

分类号: \_\_\_\_\_  
密 级: \_\_\_\_\_

单位代码: \_\_\_\_\_  
学 号: \_\_\_\_\_

# 南京信息工程大学

## 硕士学位论文



### 基于时空上下文学习的快速目标跟踪算法研究

### Fast Object Tracking via Spatio-temporal context Learning

申请人姓名: 樊佳庆

指导教师: 张开华 教授

专业名称: 系统科学

研究方向: 视频单目标跟踪

所在学院: 自动化学院

二〇一八年十一月

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的科研成果。本论文除了文中特别加以标注和致谢的内容外，不包含其他人或其他机构已经发表或撰写过的研究成果，也不包含为获得南京信息工程大学或其他教育机构的学位或证书而使用过的材料。其他同志对本研究所做的贡献均已在论文中作了声明并表示谢意。

学位论文作者签名：\_\_\_\_\_ 签字日期：\_\_\_\_\_

## 关于论文使用授权的说明

南京信息工程大学、国家图书馆、中国学术期刊（光盘版）杂志社、中国科学技术信息研究所的《中国学位论文全文数据库》有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文，并通过网络向社会提供信息服务。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权南京信息工程大学研究生院办理。

学位论文作者签名：\_\_\_\_\_ 签字日期：\_\_\_\_\_

指导教师签名：\_\_\_\_\_ 签字日期：\_\_\_\_\_

## 目 录

摘 要	I
Abstract	II
第一章 绪论	1
0.1 研究背景	1
0.2 国内外研究现状	1
0.2.1 快速相关滤波类目标跟踪	1
0.2.2 神经网络类目标跟踪	1
0.3 文章简介与结构	1
0.3.1 本文工作简介	1
0.3.2 本文结构	1
1 第二章 相关工作和评价指标	2
1.1 相关工作	2
1.1.1 基于相关滤波的目标跟踪	2
1.1.2 基于流形正则的目标跟踪	2
1.1.3 整合多类估计的跟踪	2
1.2 评价指标	3
1.2.1 OTB 数据集	3
1.2.2 VOT 数据集	3
1.2.3 Temple-Color 数据集	3
2 第三章 基于流形正则的上下文感知相关滤波跟踪	4
2.1 动机	4
2.2 做法	4
2.3 数据集上的结果	4
3 第四章 通过颜色聚类与时空正则相关学习的互补型跟踪	5
3.1 动机	5
3.2 做法	5
3.3 数据集上的结果	5
结 论	6
参考文献	8

攻读硕士学位期间发表学术论文情况 .....	9
致    谢 .....	10

## 摘 要

尽管已研究多年,快速并且高性能的目标跟踪算法依然缺失。其中一个关键问题是,视频帧中的时空上下文信息并未被充分利用,而这部分信息对于相关滤波跟踪的性能影响重大。为了解决上述问题,本文主要基于时空上下文学习对快速目标跟踪算法进行研究,使之能更有效地利用时空上下文信息建模,主要贡献概括为如下 2 点:

1. 尽管相关滤波类跟踪已经被研究多年,但是,其固有的样本循环假设也引入了严重的冗余,这样不利于学到一个有效的分类器。在本文中,我们开发了一个快速的流形正则上下文感知相关滤波跟踪算法,来挖掘不同类型样本间局部的流形结构信息。首先,不同于只利用一个基样本的传统相关滤波类跟踪,我们使用了一系列基样本周围的上下文样本,并对其进行了流形结构假设。然后,考虑到这些样本中存在的流形结构,我们在相关滤波(CF)学习中,引入了一个线性图拉普拉斯正则项。幸运的是,这个优化问题能够利用快速傅里叶变换得到一个闭式解,因而能够高效地实现。大量在 OTB100 和 VOT2016 数据集上的实验评估表明,提出的跟踪算法在准确性和鲁棒性方面,表现优于几个最先进的算法的。特别地,我们的跟踪器能够在单个 CPU 上以 28fps 的速度实时运行。

2. 最近,一种简单而高效名叫 Staple 的跟踪算法在效率和准确度方面,取得了很有潜力的结果。它配备有判别式相关滤波(DCFs)跟踪器和颜色直方图的互补学习器,这种互补特性使得它对颜色变化和形变都具有很强的适应性。然而,它也有一些缺点:(i)Staple 对于所有序列,只使用标准颜色直方图,量化标准相同,并没有考虑目标在每个序列中的具体结构信息,从而影响其区分目标和背景的能力。(ii)Staple 中使用的标准 DCFs 是有效的,但是存在多余的边界效应问题,导致一些具有挑战性的场景中的跟踪失败。为了解决这些问题,我们提出了基于颜色聚类与时空正则相关回归的互补跟踪器(CSCT)。所提出的 CSCT 包含两个互补的部分,可以自适应地处理每个序列的显著颜色变化和形变问题:首先,我们设计了一个新颖的基于颜色聚类的直方图模型,该模型首先自适应地将第一帧目标的颜色划分为几个颜色聚类中心,构建自适应颜色直方图,这样一来模型就能适应严重的目标形变问题。另外,我们提出学习时空正则化的 CFs,既能缓解边界效应,在目标剧烈变化时,又能提供一个比原 Staple 里标准 DCF 更稳健的表观模型。与 Staple 相比,我们的 CSCT 仅仅利用简单的手工特征,在 OTB100、Temple-Color 和 VOT2016 这三个数据集上,分别提升了 5.9%, 3.4% 和 1.5%。此外,我们的 CSCT 甚至超过了几个最先进的复杂深度网络跟踪器。

**关键词:** 目标跟踪; 相关滤波; 时空正则; 上下文学习

## Abstract

Despite years of research, fast and high-performance target tracking algorithms are still missing. One of the key issues is that the spatio-temporal context information in frame video is not fully utilized, and this part of information has a significant impact on correlation filtering tracking performance. In order to solve the above problems, in this thesis, we develop fast object tracking algorithms mainly based on spatio-temporal context learning. The main contribution of this study is summarized as follows:

(1) Despite the demonstrated success of numerous correlation filter (CF) based tracking approaches, their assumption of circulant structure of samples introduces significant redundancy to learn an effective classifier. In this paper, we develop a fast manifold regularized context-aware correlation tracking algorithm that mines the local manifold structure information of different types of samples. First, different from the traditional CF based tracking that only uses one base sample, we employ a set of contextual samples near to the base sample, and impose a manifold structure assumption on them. Afterwards, to take into account the manifold structure among these samples, we introduce a linear graph Laplacian regularized term into the objective of CF learning. Fortunately, the optimization can be efficiently solved in a closed form with fast Fourier transforms (FFTs), which contributes to a highly efficient implementation. Extensive evaluations on the OTB100 and VOT2016 datasets demonstrate that the proposed tracker performs favorably against several state-of-the-art algorithms in terms of accuracy and robustness. Especially, our tracker is able to run in real-time with 28 fps on a single CPU.

(2) Recently, a simple, yet effective and efficient tracker named Staple has achieved promising performance in terms of efficiency and accuracy on a series of visual tracking benchmarks. Staple is equipped with complementary learners of discriminative correlation filters (DCF) and color histograms, which are robust to both color changes and deformations. However, it has some drawbacks: (i) Staple only employs standard color histograms with the same quantization step for all sequences, which does not consider the specific structural information of target in each sequence, thereby affecting its discriminative capability to separate target from background. (ii) The standard DCFs are efficient but suffer from unwanted boundary effects, leading to failures in some challenging scenarios. To address these issues, we present a dual color clustering and spatio-temporal regularized correlation regressions based complementary tracker (CSCT). The

proposed CSCT includes two components with complementary merits to adaptively deal with significant color variations and deformations for each sequence: First, we design a novel color clustering based histogram model that first adaptively divides the colors of the target in the 1st frame into several cluster centers, and then the cluster centers are taken as references to construct adaptive color histograms for targets in the coming frames, which enable to adapt significant target deformations. Second, we propose to learn spatio-temporal regularized CFs, which not only enable to avoid boundary effects but also provides a more robust appearance model than the discriminative CFs in Staple in the case of large appearance variations. Compared with Staple, our CSCT with handcrafted features achieves a gain of 5.9%, 3.4% and 1.5% on OTB100, Temple-Color and VOT2016 benchmarks in terms of AUC and EAO scores, respectively. Moreover, our CSCT performs favorably against several state-of-the-art trackers including the deep learning based trackers.

**Key Words:** Visual tracking; correlation filter; spatio-temporal regularization; context learning





## 第一章 绪论

### 0.1 研究背景

视觉跟踪是目前国内外研究的热点之一，在计算机视觉中众多的应用，如视频监控，运动分析，自动驾驶，举几个例子<sup>[1-5]</sup>。尽管近年来取得了很大进展，但开发一种鲁棒跟踪算法仍然是一个挑战，主要由于目标外观的显著变化而引起的例如光照变化，快速运动，姿势变化，部分遮挡和背景混乱等。正因为遭受了这些挑战，一个鲁棒的表示在视觉跟踪中更显得尤为重要，因此在过去的几十年中它引起了广泛的关注。

### 0.2 国内外研究现状

#### 0.2.1 快速相关滤波类目标跟踪

相关滤波类，在线训练，速度快，效果较好。

#### 0.2.2 神经网络类目标跟踪

孪生网络类，训练慢，测试快，效果好。

### 0.3 文章简介与结构

#### 0.3.1 本文工作简介

基于时空上下文学习，我们做了 2 个工作。

- (1) 提出了基于流形正则的上下文感知相关滤波跟踪。
- (2) 提出了通过颜色聚类与时空正则相关学习的互补型跟踪。

#### 0.3.2 本文结构

二、三章分别详细介绍这两个工作。

## 1 第二章 相关工作和评价指标

### 1.1 相关工作

#### 1.1.1 基于相关滤波的目标跟踪

最近, 判别式相关滤波跟踪 (DCF) 在视觉跟踪方面引起了广泛的关注, 由于它们在效率和稳健性方面的优势。使用 DCFs 进行视觉跟踪从 MOSSE<sup>[6]</sup> 开始, 它在频率域中用几个样本学习 CFs, 又利用快速傅里叶变换 (FFTs) 高效地计算, 最终能以 669 帧/每秒 (FPS) 运行。在<sup>[7]</sup> 中, Henriques 等人首先探索了循环密集样品的结构与核化嵌入, 学习出 CFs 进行快速跟踪。在<sup>[8]</sup> 中, Henriques 等人进一步改进了<sup>[7]</sup> 中的 CF 跟踪器, 把特征表示从原始图像强度升级到梯度方向直方图。Ma 等人<sup>[9]</sup> 利用不同层次深层特征间的互补, 使用由粗至细的特征搜索策略, 学习更有效的视觉 CFs 跟踪, 显著提高了相关滤波在 OTB100 数据集上的性能。最近, Danelljan 等人在空间上提出了一系列的基于正则化 CF 的跟踪器<sup>[10-12]</sup>, 取得了令人印象深刻的性能。空间正则化 DCF (SRDCF) 跟踪器<sup>[10]</sup> 试图抑制学习 CFs 的边界效应, 它利用了具有高斯形状的空间正则化权值。基于<sup>[10]</sup>, <sup>[11]</sup> 提出了一种自适应去污方案, 因而学习到了更有效的 CFs, 它能自适应地学习可靠性并消除各训练样本中被污染的样本。在<sup>[12]</sup> 中, 是在各种特征映射的连续空间域内学习 CFs, 能够获得亚像素级别的跟踪精度。

#### 1.1.2 基于流形正则的目标跟踪

流形正则化通常用于半监督的有标记和无标记样本学习<sup>[13-15]</sup>, 它构造了一个 Laplacian 图来利用样本利用特征空间的隐式几何结构。例如, 在特征空间分析中, Chang 和 Yang<sup>[14]</sup> 利用标记和未标记的训练数据进行更多的研究可靠的特征空间选择算法。此外, 在视觉跟踪里, Yu<sup>[15]</sup> 等人利用具有时空约束的流形结构进行跟踪, 在现实世界的人员定位和监控场景中具有较好的效果。Bai 和 Tang<sup>[16]</sup> 用了一个在线拉普拉斯正则化的排序支持向量机, 来估计视觉跟踪的对象位置。为了更好地使用无标记数据和流形结构样本空间, Hu 等人<sup>[17]</sup> 提出了一个基于流形正则化 DCF 的跟踪器, 增加了循环位移的样本并利用一个块优化策略, 可以有效地通过 FFTs 计算。Zhuang 等人<sup>[18]</sup> 为视觉跟踪构建了一个判别式的稀疏相似图, 它是基于多任务的拉普拉斯正则反稀疏表示。

#### 1.1.3 整合多类估计的跟踪

有一种常用的用来减少不精确预测的策略, 就是将一系列方法的估计结合起来, 这样跟踪器的缺点就会得到相应的补偿。在<sup>[19,20]</sup> 里, Kwon 等人利用互补的基础跟踪器,

结合不同的观测模型和运动模型，然后在一个采样的框架中集成他们的估计。在<sup>[21]</sup>中，Wang 和 Yeung 通过一个因子的 HMM 结合了几个独立的跟踪器，同时建模了目标轨迹和每个跟踪器随着时间变化的可靠性。不同于使用不同类型的跟踪器，多专家最小熵跟踪器保留了一个过去模型的集合，并且根据熵判断准则选择出了一个最优的预测。在<sup>[22]</sup>中，Bertinetto 等人直接合并了两个常见的岭回归得分，这里的局部表示利用 HOG 特征，全局表示利用了颜色特征，最后两者一同互补地工作。

## 1.2 评价指标

### 1.2.1 OTB 数据集

OTB100 数据集包含了 100 个测试视频序列，使用了两个典型的评价标准。第一个是精度，即定位误差小于阈值的帧所占的百分比。定位误差定义为跟踪中心与真实边界框中心之间的距离 (以像素为单位)。另一个是成功率，即重叠率大于阈值的帧所占的百分比。重叠得分被定义为

$$OS = \frac{|G_{rec} \cap T_{rec}|}{|G_{rec} \cup T_{rec}|}, \quad (1.1)$$

这里的  $G_{rec}$  表示真实边界框， $T_{rec}$  是跟踪到的边界框。跟踪结果的得分只要高于阈值  $t$  就会被认为是一次成功的跟踪。最终的成功率图显示了重叠分数大于  $t$  的帧所占百分比，其中  $t$  在 0 到 1 之间变化。利用曲线下面积 (AUC) 对评估后的跟踪器进行排序<sup>[23]</sup>。

### 1.2.2 VOT 数据集

VOT 数据集每年都以挑战赛的形式更新，一般每年的 VOT 数据集都包含了 60 个视频序列，它使用了预期的平均覆盖率 (EAO) 作为评价标准，这个得分是根据准确性、稳健性这两个指标计算出来的。准确性是指预测到的边界框和真实边界框之间的平均覆盖率。而稳健性度量了在跟踪中跟丢目标的次数，反映了跟踪算法针对不同视频的稳定性<sup>[24]</sup>。

### 1.2.3 Temple-Color 数据集

Temple-Color 数据集包含了 128 个颜色视频序列，旨在于评价跟踪算法在彩色视频序列上的性能<sup>[25]</sup>。

## 2 第三章 基于流形正则的上下文感知相关滤波跟踪

### 2.1 动机

尽管相关滤波类跟踪已经被研究多年，但是，其固有的样本循环假设也引入了严重的冗余，这样不利于学到一个有效的分类器。

### 2.2 做法

在本节中，我们开发了一个快速的流形正则上下文感知相关滤波跟踪算法，来挖掘不同类型样本间局部的流形结构信息。

首先，不同于只利用一个基样本的传统相关滤波类跟踪，我们使用了一系列基样本周围的上下文样本，并对其进行了流形结构假设。然后，考虑到这些样本中存在的流形结构，我们在相关滤波（CF）学习中，引入了一个线性图拉普拉斯正则项。幸运的是，这个优化问题能够利用快速傅里叶变换得到一个闭式解，因而能够高效地实现。

### 2.3 数据集上的结果

大量在 OTB100 和 VOT2016 数据集上的实验评估表明，提出的跟踪算法在准确性和鲁棒性方面，表现优于几个最先进的算法的。特别地，我们的跟踪器能够在单个 CPU 上以 28fps 的速度实时运行。

### 3 第四章 通过颜色聚类与时空正则相关学习的互补型跟踪

#### 3.1 动机

最近,一种简单而高效名叫 **Staple** 的跟踪算法在效率和准确度方面,取得了很有潜力的结果。它配备有判别式相关滤波 (DCFs) 跟踪器和颜色直方图的互补学习器,这种互补特性使得它对颜色变化和形变都具有很强的适应性。然而,它也有一些缺点: (i)**Staple** 对于所有序列,只使用标准颜色直方图,量化标准相同,并没有考虑目标在每个序列中的具体结构信息,从而影响其区分目标和背景的能力。(ii)**Staple** 中使用的标准 DCFs 是有效的,但是存在多余的边界效应问题,导致一些具有挑战性的场景中的跟踪失败。

#### 3.2 做法

为了解决这些问题,我们提出了基于颜色聚类与时空正则相关回归的互补跟踪器 (CSCT)。所提出的 CSCT 包含两个互补的部分,可以自适应地处理每个序列的显着颜色变化和形变问题: 首先,我们设计了一个新颖的基于颜色聚类的直方图模型,该模型首先自适应地将第一帧目标的颜色划分为几个颜色聚类中心,构建自适应颜色直方图,这样一来模型就能适应严重的目标形变问题。另外,我们提出学习时空正则化的 CFs,既能缓解边界效应,在目标剧烈变化时,又能提供一个比原 **Staple** 里标准 DCF 更稳健的表观模型。

#### 3.3 数据集上的结果

与 **Staple** 相比,我们的 CSCT 仅仅利用简单的手工特征,在 OTB100、Temple-Color 和 VOT2016 这三个数据集上,分别提升了 5.9%, 3.4% 和 1.5%。此外,我们的 CSCT 甚至超过了几个最先进的复杂深度网络跟踪器。

## 结 论

本文从时空上下文学习的角度，对快速相关滤波目标跟踪算法进行了研究。

- (1) 提出了基于流形正则的上下文感知相关滤波跟踪。
- (2) 提出了通过颜色聚类与时空正则相关学习的互补型跟踪。

这些提出的跟踪算法有效地利用了时空上下文信息，实验证明确实提升了跟踪算法的性能。

## 参考文献

- [1] LI X, HU W, SHEN C, et al. A survey of appearance models in visual object tracking[J]. ACM transactions on Intelligent Systems and Technology (TIST), 2013, 4(4):58.
- [2] WANG H, GE H. Visual tracking using discriminative representation with  $l_2$  regularization[J]. Frontiers of Computer Science, 2018, 1–13.
- [3] ALI A, JALIL A, NIU J, et al. Visual object tracking—classical and contemporary approaches[J]. Frontiers of Computer Science, 2016, 10(1):167–188.
- [4] ZHANG K, LIU Q, YANG J, et al. Visual tracking via boolean map representations[J]. Pattern Recognition, 2018, 81:147–160.
- [5] ZHANG K, LI X, SONG H, et al. Visual tracking using spatio-temporally nonlocally regularized correlation filter[J]. Pattern Recognition, 2018, 83:185–195.
- [6] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]// Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE. 2010: 2544–2550.
- [7] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]// European conference on computer vision. Springer. 2012: 702–715.
- [8] 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.
- [9] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 3074–3082.
- [10] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Learning spatially regularized correlation filters for visual tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 4310–4318.
- [11] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1430–1438.
- [12] 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. 2016: 472–488.
- [13] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. Journal of machine learning research, 2006, 7(Nov):2399–2434.
- [14] CHANG X, YANG Y. Semisupervised feature analysis by mining correlations among multiple tasks[J]. IEEE transactions on neural networks and learning systems, 2017, 28(10):2294–2305.
- [15] YU S I, YANG Y, HAUPTMANN A. Harry potter’s marauder’s map: Localizing and tracking multiple persons-of-interest by nonnegative discretization[C]// Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE. 2013: 3714–3720.
- [16] BAI Y, TANG M. Robust tracking via weakly supervised ranking svm[C]// Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE. 2012: 1854–1861.
- [17] HU H, MA B, SHEN J, et al. Manifold regularized correlation object tracking[J]. IEEE transactions on neural networks and learning systems, 2017.
- [18] ZHUANG B, LU H, XIAO Z, et al. Visual tracking via discriminative sparse similarity map[J]. IEEE

- Transactions on Image Processing, 2014, 23(4):1872–1881.
- [19] KWON J, LEE K M. Visual tracking decomposition[C]// Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE. 2010: 1269–1276.
  - [20] KWON J, LEE K M. Tracking by sampling trackers[C]// Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE. 2011: 1195–1202.
  - [21] WANG N, YEUNG D Y. Ensemble-based tracking: Aggregating crowdsourced structured time series data[C]// International Conference on Machine Learning. 2014: 1107–1115.
  - [22] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary learners for real-time tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1401–1409.
  - [23] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1834–1848.
  - [24] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking vot2017 challenge results[C]// Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on. IEEE. 2017: 1949–1972.
  - [25] LIANG P, BLASCH E, LING H. Encoding color information for visual tracking: Algorithms and benchmark[J]. IEEE Transactions on Image Processing, 2015, 24(12):5630–5644.



## 攻读硕士学位期间发表学术论文情况

- [1] Kaihua Zhang, **Jiaqing Fan**, *et.al.* Parallel Attentive Correlation Tracking. IEEE Transactions on Image Processing, (CCF A 类, SCI 二区).
- [2] Qingshan Liu, **Jiaqing Fan**, *et.al.* Visual Tracking via Nonlocal Similarity Learning. IEEE Transactions on Image Processing, (CCF B 类, SCI 二区).
- [3] **Jiaqing Fan**, Huihui Song, *et.al.* Real-time Manifold Regularized Context-aware Correlation Tracking. Frontiers of Computer Science, (CCF C 类, SCI 四区).
- [4] **Jiaqing Fan**, Huihui Song, *et.al.* Complementary Tracking via Dual Color Clustering and Spatio-Temporal Regularized Correlation Learning. IEEE Access, (SCI 三区).

## 致 谢

感谢宋慧慧老师，张开华老师，刘青山老师及在研究过程中帮助过我的朋友们。感谢南京信息工程大学自动化学院提供的研究平台。

特别致谢：孙师兄。