

## 车载视频下改进的核相关滤波跟踪算法

黄立勤 朱 飘\*

(福州大学物理与信息工程学院 福州 350116)

**摘要:** 针对相关滤波跟踪算法在车载视频下由于环境复杂及目标尺度变化等情况下容易跟踪失败的问题, 该文提出一种基于背景信息的尺度自适应相关滤波跟踪算法。首先利用背景感知相关滤波跟踪器融合方向梯度直方图特征预测目标下一帧位置, 然后根据预测位置选取图像块进行检测, 最后结合动态尺度比例金字塔模型对目标进行尺度估计。实验选取了 KITTI 数据库中 23 段车载视频和标注国内的 4 段车载视频进行测试, 实验结果表明, 该算法能有效降低车载环境的复杂背景、目标尺度变化等因素干扰, 整体性能优于 KCF, DSST, SAMF, SATPLE 等主流相关滤波算法, 对车载环境下复杂背景和尺度变化的目标跟踪具有鲁棒性。

**关键词:** 目标跟踪; 核相关滤波; 车载视频; 背景感知; 尺度估计

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2018)08-1887-08

DOI: 10.11999/JEIT171109

## Improved Kernel Correlation Filtering Tracking for Vehicle Video

HUANG Liqin ZHU Piao

(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China)

**Abstract:** For videos captured by in-car cameras, the filter-based tracking is a challenging task due to complex environments and mutable object scales. A scale adaptive tracking filter is proposed based on the background information. Firstly, the relative motion of each object is estimated by extracting features from gradient histograms between frames. Then, the object location on the next frame is determined and utilized to delimit an image block. Finally, the object scale is obtained through dynamic scaling pyramid model within image block. The proposed algorithm is examined by 27 in-car videos including 23 KITTI videos and 4 domestic videos. In experiments, the proposed algorithm suppresses effectively the interferences of environments and objects. It achieves more accurate and more robust object tracking than several popular benchmarks including KCF, DSST, SAMF, SATPLE.

**Key words:** Object tracking; Kernelized correlation filters; Videos captured by in-car cameras; Context aware; Scale estimation

### 1 引言

目标跟踪是计算机视觉领域一个核心的研究课题, 它在视频监控、智能交通、室外监控和人机交互等领域有着广泛的应用<sup>[1]</sup>。视觉跟踪任务一般指算法根据已知目标的初始位置信息来预测整个视频序列中该目标的运动轨迹。随着智能驾驶时代来临, 在车载视频序列中跟踪目标成为计算机视觉领域中一个重要问题。车载视频与其他固定摄像头拍摄的视频(如: 监控视频)相比, 背景更复杂多变, 同时

目标形变更明显且有剧烈的相对运动。然而现有的目标跟踪方法虽然能在固定摄像头下成功应用, 但是在车载摄像头下表现却不尽人意。可见影响跟踪有内外在诸多因素, 目标跟踪依然是计算机视觉中具有挑战性的问题。

相关滤波(Correlation Filter, CF)跟踪器具有较好的跟踪性能和计算效率而成为当前目标跟踪领域的研究热点之一。Bolme 等人<sup>[2]</sup>提出的最小输出平方误差和(Minimum Output Sum of Squared Error, MOSSE)算法将相关滤波思想引入跟踪, 跟踪速度可达数百帧每秒。Henriques 等人<sup>[3]</sup>通过训练正则化最小二乘(Regularized Least Squares, RLS)分类器进行跟踪判定, 之后他们又提取目标的方向梯度直方图(Histogram of Oriented Gradient, HOG)特征并建立了多通道特征融合模型<sup>[4]</sup>, 取得了较好的跟踪效果。随后毕笃彦等人<sup>[5]</sup>在颜色空间中引入加权的背景约束, 能较好抑制相似背景的干扰。Qi 等

收稿日期: 2017-11-27; 改回日期: 2018-04-18; 网络出版:

\*通信作者: 朱飘 piao@foxmail.com

基金项目: 国家自然科学基金(61471124), 福建省重大重点科技项目(2017H6009, 2018H0018), 赛尔网络创新项目(NGII20160208, NGII20170201)

Foundation Items: The National Natural Science Foundation of China (61471124), The Major Science and Technology Projects in Fujian Province (2017H6009, 2018H0018), The Cernet Innovation Projects (NGII20160208, NGII20170201)

人<sup>[6]</sup>充分利用 CNN 层级所抽取的特征建立多个弱相关滤波跟踪器, 并采用自适应的对冲思想将多个弱跟踪器整合成强跟踪器。但上述算法<sup>[4-6]</sup>在两个问题: 一是采用固定尺度的滤波模板, 不能适应目标尺度变化; 二是没有充分利用上下文信息, 在目标快速运动、遮挡和背景混淆时易引起跟踪漂移。在解决尺度变化问题上, Danelljan 等人<sup>[7]</sup>和 Li 等人<sup>[8]</sup>采用尺度金字塔思想, 但仍然存在两个问题: 一是该策略以目标原始大小为中心层, 向上下层进行权重递减叠加, 当目标尺度发生变化, 固定的权重叠加会导致真实目标尺度响应的减弱; 二是各层间具有一定间隔, 由金字塔对应层直接得到目标尺度, 则会出现估计偏差较大问题。因此本文引入动态尺度比例的金字塔模型, 形成可变的尺度搜索空间, 更加贴近目标尺度变化前后的规律, 从而更加精确估计尺度。与基于分块模型估计目标尺度的方法<sup>[9-11]</sup>不同, 不需要执行目标分块、计算、统计计算结果及判定等复杂操作步骤。针对复杂环境和快速运动导致的跟踪漂移问题, Camplani 等人<sup>[12]</sup>和 Ma 等人<sup>[13]</sup>、Danelljan 等人<sup>[14]</sup>引入了更强的景深信息和卷积特征来加强图像的表征力, 减少跟踪漂移, 但是由于特征维度高, 导致计算量增加, 跟踪效率下降; Danelljan 等人<sup>[15]</sup>和 Galoogahi 等人<sup>[16]</sup>则对滤波器空间分布进行正则约束和空间裁剪来提高训练样本的质量, 优化了由于目标快速运动而产生的边界效应, 提高了跟踪精度, 但是却破坏了 DCF (Dual Correlation Filter) 的闭合解从而采用高斯-塞德尔方法迭代求解最优化, 导致算法复杂度升高。本文通过灵活使用目标周围的背景信息来强化相关滤波的判别能力, 解决目标在复杂环境下快速运动的问题。本文方法与文献<sup>[14~17]</sup>的方法相比, 既不使用复杂的特征, 也不破坏岭回归的优化形式, 从而保留了高效的计算效率。

鉴于上文分析, 为增强 CF 模型在车载情况下对复杂环境及目标尺度变化的鲁棒性, 本文在相关滤波跟踪框架下提出基于加强背景感知和尺度自适应的相关滤波跟踪算法, 主要做了以下 3 方面工作: 一是提出一种动态尺度比例的金字塔模型, 解决了固定尺度金字塔策略中权重配比和尺度估计不精确的问题; 二是给相关滤波跟踪器加入上下文信息, 增强 CF 模型的判别力; 三是针对车载环境下, 使用公开车载视频数据库和自制车载视频数据库对本文算法进行多角度的对比实验。

## 2 本文算法

本文跟踪算法利用两个独立的相关滤波器分别实现目标定位和目标尺度估计, 本文与之前工作<sup>[7]</sup>

不同的是, 在估计位置滤波器中考虑了背景信息, 同时在估计尺度滤波器中使用动态尺度比例的金字塔模型。对于目标尺度估计, 使用独立的尺度滤波器明显优于只使用一个相关滤波器进行跟踪, 因为两个滤波器可灵活采用不同的方法和特征。本文算法在车载情况下, 利用背景信息辅助跟踪, 同时能够适应目标尺度变化的规律, 具有较高的跟踪性能。

### 2.1 背景感知的多特征融合相关滤波跟踪器

本文算法的基础是 MOSSE 跟踪器, MOSSE 跟踪器通过学习一个判别式相关滤波器来对下一帧图像目标定位。目前, 判别式相关滤波器使用多维特征已扩展到各种应用, 包括视觉跟踪<sup>[4,17]</sup>, 对象检测<sup>[18,19]</sup>和对象对齐<sup>[20]</sup>。与文献<sup>[4]</sup>类似, 本文在位置滤波器使用 HOG 特征并联系了图像强度特征。然而, 任何高维特征描述都可进行结合, 所以, 本文方法也同样能结合搜索区域上的高维特征。

考虑输入信号  $f$  (图像中的某一个图像块 patch) 具有  $d$  维特征描述, 为了构造最佳相关滤波器  $h$ , 可以通过最小化代价函数得到:

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

其中, 上标  $l$  表示特征的某一维度,  $\lambda$  为正则项系数。式(1)在频域求解得到式(2), 其中  $\lambda$  可以处理信号  $f$  中零频分量, 从而避免式(2)中分母为零。

$$H^l = \frac{\overline{GF}^l}{\sum_{k=1}^d \overline{F}^k F^k + \lambda} = \frac{A_t^l}{B_t} \quad (2)$$

由于图像块中每个像素点需要求解  $d \times d$  维的线性方程, 计算非常耗时, 为了得到鲁棒的近似结果, 对式(2)中  $H^l$  的分子  $A_t^l$  和分母  $B_t$  分别进行更新:

$$\left. \begin{aligned} A_t^l &= (1 - \eta)A_{t-1}^l + \eta \overline{G}_t F_t^l \\ B_t &= (1 - \eta)B_{t-1} + \eta \sum_{k=1}^d \overline{F}_t^k F_t^k \end{aligned} \right\} \quad (3)$$

其中,  $\eta$  为学习速率。在新一帧中, 目标位置通过求解最大相关滤波器响应得分  $y$  来确定:

$$y = F^{-1} \left\{ \frac{\sum_{l=1}^d \overline{A}^l Z^l}{B + \lambda} \right\} \quad (4)$$

上面主要讨论多维特征融合的相关滤波如何进行跟踪, 下面讨论在多维特征融合的条件下如何有效地加入背景信息到相关滤波器中。在目标跟踪中, 目标周围的环境对目标跟踪的性能有很大的影响, 在交通环境下影响更甚。因此, 本文在相关滤波器中加入有效的背景信息来辅助跟踪车载视频中的目标, 它能学习到目标周围的环境信息以提高在复杂

环境中的目标跟踪性能，尤其是在车载情况该相关滤波器也有着良好适用性。

首先，背景信息的选择对于跟踪性能非常重要，下面有两种背景选择策略<sup>[21]</sup>。第 1 种，围绕目标采取固定数目和固定大小的背景，这种方法简单通用；第 2 种，提取响应较大且离目标距离较远的图像块作为背景，这种策略可作为第 1 种方法的补充，也可单独作为一种背景选择策略。本文选择的是第 2 种背景选择策略。因为第 2 种对复杂环境，尤其是交通环境下有着更好的背景信息表达效果。两种策略均有助于滤波器在下一帧或者几十帧中更好地区分背景和遮挡物，下面介绍如何利用背景信息来提升跟踪性能。

在每一帧图像中，围绕感兴趣的目標  $\mathbf{a}_0 \in \mathbb{R}^n$  抽  $K$  个背景图像块  $\mathbf{a}_i \in \mathbb{R}^n$ 。它们分别对应的循环矩阵是  $\mathbf{A}_0 \in \mathbb{R}^{n \times n}$  和  $\mathbf{A}_i \in \mathbb{R}^{n \times n}$ 。这些背景图像块被认为是负样本，里面包含着两种全局背景信息(一是干扰物，二是各种各样的背景)，目标块回归到  $\mathbf{y}$  的问题，如公式(5)。经过学习后的滤波器  $\mathbf{w} \in \mathbb{R}^n$  能够对目标图像块有着高响应得分，而对于背景图像块的响应得分要尽可能地接近于 0。该方法把背景信息放在标准公式的正则项中，对所有背景块加约束，使得他们回归到 0。背景信息回归到 0，参数  $\lambda_2$  和  $\lambda_1$  是正则项系数，见式(6)。

$$\min_{\mathbf{w}} \|\mathbf{A}_0 \mathbf{w} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2 \quad (5)$$

$$\min_{\mathbf{w}} \left\{ \|\mathbf{A}_0 \mathbf{w} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2 + \lambda_2 \sum_{i=1}^k \|\mathbf{A}_i \mathbf{w}\|_2^2 \right\} \quad (6)$$

当然，对于公式的背景信息项还有其他可能的正则化方法，这里本文不做深入研究。

## 2.2 多尺度跟踪器

现有的引入尺度估计跟踪器中<sup>[8]</sup>，计算成本高且不能满足实时性。而在车载视频情况下实时性很重要，所以理想的尺度估计方法应该是鲁棒的，高效的。本文使用一个独立的滤波器进行尺度估计，将搜索区域限制在较小的固定尺度空间，可进行高效的尺度估计。在相关滤波跟踪框架中加入了一个独立的 1 维尺度相关滤波器，它会计算图像某个位置在尺度维度上的相关响应得分，然后取最大响应得分来估计目标尺度大小。具体做法是在新的一帧图像中，先利用 2 维位移滤波器(Translation Filter, TF)确定目标的新位置，再利用 1 维尺度滤波器(Scale Filter, SF)精确评估目标当前的尺度。假设当前帧为第  $t$  帧，使用 TF 预测下一帧位置  $m_{t+1}, n_{t+1}$ ，以该位置为中心提取  $S$  个不同尺寸的图像块，每个图像块的尺度为

$$a^n P_t \times a^n R_t, n \in \left\{ \left\lfloor -\frac{s-1}{2} \right\rfloor, \left\lfloor -\frac{s}{2} \right\rfloor, \dots, \left\lfloor \frac{s-1}{2} \right\rfloor \right\} \quad (7)$$

其中， $P$  和  $R$  分别为目标在当前帧的宽、高， $a$  是尺度因子，初始值为 1.01， $s = 33$  为尺度滤波器的长度，即尺度总级数。尺度估计过程使用方向梯度直方图(HOG)特征。最后用尺度相关滤波方法<sup>[7]</sup>找到输出响应最大值  $n'$  对应的尺度，计算下一帧目标尺度的估计值：

$$P_{t+1} \times R_{t+1} = a^{n'} P_t \times a^{n'} R_t \quad (8)$$

估计出第  $t+1$  帧的目标尺度  $P_{t+1} \times R_{t+1}$  后，以下一帧目标位置  $m_{t+1}, n_{t+1}$  为中心，选取大小  $3P_{t+1} \times 3R_{t+1}$  的图像块，用它来训练分类器，并根据式(3)更新两个滤波器参数。

由于尺度系数为底数大于 1 的指数函数，所以 33 种尺度不是线性增长的。比当前尺度越大的尺度级，步长越大，比当前尺度越小的尺度级，步长越小，即对较大的尺度进行粗检测，对较小的尺度进行细检测。所以为了贴近目标尺度的变化规律，本文将尺度因子根据当前帧和下一帧尺度估计两者的差值进行动态地调整，具体公式如式(9)：

$$a = a + \left| \frac{S_{t+1} - S_t}{2} \right| \times 0.01, a \in 1.01, 1.05 \quad (9)$$

式中， $S_{t+1}$  和  $S_t$  分别表示的是  $t+1$  和  $t$  时刻的估计尺度级，范围为  $[1, 33]$ ，尺度因子  $a$  范围为  $[1.01, 1.05]$ 。

本文算法跟踪流程如图 1 所示。

## 3 实验结果及分析

本文实验针对车载情况设计了位置估计和尺度估计两个实验和一个综合性实验以对算法的有效性进行评估。一是对算法的目标位置跟踪精度进行对比测试；二是对算法的尺度适应性进行对比测试；三是与现阶段性能较好的跟踪算法进行综合实验对比。实验所用对比算法均采用作者公开的源码设置，测试序列来自公开车载数据集 KITTI<sup>[22]</sup>和自制车载跟踪数据。评价标准采用文献<sup>[23]</sup>所提到的 4 个指标：中心位置误差(Center Location Error, CLE)、距离精度(Distance Precision, DP)、重叠精度(Overlap Precision, OP)和跟踪速度(Frames Per Second, FPS)。CLE 是指跟踪结果位置与目标基准位置之间的欧式距离，单位为 Pixel(像素)；DP 是 CLE 小于某个阈值(通常为 20 Pixel)<sup>[23]</sup>的帧数占序列总帧数的百分比；重叠率  $O$  表示预测目标和真实目标重叠区域占两者总区域的百分比。OP 表示重叠率  $O$  大于 0.5 的帧数占视频总帧数的百分比，在所有评价标准中，OP 指标是衡量算法多尺度性

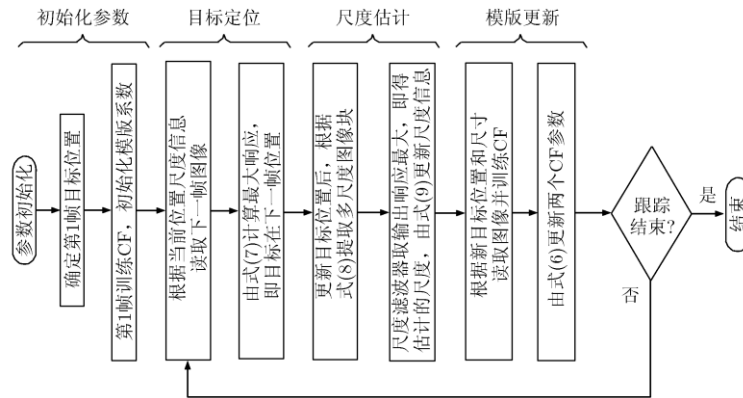


图 1 本文算法跟踪流程图

能最重要的标准。值得注意的是，CLE 指标越小跟踪精度越高，DP, OP 和 FPS 指标越大跟踪性能越高。

3.1 实验环境及参数

本文实验环境：CPU 型号为 2.8 GHz Intel (R) Core(TM) i5, 内存为 16 GB, 操作系统为 Windows7 64 bit, 软件平台为 MATLAB R2014a。本文算法参数设置如下：相关滤波跟踪器中，训练检测区域为目标大小的 3 倍, HOG 特征的元细胞大小为  $4 \times 4$  Pixel, 统计梯度方向数为 9, 正则化参数  $\lambda$  为  $10^{-4}$ ,  $\delta = 0.5$ ,  $\delta_{\text{label}} = 0.1$ , 学习速率  $\eta = 0.025$ ；尺度模型中，尺度数  $S=33$ , 尺度变化因子初始值  $a = 1.01$ 。

3.2 数据库

本文实验数据包含 27 个视频序列, 其中 KITTI 数据库有 24 段视频序列共 4818 帧, 自行标注车载视频序列 3 段共 2201 帧。这些视频序列包含了车载视频下目标跟踪可能遇见的主要问题, 如光照变化、尺度变化、目标遮挡、目标丢失、快速运动、旋转、小目标、背景混淆、抖动和视角变化等。

3.3 实验部分

实验 1 尺度优化后与现有跟踪器结果比较

为了评估在跟踪过程中使用动态尺度比例的金字塔模型的有效性, 实验对比分析了 KCF, DSST, SAMF 3 种跟踪方法, 它们是带有固定尺度比例的金字塔模型的尺度自适应跟踪滤波器。如表 1 所示,

使用 4 个算法对 12 段车载视频进行测试, 车载视频中目标尺度变化且相对运动较为剧烈, 适合做算法做尺度自适应的对比试验。表中数据指的是重叠率  $O$ , 即 12 个视频序列中每个视频重叠精度(OP)占整个视频的百分比, 每个序列下重叠率精度最高的用粗体表示。可以看出, 在 12 段视频中, 本文算法有 8 段视频表现好于其他三者, 这表明本文所提出的动态尺度比例的金字塔模型能够较好的处理目标尺度的变化。

实验 2 复杂环境下跟踪性能测试 为验证本文算法在复杂交通环境下的跟踪性能, 本次实验选择 2 种代表性的相关滤波跟踪算法, 他们是 DCF 和 STAPLE<sup>[24]</sup>, 并在 15 段典型的交通环境视频序列下进行测试对比。表 2 为 2 个算法及 2 个算法加入背景信息后(算法名后加上 “\_AB”)在车载视频中的目标跟踪精度的对比。该实验在 15 个序列上测试, 表中数据指的是跟踪精度, 即 15 个视频序列中每个视频中心位置误差(CLE)小于 20pixel 的帧数占整个视频帧数的百分比, 每个序列下最好的结果用粗体表示。因为 15 段视频中 4 个算法在其中 3 个序列上的跟踪精度都是 1, 没有比较的意义, 故表中将它们舍抹去。从表 2 中上下数据对比上可以看出, 在交通环境下加上背景信息能够有效提高目标跟踪精度, 所以在跟踪器加入背景信息是非常有必要的。

表1 重叠率精度结果对比(%)

方法	视频段序号											
	1	2	3	4	5	6	7	8	9	10	11	12
本文算法	<b>100</b>	<b>100</b>	<b>34.7</b>	24.7	<b>69.3</b>	30.5	<b>5.76</b>	<b>58.3</b>	<b>35.2</b>	6.2	<b>15.5</b>	94.8
DSST	<b>100</b>	<b>100</b>	26.7	22.4	66.2	28.6	4.32	50.0	30.8	5.0	11.1	93.9
KCF	61.0	34.6	18.7	43.1	48.5	<b>33.8</b>	1.44	45.9	26.9	35.0	5.0	<b>95.5</b>
SAMF	98.3	71.4	21.3	<b>80.1</b>	66.2	27.9	2.16	32.4	34.6	<b>37.5</b>	5.56	94.7

表 2 跟踪精度对比

方法	视频段序号											
	16	17	18	19	20	21	22	23	24	25	26	27
DCF_AB	<b>0.824</b>	0.680	<b>0.657</b>	0.902	1.00	<b>0.736</b>	0.061	0.135	<b>0.548</b>	<b>0.655</b>	0.125	0.047
DCF	0.706	<b>0.772</b>	0.597	0.902	1.00	0.610	0.061	<b>0.635</b>	0.405	0.069	<b>1.00</b>	0.047
STAPLE_AB	<b>0.882</b>	<b>0.567</b>	<b>0.874</b>	0.927	1.00	<b>0.902</b>	0.061	<b>0.615</b>	0.500	0.069	0.141	0.023
STAPLE	0.529	0.552	0.845	0.927	1.00	0.683	0.061	0.423	0.500	<b>0.103</b>	0.141	<b>0.047</b>

**实验 3 尺度优化和加入背景后与现有跟踪器进行比较** 如表 3 所示, 在 27 段车载视频序列上测试本文算法与各主流相关滤波跟踪算法的性能表现。表中显示的数据是各个算法在 27 段视频中性能表现的中位数, 即每个算法在 27 段序列上测试, 那么每个指标都会有 27 个性能值, 最终选择性能值排序后第 14 个性能值作为对比值, 每个指标中最好的结果用粗体表示。与 DSST, KCF, SAMF, SRDCF 和 STABLE 5 个算法相比, 本文算法除了跟踪速度外各个性能指标均有提升, 所以在处理交通环境复杂多变和尺度剧烈变化的目标时本文算法要明显优于其他 5 个算法, 且跟踪速率仍快于 SAMF 算法, 完全满足在车载视频下的实时跟踪。

为了便于直观突出本文算法在处理交通环境下目标跟踪问题的优越性, 图 2 和图 3 给出了 6 个跟踪算法在 3 段典型的车载视频下的成功率曲线和距离精度曲线, 红色曲线代表本文算法在视频中的跟踪表现。图 2 是重叠率阈值-重叠率精度曲线, 即成功率曲线。成功率曲线表示不同重叠阈值下, 目标

跟踪成功帧数占整段视频帧数的比例, 成功率曲线与坐标轴围成的面积越大, 代表跟踪算法在目标尺度变化上处理得越好。从图 2 可以看出, 红色曲线所围成的面积均大于其他的曲线, 说明本文算法在处理尺度变化问题的有效性。图 3 是 CLE-帧序号曲线, 即距离精度曲线。它表示不同的帧序号下跟踪框中心位置与真实框中心位置的距离差, 单位为 pixel。距离精度曲线和成功率曲线不同, 与坐标轴围成的面积越小, 代表跟踪算法在目标位置变化上处理得越好。从图 3 中可以看出, 红色曲线围成的面积最小, 说明本文算法能够紧跟目标的位置变化。综上所述, 本文通过引入动态尺度比例的金字塔模型和考虑背景信息使得尺度自适应性能和跟踪精度均优于比较方法。

车载视频下目标的跟踪速度是一项重要指标, 本文随机抽取了 12 段视频序列作为速度测试数据以分析跟踪速度的变化。图 4 为 MOSSE 跟踪器加入背景信息和尺度优化前后跟踪速度变化柱状图; 图中横坐标为视频段的序号, 纵坐标为跟踪速度,

表 3 各个算法性能对比

	本文算法	DSST	KCF	SAMF	SRDCF	STABLE
OP(%)	<b>50.3</b>	42.6	35.9	44.6	47.6	48.1
DP(%)	<b>62.5</b>	48.6	53.8	53.8	57.7	59.9
CLE(pixel)	<b>35.7</b>	47.1	45.3	44.9	39.6	37.0
FPS(fps)	35.3	50.8	<b>437.6</b>	23.4	13.9	44.3

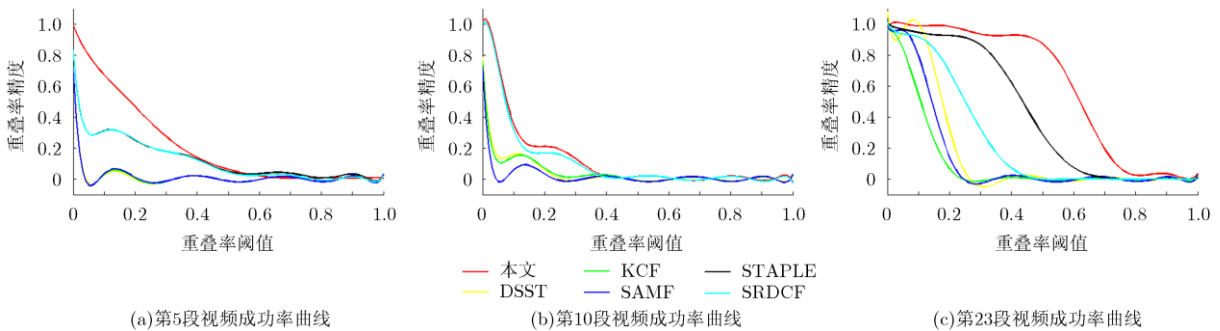


图 2 成功率曲线

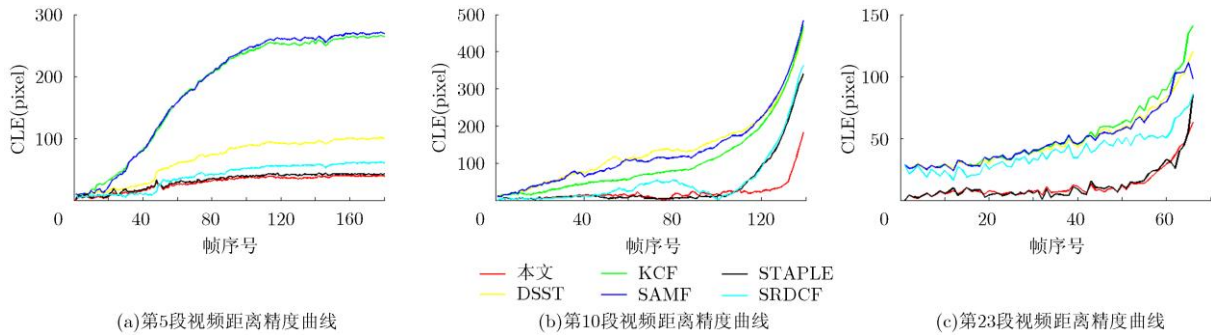


图3 距离精度曲线

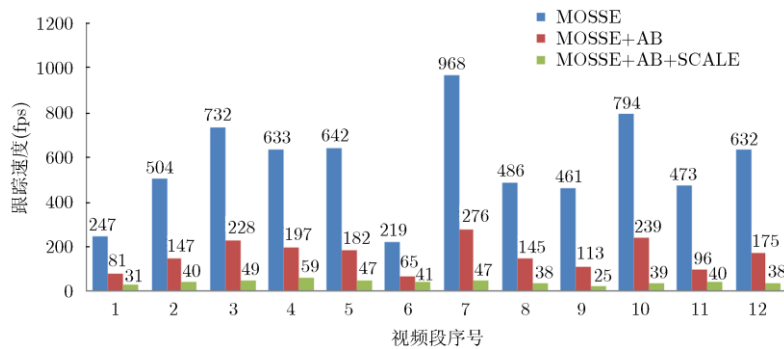


图4 优化尺度和加入背景前后的跟踪速度变化

单位为 fps; 图中共有 12 组由蓝、红、绿 3 种颜色组成条形柱组, 每一组中 3 条条形柱分别代表 3 种跟踪器在同一视频段上的跟踪速度。图中蓝色代表的是 MOSSE 跟踪器, 在每一组中速度都远高于其他两者; 红色代表的是 MOSSE 跟踪器加入背景信息后的跟踪器(MOSSE+AB), 因为跟踪器增加了筛选背景块和提取背景样本的步骤、训练过程数据量增加等因素, 每一组中 MOSSE+AB 跟踪速度较 MOSSE 都下降明显, 当然跟踪速度下降的程度也与目标的大小及背景复杂度有关; 绿色代表 MOSSE 跟踪器加入背景信息和尺度优化后的跟踪器(MOSSE+AB+SCALE), 基于 MOSSE+AB 加入尺度优化后跟踪速度下降的原因主要是增加了一个独立的尺度滤波器, 它使用灵活的尺度调整策略, 尺度搜索空间随着目标变化而变化, 导致计算步骤和计算量增加, 但是相对于其他两者, 各段视频序列跟踪速度较为平稳, 没有出现剧烈的波动且总体来看, 算法满足实时性要求。

### 3.4 跟踪结果实例分析

图 5 给出了本文跟踪方法和主流的 6 种核相关滤波跟踪方法在 3 段具有代表性的车载视频序列中的跟踪结果, 每段视频选择 3 帧, 其中不同方法的跟踪结果使用不同颜色的跟踪框进行区分。这些车载视频序列中的跟踪目标存在遮挡、尺度变化、快

速运动、抖动、小目标、视角变化、背景混淆、光照变化等问题。从图 5(a)目标尺度的跟踪中可以看出, 本文算法能够一直能够紧贴目标尺寸, 而其他算法则出现较大偏差; 从图 5(b)小目标的跟踪可以看出, 本文算法没有出现跟踪漂移且能够精确估计目标的尺寸, 而其他算法基本都出现了跟踪丢失; 从图 5(c)快速运动的目标跟踪中可以看出, 本文算法跟踪目标能够兼顾位置和尺寸, 而其他算法都出现了一定的跟踪漂移和尺度偏差。综上, 本文方法在处理车载视频下目标跟踪的常见问题均有一定的成效, 能够有效降低环境和目标运动等干扰因素, 达到了很好的跟踪效果, 且明显要优于现有主流的相关滤波算法。

## 4 结论

本文提出一种在车载情况下实现实时性和高精度的尺度自适应目标跟踪方法。本文方法通过加入更多的背景信息和更加合理的尺度调整策略来处理目标尺度变化和背景快速变化的问题。大量的实验表明: 与现有跟踪相比, 本文跟踪器在同样的计算效率下有更高的性能表现。在交通环境下本文方法保证了目标实时跟踪, 跟踪器性能优于主流跟踪器同时更具鲁棒性。



图5 6种算法在车载视频序列下的跟踪结果图

### 参考文献

- [1] 刘红亮, 周生华, 刘宏伟, 等. 一种航迹恒虚警的目标检测跟踪一体化算法[J]. 电子与信息学报, 2016, 38(5): 1072-1078. doi: 10.11999/JEIT150638.
- LIU Hongliang, ZHOU Shenghua, LIU Hongwei, et al. An integrated target detection and tracking algorithm with constant track false alarm rate[J]. *Journal of Electronics & Information Technology*, 2016, 38(5): 1072-1078. doi: 10.11999/JEIT150638.
- [2] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]. *Computer Vision and Pattern Recognition*, San Francisco, 2010: 2544-2550.
- [3] HENRIQUES J F, RUI C, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]. *European Conference on Computer Vision*, Florence, 2012: 702-715.
- [4] HENRIQUES J F, RUI C, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 37(3): 583-596. doi: 10.1109/tpami.2014.2345390.
- [5] 毕笃彦, 库涛, 查宇飞, 等. 基于颜色属性直方图的尺度目标跟踪算法研究[J]. 电子与信息学报, 2016, 38(5): 1099-1106. doi: 10.11999/JEIT150921.
- BI Duyan, KU Tao, ZHA Yufei, et al. Scale-adaptive object tracking based on color names histogram[J]. *Journal of*

- Electronics & Information Technology*, 2016, 38(5): 1099–1106. doi: 10.11999/JEIT150921.
- [6] QI Yuankai, ZHANG Shengping, QIN Lei, *et al.* Hedged deep tracking[J]. *Computer Vision and Pattern Recognition*, 2016, 4303–4311. doi: 10.1109/cvpr.2016.466.
- [7] DANELLJAN M, HÄGER G, KHAN F S, *et al.* Accurate Scale Estimation for Robust Visual Tracking[C]. British Machine Vision Conference, Nottingham, 2014: 61–65.
- [8] LI Yang and ZHU Jianke. A scale adaptive kernel correlation filter tracker with feature integration[C]. European Conference on Computer Vision, 2014, 8926: 254–265. doi: 10.1007/978-3-319-16181-5\_18.
- [9] XU Yulong, WANG Jiabao, LI Hang, *et al.* Patch-based scale calculation for real-time visual tracking[J]. *IEEE Signal Processing Letters*, 2015, 23(1): 40–44. doi: 10.1109/wesp.2015.7341015.
- [10] AKIN O, ERDEM E, ERDEM A, *et al.* Deformable part-based tracking by coupled global and local correlation filters[J]. *Journal of Visual Communication & Image Representation*, 2016, 38(C): 763–774. doi: 10.1016/j.jvcir.2016.04.018.
- [11] YAO Rui, XIA Shixiong, SHEN Fumin, *et al.* Exploiting spatial structure from parts for adaptive kernelized correlation filter tracker[J]. *IEEE Signal Processing Letters*, 2016, 23(5): 658–662. doi: 10.1109/lsp.2016.2545705.
- [12] CAMPLANI M, HANNUNA S, MIRMEHDI M, *et al.* Real-time RGB-D tracking with depth scaling kernelized correlation filters and occlusion handling[C]. British Machine Vision Conference, SWANSEA, 2015. 2015: 141–145. doi: 10.5244/c.29.145.
- [13] MA Chao, HUANG Jiabin, YANG Xiaokang, *et al.* Robust Visual Tracking via Hierarchical Convolutional Features[J]. *Computer Vision and Pattern Recognition*, 2017, (2017): 425–434. doi: 10.1007/978-3-319-70090-8\_44.
- [14] DANELLJAN M, BHAT G, KHAN F S, *et al.* ECO: Efficient convolution operators for tracking[C]. *Computer Vision and Pattern Recognition*, Honolulu, 2017: 21–26. doi: 10.1109/cvpr.2017.733.
- [15] DANELLJAN M, HAGER G, KHAN F S, *et al.* Learning spatially regularized correlation filters for visual tracking[C]. International Conference on Computer Vision, Santiago, 2015: 4310–4318. doi: 10.1109/iccv.2015.490.
- [16] GALOOGAHI H K, FAGG A, and LUCEY S. Learning background-aware correlation filters for visual tracking[J]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017: 21–26. doi: 10.1109/iccv.2017.129.
- [17] DANELLJAN M, KHAN F S, FELSBERG M, *et al.* Adaptive color attributes for real-time visual tracking[C]. *Computer Vision and Pattern Recognition*, Washington, 2014: 1090–1097.
- [18] GALOOGAHI H K, SIM T, and LUCEY S. Multi-channel correlation filters[C]. IEEE International Conference on Computer Vision, Sydney, 2013: 3072–3079.
- [19] RUI C and BATISTA J. Beyond hard negative mining: efficient detector learning via block-circulant decomposition[C]. IEEE International Conference on Computer Vision, Sydney, 2013: 2760–2767.
- [20] BODDETI V N, KANADE T, and KUMAR B V K V. Correlation filters for object alignment[C]. *Computer Vision and Pattern Recognition (CVPR)*, Portland, 2013: 2291–2298.
- [21] MUELLER M, SMITH N, and GHANEM B. Context-aware correlation filter tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017: 1387–1395.
- [22] GEIGER A, LENZ P, STILLER C, *et al.* Vision meets robotics: the KITTI dataset[J]. *International Journal of Robotics Research*, 2013, 32(11): 1231–1237. doi: 10.1177/0278364913491297.
- [23] WU Yi, LIM Jongwoo, and YANG Minghsuan. Online object tracking: A benchmark[C]. *Computer Vision and Pattern Recognition*, Portland, 2013: 2411–2418.
- [24] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al.* Staple: Complementary learners for real-time tracking[C]. *Computer Vision and Pattern Recognition*, Las Vegas, 2016: 1401–1409.
- 黄立勤: 男, 1973年生, 教授, 主要研究方向为图像处理与计算机视觉、机器学习算法在视频数据分析、医学图像处理及辅助诊断中的应用。
- 朱 飘: 男, 1992年生, 硕士生, 研究方向为机器学习、模式识别、图像处理。