时空上下文融合的无人艇海面目标跟踪

彭艳, 陈加宏, 李小毛, 罗均, 谢少荣, 刘畅, 蒲华燕*

上海大学机电工程与自动化学院, 上海 200444

* E-mail: phygood_2001@shu.edu.cn

收稿日期: 2018-06-29; 接受日期: 2018-09-03; 网络版发表日期: 2018-11-16

国家自然科学基金(批准号: 91648119, 61773254, 61673254)资助项目

摘要 视觉感知作为无人艇环境感知的重要组成部分, 为无人艇智能控制提供十分重要的外界信息. 准确、鲁棒且实时的海面目标跟踪算法可以有效提升无人艇的智能化水平. 海面目标跟踪面临着几个独特的挑战: 目标尺度变化极大、目标视角变化极大、目标抖动剧烈. 为了解决上述海面目标跟踪任务中面临的三个难点, 本文提出了一种时空上下文融合的目标跟踪算法, 该算法结合了深度网络多尺度检测的图像空间上下文信息和相关滤波跟踪的视频时间上下文信息, 实现了实时、精确且鲁棒的目标跟踪效果. 最后, 以无人艇在东海、南海实际作业场景中采集的图像数据作为跟踪算法的测试集, 验证了该跟踪算法的有效性.

关键词 海面目标跟踪, 时空上下文, 相关滤波跟踪, 多尺度检测

1 引言

水面无人艇是在海面上航行的一种无人智能作业平台, 主要用来执行一些危险、枯燥以及不适合人类操作的海上任务^[1], 图1展示的是上海大学无人艇工程研究院自主开发研制的“精海”1号至8号系列水面无人艇的工作照. 水面无人艇需要在各种不同的海况下安全航行, 并完成相关任务, 特别是在超视距的情况下, 远程遥控不能发挥作用, 要求无人艇必须具备自主环境感知、自主避障、实时路径规划以及自主作业的能力. 其中, 自主环境感知功能是自主避障、实时路径规划以及完成具体任务的技术基础, 也是实现无人艇智能化的重要前提之一. 水面无人艇的环境感知技术主要包括: 光视觉感知技术、红外感知技术、雷达感知

技术以及水声感知技术. 与其他感知技术相比, 由于光学图像分辨率更高、更直观、包含更丰富的目标细节信息, 基于视觉的感知可以对水面目标进行更有效的辨别^[2]. 同时, 视觉感知到的信息不仅包含目标物的位置信息还包括目标物的类别信息, 可以与激光、超声等传感器感知到的信息互相补充, 以实现自主航行、安全避障和三维环境感知. 海面目标跟踪任务是视觉环境感知技术中重要的组成部分, 开展基于光视觉的海面目标跟踪技术的研究, 有利于水面无人艇在自主作业过程中对航行的舰艇、水面障碍物等各种类型的目标进行感知, 协助实现自主避障功能, 提高无人艇的生存与作业能力.

在计算机视觉领域, 视觉目标跟踪一直是最活跃的研究范畴之一. 它被广泛应用于人机交互、智能监

引用格式: 彭艳, 陈加宏, 李小毛, 等. 时空上下文融合的无人艇海面目标跟踪. 中国科学: 技术科学, 2018, 48: 1357–1372

Peng Y, Chen J H, Li X M, et al. Sea surface object tracking for USV with spatio-temporal context fusion (in Chinese). Sci Sin Tech, 2018, 48: 1357–1372, doi: 10.1360/N092018-00218

图1 (网络版彩图)上海大学“精海”系列水面无人艇

控、运动分析等方面^[3]。视觉目标跟踪任务往往指的是单目标跟踪,定义如下^[4]:只给定视频图像序列首帧中目标的位置信息,即目标在图像坐标系中的位置信息。该任务的研究目的是在接下来的所有图像序列中得到该目标的具体位置信息,即根据首帧目标的信息得到该目标在该时段完整的运动轨迹。具体的目标跟踪示意过程如图2所示。

最近几年,单目标视觉跟踪技术取得了显著进步,但是目前仍没有一种跟踪算法能很好地适用于所有复杂的应用场景。其原因包括三个方面:(1)图像是通过摄像机采集的,本质上是对现实三维场景的一种二维映射,这一映射过程丢失了大量有用的视觉信息,给大多数计算机视觉任务的实施带来了巨大的影响;(2)现实场景中的目标会受到环境的各种干扰,比如光照变化、雨天雾天的视线干扰、目标相似物干扰等。同时,目标本身也会有尺度大小的变化、刚性或非刚性变形、快速运动,目标间会互相遮挡;(3)成像设备的自身原因,会存在图像分辨率低、目标运行模糊等降低图像质量的客观原因,进一步增加了目标跟踪任务的难度。因此,视觉目标跟踪研究会遇到如下挑战^[4]:目标形变、运动模糊、快速运动、光照变化、背景杂波、尺度变化、低分辨率、目标部分遮挡。

经过数十年的发展,目标跟踪在跟踪精度和跟踪速度方面已经取得了巨大的进步^[5]。与通用的目标跟踪算法的快速发展相比,无人艇海面目标跟踪的研究却是十分匮乏的。无人艇视野下的海上目标具有如下特点:(1)目标物在视野中抖动明显,在图像中表现为急剧、快速且无规律的运动;(2)海面目标尺度变化极大;(3)海面目标视角变化极大。造成海面目标剧烈抖动的主要原因是无人艇在运动过程中受到了海浪、风浪等作用力的影响,使其俯仰、横滚、偏航运动。为了形象的表示出海面目标在视频图像序列中的抖动程度,我们选择了无人艇目标跟踪数据集中的一个图像序列和基准跟踪数据集中的一个图像序列进行对比,对比结果如图3所示。图中的横坐标表示帧序号,纵坐标表示目标在图像坐标系中 x 和 y 轴的坐标值的相对变化率即 $\Delta x/w$ 和 $\Delta y/h$, w 表示目标的宽度, h 表示目标的高度。从图中可以明显的看出,海面目标的位移变化比一般场景下的目标要剧烈。造成海面目标尺度变化更大的原因是:一般场景(普通视频监控或无人车中摄像机采集的道路上的场景)感知距离在500 m以内,但是海面环境更加开阔,无人艇搭载的摄像机的感知距离更远,一般能够达到1 km以上。如图4所示,在相同视角的前提下,海面目标的尺度变化是11倍,普通目标的

图2 (网络版彩图)视觉目标跟踪过程示意图

图 3 (网络版彩图)视觉海面目标抖动明显. (a) 海面目标; (b) 普通目标

图 4 (网络版彩图)海面目标尺度变化极大

尺度变化是6倍。海面目标的运动路径相对更加自由,使得目标在图像中的视角变化更大,海面目标视角变化示意过程如图5所示。

为了解决上述海面目标跟踪任务中面临的三个难点,本文提出了一种时空上下文融合的目标跟踪算法。在信号处理领域,用相关性概念来描述两个信号之间的联系。相关是衡量两个信号相似值的度量,如果两个信号越相似,那么其相关值就越高。相关滤波技术源于信号处理领域,后被广泛应用于图像分类、目标检测与识别等领域^[6]。相关滤波技术在目标跟踪任务中应用时最基本的思路是:在目标跟踪过程中,已知第一帧的目标,在接下来的每一帧图像的搜索区域中寻找与前面各帧中目标相关性最高的图像块作为当前帧的跟踪目标。以上跟踪思路充分利用了视频图像序列中目标的时间上下文信息。但是这种跟踪策略应用

图 5 (网络版彩图)海面目标视角变化极大

在海面目标跟踪任务中存在如下缺陷: (1) 当目标在图像序列中发生剧烈运动时,目标容易跳出跟踪算法的搜索区域,造成跟踪失败; (2) 该方法无法应对目标尺度和视角剧烈变化的情景,容易造成跟踪漂移乃至失败。所以,相关滤波跟踪框架对海面目标的跟踪精度较低。而近些年来基于深度学习的目标跟踪方法发展迅速,虽然其跟踪精度较高,但是跟踪速度较慢,不能满足无人艇实时环境感知的需求;其次,深度特征的提取依赖于高功耗的计算平台,这也不符合无人艇等嵌入式控制平台的应用需求。为了解决上述问题,我们在上述相关滤波跟踪框架中融合了图像空间上下文的信息,即充分利用深度神经网络的多尺度学习能力对海面目标进行自主检测以实现目标尺度和视角变化的自适应调整,通过扩大搜索区域以解决海面目标剧烈运动的难题。为了验证该跟踪算法的有效性,我们采集了无人艇在中国东海、南海实际作业场景中的大量视频图像数据作为跟踪算法验证的测试集。实验结果表明,该算法结合了深度网络多尺度检测的图像

空间上下文信息和相关滤波跟踪的视频时间上下文信息后, 实现了实时、精确且鲁棒的目标跟踪效果。

2 相关工作

早在20世纪60年代, 国外就已经开展了对运动目标的检测与跟踪技术的研究^[7]。20世纪90年代开始, 一些基于可见光成像的图像处理算法就已经用于海上小目标的检测^[8]。现在基于光视觉的目标检测与跟踪算法已越来越多的搭载到无人操作水面平台上, 进行海面目标的检测、识别与跟踪, 从而协助水面无人平台更好的感知和监测周围环境。本节根据对视频图像数据时空上下文信息利用层面的不同, 分别对与本文研究相关的目标检测算法和目标跟踪算法的研究历史做简单的回顾。

2.1 基于空间上下文的目标检测算法

目标检测一直是计算机视觉的研究热点。特殊类别的目标检测, 例如人脸和行人, 其检测技术已经较为成熟。Viola等人于2001年提出基于AdaBoost算法的检测框架^[9], 使用Haar-like小波特征进行分类, 采用滑动窗口搜索策略对人脸目标实现准确有效的定位。它是第一种能实时处理并给出很好检测率的物体类别检测算法, 但是主要适用于人脸检测领域。随后, Dalal等人^[10]使用图像的局部梯度方向直方图HOG作为特征表示方法, 利用支持向量机SVM作为分类器进行行人检测, 其大幅度的提升了算法对行人的检测能力。更为普遍的目标检测工作关注的是自然图像中一般类别的检测。自然界的大部分物体具有运动能力, 会发生非刚体形变, 为此, Felzenszwalb等人^[11]提出了多尺度形变部件模型DPM, 它是目标类别检测最具影响力的方法之一, 继承使用HOG特征和SVM分类器的优点。DPM目标检测器由一个根滤波器和一些部件滤波器组成, 组件间的形变通过隐变量进行推理。由于目标模板分辨率固定, 算法采用滑动窗口策略在不同尺度和纵横比的图像上搜索目标。后续工作采用不同策略加速了DPM的穷尽搜索策略。Malisiewicz等人^[12]提出一种简单高效的集成学习算法用于提升目标检测效果, 该方法分别为每个正样本训练一个使用HOG特征的线性SVM, 通过集成每个样本的线性SVM结果达到优良的泛化性能。传统的目标检测算法往往由于目标特征的

表征能力不足, 检测的精度不能满足实际的工程需求。

在2012年之前, 目标检测中分类任务的框架都是采用人为设计的特征训练浅层分类器完成分类任务, 但是人为设计特征的泛化能力和浅层分类器的分类能力有限, 会造成目标检测的精度不高。所以, 最佳的算法是基于DPM框架的各种改进算法。2012年, Krizhevsky等人^[13]提出了基于深度学习理论的深度卷积神经网络的图像分类算法, 图像分类的准确率大幅提升, 同时也带动了目标检测准确率的提升。随后, Sermanet等人^[14]提出了一种基于深度卷积神经网络的目标识别框架OverFeat, 其集成了目标识别、定位和检测任务。OverFeat对输入图像采用滑动窗口策略进行目标搜索, 并用分类模型确定每个窗口中目标的类别, 然后使用对应类别的定位模型预测目标的候选框, 根据分类得分为每个类选出候选目标框并进行合并, 得到最终的检测结果。与OverFeat不同, 基于区域搜索的卷积神经网络算法R-CNN^[15]采用选择性搜索策略而不是滑动窗口来提高检测效率。R-CNN算法利用选择性搜索方法在输入图像上选择若干候选目标框, 对每个目标候选框利用卷积网络进行特征提取, 以提取到的目标特征作为输入为每个类训练SVM分类器, 得到候选目标框属于每个类的得分。最后, R-CNN使用非极大值抑制方法舍弃部分目标候选框, 得到检测结果。上述方法使用的深度卷积神经网络结构基本来自于Krizhevsky的7层网络结构设计。为了提高深度卷积神经网络的分类性能以及检测的准确率, Simonyan等人^[16]和Szegedy等人^[17]设计了层数为22层的深度卷积神经网络, 采用的检测框架与R-CNN类似。为了提升深度学习算法的检测效率, Liu等人^[18]于2016年提出了基于位置回归的检测框架, 但是由于公开数据集中训练数据与海面目标的特点相差较大对海面目标的检测效果不佳。本文在其思想的基础上增加了海面目标训练环节, 对模型参数进行微调以达到对海面目标更好的检测效果。以上所有的目标检测算法都是针对静态图片进行的, 仅仅利用的是图像序列的空间上下文信息, 缺乏对视频图像时间上下文信息的利用, 这容易产生目标漏检现象。而本文利用目标跟踪算法对视频图像序列在时间维度的关联性来弥补目标检测算法不能对特定目标进行持续关注的难点, 在保持原有跟踪和检测框架不变的前提下, 融合两者的结果来完成目标跟踪的任务。

2.2 基于时间上下文的目标跟踪算法

视觉目标跟踪算法与目标检测算法相比, 对算法运行的实时性要求更高. 相关滤波方法由于其计算的高效性, 以及能满足跟踪的实时性要求, 自2010年以来, 得到迅速发展. Bolme等人在2010年率先提出基于误差平方和最小原则训练滤波器的相关滤波跟踪算法MOSSE^[19], 充分发挥了相关滤波技术的计算高效性, 能够实现600帧每秒以上的跟踪速度. Heriques等人提出了一种基于核空间的相关滤波算法^[20], 实现了岭回归从线性到非线性的扩展. 与此同时, 他们还使用HOG特征代替原来的灰度特征应用到跟踪器上^[21], 使跟踪过程对目标受光照变化更加鲁棒. 文献^[22]的研究中将颜色特征应用到相关滤波跟踪框架, 有效地增强了跟踪算法对目标形变的鲁棒性. Danelljan等人提出将多通道的特征加入到高斯核空间中, 同时利用PCA主成分分析降维来实现高效、鲁棒且实时性好的目标跟踪效果. 同年, Zhang等人提出的时空上下文跟踪STC算法以概率分布的角度解释了相关滤波的跟踪原理, 由于该跟踪过程中分类器的判别能力有限, 所以跟踪精度不高^[23]. 为了增强算法对目标受遮挡的鲁棒性, 基于分块策略和随机采样策略的RPT (Reliable Patch Trackers)算法^[24]被提出. 在尺度自适应方面, 具有杰出贡献的是DSST (Discriminative Scale Space Tracking)算法^[25]和SAMF(Scale Adaptive with Multiple Features tracker)算法^[26], 这两个算法分别使用一维滤波器和尺度搜索池对最佳的尺度进行搜索. KCFDP (Kernelized Correlation Filter with Detection Proposal)算法^[27]则是利用目标检测中生成目标候选框的方法来完成对跟踪目标的尺度和纵横比的自适应. 在跟踪的融合策略方面, STAPLE (Sum of Template And Pixel-wise LEarners)^[28]融合了HOG特征和颜色直方图特征来实现高效的跟踪; MUSTer (Multi-Store Tracker)算法^[29]则结合了相关滤波和关键点跟踪的优势实现了长时间的目标跟踪. 与上述基于不同特征融合的算法相比, 本文所提出的时空上下文融合策略将原有的基于空间上下文的检测算法的检测结果和基于时间上下文的跟踪算法的跟踪结果进行融合, 以修正跟踪结果提升跟踪算法的鲁棒性, 通过提升对目标的搜索维度得到更稳定的跟踪结果. 相关滤波跟踪算法的高效性是因为利用了样本的周期性循环假设, 但同时也

会出现边界效应, 影响跟踪算法的鲁棒性. 为了增强相关滤波跟踪算法的稳定性, 尽量避免漂移现象的发生, SRDCF (Spatially Regularized Discriminative Correlation Filters)^[30]采用了一种空间正则化策略, 降低样本框边界对跟踪效果的影响; CFLB (Correlation Filters with Limited Boundaries)算法^[31]则是通过减少受边界影响样本的个数来克服边界效应的影响. 以此实现跟踪鲁棒. 以上所有的目标跟踪算法都充分利用了视频序列中的时间上下文信息, 但是对于图像数据中空间上下文信息的利用较差, 很难精准地确定目标位置和尺度. 同时, 基于时间上下文信息设计的算法在跟踪过程中容易发生误差的积累, 造成跟踪漂移现象, 最终导致跟踪失败.

以上基于空间上下文的目标检测算法和基于时间上下文的目标跟踪算法虽然在实际应用过程中各有利弊, 但同时它们也能够互相弥补弱点. 本文提出的时空上下文跟踪算法能够充分利用目标检测算法对目标尺度和视角的自适应能力以及目标跟踪算法对目标时间维度上的搜索能力, 有效提升跟踪精度和鲁棒性. 与此同时, 高效的多尺度目标位置回归检测方法和相关滤波跟踪框架的使用保证了算法执行的实时性.

3 时空上下文融合的海面目标跟踪

本文提出的时空上下文融合的海面目标跟踪算法是对深度网络多尺度检测和相关滤波跟踪技术的融合. 本节首先简要的描述相关滤波跟踪框架以利用图像序列的时间上下文信息; 接着提出深度网络多尺度检测框架对海面目标进行检测以利用视频图像的空间上下文信息; 最后提出时空上下文信息融合的跟踪策略以构建出完整的海面目标跟踪框架.

3.1 相关滤波跟踪框架

从宏观角度来看, 目标跟踪问题可以看作是一个关于目标与背景的二分类问题, 但是在特定的图像帧中真正的正样本只有一个, 而负样本的个数则远多于正样本, 很明显存在样本数量不均衡的学习问题. 这个问题的存在使得训练得到的分类器的分类性能不高, 容易出现模型的过拟合现象. 为了解决目标跟踪问题中训练样本极度缺乏的问题, 相关滤波跟踪框架将目标跟踪任务看作是回归问题, 岭回归模型作为机器

学习领域中经典的回归模型, 因为其求解便利, 被广泛应用。

具体地, 给定一个二维高斯分布用于表示各个训练样本属于目标正样本的置信度, 高斯分布的中心就是初始目标框的中心。这样的处理不仅更加符合模型训练的要求, 同时也更加符合现实场景中关于目标分布的先验知识, 即越靠近目标中心的图像块, 属于目标的可能就越大。这是相关滤波技术应用到目标跟踪任务中的一个基本假设。

岭回归的目的是通过训练数据集来建立样本变量 x 和响应输出 y 之间线性映射关系。其中训练集的数据格式表示如下^[19]:

$$\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}, x^{(i)} \in R^n, y^{(i)} \in R. \quad (1)$$

建立的变量 x 和 y 之间线性映射关系为

$$y = f_w(x) = w^T x, w \in R^n. \quad (2)$$

线性岭回归建立如下目标函数, 并关于参数 w 进行极小化求解:

$$\begin{aligned} \min_w J(w) &= \min \sum_i (f_w(x^{(i)}) - y^{(i)})^2 + \lambda w^2 \\ &= \min (w^T X^{(i)} - y^{(i)})^2 + \lambda w^2, \end{aligned} \quad (3)$$

其中, 公式第一项是一个均方差, 第二项是一个规则化项, 也叫权重衰减, 其目的是减小权重的幅度, 防止训练得到的模型出现过度拟合。 λ 是控制规则化比例的权重。回归参数 w 有闭合求解公式:

$$w = (X^T X + \lambda I)^{-1} X^T y, \quad (4)$$

其中,

$$X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{bmatrix}, y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}. \quad (5)$$

对于上述问题的求解存在如下问题: 当样本个数和特征维数很大, 需要对一个大维数矩阵求逆, 才能获取参数 w 。如果在密集样本条件下, 训练样本的数量极大, 那么线性岭回归过程就不能满足实时性要求。利用傅里叶变换可以将循环矩阵对角化, 对原始图像样本进行循环位移操作。 x 是基准样本, X 是对基准样本进行循环位移得到的循环矩阵, 则可以根据下面的公

式对其进行对角化:

$$X = C(x) = F \text{diag}(\hat{x}) F^H. \quad (6)$$

那么可以对岭回归的解进行进一步的推导, 得到

$$\hat{w} = \frac{\hat{x} \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda}. \quad (7)$$

对上述问题进行多通道扩展, 可得: 假设有 C 个通道的特征表示, 每个基本样本 x_i 可以构成一个循环矩阵 $X_i = C(x_i)$ 。则多通道参数 w 学习公式为

$$\hat{w} = \frac{\sum_i \hat{x}_i^* \odot \hat{y}_i}{\sum_i \hat{x}_i^* \odot \hat{x}_i + \lambda}. \quad (8)$$

根据上述的数学模型和优化求解过程, 构建出基于相关滤波技术的目标跟踪框架, 如图6所示。在图像序列的首帧处进行相关滤波模板的初始化, 具体过程如下: 对初始搜索区域的图像块进行特征提取得到目标搜索区域的特征表示。为了降低循环位移假设带来的边界效应的影响, 对图像特征矩阵加入汉宁窗的约束, 从而达到突出目标中心, 忽略目标边缘的目的。接下来根据加入汉宁窗的特征表示和期望的高斯输出就可以对初始化的相关滤波模板进行训练, 以上的训练过程都是在傅里叶频域中进行的, 在该训练过程中计算效率更高, 训练消耗时间更短。图6上半部分显示了上述跟踪初始化过程, 其中展示的特征表示和汉宁窗都是二维平面的形式。图6下半部分显示了跟踪算法在跟踪过程中获取跟踪结果以及更新相关滤波器的过程, 都是以三维输出图谱的形式展现。在目标跟踪过程中, 对当前图像帧中搜索区域进行特征提取, 紧接着加入汉宁窗, 然后根据前一帧训练得到的滤波器模板找到图像搜索区域中与上一帧目标最为相关的也就是最相似的目标的位置, 即为我们需要找到的该帧目标的跟踪位置。最后不断的执行上述的滤波器训练、跟踪、滤波器更新的循环过程, 直至结束跟踪过程。

3.2 深度网络多尺度检测

无人艇在执行任务时, 其视觉环境中包括的目标类别主要有船只、岛礁以及浮标。海面目标种类相对地面目标少了很多, 但是除了每种类别的外观变化、光照变化、尺度变化、视角变化等共有的干扰因素之外, 还面临着诸如图像清晰度不高、背景杂波多、目标尺度小等特殊挑战, 这些干扰因素对海面目标检测

图 6 (网络版彩图)相关滤波跟踪框架

算法提出了更高的要求。

目前流行的、优秀的目标检测算法步骤大致如下: 先生成一些假设的目标候选框, 然后在这些目标框中提取特征, 再经过一个分类器, 来判断里面是不是有物体, 是什么物体。这类流程的代表算法为R-CNN系列^[15,32,33]。但这类方法对于嵌入式系统(如无人艇等离线无人控制系统)而言, 计算时间太长, 不能做到实时检测, 也不能满足对实时环境感知的需求。所以, 针对R-CNN系列的改进工作有许多是以实时目标检测为方向, 但到目前为止, 大多数都是以牺牲检测精度来换取检测时间。

本文采用的检测算法借鉴了SSD算法^[18], 是基于目标位置回归的思路实现的, 消除了先前检测算法框架中的候选目标框提取、特征重采样等过程, 在保证检测实时性的同时, 也可以获得很高的检测精度。据相关研究表明^[34]: 越低层的特征图保留的图像细节越多, 越高层的特征图保留的图像语义信息越多, 所以, 融合多层特征图信息的识别结果更加可靠。一个卷积神经网络中不同的层有着不同尺寸的感受野。低层的特征图, 感受野比较小可以检测小尺度目标, 高层的感受野比较大可以检测大尺度目标, 对不同的特征图进行卷积检测, 可以达到对目标多尺度检测的目的。

本文检测算法采用的基础网络是VGG^[16], 使用其前面的五段卷积层和最后一个池化层, 然后利用atrous^[34]算法将原始VGG网络的第六层和第七层的全连接层替换成两个卷积层, 最后额外增加三个卷积层和一个平均池化层, 改进后的详细结构框架如图7所示。图中蓝色框表示卷积层输出的特征图, 在这些特征图上会选择很多默认框, 用于寻找目标真实的位置和尺度, 总共选择了六个特征图上的判断结果参与到最后的非极大值抑制过程, 最后得到目标框的位置、尺度和类别信息反馈。

该检测算法对每一个特征图中默认框的尺寸大小选择的计算如下:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), k \in [1, m], \quad (9)$$

其中, k 表示默认框的个数, $s_{\min}$ 取值0.2, $s_{\max}$ 取值0.95, 这意味着最低层的尺度是0.2, 最高层的尺度是0.95, 不同纵横比的默认框用 α_r 来表示, 分别可以取值为1, 2, 3, 1/2, 1/3, 那么每一个默认框的宽度和高度就可以计算出来。

本文算法训练阶段的目标损失函数有两部分组成, 分别用于对目标框位置的回归和对目标类别的分类。 x_{ij}^p 表示第 i 个默认框与类别为 p 的第 j 个标准框是否

图7 目标检测算法网络结构图

匹配,以重叠率阈值为匹配标准.本文取重叠率阈值为0.6,重叠率超出该阈值则认为该默认框与对应类别的标准框匹配,则 $x_{ij}^p=1$;否则为不匹配,即 $x_{ij}^p=0$.根据上面的匹配策略,一定有 $\sum_j x_{ij}^p \geq 1$,意味着对于第 j 个标准框有可能有多个默认框与其相匹配.那么总的目标损失函数为

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)), \quad (10)$$

其中, N 是与标准框相匹配的默认框个数.位置损失 L_{loc} 采用类似于Fast R-CNN中的Smooth L1损失,用于回归预测框 l 与标准框 g 的四个参数包括中心坐标位置、宽和高.分类置信度损失 $L_{\text{conf}}(x, c)$ 是Softmax损失,输入为每一类的置信度 c .两种损失的权重项 α 设置为1.

基于上述网络结构以及目标损失函数的设计,我们要训练出满足无人艇海上目标检测任务的网络模型.基于特定训练数据集的模型训练过程分为三个阶段,如图8所示.第一步,对于VGG基础网络,使用公开的ImageNet图像分类数据集进行网络参数的预训练过程,这一步骤主要是为了训练出网络对于图像分类的特征提取能力;第二步,基于前一阶段训练后网络的参数,构建出本文改造后的模型架构.如图8所示,使用公开的目标检测数据集VOC对改造后的网络参数进行预训练,这一步骤主要是为了训练出网络的目标检测能力,包括目标位置的回归和目标类别的判断;第三步,基于无人艇实际作业场景中目标的类别,构

图8 标检测算法的网络参数训练过程

建实用的数据集对前两步得到的网络参数进行微调,以满足海面目标检测的需求.

3.3 时空上下文融合策略

基于深度网络多尺度的目标检测算法对海面目标进行检测后,可以得到图像中目标的位置、尺度和类别,但是这种基于静态图像的目标检测方法无法利用图像序列中的时间上下文信息,会造成两帧图像之间出现目标漏检的情况.同时,目标检测算法主要用于区分目标的类别,缺乏对特定目标的分类能力,不能直接应用于目标跟踪任务中.而基于相关滤波技术的目标跟踪框架则善于对特定目标进行持续的跟踪,能够充分利用视频图像序列的时间上下文信息,对目标跟踪模型进行实时调整以达到自适应跟踪的目的.然而,该框架无法做到对目标尺度变化的自适应跟踪,同时对于目标视角变化的鲁棒性极差.因此,对上述两种技术的目标跟踪算法进行结合,可实现鲁棒性和精度都极高的跟踪效果.

本文提出的时空上下文算法包括目标检测模块、目标跟踪模块以及融合策略模块.目标检测模块包含了深度网络多尺度检测算法,负责检测特定类别的目标,在无人艇作业领域包括的类别有船舶、岛礁和浮标,主要是通过视频图像的空间上下文信息实现目标类间分类;目标跟踪模块包含了相关滤波跟踪算法,负责对目标进行持续的跟踪,主要是利用视频图像序列的上下文信息实现区别类内目标的目的.基于上述两个模块的实现及融合,即对图像时间和空间上下文信息的融合,该无人艇海面目标跟踪算法可以对海面目标实现实时、鲁棒且精度高的跟踪效果.

目标检测模块使用深度网络多尺度检测算法,利用3.2节训练得到的检测模型对序列图像进行目标检测过程,该过程在图9中用蓝色虚线所示.算法的执行结果包括图像画面中目标的种类、置信度以及目标物

图9 时空上下文融合流程图

的位置. 以图10所示的真实场景数据为例可知, 基于图像空间上下文信息进行的检测过程会存在目标漏检的现象, 但是对目标尺度的自适应更好, 对目标视角变化也更加鲁棒. 目标跟踪模块使用相关滤波跟踪算法, 根据首帧目标的位置信息, 利用公式(8)计算滤波器模板, 用于在后续图像帧中搜寻目标的位置, 在跟踪过程中通过二次插值法不断更新滤波器模板参数, 以上目标跟踪过程如图10紫色虚线所示. 在视频序列首帧处, 已知目标的位置信息, 能够初始化相关滤波模板, 同时将该目标物大小的2.5倍设置为目标检测模块的搜索区域. 如果在下一帧图像的搜索区域内检测到同类目标物, 则以该目标检测模块检测到的目标作为最终的目标跟踪结果. 如果在下一帧图像中检测到的同类目标在搜索区域之外或者没有检测到同类目标, 则以目标跟踪模块跟踪到的目标为最终的目标跟踪结果. 在确定了目标跟踪位置之后, 更新滤波器模板以及检测搜索区域, 然后循环执行上述过程, 详细的时空上下文融合过程如图9所示.

4 实验分析与评价

为了验证本文构建的海面目标跟踪算法的有效性, 我们在中国东海和南海收集了具有船只目标的大

量的图像序列. 同时, 为了可以定量的评估该跟踪框架对于船只目标的跟踪成功率, 我们对所有收集到的图像序列中的船只目标进行人工标注, 详细信息如表1所示. 本文采用Matlab和C语言混合编写, 实验均在AMD A8处理器、主频为2.4 GHz、内存为4 G配置的主机上通过Matlab2014a软件进行. 实验中对于所有测试图像序列, 算法中的参数保持不变.

4.1 海面目标数据集

为了提升本文检测算法在无人艇海上目标检测任务中的检测精度和成功率, 我们制作了船舶数据集, 以下称为无人艇船舶检测训练数据集. 该训练数据集有两个部分组成: 网络爬虫数据和实际应用场景数据. 网络爬虫数据是根据网络爬虫技术, 对互联网上各种类型的船只图像进行抓取; 实际应用场景数据是通过摄像机对海面船只进行拍摄, 然后截取其中包含船只部分的图像, 该部分的数据与检测算法性能的测试集数据完全不重复. 通过上述两种途径获取的数据集需要通过人工标注目标位置和种类才能用于网络模型参数的训练. 最终, 我们制作的目标检测训练数据集包含758张网络爬虫图片数据和587张实际工作场景图片数据, 图11显示了该数据集中部分图像数据以及人工标注结果.

图 10 时空上下文融合效果图

表 1 目标跟踪数据集的参数说明

序号	跟踪序列名称	总帧数	主要挑战
1	OcclusionShip2	318	剧烈抖动、尺度变化
2	VID001	1868	尺度变化、视角变化
3	VID008	2296	尺度变化
4	BigshipBB	947	尺度变化
5	Nearly2far3	1598	尺度变化
6	Nearly2far4	1460	尺度变化、视角变化
7	Nearly2far6	941	尺度变化、视角变化
8	Nearly2far7	1750	尺度变化、视角变化、剧烈抖动
9	Nearly2far8	985	尺度变化、视角变化
10	Othership1	728	尺度变化、视角变化
11	Rotate	2053	尺度变化、视角变化
12	Towardbigship	1295	尺度变化、视角变化

为了验证各种跟踪算法对无人艇海面目标跟踪的效果，我们采集了12个具有代表性的图像序列组成了无人艇海面目标跟踪数据集。表1展示了在无人艇实际作业场景下采集到的这12个图像跟踪序列的具体参数，包括该跟踪序列的名称、总帧数以及跟踪序列中

目标存在的主要跟踪挑战。

4.2 评价方式

评估方法包括对目标跟踪的精度、跟踪鲁棒性以及跟踪实时性的评估^[4]。目标跟踪精度的评估往往有

图 11 深度多尺度网络训练数据集

两个标准, 分别为位置误差和尺度重合率. 位置误差更确切的表述是中心位置误差 (Center Location Error, CLE), 其被定义为跟踪目标的中心位置和标定的准确位置之间的平均欧氏距离, 具体如图12中的左图所示. 绿色框表示目标的标注位置, 红色框表示跟踪算法的跟踪位置, a 点表示跟踪框的中心位置, b 点表示标注框的中心位置, L_{ab} 表示 a 与 b 在图像坐标系中的欧氏距离, 计算公式如下:

$$CLE = L_{ab} = \sqrt{(a.x - b.x)^2 + (a.y - b.y)^2}. \quad (11)$$

为了直观地展示位置误差, 可以采用精确度图来形象地表示, 精确度图能够显示出评估位置在给定的准确值的阈值距离之内的帧数占总帧数的百分比. 为了定量地表示各跟踪算法的跟踪精确度, 通常认为 CLE 小于等于 20 个像素时目标被正确跟踪. 一个序列中所有帧的平均中心位置误差被用于概括跟踪算法对该序列的总体性能. 当跟踪器丢失目标时, 输出的跟踪位置是随机的, 此时的平均误差值无法正确估量跟踪的性能.

另一个跟踪精度评估标准是目标框的重叠率. 假

设跟踪的目标框区域为 A , 准确的边界框是 B , 重叠率被定义为

$$S = \frac{|A \cap B|}{|A \cup B|}, \quad (12)$$

其中, $\cap$ 和 $\cup$ 分别表示两个区域的交集和并集, $|\dots|$ 表示该区域内的所有像素点的个数. 为了估量算法在一系列帧中的性能, 设定重叠率 S 大于给定的阈值 T 即表示跟踪成功的图像帧的数量. 成功率图给出了此阈值从 0 到 1 变化时成功帧所占的比例 (Success Rate, SR), 与精确度评估类似的是使用某一特定阈值 (如 0.5). 但是仅仅根据一个成功率数值来评估跟踪器的性能并不公平也并不具有足够的代表性, 所以通常使用成功率图曲线下的面积作为定量评估标准对各个跟踪算法进行跟踪性能排序.

根据第一帧中的标注位置进行初始化, 然后在整个测试序列中运行算法, 最后得出平均精确度或成功率的结果. 这种方法称为一次通过的评估 (One-Pass Evaluation, OPE).

跟踪实时性的评估是最重要的评估方式之一. 如果一个跟踪算法做不到跟踪实时性, 意味着该算法只

图 12 跟踪精度性能评估方式

存在理论研究价值, 不能满足实际的工程需求. 在计算机视觉领域对于算法处理速率的评估是以指定算法对于单帧图像处理的平均速率即每秒处理的帧数(Frame Per Second, FPS)为依据的.

为了能够满足无人艇实际的工程需求, 图像中算法跟踪的目标框位置与标准目标框的重合率高于90%才认为是跟踪成功的, 否则判定跟踪失败. 那么, 检测成功率的计算公式可表示如下:

$$T = \frac{N_{\text{success}}}{\text{Frames}}. \quad (13)$$

跟踪成功率 T 等于跟踪成功的帧数 N_{success} 与图像序列的总帧数Frames的比值, 跟踪成功的帧数的计算要满足目标框与标准框的重合度达到90%以上.

4.3 对比方法

本文选取了目前为止世界上最出色的10个相关滤波跟踪算法与本文提出的时空上下文跟踪算法进行对比. 这些跟踪方法包括背景感知相关滤波器(BACF)、自适应颜色特征跟踪器(CN)、尺度估计跟踪器(DSST)、快速尺度估计跟踪器(fDSST)、核相关滤波跟踪器(KCF)、候选框跟踪器(KCFDP)、尺度自适应多特征跟踪器(SAMF)、空间正则化相关滤波跟踪器(SRDCF)、随机块跟踪器(RPT)、颜色纹理特征融合跟踪器(Staple), 它们都可以在CPU平台上实时运行, 且这些算法的参数以原论文的设置为标准.

4.4 实验结果及分析

(1) 跟踪性能方面

图13是无人艇海面目标跟踪数据集中精确度和成功率的对比方法曲线. 图中的实验数据结果表明: 海面目标跟踪存在着引言中提及的目标尺度变化极大、目标视角变化极大、目标抖动剧烈的难点, 这三个难点会对跟踪精度和成功率产生显著的影响. 传统的单一相关滤波跟踪算法的跟踪精度最高为0.824, 最高的跟踪成功率为0.662, 都远小于实际工程上需要的0.9的阈值. 上述难点也会导致传统跟踪方法跟踪性能上的不足, 使其很难满足无人艇实际作业的工程需求. 根据4.2节的跟踪性能评估方法, 对传统相关滤波跟踪算法与本文算法的跟踪性能进行比较. 精确度上, 以20个像素阈值得到的实验结果如图14(a)所示; 成功率图上, 以重叠率为0.5的阈值得到的实验结果如图14(b)所示. 从上述定量结果来看, 本文跟踪算法的精确度达到0.976, 相较于Staple的0.824提升了18.4%; 成功率为0.921, 与SRDCF的0.662相比, 提升幅度达38.1%. 以上实验结果表明: 在相关滤波跟踪的基础上融合目标检测结果可以有效地提升对海面目标的跟踪性能.

(2) 跟踪速度方面

跟踪速度如表2所示, 本文的跟踪系统可以实现实时的目标跟踪效果, 平均速度达到了17 FPS. 以上跟踪速度是在一个普通主机上获得的, 其配置为2.4 GHz主频的A8处理器, 内存4 GB. 在实际的无人艇作业场景中可以配置性能较高的处理主机, 实现30 FPS以上的目标实时跟踪效果.

表2中关于跟踪成功率的实验结果表明: 该目标跟踪系统对于船只目标的跟踪是准确且鲁棒的, 平均跟踪成功率达到了97.8%, 这是在要求目标跟踪重叠率

图 13 对比方法的整体精确度和成功率曲线

图 14 (网络版彩图) (a) 跟踪精确度对比结果; (b) 跟踪成功率对比结果

表 2 用于验证本文跟踪系统有效性的测试数据集及跟踪成功率

数据集名称	总帧数	检测帧数	跟踪帧数	成功率	处理速度/FPS
OcclusionShip2	318	172	146	93%	17.049
VID001	1868	1748	120	100%	18.512
VID008	2296	132	2164	100%	18.061
BigshipBB	947	423	524	98%	14.789
Nearly2far3	1598	499	1099	100%	19.363
Nearly2far4	1460	1329	130	100%	17.121
Nearly2far6	941	332	609	100%	14.792
Nearly2far7	1750	1206	544	100%	15.423
Nearly2far8	985	727	258	97%	18.213
Othership1	728	173	555	91%	13.284
Rotate	2053	544	1509	96%	15.468
Towardbigship	1295	764	531	99%	17.571

是90%以上获得的, 可以认为该跟踪算法对于海面目标的跟踪是完全有效的, 其为无人艇控制系统提供的目标位置信息是完全可信的. 与单独使用传统跟踪算法对于船只目标的跟踪成功率对比可以发现, 将深度网络多尺度检测结果与相关滤波跟踪结果进行融合可以明显地提高目标跟踪的成功率. 表2中检测帧数和跟踪帧数的详细统计信息可以更加充分地验证本文融合

策略的有效性, 当检测失败时跟踪算法可以跟踪住目标, 当检测成功后采用检测的结果来避免跟踪过程中出现漂移现象.

从定性结果来看, 本文提出的海面目标跟踪算法在无人艇实际工作场景中对于目标跟踪结果如图15、16、17所示, 分别对应目标剧烈抖动场景即Nearly2-far7图像序列、目标尺度剧烈变化场景即Towardbig-

图 15 目标抖动跟踪效果

图 17 目标视角变化跟踪效果

图 16 目标尺度变化跟踪效果

ship图像序列、目标视角剧烈变化场景即Rotate图像序列。实验结果截取了相同时刻下各个跟踪算法的跟踪结果, 图片中的白色框表示本文算法的跟踪结果, 其他颜色表示传统相关滤波跟踪算法的跟踪结果, 详细的颜色表示参见各图。图中实验结果表明: 本文提出的海面目标跟踪算法能够稳定地跟踪住感兴趣目标, 跟踪过程中合理地融合了传统检测和跟踪算法的时空上下文信息, 在目标发生剧烈抖动、尺度变化以及视角变化等关键节点时通过更强的目标搜索能力以达到稳定跟踪的效果。

5 结论

海面目标跟踪任务面临着几个独特的挑战: 目标尺度变化极大、目标视角变化极大、目标抖动剧烈。本文从解决上述难题的角度出发, 提出了一种融合视频图像序列时空上下文的海面目标跟踪算法。不同于只利用视频序列中的部分时间或者空间信息的传统的目标检测和跟踪算法, 本文通过合理的融合策略将深度网络多尺度检测的图像空间上下文信息和相关滤波的视频时间上下文信息进行融合。融合后的算法既能利用相关滤波跟踪框架对于单一目标的持续跟踪能力, 又能利用深度检测模块提升目标跟踪对于目标尺度和视角变化的自适应能力以及对于图像的全局搜索能力, 从而较好地解决了上述三个难题。为了验证该跟踪算法的有效性, 我们以无人艇在东海、南海实际作业场景中的图像数据作为跟踪算法的测试集, 通过对比融合前后的跟踪效果实验可以看出本文提出的上下文融合的跟踪算法对于无人艇场景下的海面目标跟踪是有效的。虽然本文的融合策略已经能够满足无人艇视野下的大部分应用场景, 而一些特殊情况(如目标检测搜索区域内存在多个同类目标时如何鉴定哪一个待跟踪目标)仍是未来亟待解决的工作之一。

参考文献

- 1 宋磊. 全球自主: 国外无人水面艇未来发展及关键技术. 军事文摘, 2015, 13: 26-28
- 2 Kristan M, Sulic V, et al. Fast image-based obstacle detection from unmanned surface vehicles. IEEE T Cybernetics, 2016, 46: 641-654
- 3 Yilmaz, Javed O, Shah M. Object tracking: A Survey. ACM Comput Surv, 2006, 38: 1-45
- 4 Wu Y, Lim J, Yang M H. Online object tracking: A benchmark. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, 2013. 2411-2418

- 5 Smeulders A W M, Chu D M, Cucchiara R, et al. Visual tracking: An experimental survey. *IEEE Trans Pattern Anal Mach Intell*, 2014, 36: 1442–1468
- 6 Bolme D S, Draper B A, Beveridge J R. Average of synthetic exact filters. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Kyoto, 2009. 2–3
- 7 Worst M T. Aircraft warning system: US. US 3053932 A. 1962
- 8 Wang G, Sang N. Efficient small-target detection algorithm. *International Society for Optics and Photonics*, 1995, 321–329
- 9 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Kauai, 2001. I-511–I-518
- 10 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. San Diego, 2005. 886–893
- 11 Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models. *IEEE Trans Pattern Anal Mach Intell*, 2010, 32: 1627–1645
- 12 Malisiewicz T, Gupta A, Efros A A. Ensemble of exemplar-svms for object detection and beyond. In: *Proceedings of IEEE International Conference on Computer Vision*. Barcelona, 2011. 89–96
- 13 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: *Proceedings of Advances in Neural Information Processing Systems*. 2012. 1097–1105
- 14 Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014
- 15 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014. 580–587
- 16 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Boston, 2015
- 17 Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Boston, 2015. 1–9
- 18 Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. In: *Proceedings of the European Conference on Computer Vision*. Amsterdam, 2016. 21–37
- 19 Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, 2010. 117–127
- 20 Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels. In: *Proceedings of the 12th European Conference on Computer Vision*. Florence, 2012. 702–715
- 21 Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. *IEEE Trans Pattern Anal Mach Intell*, 2015, 37: 583–596
- 22 Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014. 156–166
- 23 Zhang K, Zhang L, Liu Q, et al. Fast visual tracking via dense spatio-temporal context learning. In: *Proceedings of the European Conference on Computer Vision*. Zurich, 2014. 127–141
- 24 Li Y, Zhu J K. Reliable patch trackers: Robust visual tracking by exploiting reliable patches. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Boston, 2015. 353–361
- 25 Danelljan M, Hager G, Khan F S, et al. Accurate scale estimation for robust visual tracking. In: *Proceedings of British Machine Vision Conference*. Nottingham, 2014
- 26 Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration. In: *Proceedings of the European Conference on Computer Vision*. Zurich, 2014. 254–265
- 27 Huang D, Luo L, Wen M, et al. Enable scale and aspect ratio adaptability in visual tracking with detection proposals. In: *Proceedings of British Machine Vision Conference*. Swansea, 2015. 185.1–185.12
- 28 Bertinetto L, Valmadre J, et al. Staple: Complementary learners for real-time tracking. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 2016. 1401–1409

- 29 Hong Z B, Chen Z, Wang C H, et al. MUlti-Store Tracker (MUSTer): A cognitive psychology inspired approach to object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015. 749–758
- 30 Danelljan M, Hager G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of IEEE International Conference on Computer Vision. Washington, 2015. 4310–4318
- 31 Galoogahi H K, Sim T, Lucey S. Correlation filters with limited boundaries. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015. 4630–4638
- 32 Girshick R. Fast R-CNN. In: Proceedings of IEEE International Conference on Computer Vision. Washington, 2015. 1440–1448
- 33 Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proceeding of the 28th International Conference on Neural Information Processing System. Montreal, 2015. 91–99
- 34 He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Proceedings of the European Conference on Computer Vision. Zurich, 2014. 346–361
- 35 Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell*, 2018, 40: 834–848

Sea surface object tracking for USV with spatio-temporal context fusion

PENG Yan, CHEN JiaHong, LI XiaoMao, LUO Jun, XIE ShaoRong, LIU Chang & PU HuaYan

School of Mechanical Engineering and Automation, Shanghai University, Shanghai 200444, China

As an important component of unmanned surface vessel environment awareness, visual perception provides very important external information for intelligent control of USV. Accurate, robust and real-time sea surface object tracking algorithms can effectively improve the intelligence level of USV. Sea surface object tracking faces several unique challenges: the object scale changes greatly, the object perspective changes greatly, and the object jitters drastically. In order to solve the three difficulties encountered in the above sea object tracking task, this paper proposed an object tracking algorithm with spatio-temporal context fusion. This algorithm combines the image space context information of the multi-scale detection of deep network and the correlated filter tracking video time context information. It is a real-time, accurate and robust object tracking. Finally, using the image data of the USV in the actual operation scene in the East China Sea and South China Sea as the test set of the tracking algorithm, the effectiveness of the tracking algorithm is verified.

sea surface object tracking, spatio-temporal context, correlation filter tracking, multi-scale detection

doi: [10.1360/N092018-00218](https://doi.org/10.1360/N092018-00218)