



西安电子科技大学
XIDIAN UNIVERSITY

基于优化集成学习与空间相关 滤波的视觉目标跟踪

学 生: 赵素杰
指导老师: 田小林 副教授

西安电子科技大学
“智能感知与图像理解”实验室
2017年5月31日

○○
○○○
○○○○
○○○○○

○
○○○
○○
○○○○

○
○○○○
○○○
○○○○

○○

目录

研究背景及意义

基于逻辑回归的集成学习跟踪

自适应目标响应的长时相关滤波跟踪

结构化输出的相关滤波跟踪

评审问题

●○

○

○○

○○

○

○○○○

○

○○○

○○

○

○○○○

○

○○○○

○○

○○○○

○○

视觉跟踪

视觉跟踪的研究内容

视觉跟踪即连续不断地定位运动目标。给定第一帧中的目标位置，跟踪算法能够在视频中预测出目标运动。



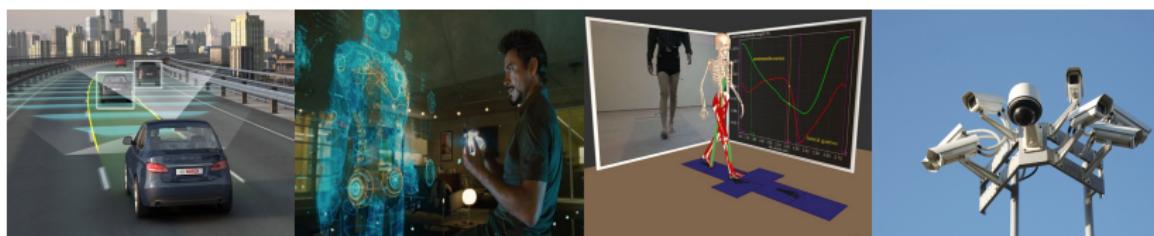
图 1: 单目标视觉跟踪 (无模型)



视觉跟踪

视觉跟踪的应用

跟踪是视觉监控系统的主要组成部分；在自动驾驶系统中，实时跟踪车辆并预测其运动是非常重要的；自主机器人跟踪它们周围的对象，以便识别别人的意图；医学数据分析需要使用可变形模板来跟踪非刚性结构。



(a) 自动驾驶

(b) 人机交互

(c) 运动分析

(d) 智能监控

图 2: 视觉跟踪的实际应用

○○
○○○
○○○○
○○○○○

○
○○○
○○
○○○○

○
○○○○
○○○
○○○○○

○○

创新点 1

基于逻辑回归的集成学习跟踪

算法简介

逻辑回归分类器

弱分类器

逻辑回归集成学习框架

实验结果



算法简介

算法简介

基于逻辑回归的集成学习跟踪算法借助简单快速的弱分类器预测目标。为克服弱分类器的性能缺陷，以逻辑回归对其进行选取和集成。该算法以简单特征（Haar-like 特征）和粗糙分类器为基础，并显著提高了跟踪的准确性。

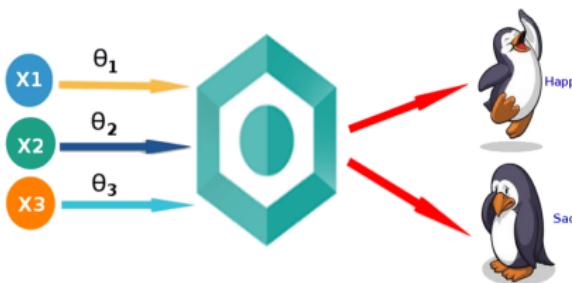
○○

○
●○
○○
○
○○○○○
○○○
○
○
○○○○○
○○○○
○○
○○○○

○○

逻辑回归分类器

逻辑回归



逻辑回归为概率型非线性回归模型，是研究分类观察结果与影响因素之间关系的一种多变量分析方法。

$$P(y|x, w) = \frac{1}{1 + \exp(-yx^T w)} \quad (1)$$

其中 $x \in \mathbb{R}^N$ 为一组解释变量或者特征变量， $y \in \{-1, +1\}$ 为相关的二进制输出。

○○

○

○●

○○

○

○○○○

○

○○○

○

○

○○○○

○

○○○○

○○

○○○○

逻辑回归分类器

模型参数

逻辑回归试图寻找特征空间中的以法向量 $w \in \mathbb{R}^N$ 为参数的分类超平面，将样本分为两类。假设给定一个训练或观测样本集合 $x = \{x_1, x_2, \dots, x_M\}$ 与其对应标签 $y = \{y_1, y_2, \dots, y_M\}$ ，模型参数 w 可通过样本的最大似然估计得到。最大似然估计函数最小化平均损失：

$$l_{avg}(w) = \frac{1}{M} \sum_{i=1}^M \log (1 + \exp (-y_i w^T x_i)) \quad (2)$$

Haar-like 特征弱分类器

本章算法使用 Haar-like 特征。每个弱分类器 h_k 由 1 个 Haar-like 特征 f_k 和 4 个在线估计的参数 $(\mu_+, \sigma_+, \mu_-, \sigma_-)$ 组成。分类器返回对数比值比：

$$\begin{aligned}
 h_k(x) &= \log \left[\frac{P(y = +1 | f_k(x))}{P(y = -1 | f_k(x))} \right] \\
 &= \log \left[\frac{P(f_k(x) | y = +1) P(y = +1)}{P(f_k(x) | y = -1) P(y = -1)} \right]
 \end{aligned} \tag{3}$$

其中 $P(f_k(x) | y = +1)$ 和 $P(f_k(x) | y = -1)$ 服从正态分布 $N(\mu_+, \sigma_+)$ 。这里令 $P(y = +1) = P(y = -1)$ ，根据贝叶斯规则可以求解上述等式。

○○

○

○○

○●

○

○○○○

○

○○○

○

○○○○

○

○○○○

○○

○○○○

弱分类器

弱分类器更新

当弱分类器获得新数据时，使用如下方法进行更新：

$$\mu_+ \leftarrow \gamma\mu_+ + (1 - \gamma) \frac{1}{M} \sum_{i|y_i=+1} f_k(x_i) \quad (4)$$

$$\mu_- \leftarrow \gamma\mu_- + (1 - \gamma) \frac{1}{M} \sum_{i|y_i=-1} f_k(x_i) \quad (5)$$

$$\sigma_+ \leftarrow \gamma \log \sigma_+ + (1 - \gamma) \sqrt{\frac{1}{M} \sum_{i|y_i=+1} (f_k(x_i) - \mu_+)^2} \quad (6)$$

$$\sigma_- \leftarrow \gamma \log \sigma_- + (1 - \gamma) \sqrt{\frac{1}{M} \sum_{i|y_i=-1} (f_k(x_i) - \mu_-)^2} \quad (7)$$



逻辑回归集成学习框架

跟踪流程

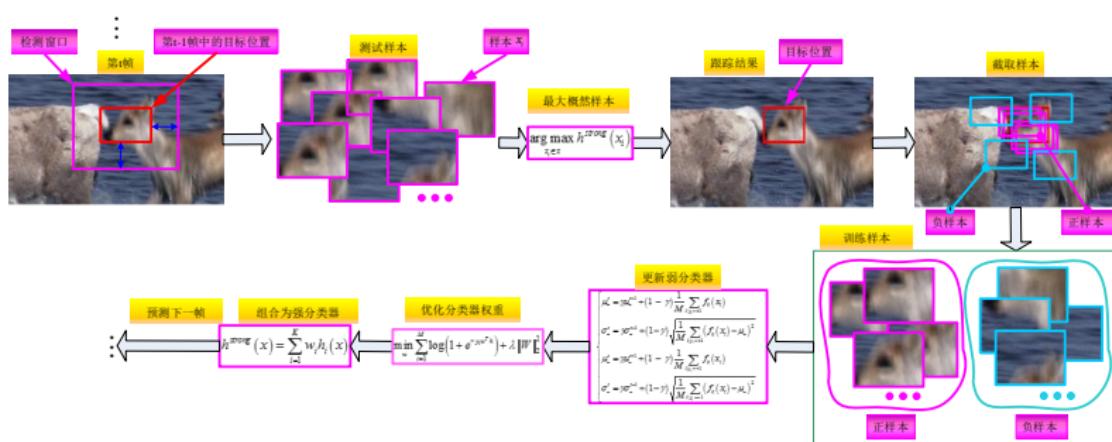


图 3: 逻辑回归集成学习跟踪模型

○○

○

○○

○○

○

●○○○

○

○○○

○

○

○○○○

○

○○○○

○○

○○○○

实验结果

中心位置误差 (CLE)

表 1: 中心位置误差 (像素)

Sequence	CT	CXT	DF	MIL	SCM	Struck	TLD	VTD	Ours
Basketball	89	215	18	92	53	118	269	6	10
David3	89	222	51	30	73	107	281	67	13
Football	12	13	9	12	17	17	14	14	12
Jogging	92	6	31	96	132	62	7	83	5
Liquor	186	132	221	142	99	91	100	60	57

中心位置误差 (Center Location Error, CLE) , 即检测到的目标框中心与真实目标框中心的平均欧式距离 , 其值越小代表跟踪结果准确性越高。

○○

○

○

○

○○

○○

○○○

○○○○

○○○

○○○

○○○

○○○○

○○○○

○○○○

○○○○

○○○○

实验结果

重叠精度 (OP)

表 2: 阈值为 0.5 的重叠精度 (%)

Sequence	CT	CXT	DF	MIL	SCM	Struck	TLD	VTD	Ours
Basketball	25.93	2.48	71.59	27.45	60.28	10.21	2.48	92.41	81.51
David3	34.92	13.89	74.21	68.25	48.02	33.73	10.32	48.41	84.52
Football	78.45	65.19	84.25	73.76	57.18	66.02	41.16	76.80	78.72
Jogging	22.48	95.44	21.50	22.48	21.17	22.48	96.74	21.50	95.11
Liquor	20.85	20.96	22.92	20.10	32.45	40.61	56.17	57.96	69.79

阈值为 0.5 的重叠精度 (Overlap Precision, OP) , 即所得目标边界框与真实边界框重叠率超过给定阈值的帧数占总视频的百分比 (本文实验中阈值为 0.5) , 其值越大代表跟踪结果越好。

○○

○
○○
○○
○
○○●○○
○○○
○
○
○○○○○
○○○○
○○
○○○○

○○

实验结果

代表性序列帧及对比算法跟踪结果

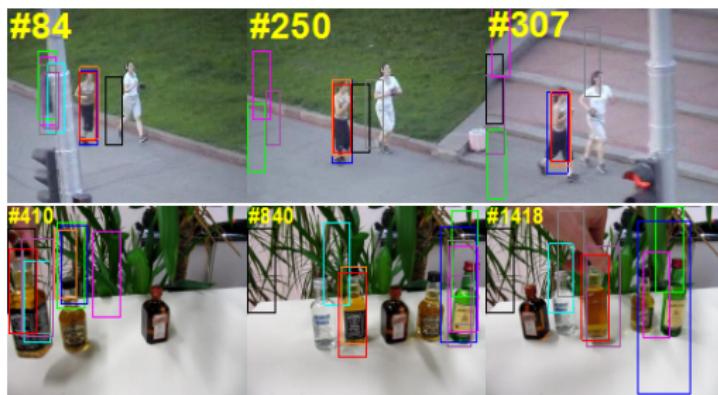


— CT — CXT — DT — MIL — SCM — Struck — TLD — VTD — Ours

○○

○
○○
○○
○
○○○●○
○○○
○
○○○○○
○○○○
○○○
○○○○

代表性序列帧及对比算法跟踪结果



— CT — CXT — DT — MIL — SCM — Struck — TLD — VTD — Ours

○○
○○○
○○○○
○○○○○

○
○○○
○○
○○○○

○
○○○○
○○
○○○○

○○

创新点 2

自适应目标响应的长时相关滤波跟踪

算法简介

位移滤波器

尺度滤波器

SVM 检测器

实验结果



算法简介

着眼问题

已有算法先估计位移再估计尺度，这使得后续步骤对目标定位精确度的要求较高。目标定位时存在的偏差会作用于特征金字塔，导致误差在跟踪过程中不断累积。

解决方案

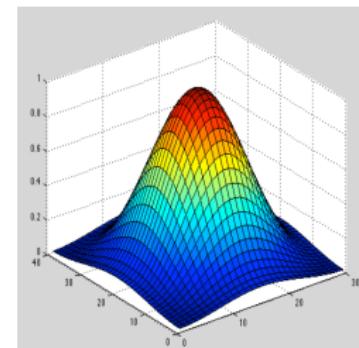
本章所提算法使用一个自适应目标响应的相关滤波器预测目标位置，然后使用一个一维的相关滤波器快速获取目标尺度。SVM 检测器用于处理跟踪中的异常情况。



滤波器的目标响应

基于相关滤波的跟踪器利用两个主要步骤：检测和训练，而且在训练期间使用的目标响应独立于帧，取其为中心在窗口中心的高斯分布。

这种方法实质是假定检测到的窗口的位置是正确的。当检测中出现误差（甚至小至几个像素）时，目标响应 y 不能正确居中，并且这些误差会传播到滤波器估计中。如果后续遇到多个检测错误，则该错误传播通常会导致跟踪器漂移。





自适应目标响应滤波器

自适应目标响应滤波器假设目标响应 y 是未知的并且在每一帧通过使用目标响应先验 y_o 来估计它，即利用在下一帧中实际平移的相关值来帮助过滤器更新回归到更真实的目标值。为实现此目的，进行如下联合优化求解：

$$\min_{w,y} \|\tilde{X}w - y\|_2^2 + \lambda_1 \|w\|_2^2 + \lambda_2 \|y - y_o\|_2^2 \quad (8)$$

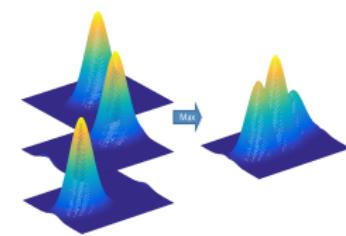
其中， $\tilde{X}^\top = [X_1^\top \ X_2^\top \ \dots \ X_k^\top] \in \mathbb{R}^{kn \times n}$ 由从全部模板生成的循环矩阵连接而成。



构造响应先验

在新的一帧中，以原目标位置为中心截取 p 个样本；计算每个样本 x_i 到目标 c 的欧几里得距离 $f_d(x_i, c)$ ；根据公式 (9) 得到背景与目标的相似度 u_i ；最后，生成以 u_i 为峰值的高斯分布，依据取样位置平移后对 p 个分布取最大值得到 y_o 。

$$u_i = \frac{1}{\sum_{j=1}^p \frac{1}{f_d(x_j, c)+1}} \quad (9)$$

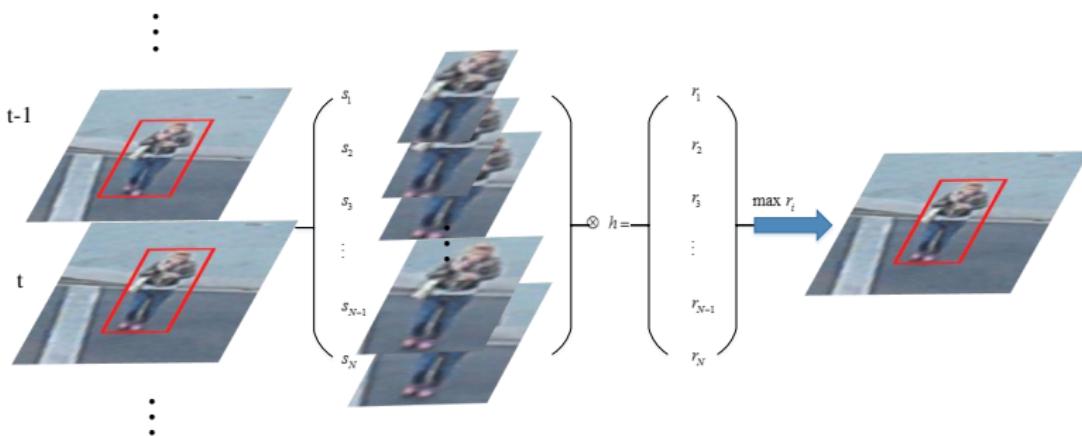


其中， $f_d(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - B_i)^2}$ 。



目标尺度估计

算法使用一个独立的一维相关滤波器预测目标尺寸。可以通过在采样窗口中以最大相关分数搜索窗口来估计当前目标的大小。



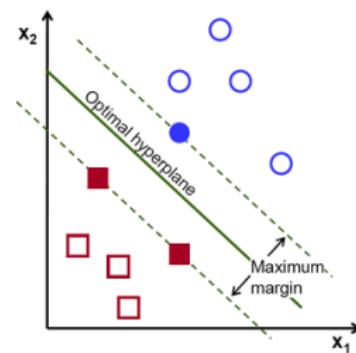


检测模块

鲁棒的长期跟踪算法需要一个重检测模块，以应对例如长期遮挡和重新进入视场等跟踪异常的情况。

无论是 Online Boosting 分类器还是相关滤波器，均采用学习率参数进行更新，这导致先前目标特征的作用会随时间呈指数级降低。这虽然有利于适应目标和背景的变化，但却无法保证跟踪的鲁棒性。

SVM 线性整合支持向量，保留了目标的初始特征。鉴于 SVM 在检测和跟踪中的良好性能，本算法采用 SVM 方法作为检测器。



○○

○

○○

○○

○

○○○○

○

○○○

○○

○

●○○○

○

○○○○

○○

○○○○

○○

实验结果

距离精度 (DP)

表 3: 20 个像素偏差内的距离精度

Sequence	TLD	DSST	SAMF	LCT	MUSTer	Ours
BlurOwl	0.686	0.233	0.222	0.891	0.656	0.926
Coke	0.684	0.918	0.907	0.914	0.831	0.983
DragonBaby	0.248	0.062	0.681	0.549	0.602	0.814
Human2	0.257	0.384	0.527	0.595	0.301	0.801
Lemming	0.858	0.421	0.275	0.846	0.857	0.941
Sylvester	0.949	0.754	0.842	0.975	0.959	0.967
Trellis	0.529	1.00	0.842	1.00	1.00	1.00
Walking2	0.426	1.00	0.570	0.404	1.00	1.00

○○

○

○○

○○

○

○○○○

○

○○○

○

○

○○○○

○

○○○○

○○

○○○○

○○

实验结果

重叠精度 (OP)

表 4: 阈值为 0.5 的重叠精度

Sequence	TLD	DSST	SAMF	LCT	MUSTer	Ours
BlurOwl	0.686	0.235	0.222	0.894	0.661	0.927
Coke	0.289	0.863	0.780	0.914	0.546	0.969
DragonBaby	0.248	0.062	0.628	0.310	0.522	0.779
Human2	0.525	0.423	0.767	0.945	0.337	0.953
Lemming	0.594	0.418	0.266	0.879	0.845	0.871
Sylvester	0.928	0.714	0.819	0.929	0.885	0.932
Trellis	0.473	0.838	1.00	0.923	0.991	1.00
Walking2	0.340	0.414	0.406	0.406	1.00	0.988

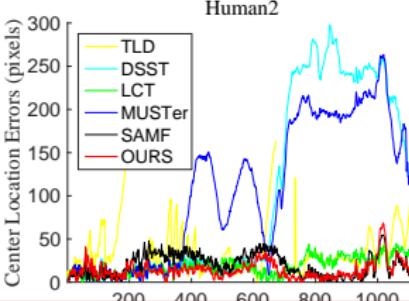
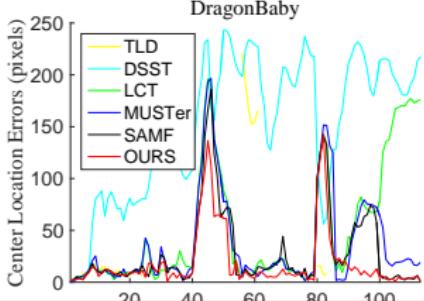
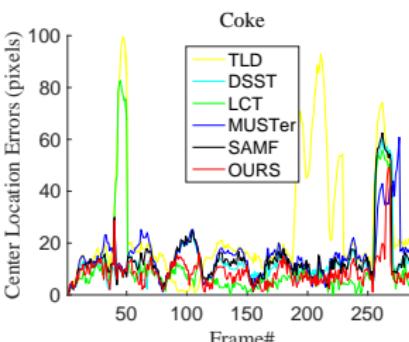
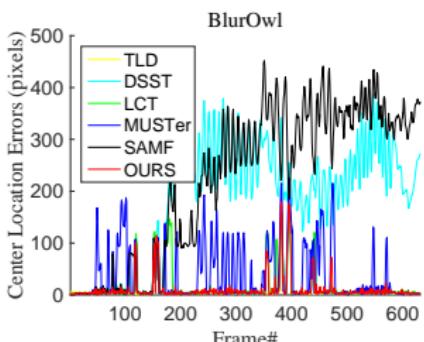
○○

○
○○
○○
○○○
○○○○○
○○○
○○○○
○○●○○
○○○○
○○○○
○○○○

○○

实验结果

中心位置误差对比



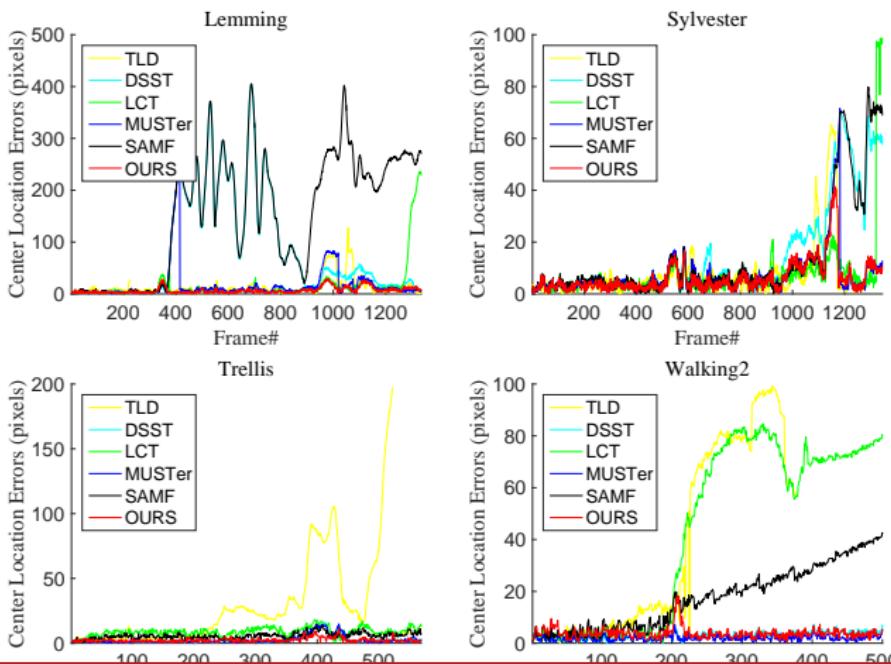
○○

○
○○
○○
○○○
○○○○○
○○○
○○○○
○○○○●○
○○○○
○○○○
○○○○

○○

实验结果

中心位置误差对比



○○
○○○
○○○○
○○○○○

○
○○○
○○
○○○○

○
○○○○
○○
○○○○

○○

创新点 3

结构化输出的相关滤波跟踪

算法简介

有限边界的相关滤波器

对于有限边界滤波器的改进

实验结果



算法简介

算法简介

着眼问题

相关滤波器利用傅里叶域特性能够高效地预测，但边界效应对跟踪性能有显著的影响。

解决方案

本章提出一种结构化输出的相关滤波跟踪方法，在利用相关滤波密集取样优势的同时，显著减少边界效应带来的性能损失。

○○

○

○○

○○

○

○○○○

○

○○○

○

○

○○○○

○

●○○○

○○

○○○○

有限边界的相关滤波器

MOSSE 相关滤波器

MOSSE 相关滤波器可以在空间域中表示为求解岭回归问题：

$$E(h) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^D \|y_i(j) - h^\top x_i[\Delta\tau_j]\|_2^2 + \frac{\lambda}{2} \|h\|_2^2 \quad (10)$$

其中 $y_i \in \mathbb{R}^D$ 是第 i 个观测 $x_i \in \mathbb{R}^D$ 的期望响应， λ 是正则项参数。 $\mathbb{C} = [\Delta\tau_1, \dots, \Delta\tau_D]$ 表示长度为 D 的信号的所有循环移位集合。



有限边界的相关滤波器



边界效应

相关滤波器从由 1 个真实示例与其他合成示例组成的非平衡集合中估计出判别性模板。这些合成的样本是通过对真实样本应用循环移位来创建的。

如右图所示，循环移位的样本都受循环边界效应的影响，并不能代表真实的移动。边界效应可以显著影响所得到的估计模板，使得相关滤波器对平移中的偏差特别敏感。

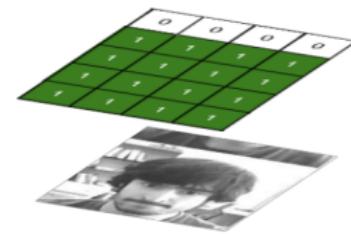




有限边界的相关滤波器

引入掩蔽矩阵

有限边界滤波器能够在空间上避免边界效应。其训练信号 $x \in \mathbb{R}^T$ 的大小比滤波器 $h \in \mathbb{R}^D$ 大。通过使用掩蔽矩阵 $P \in \mathbb{R}^{D \times T}$ ，可以将等式 (10) 表示为：



$$E(h) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^T \|y_i(j) - h^\top P x_i[\Delta\tau_j]\|_2^2 + \frac{\lambda}{2} \|h\|_2^2 \quad (11)$$

掩蔽矩阵 P 封装了信号，其中的 1 和 0 决定哪一部分应该是有效的，哪一部分是无效的。

○○

○

○○

○○

○

○○○○

○

○○○

○○

○

○○○○

○

○○○●

○○○

○○○○

有限边界的相关滤波器

样本对比



(a) 传统滤波器样本



(b) 有限边界滤波器样本

○○
○○○
○○○○
○○○○○

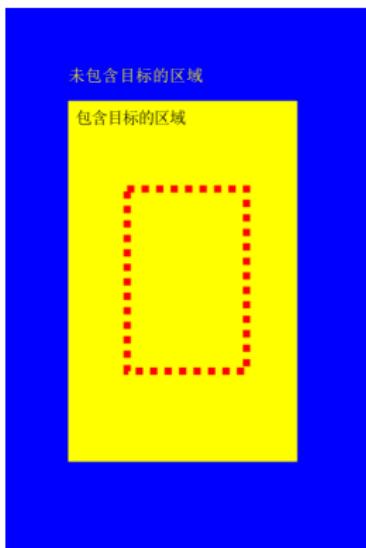
○
○○○
○○○○
○○○○○

○
○○○○
●○
○○○○

○○

对于有限边界滤波器的改进

标签图



有限边界相关滤波器只对样本进行了加工，而未注意到样本与标签的匹配问题。由于扩大滤波器尺寸的同时仍采用了传统相关滤波器的平滑高斯函数生成标签，未包含目标的样本块同样被赋予了正标签。因而，滤波器无法学习到区分度高的特征。

图 5: 有限边界滤波器标签图

○○

○

○○

○○

○

○○○○

○

○○○

○○○

○

○○○○

○

○○○○

○●

○○○○

对于有限边界滤波器的改进

以结构化方式定义样本

借鉴图像检测中的结构化输出方法，以样本框的坐标作为样本标签，这使得样本描述与实际问题相一致。

$$E(h) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^T \|\widehat{l(y_i(j))} - \text{diag}(\hat{x}_i)^\top \sqrt{DFP^\top} h\|_2^2 + \frac{\lambda}{2} \|h\|_2^2 \quad (12)$$

其中，评价函数 $l(y) = \frac{\text{Area}(y \cap y^*)}{\text{Area}(y \cup y^*)}$ 表示样本框与目标框的重叠率， y^* 为目标框的坐标表示，Area 表示求两个矩形框的面积。

○○

○

○○

○○

○

○○○○

○

○○○

○○

○

○○○○

○

○○○○

○○

●○○○

实验结果

距离精度 (DP)

表 5: 20 个像素偏差内的距离精度

Sequance	Frag	OAB	SBT	MIL	Struck	CSK	CFwLB	Ours
Coke	0.034	0.168	0.048	0.117	0.942	0.739	0.918	0.959
David	0.121	0.151	0.204	0.229	0.236	0.236	0.144	0.396
Dog	0.173	0.157	0.079	0.197	0.157	0.144	0.858	0.992
Doll	0.663	0.663	0.149	0.433	0.688	0.218	0.947	0.986
Gym	0.369	0.016	0.046	0.329	0.219	0.091	0.113	0.801
KiteSurf	0.143	0.381	0.369	0.381	0.905	0.321	0.274	0.964
Surfer	0.176	0.045	0.133	0.088	0.157	0.005	0.468	0.997
Sylvester	0.685	0.680	0.430	0.546	0.929	0.717	0.921	0.947
Vase	0.166	0.155	0.129	0.166	0.140	0.166	0.181	0.657
mean	0.281	0.268	0.176	0.276	0.486	0.293	0.536	0.855

○○

○

○○

○○

○

○○○○

○

○○○

○○

○

○○○○

○

○○○○

○○

○●○○

○○

实验结果

中心位置误差 (CLE)

表 6: 平均中心位置误差

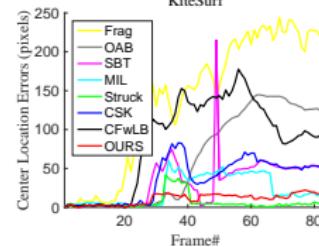
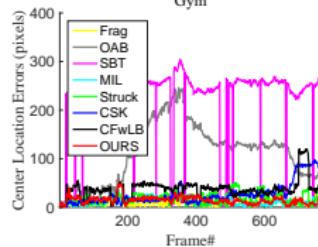
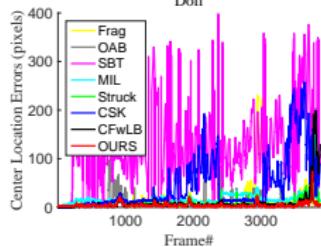
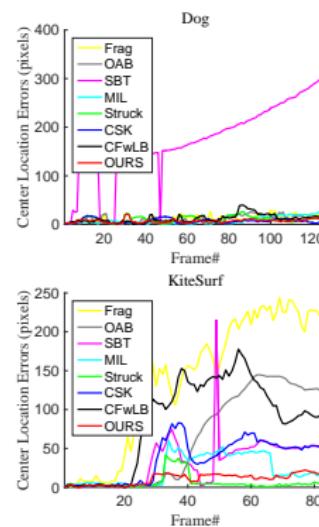
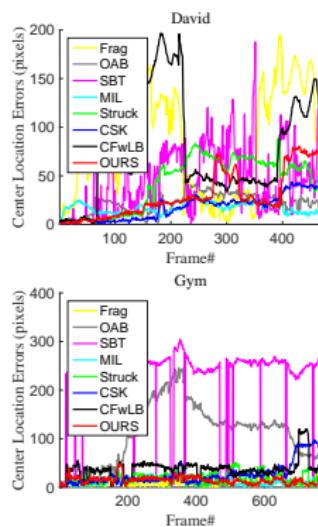
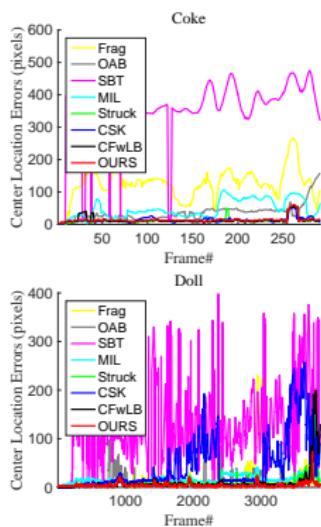
Sequance	Frag	OAB	SBT	MIL	Struck	CSK	CFwLB	Ours
Coke	124.8	35.9	365.3	46.7	12.1	13.6	13.2	12.9
David	82.1	21.7	47.1	16.9	42.8	17.7	73.6	28.9
Dog	12.2	10.7	172.5	8.2	10.4	7.0	11.8	6.8
Doll	13.7	12.4	113.9	16.7	8.9	44.7	9.5	4.6
Gym	10.0	111.0	230.4	11.8	18.5	27.1	9.5	4.6
KiteSurf	141.1	64.6	32.9	22.6	6.1	36.5	85.7	10.6
Surfer	51.6	72.1	218.5	17.0	9.0	161.7	25.3	6.5
Sylvester	15.0	14.8	101.5	15.2	6.3	9.9	8.8	7.0
Vase	18.2	34.7	172.3	19.0	24.3	12.9	59.7	16.7
mean	52.1	52.0	161.6	19.3	15.4	36.8	33.0	11.0

○○

○
○○
○○
○○○○○
○○○
○○○○○
○○○○
○○○○
○○●○

实验结果

中心位置误差对比

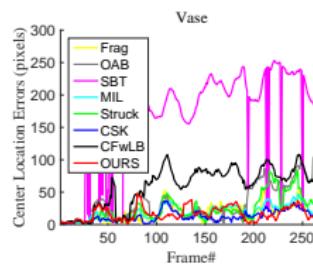
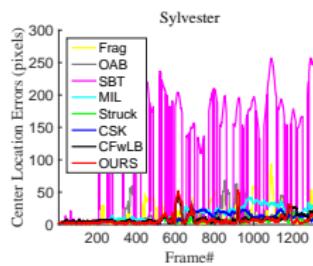
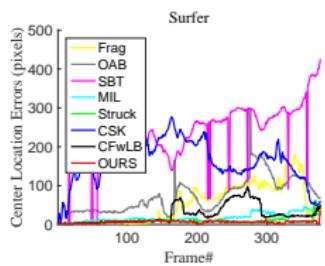


○○

○
○○
○○
○○○○○
○○○
○○
○○○○○
○○○○
○○○
○○○●

实验结果

中心位置误差对比



○○

○

○○

○○

○

○○○○

○

○○○

○

○

○○○○

○

○○○○

○○

○○○○

●○

三种模型的适用范围

评审问题

所提出的三个算法分别是解决跟踪算法中的哪些对应的技术难点？请分析说明三种模型的适用范围

基于逻辑回归模型的集成学习跟踪从传统方法角度着眼，解决跟踪问题；

自适应目标响应的长时相关滤波跟踪多模块协作，适用于长时跟踪，能够应对复杂场景；

结构化输出的相关滤波跟踪以原始像素为特征，速度约 80fps，适用于实时性要求高的环境。



西安电子科技大学
XIDIAN UNIVERSITY

谢谢！