 相关滤波跟踪算法是对 MOSSE 算法的改进，实现了对位置和尺度的良好估计。其主要思想是使用两个独立的相关滤波器：用位置滤波器对跟踪目标的位置进行预测，随后在预测的位置上使用尺度滤波器预测目标的最佳尺度。

2.1　位置估计

首先，在当前帧，以目标确定位置为中心、给定目标边框的一定比例为大小提取样本并用来训练位置滤波器。为了使样本包含时空上下文信息，一般将目标边框的两倍作为目标样本的大小。因为傅里叶变换能够大大提高卷积运算的速度，所以滤波器的求解均在频域进行。但是由于傅里叶变换中存在边缘效应，通常将样本特征乘以余弦窗口以减轻边缘效应。提取的样本特征包含 HOG 特征和灰度特征。为了构造一个最佳的滤波器，该算法通过最小化式(1)的损失函数进行训练：

$$\varepsilon=\left\|\sum_{l=1}^{d}h^l * f^l-g\right\|^2+\lambda\sum_{l=1}^{d}\|h^l\|^2 \tag{1}$$

其中，h 表示训练的滤波器，f 表示提取的 d 维特征，$*$ 表示循环相关，g 表示理想二维高斯响应，l 表示特征维数，$l\in\{1,\cdots,d\}$，λ 为正则化系数。损失函数的前半部分表示实际响应与理想响应之间的误差，后半部分为防止滤波器过拟合和式(2)分母为零而添加的正则化项。相关滤波跟踪算法最大的特点就是可以使用傅里叶变换加速相关运算。由于时域卷积计算极为耗时，因此可以转换到频域快速求解。式(1)滤波器的解为：

$$H^l=\frac{\overline{G}F^l}{\sum_{k=1}^{d}\overline{F^k}F^k+\lambda} \tag{2}$$

其中，大写字母均表示经过离散傅里叶变换后的频域描述，上划线表示对应的共轭复数。为了减少计算量，DSST 算法按照下列方法更新滤波器：

$$A_t^l=(1-\eta)A_{t-1}^l+\eta\overline{G_t}F_t^l \tag{3}$$

$$B_t=(1-\eta)B_{t-1}+\eta\sum_{k=1}^{d}\overline{F_t^k}F_t^k \tag{4}$$

其中，A，B 表示所训练滤波器的分子与分母，t 表示视频帧数，η 表示学习率，用于赋予滤波器历史“记忆”。滤波器训练完成后即在下一帧中确定目标位置。在下一帧中以当前尺度两倍大小和位置提取样本 Z。由滤波器和待检测样本卷积得到最终响应 y。新的位置由响应 y 的最大值确定，其中 $\mathcal{F}^{-1}$ 表示傅里叶反变换：

$$\mathrm{y}=\mathcal{F}^{-1}\left\{\frac{\sum_{l=1}^{d}\overline{A^l}Z^l}{B+\lambda}\right\} \tag{5}$$

2.2　尺度估计

尺度估计的原理与本文提出的算法相同，因此将在本文算法中进行详述。首先取当前帧目标的 HOG 特征，并按照理想一维高斯响应输出训练一维尺度滤波器。然后按照位置滤波器的预测位置在下一帧中提取样本，并将样本与尺度滤波器卷积，得到尺度响应。当前估计的目标尺度即为尺度响应最大值所表示的尺度。

3 融合多层卷积特征的相关滤波运动目标跟踪算法

本文在 DSST 相关滤波框架的基础上提出了一种融合多层卷积特征的相关滤波算法。针对 DSST 算法中 HOG 特征对目标表达能力不足的缺点，通过提取卷积特征，融合多层卷积层的跟踪结果的方式，从粗到细地确定目标位置。由于卷积特征存在维度复杂、信息冗余的缺点，为了减少计算量，我们简化了每个卷积层的特征以提高跟踪速度。图 1 是本文算法的流程图，我们将跟踪算法分为位置预测和尺度预测。使用三层卷积特征融合预测位置；同时使用 HOG 特征预测目标最佳尺度。

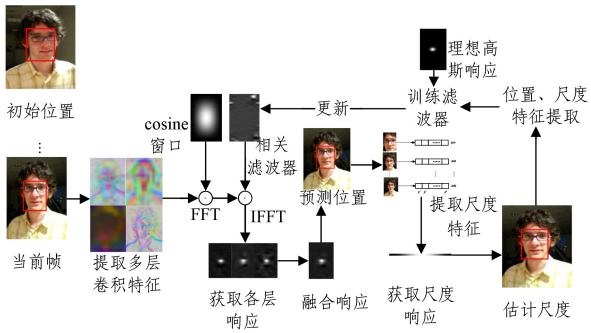

图 1 本文算法流程图

Fig. 1 Flowchart of proposed algorithm

3.1 特征选取

在图像分类、目标检测、人脸识别等领域，卷积神经网络都获得了极大的成功。得益于卷积神经网络丰富的特征表示，卷积特征能够克服传统手工特征对于目标表示不足的缺陷。在 VGG 网络中，随着网络深度的不断增加，空间信息越来越少，语义信息愈加丰富。图 2 为 MotorRolling 视频序列在不同卷积层中的表达，我们将各层卷积特征前 3 个主要分量经过双线性插值到相同大小后进行可视化，其中红色框为本算法的跟踪结果。可以看到：随着网络深度的增加，图像细节逐渐减少，语义信息逐渐增多。然而，随着深度的增加，对目标定位极其重要的分辨率却急剧减小。我们使用在 ImageNet 上训练得到的 VGG-19 网络来提取目标特征，以最后一个池化层的输出特征为例，输入网络的图像尺寸为 224×224×3，池化层输出尺寸为 7×7×512。输出图像尺寸仅为输入的约 0.1%。如此低的分辨率显然不适用于精确定位目标。然而，作为语义信息最丰富的卷积层，即便目标外观发生较大变化，第五层仍然能够有效识别目标。相比于第一层和第二层，第三层和第四层包含了一些语义信息，简化了特征细节，同时有着较高的分辨率，能够提供信息协助第五层确定目标位置。为了缓解分辨率不一致的问题，我们将第三层到第五层得到的卷积特征通过双线性插值到样本尺寸以获得精确的估计位置。如图 3 所示，在实验中发现，得到的卷积特征包含不需要的特征，或者对目标定位贡献极小，因此考虑到速度和鲁棒性，将特征按方差进行筛选。方差的公式为：

$$\sigma^2=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(F_{i,j}-F_{\text{ave}})^2 \tag{6}$$

其中，m,n 表示每层特征的长和宽，$F_{i,j}$ 表示求和过程中卷积特征某一通道的特征值，F_{ave} 表示该通道特征的平均值。每一层按方差大小取前 30 通道特征用于训练滤波器。为了减轻傅里叶变换中边缘效应的影响，我们将使用余弦窗口处理筛选得到的卷积特征，至此获得了每一层用于跟踪的目标卷积特征。

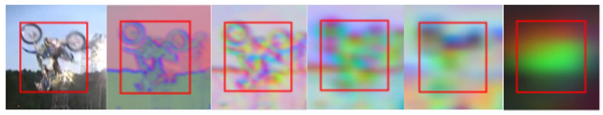

图 2 目标在不同卷积层中的特征表达(电子版为彩色)

Fig. 2 Feature expression for objects in different layers

图 3 第五卷积层中的不同特征

Fig. 3 Different features in fifth convolutional layer

3.2 位置估计

首先按照式(1)训练滤波器，分别在每一层训练一个滤波模板。再以当前位置为中心，目标边框的一定比例为大小，在下一帧中提取目标样本。目标样本包含了目标的一些背景信息，提高了算法的鲁棒性。然后按上述特征提取方式提取样本的卷积特征。为了提高卷积的运算速度，将目标特征用傅里叶变换转换到频域与训练得到的滤波器进行计算以得到目标的空间响应。为了解决不同卷积层的响应值范围不一致、不便于融合的问题，我们将每一层得到的响应归一化到[0,1]的范围。高层卷积特征语义信息丰富，对目标外观变化效果较好，应该给予较高的融合权重；低层分辨率较高，能够精确定位目标，但是对遮挡、形变比较敏感，应该给予较低的融合权重。按照由粗到细、从高到低的原则，我们给每层的响应以固定权重，将这些固定权重相加得到最终响应[16-17]。图 4 给出了确定位置的过程。用 α_i 表示每层对应的融合权重，y_i 表示每层对应的归一后的空间响应。最终响应 y 为每层响应加权后的线性叠加，加权后得到的最终响应的最大值所在位置即为滤波器估计的目标所在的位置。

$$y=\sum_{i=1}^{3}\alpha_i y_i \tag{7}$$

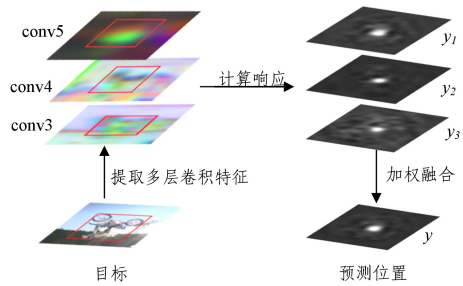

图 4 目标位置估计的主要过程

Fig. 4 Main steps of target position prediction

3.3 尺度估计

由于 HOG 特征具有远小于卷积特征的计算量，出于实时性考虑，我们选取 HOG 特征而非卷积特征来估计目标尺度。首先，在估计的目标位置上以多个尺度提取目标样本，每个尺度与目标大小成比例。以 P,R 表示上一帧的目标尺度，S 表示尺度样本个数，a 表示尺度因子。以 $a^nP\times a^nR$ 为尺度来提取样本，其中，$n\in\{[-(S-1)/2],\cdots,[(S-1)/2]\}$。将得到的尺度样本统一插值到 $P\times R$ 的大小，提取尺度特征并

将其与一维尺度滤波器进行卷积以获得尺度响应，尺度响应最大值表示的尺度就是估计的当前目标的大小。如图 5 所示，每一行为提取尺度样本的 HOG 特征，每一列为用于训练一维尺度滤波器的尺度特征。

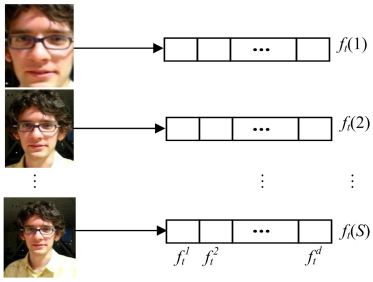

图 5　尺度特征的提取

Fig. 5　Extraction of scale features

3.4　模型更新

通常通过卷积神经网络得到的卷积特征都特别多，在本算法中我们需要给每个特征训练一个滤波器，因此计算量较大。在得到目标新的尺度和位置后，分别提取用于位置滤波器更新的卷积特征和尺度滤波器更新的尺度特征。为了减少计算量，得到一个鲁棒的近似，按照式(3)和式(4)更新位置和尺度滤波器。

4　实验结果与分析

为了验证本文算法的有效性，实验选取标准数据集 OTB100 测试集中的 20 个视频序列进行实验，测试视频包括 Basketball，BlurOwl，Bolt2，Car1，CarScale，DragonBaby，Football，Freeman4，Human3，Human5，Human6，Human7，Jumping，KiteSurf，Lemming，MotorRolling，Skiing，Shaking，Surfer，Sylvester，包含光照变化(Illumination Variation，IV)、尺度变化(Scale Variation，SV)、遮挡(Occlusion，OCC)、形变(Deformation，DEF)、运动模糊(Motion Blur，MB)、快速运动(Fast Motion，FM)、平面外旋转(Out-of-Plane Rotation，OPR)、平面内旋转(In-Plane Rotation，IPR)、超出视野(Out-of-View，OV)和背景杂乱(Background Clutters，BC)等大量属性。

本文采用距离精度(Distance Precision，DP)、重叠精度(Overlap Precision，OP)和每秒的帧数(Frames Per Second，FPS)来评估本文算法。距离精度的定义为目标中心位置与真实位置的欧氏距离误差小于一定阈值的帧数的百分比。重叠精度定义为：$OL=\text{area}(R_t \cap R_g) / \text{area}(R_t \cup R_g)$，其中，$R_t$ 表示跟踪器预测的目标框，R_g 表示跟踪视频目标真实位置的目标框。如果某一帧跟踪器预测的目标框和真实位置的目标框重叠率 OL 高于一定的阈值，则认为该帧跟踪成功。最终结果为跟踪成功帧数所占总帧数的百分比。

为了证明该算法的有效性，我们选取 4 种其他算法进行比较分析，包括 CN，KCF，DSST，HCFT 算法。其中 CN，KCF，DSST 3 种算法是使用手工特征的经典相关滤波跟踪算法，有较好的准确性和跟踪速度。HCFT 是一种结合多层卷积特征的相关滤波跟踪算法，在 VOT2016 中有良好的表现。另外，CN，KCF 算法中无尺度处理，DSST，HCFT 算法对尺度进行了估计预测。

4.1　实验环境及和参数设置

本实验基于 MatConvNet toolbox 框架[18]，VGG 网络为预训练好的 imagenet-vgg-verydeep-19，使用 GPU 来提取卷积特征。实验平台为 win10 系统，使用 MATLAB 2017b，实验均在 Intel(R) Core(TM) i5-7300HQ(主频 2.5GHz)，GPU 为 NVIDIA GeForce GTX 1060，内存为 8GB 的电脑上完成。根据本实验中由粗到细、从高到低的搜索策略，第五层到第三层中卷积特征所占权重设置为 $\alpha=\{1, 0.5, 0.3\}$，为了适应目标尺度变化，尺度样本个数 S 设置为 33，尺度因子 a 设定为 1.02，模型更新中学习率 η 设定为 0.0075，正则化参数 λ 设定为 10^{-4}，考虑到计算量和鲁棒性，每一层的特征按方差大小取前 30 通道，距离精度阈值取 20 像素，重叠精度阈值取值为 50%，即跟踪结果的距离精度误差小于 20 像素时认为该帧跟踪成功，重叠率高于 50%时认为该帧跟踪成功。

4.2　定量分析

表 1 是 20 组视频序列的跟踪结果。其中最优结果用加粗字体表示，次优结果用下划线表示。可以看出，本文算法的平均 DP、平均 OP 均为最优。同次优算法 HCFT 相比，平均 DP 提高了 19.1%，平均 OP 提高了 25.2%。但是，本文算法的帧率(FPS)仅有 6.1 fps，在所有算法中处于最低。

表 1　5 种算法的平均 DP，OP 与 FPS

Table 1　Average DP，OP and FPS of 5 tracking algorithms

	平均 DP/%	平均 OP/%	*FPS*/fps
KCF	44.4	31.8	**343.8**
CN	42.4	31.9	337.9
DSST	47.1	38.1	59.1
HCFT	76.9	64.8	13.1
本文算法	**96.0**	**90.0**	6.1

表 2　5 种算法在 8 个视频中的距离精度

Table 2　Distance precision of five tracking algorithms in eight video sequences

(单位：%)

	KCF	CN	DSST	HCFT	本文算法
Lemming	48.70	30.80	43.00	41.70	**95.70**
Human5	26.40	39.00	24.40	24.50	**100**
Human6	28.90	28.50	44.80	32.70	**91.70**
Bolt2	2.05	1.71	2.05	94.50	**94.90**
CarScale	**80.60**	72.20	75.80	56.30	78.20
DragonBaby	33.60	32.70	6.19	**88.50**	86.70
MotorRolling	4.88	4.88	4.88	**94.50**	89.00
Sylvester	84.30	92.90	84.20	85.30	**100**

表 3　5 种算法在 8 个视频中的重叠精度

Table 3　Overlap precision of five tracking algorithms in eight video sequences

(单位：%)

	KCF	CN	DSST	HCFT	本文算法
Lemming	43.20	29.30	27.20	45.70	**93.70**
Human5	23.30	24.00	24.30	12.30	**100**
Human6	22.60	22.60	45.60	28.00	**96.20**
Bolt2	1.02	1.02	1.02	84.60	**90.80**
CarScale	44.40	44.80	84.90	50.80	**92.10**
DragonBaby	31.00	30.10	6.19	81.40	**82.30**
MotorRolling	7.93	7.32	6.71	67.10	**75.00**
Sylvester	82.10	74.20	73.80	81.60	**98.20**

表 2 列出了 5 种算法在其中 8 个视频的平均 DP，表 3 列出了这 8 个视频的平均 OP。可以看出，本文算法不管是距离精度(DP)还是重叠精度(OP)均处于较高水平。在大部分视频上，本算法优于使用传统手工特征的 KCF，CN，DSST 算法，与同样使用深度特征的 HCFT 相比也有一定提升。

图 6 为本文算法在 8 个测试视频上的距离精度曲线。横坐标表示距离精度阈值，纵坐标表示距离精度。距离精度值越大，表示跟踪器性能越好。可以看到，本文算法在大部分视频上处于最优，且均优于其他使用手工特征的算法。

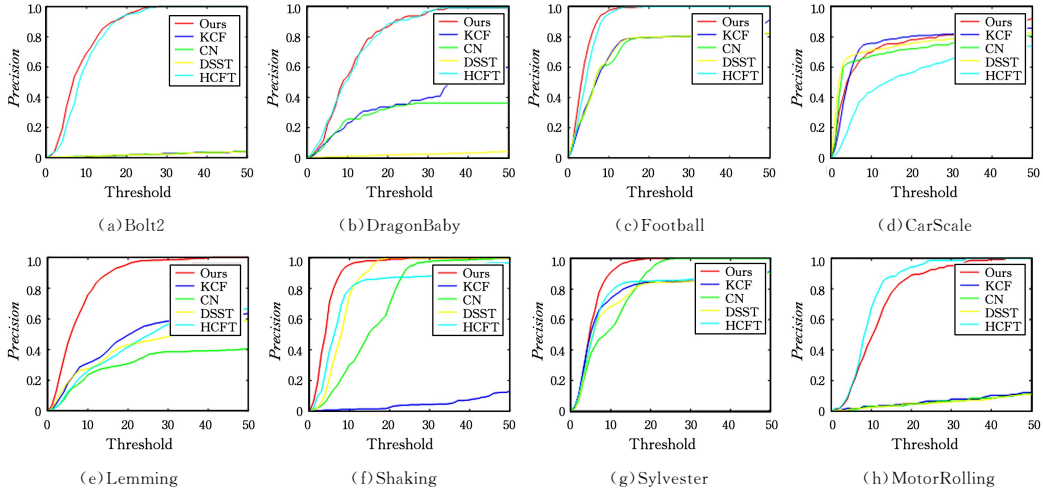

图 6 5 种算法在 8 个视频中的距离精度曲线

Fig. 6 DP curves of five tracking algorithms in eight video sequences

4.3 定性分析

本文选取了 8 段具有挑战性的视频，其中包括光照变化(图 7(e)，图 7(f)，图 7(g)，图 7(h))、尺度变化(图 7(d)，图 7(e)，图 7(f)，图 7(g))、遮挡(图 7(b)，图 7(c)，图 7(d)，图 7(e))、形变(图 7(a))、运动模糊(图 7(b)，图 7(h))、快速运动(图 7(b)，图 7(d)，图 7(e))、旋转(图 7(b)，图 7(c)，图 7(d)，图 7(e)，图 7(f)，图 7(g)，图 7(h))、背景杂乱(图 7(a)，图 7(c)，图 7(f)，图 7(h))、超出视野(图 7(b)，图 7(e))等复杂属性。

图 7 8 个视频序列的跟踪结果

Fig. 7 Tracking results of eight video sequences

形变和相似目标：图 7(a)中，目标发生了明显形变，且周围存在相似目标。DSST，KCF，CN 由于采用了传统手工特征不能跟上目标，本文算法和 HCFT 算法则能够正常跟踪；在 114 帧时由于存在相似目标干扰，HCFT 发生了定位偏差，而本文算法能够正确分辨目标。在图 7(c)中，由于存在相似物体的干扰，CN，DSST，KCF 等均发生了跟踪错误，本算法及 HCFT 则能正确区分目标，完成跟踪。

快速运动和运动模糊：图 7(b)中，40 帧左右，目标快速运动并发生了模糊。CN，KCF，DSST 发生漂移跟踪失败，只有 HCFT 和本文算法能够跟上目标的快速运动。

尺度变化：图 7(d)中，CarScale 视频序列中目标发生了明显的尺度变化。CN，KCF 算法由于没有尺度适应，不能跟上目标；DSST 和 HCFT 定位发生了偏移，不能准确覆盖目标。本算法由于结合了目标语义信息，在目标外观变化的情况下依然能够正确识别目标完成跟踪。

遮挡和超出视野：图 7(e)中，目标发生了明显的遮挡。由于目标信息过少，其他算法发生了跟踪错误，学习了错误的目标信息。本文算法能够识别目标并继续跟踪。在 555 帧左右，目标超出视野，本文算法仍然能够正确跟踪目标。

光照变化：图 7(f)中，视频 Shaking 由于存在光照变化，HCFT 算法发生了跟踪漂移，KCF 算法则跟踪错误。本算法由于选取了有效的卷积特征，能够适应目标外观变化，可以正常跟踪目标。

旋转和背景杂乱：图 7(g)中，目标主体发生了旋转，存在背景变化，KCF，DSST，HCFT 均发生了目标丢失，而本文算法依旧能够跟上目标；图 7(h)中，目标发生了旋转且背景中存在光源干扰，传统特征如 HOG 和灰度特征对旋转不敏感，鲁棒性差，无法跟上目标，而本文算法和 HCFT 采用的卷积特征对旋转适应性好，能够顺利跟踪目标。

实验表明，本算法对光照变化、尺度变化、遮挡、形变、运动模糊、快速运动、旋转、背景杂乱、超出视野等均有良好的鲁棒性。相比使用传统特征的跟踪算法，本文算法能够更准确和稳定地确定目标。

结束语　本文在传统 DSST 跟踪算法的基础上进行了两点改进，提出了一种融合多层卷积特征的相关滤波运动目标跟踪算法，融合了多层卷积特征并简化了计算量。通过预训练好的 VGG 网络提取多层卷积特征来取代表达能力较弱的手工特征。鉴于卷积特征分辨率不一、冗余繁多的问题，我们筛选出可用于跟踪的特征，并在每个卷积层上对目标位置进行估计。根据卷积神经网络语义从多到少、分辨率由粗到细的特点将每个卷积层的结果进行线性加权融合。同时提取目标多个尺度的 HOG 特征来构建滤波器，选择目标的最佳尺度。实验表明，相比于使用传统手工特征的相关滤波跟踪算法，本文算法使用的卷积特征能有效提高相关滤波算法在复杂场景下的有效性。但本文算法中也有一定的不足之处，例如：相比于传统算法，其运行速度仍然较慢，在实验中其 FPS 仅有 6.1fps，其中卷积特征的计算耗时较多。将来可以在上述方面进行优化，提高本算法的性能。

参 考 文 献

[1] BOLME D S，BEVERIDGE J R，DRAPER B A，et al. Visual object tracking using adaptive correlation filters[C]// Computer Vision and Pattern Recognition IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE，2010：2544-2550.

[2] DANELLJAN M，HÄGER G，KHAN F S，et al. Discriminative scale space tracking[J]. IEEE transactions on pattern analysis and machine intelligence，2017，39(8)：1561-1575.

[3] 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.

[4] ZHANG W，KANG B S. Recent advances in correlation filter-based object tracking：a review[J]. Journal of Image and Graphics，2017，22(8)：1017-1033.(in Chinese)
张微，康宝生. 相关滤波目标跟踪进展综述[J]. 中国图象图形学报，2017，22(8)：1017-1033.

[5] SMEULDERS A W M，CHU D M，CUCCHIARA R，et al. Visual tracking：An experimental survey[J]. IEEE transactions on pattern analysis and machine intelligence，2014，36(7)：1442-1468.

[6] DANELLJAN M，SHAHBAZ KHAN F，FELSBERG M，et al. Adaptive color attributes for real-time visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2014：1090-1097.

[7] LI H，LI Y，PORIKLI F. Robust online visual tracking with a single convolutional neural network[C]// Asian Conference on Computer Vision. 2014：194-209.

[8] WANG L，OUYANG W，WANG X，et al. Visual tracking with fully convolutional networks[C]// Proceedings of the IEEE international conference on computer vision. Springer-Verlag，2015：3119-3127.

[9] DANELLJAN M，ROBINSON A，KHAN F S，et al. Beyond correlation filters：Learning continuous convolution operators for visual tracking[C]// European Conference on Computer Vision. Springer-Verlag，2016：472-488.

[10] RUSSAKOVSKY O，DENG J，SU H，et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision，2015，115(3)：211-252.

[11] DANELLJAN M，HAGER G，SHAHBAZ KHAN F，et al. Learning spatially regularized correlation filters for visual tracking[C]// Proceedings of the IEEE International Confe-rence on Computer Vision. IEEE，2015：4310-4318.

[12] MA C，HUANG J B，YANG X，et al. Hierarchical Convolutional Features for Visual Tracking[C]// IEEE International Conference on Computer Vision. IEEE，2015：3074-3082.

[13] XU Y，LI Y，WANG J，et al. Multi-scale correlation tracking with convolutional features[C]// Signal and Image Processing (ICSIP)，IEEE International Conference onIEEE International Conference on Signal and Image Processing. IEEE，2016：310-314.

[14] WANG L，LIU T，WANG G，et al. Video tracking using learned hierarchical features[J]. IEEE Transactions on Image Processing，2015，24(4)：1424-1435.

[15] CHEN H，FAN B. Hierarchical Convolutional Features for Long-Term Correlation Tracking[C]// CCF Chinese Conference on Computer Vision. Springer-Verlag，2017：677-686.

[16] YANG T，SHI Z，WANG G. Visual Tracking by Assembling Multiple Correlation Filters[C]// Pacific Rim Conference on Multimedia. Springer-Verlag，2015：310-320.

[17] HE Z，FAN Y，ZHUANG J，et al. Correlation Filters with Weighted Convolution Responses[C]// ICCV Workshops. IEEE，2017：1992-2000.

[18] VEDALDI A，LENC K. Matconvnet：Convolutional neural networks for matlab[C]// Proceedings of the 23rd ACM international conference on Multimedia. ACM，2015：689-692.