

基于图像处理的海上船只姿态测量方法

闫威, 郑智辉, 郭宸瑞, 王硕, 董昊天
(北京航天自动控制研究所 北京 100854)

摘要: 船只在海面上航行时, 由于船只自身的运动、风力和海浪对船只运动的影响, 船只的姿态是实时变化的, 要在飞机上实时、准确测量出船只的姿态是很困难的。针对上述问题, 设计了一种融合传统的模板匹配方法和深度学习方法的姿态估计方法, 深度学习提高了姿态估计的准确性、鲁棒性和环境适应性, 结合基于轮廓特征的模板匹配方法提高了姿态估计的实时性。首先采用目标船只的三维模型, 渲染出船只的多姿态图像, 通过实例分割算法建立船只姿态模板库。然后采集目标船只的可见光图像, 通过目标检测、实例分割算法获得船只匹配图像, 将船只匹配图像与船只姿态模板库中的图像进行匹配, 匹配成功的船只姿态模板图像对应的姿态即为船只的姿态。通过仿真验证, 三维姿态估计精度可达到 1° 。

关键词: 三维建模; 实例分割; 模板匹配; 三维姿态

中图分类号: TP751 **文献标志码:** A **文章编号:** 2095-1000(2025)02-0116-08

DOI: 10.12347/j.ycyk.20240902001

引用格式: 闫威, 郑智辉, 郭宸瑞, 等. 基于图像处理的海上船只姿态测量方法[J]. 遥测遥控, 2025, 46(2): 116-123.

Method for Measuring the Attitude of Ships at Sea Based on Image Processing

YAN Wei, ZHENG Zhihu, GUO Chenrui, WANG Shuo, DONG Haotian
(Beijing Automatic and Control Institute, Beijing 100854, China)

Abstract: When a ship is sailing on the sea, the attitude of the ship changes in real-time due to the influence of the ship's own movement, wind, and waves. And it is difficult to accurately measure the ship's attitude in real-time from an aircraft. In order to solve the above problems, a pose-estimation method integrating the traditional template-matching method and the deep-learning method was designed. The deep-learning method improves the accuracy, robustness, and environmental adaptability of pose estimation, while the real-time performance of pose estimation is enhanced by combining it with the template-matching method based on contour features. Firstly, the three-dimensional model of the target ship was used to render the multi-pose images of the ship, and the ship attitude template library was established through the instance segmentation algorithm. Then, the visible-light images of the target ship were collected. The ship-matching images were obtained through the target-detection and instance-segmentation algorithms. These ship-matching images were matched with the images in the ship-attitude template database, and the attitude corresponding to the successfully-matched ship-attitude template image was the attitude of the ship. Through simulation verification, the accuracy of 3D attitude estimation could reach 1° .

Keywords: 3D modeling; Instance segmentation; Template matching; 3D attitude

Citation: YAN Wei, ZHENG Zhihu, GUO Chenrui, et al. Method for Measuring the Attitude of Ships at Sea Based on Image Processing[J]. Journal of Telemetry, Tracking and Command, 2025, 46(2): 116-123.

0 引言

船只在海面上航行时, 由于船只自身的运动、风力和海浪对船只的影响, 船只的姿态是实时变化的。对任意的、姿态快速变化的船只进行实时、准确的姿态测量需要一种实时的、非直接的姿态测量方法。姿态测量方法可分为直接测量法和间

接测量法。直接测量法是在测量目标上安装传感器, 如陀螺仪、加速度计、H轴(Heading axis, 航向轴)磁阻传感器等惯性器件来测量三维姿态^[1]或安装多个GPS(Global Positioning System, 全球定位系统)设备, 根据相位干涉原理及天线布阵技术, 利用GPS进行船只的方向定位和姿态测量^[2]或采用GPS与INS(Inertial Navigation System, 惯

性导航系统)组合进行船只的位置和姿态测量^[5]。间接测量方法采用单/双/多目相机或激光雷达测量出目标的2D图像或3D点云,应用姿态估计方法得出目标姿态。直接测量方法需要在目标船只上安装设备,不能对任意船只进行姿态测量。间接测量法的双/多目相机重建方法需要的相机数量多,使得整体设备复杂度增加,成本高。激光雷达可精确测量深度信息,但其成本高、测量距离有限。单目相机可高分辨率、实时地捕获目标图像,获取目标准确详细信息,且重量较轻、占用空间小、测量距离远,成为紧凑型航空器的理想成像工具^[4]。

传统的姿态估计方法主要是依据图像特征构建二维图像与三维目标位置点的映射关系来计算出目标姿态,通常分为基于特征点匹配的方法和基于模板匹配的方法。基于特征点匹配的方法通过提取图像的像素局部特征与目标三维模型的特征进行匹配。常用SIFT (Scale-Invariant Feature Transform, 尺度不变特征变换)^[5]、SURF (Speeded Up Robust Features, 加速稳健特征)^[6]、ORB (Oriented FAST and Rotated BRIEF, 快速特征点提取和描述)^[7]等特征匹配算法建立图像的2D-3D坐标对应关系,然后利用Perspective-n-Point (PnP, 多点透视成像)^[8]算法得出当前视角下的目标姿态。此方法需要目标有丰富的纹理来计算局部特征,因此对弱纹理目标的计算效果不佳。基于模板匹配的方法通过模板匹配得到目标的姿态信息,利用渲染工具获取目标三维模型各视角下的轮廓和边缘等信息,形成模板库。在计算姿态时,逐一将目标与模板库中模板匹配。此类算法有LineMod (Line - based 3D Object Detection and Pose Estimation, 基于直线的3D目标检测与姿态估计)^[9]、Super4PCS (Super 4 - Point Congruent Sets, 超级4点一致集)^[10]、Go-ICP (Generalized Over complete - Iterative Closest Point, 广义过完备迭代最近点)^[11]等。此方法适用于弱纹理目标,但需要建立模板库,模板匹配的搜索空间大,对光照和有遮挡目标很敏感。

传统的姿态估计方法是利用了目标的几何或纹理特征,在实际使用中由于光照变化、复杂的背景、物体遮挡会导致目标特征提取困难,导致环境适应性差,优点是计算量小,实时性好。深度学习方法因学习能力强、环境自适应好的优点

逐渐成为姿态估计的新方向。基于深度学习的姿态估计方法可分为基于关键特征点、基于像素投票和基于回归的方法。基于关键特征点的方法通过深度神经网络进行图像特征提取,得到关键特征信息,建立2D与3D对应关系,再利用PnP等算法进行姿态估计。YOLO-6D (You Only Look Once-6D, 一次看6D)^[12]是应用YOLOV2 (You Only Look Once Version 2, YOLO二代)网络将目标顶点投影到二维图像上,预测了一个中心点和8个角点,得到3D与2D的对应关系。BB8^[13]使用两阶段深度学习网络,第一阶段预测出目标的中心点,第二阶段预测目标顶点对应的2D投影坐标,最后进行姿态解算。基于投票的方法中每个像素或三维点都对结果产生贡献,关键点及其概率由一致性投票计算得出,概率分布有助于提高PnP姿态估计准确度。典型的投票方法有PVNet (Pixel-wise Voting Network, 逐像素投票网络)^[14]、DenseFusion (密集融合)^[15]、DPVL (Dense Perspective Voting for 6D Object Pose Estimation, 用于6D目标姿态估计的密集透视投票)^[16]、Pix2Pose (Pixel-Wise Coordinate Regression of Objects for 6D Pose Estimation, 基于像素级坐标回归的物体6D姿态估计)^[17]等。此方法能提高被遮挡目标的姿态估计精度。基于回归的方法通过深度神经网络直接回归目标姿态信息。PoseNet (基于CNN的实时6自由度机位定位算法)^[18]采用GoogLeNet作为骨干网络,用回归器替代其分类器,用全连接层替代Softmax (归一化指数函数)层,得到具有一定鲁棒性的姿态估计结果。PoseNet2^[19]通过对PoseNet网络进行改进,提高了场景泛化能力。PoseCNN (A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes, 用于杂乱场景中6D物体姿态估计的卷积神经网络)^[20]、Deep6D (Deep Learning based 6D Pose Estimation, 基于深度学习的6D姿态估计)^[21]网络从单个彩色图像中联合深度学习目标检测和分割网络直接回归出目标的姿态。基于深度学习的姿态估计方法具有泛化能力强、受环境影响小、姿态精度高的优点,但也存在需要大量数据进行模型训练、模型训练时间长、模型计算量大的问题。

针对海面任意动态船只的实时、准确姿态测量需求,本文提出的姿态测量方法应用单目相机采集图像数据,利用融合传统的模板匹配方法和

深度学习方法的姿态估计方法估计船只三维姿态; 通过采用深度学习的目标检测和实例分割网络, 提高了姿态估计的准确性、鲁棒性和环境适应性; 提出基于轮廓特征的掩码图像匹配方法, 避免了光照和目标纹理的影响, 提高了匹配算法的实时性和环境适应性。其工作流程是首先采用目标船只的三维模型, 渲染出船只的多姿态图像, 建立船只姿态模板库; 然后, 采集目标船只的可见光图像, 应用深度学习目标检测算法准确估计出目标位置坐标, 基于检测框长边方向得出粗偏航角, 用于降低模板匹配时的模板范围; 最后, 应用深度分割网络从环境中精确分割出目标, 获得目标和模板的掩码图像, 选择 ReLU 函数作为掩码图像匹配函数, 实现快速模板匹配, 匹配成功的船只姿态模板图像对应的姿态即为船只的姿态。

1 船只姿态渲染

基于先验知识, 设计目标船只的三维模型, 建立船只姿态模板库。可采用 3ds Max 等建模软件设计船只的三维模型, 构建虚拟环境, 控制船只的姿态 (偏航角、俯仰角和横滚角)。以 1° 为间隔, 在偏航角 $0^\circ\sim 360^\circ$ 、俯仰角 $-30^\circ\sim 30^\circ$ 、横滚角 $-20^\circ\sim 20^\circ$ 范围内, 遍历生成船只在不同姿态时的渲染图像。船只三维模型渲染图像示例如图 1 所示。

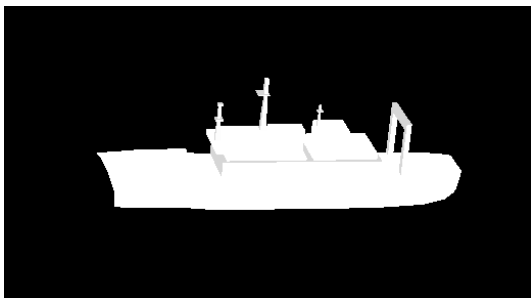


图 1 船只三维模型渲染图像

Fig. 1 Render image of 3D model of ship

2 目标检测

目标检测方法通常分为传统目标检测方法和深度学习目标检测方法。由于传统目标检测方法普遍存在适应性不足、泛化性差等问题, 其逐渐被以卷积神经网络为核心的深度学习目标检测方法所替代。深度学习目标检测方法通常可分为单阶段目标检测方法和双阶段目标检测方法。双阶

段目标检测方法主要代表有 SPPNet (Spatial Pyramid Pooling Network, 空间金字塔池化网络)^[22]、R-CNN (Region-based Convolutional Neural Networks, 区域卷积神经网络)^[23]等, 基于 R-CNN 算法又衍生出多种改进的算法, 主要有 Fast R-CNN (快速区域卷积神经网络)^[24]、Faster R-CNN (更快的区域卷积神经网络)^[25]、Mask R-CNN (掩码区域卷积神经网络)^[26]等。双阶段算法尽管在精度上表现出色, 其基于锚框的设计和需要生成候选区的弊端导致处理时间较长, 不适合实时性要求高的应用场景。单阶段目标检测方法省略了候选区域生成阶段, 极大地提高了算法的检测算法速度, 只是检测精度有一定的降低, 主要代表算法是 SSD (单阶段多框检测器)^[27]、RetinaNet (视网膜网络)^[28]、YOLO^[29]、CenterNet (中心网络)^[30]、ExtremeNet (极值点网络)^[31]、CornerNet (角点网络)^[32]等。2023 年 Ultralytics 团队提出了 YOLOv8 目标检测算法, 其作为一种单阶段无锚框目标检测算法, 其精度和检测速度均为最优。通过不断优化和改进网络结构, YOLO 系列目标检测算法在速度和精度之间取得了平衡, 并被持续优化, 广泛应用于各种目标检测场景^[33]。

对采集的船只图像, 需要利用目标检测算法检测出目标的位置和粗偏航角。船只目标轮廓通常是长方形的, 船头在长方形的长边方向, 如果能用任意角度的旋转框来标注船只的边界, 就可以通过长方形的长边轴方向来初步确定船只的粗偏航角。由于有实时性的要求, 本文采用 YOLOv8 目标检测算法的旋转目标检测功能 (YOLOv8-ORB, You Only Look Once version 8 with Oriented Bounding Boxes) 对图像中的船只进行目标检测, 得出目标切片。依据检测结果旋转框的角度, 得出粗偏航角。

YOLOv8-ORB 是一种先进的对象检测算法, 它在传统的 YOLOv3 和 YOLOv4 基础上进行了优化, 加入了 ORB (Oriented Bounding Box, 有向包围盒) 旋转框检测, 能够更精确地检测并定位出目标物体的位置。YOLOv8-ORB 通过引入 ORB 旋转框检测, 它允许边界框以任意角度存在, 更能适应不同方向的目标物体。图 2 所示为应用 YOLOv8-ORB 检测算法对船只图像进行目标检测的结果, 此算法输出的旋转检测框可准确地框在目标四周, 进而得出船只的位置和粗偏航角。

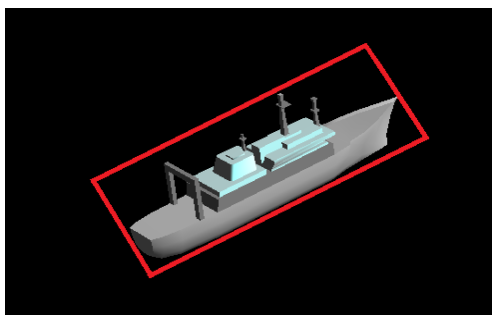


图2 YOLOv8-OBb的检测结果

Fig. 2 YOLOv8-OBb detection results

3 图像匹配

图像匹配算法可分为基于特征的图像匹配算法、基于灰度信息的图像匹配算法和基于变换域的图像匹配算法。基于灰度的图像匹配算法直接利用图像中的灰度变化特征进行相似度匹配，这类方法具有匹配速度快、匹配精度高的优点，但存在对光照及噪声敏感度高的缺点。基于变换域的图像匹配算法将图像灰度信息从空间域变换到频率域，提取频率域的幅度和相位特征进行相似度匹配，这类方法具备较强的抗噪能力和较快的计算速度。基于特征的图像匹配算法对图像进行数学分析，然后提取图像中具有代表性的特征信息，如几何、纹理、梯度特征等，然后对特征信息进行描述和相似性度量，从而完成匹配。这类方法计算量小，具备一定的抗噪、抗光照变化和抗形变能力，可以对复杂的图像进行匹配，但对弱纹理图像匹配效果不好。

根据特征的种类，基于特征的图像匹配算法可以分为点特征匹配、边缘特征匹配和面特征匹配。点特征匹配方法通过特定函数获取灰度极值点，具有良好的尺度、旋转不变性，同时具备一定的抗形变能力。边缘特征匹配算法，利用物体基础的轮廓形状特征进行相似度度量，此方法对于刚性物体有效，不能适用于轮廓变形的物体。面特征匹配算法适用于图像中的物体具有大面积颜色、纹理特征的情况，匹配精度高。颜色和纹理是常见的特征，具有一定的不变性，但不同相机拍摄的图像间可能存在较大的颜色和纹理差异，会影响特征匹配精度^[34]。

实测船只图像与三维渲染图像的颜色、纹理和强度信息不同，实测船只图像的强度是由测量

时的光照强度、角度和可见光相机的参数决定的。三维模型图像强度表征的是船体的明暗关系，甚至与船体的涂装有关。因此，使用原始的测量图像与三维渲染图像进行图像匹配或特征点匹配是很困难的。虽然两幅图像的颜色、纹理和强度不同，但船体和上层建筑的轮廓是相同的，因此本文采用船只的轮廓特征进行匹配，提取船只图像的掩码图像进行图像匹配。

将实测船只目标掩码图作为匹配图像，船只三维姿态掩码图作为匹配模板。因为二者都是简单的二值图像，所以本文采用ReLU函数作为图像匹配函数^[35]。

ReLU函数的定义为：

$$y = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

设匹配模板图像为 $T(x,y)$ ，图像尺寸为 $m \times n$ 。设匹配图像为 $I(x,y)$ ，图像尺寸为 $M \times N$ ，且 $m \leq M$ ， $n \leq N$ 。定义相关度计算的代价函数为 $\text{ReLU}(|T(x,y) - I(x,y)|)$ 公式为：

$$\text{ReLU}(|T(x,y) - I(x,y)|) = \max(0, |T(x,y) - I(x,y)|) \quad (2)$$

由于匹配图像采用掩码图，背景是很纯净的，这种代价函数可以很好的适应这种情况。匹配的相关系数可以表示为：

$$R(x,y) = \sum_{x_1,y_1} \text{ReLU}(|T(x_1,y_1) - I(x+x_1,y+y_1)|) \quad (3)$$

其中， x,y 是匹配模板图的左上角像素在匹配图像中的像素坐标， x_1,y_1 是匹配模板图的像素坐标。

4 实例分割

本文采用实例分割算法来获得目标的掩码图，实例分割主流的方法分为三大类：一阶段实例分割、二阶段实例分割和 Query-based（基于查询）实例分割。一阶段实例分割，如 BlendMask（一种分割模型）^[36]、CondInst（动态实例特征驱动的一阶段实例分割模型）^[37]、YOLOv8-seg^[38]直接在单一的网络中同时进行检测和分割，可以在速度和精度之间取得平衡，并且实时性好。二阶段实例分割，如 Mask R-CNN（掩码区域卷积神经网络）^[39]、HTC（混合任务级联）^[40]、Cascade Mask R-CNN（级联掩码区域卷积神经网络）^[41]分为两个阶段，第一个阶段使用目标检测器来检测图像

中的对象, 第二阶段使用一个分割器来对每一个检测到的对象进行像素级的分割, 可以实现高精度的分割, 但实时性不好。Query-based 实例分割 (如 SOLOv2^[42]) 需要用户指定一个查询对象, 然后自动检测和分割该对象的所有实例。

本文选用 YOLOv8-seg 对实测船只图像和船只三维渲染图进行实例分割, 分别获得实测船只目标掩码图和船只三维姿态掩码图。对多种姿态下的船只三维渲染图进行实例分割, 得到船只姿态模板库。YOLOv8 的实例分割示例如图 3 所示。

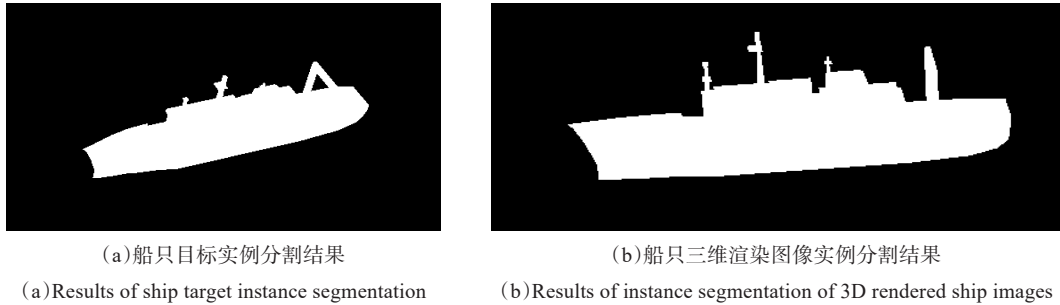


图 3 YOLOv8 的实例分割结果

Fig. 3 Instance segmentation results of YOLOv8

5 工作流程

姿态估计方法的工作流程可分为船只姿态模板库构建和船只姿态估计, 如图 4 所示。构建多种姿态下船只的掩码图, 作为船只姿态估计的图像匹配模板, 主要步骤为: 建立船只三维模型, 构建虚拟环境, 渲染不同姿态下的船只图像, 对渲染图进行实例分割生成掩码图像, 从而构成船只姿态模板库。

船只姿态估计首先对实测船只图像进行目标检测, 获得船只的旋转角度, 将旋转角度作为粗偏航角。根据粗偏航角, 从船只姿态模板库中选取相近姿态的掩码图像模板, 作为船只姿态模板组。将实测船只图像掩码图与船只姿态模板组中的图像依次进行图像匹配, 计算出每次图像匹配的相关系数, 选取相关系数最小的掩码图像模板的姿态作为实测船只的姿态。

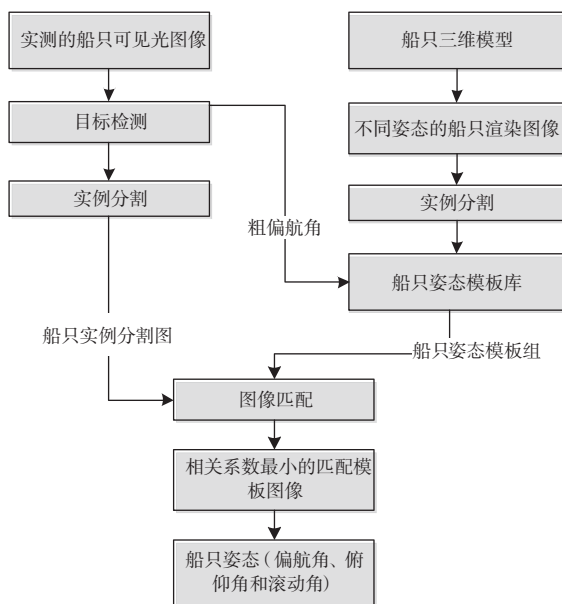


图 4 姿态估计方法工作流程图

Fig. 4 Workflow diagram of pose estimation method

6 仿真验证

6.1 构建仿真数据

使用 MATLAB 的 3D World Edit 软件, 通过添加 Container_Ship.x3d 船只三维模型节点、视点、光源等节点, 建立船只三维虚拟环境。通过 MATLAB 代码控制虚拟环境中的船只多角度旋转, 每旋转一次, 截取一幅虚拟画面图像, 构建此姿态下的船只图像模板。

本仿真验证选取偏航角 $85^{\circ}\sim 95^{\circ}$ 、俯仰角 $0^{\circ}\sim 5^{\circ}$ 、横滚角 $0^{\circ}\sim 5^{\circ}$ 的范围内, 以 1° 为间隔截取虚拟画面图像, 采用 YOLOv8 实例分割模型对虚拟画面图像进行实例分割获得掩码图像, 从而建立船只姿态模板库。船只姿态模板库的部分图像如图 5 所示。

6.2 构建船只实例图像

在上述虚拟环境中, 模拟现实场景, 调整船只表面材料和光照, 并截取 4 个显示画面作为船只

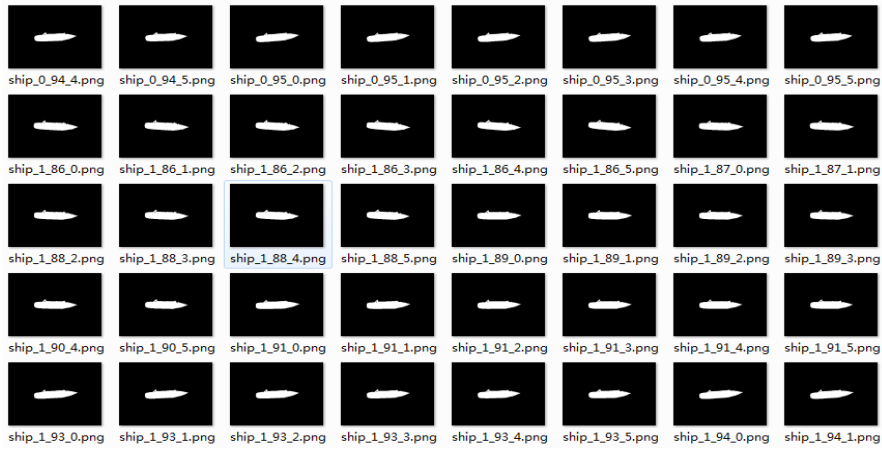


图 5 船只姿态模板库部分图像

Fig. 5 Some images of the ship attitude template library

采集图像，再通过目标检测和实例分割，获得实测船只图像掩码图。结果如图 6 和图 7 所示。

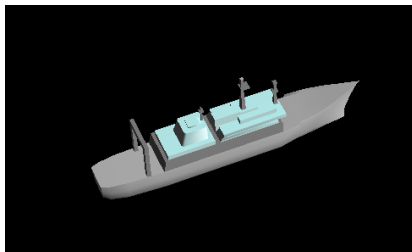


图 6 采集的船只图像

Fig. 6 Collected ship images



图 7 实测船只掩码图像

Fig. 7 Ship mask image

表 1 匹配结果

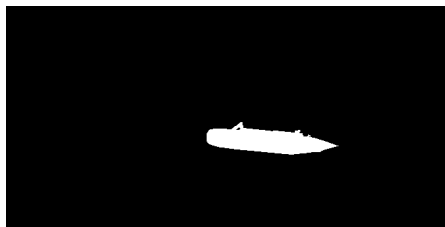
Table 1 Match result

序号	匹配图像文件名	匹配结果文件名	最小的匹配相关系数
1	cur_ship_0_89_4.bmp	ship_0_89_4.png	24
2	cur_ship_1_85_4.bmp	ship_1_85_4.png	31
3	cur_ship_2_90_2.bmp	ship_2_90_2.png	28
4	cur_ship_3_93_1.bmp	ship_3_93_1.png	19

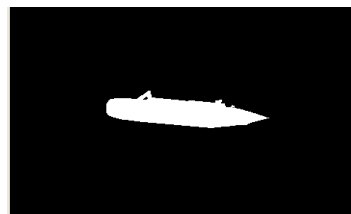
6.3 图像匹配验证

将实测船只图像掩码图作为匹配图像，与船只姿态模板库中的图像分别进行匹配。取相关系数最小的匹配模板图像作为匹配结果，匹配结果见表 1。图像文件名中标识了船只的姿态角，对比表 1 中匹配图像文件名和匹配结果文件名，可以看出匹配结果全部正确。

其中一对匹配结果图像如图 8 所示。



(a) 船只图像掩码图: cur_ship_1_85_4.bmp
(a) Ship image mask map: cur_ship_1_85_4.bmp



(b) 匹配结果文件: ship_1_85_4.png
(b) Matching result file: ship_1_85_4.png

图 8 匹配结果图像

Fig. 8 Match result image

7 结束语

在海洋监测与船舶运行管理领域, 精确获取海上船只动态姿态信息具有至关重要的意义。本文创新性地提出一种海上船只动态姿态测量方法, 该方法借助单目相机采集可见光图像, 将传统模板匹配算法与深度学习算法进行有机融合, 以实现对运动船只三维姿态的精确估计。仿真实验结果表明, 该方法展现出了极高的精度。

此方法具备显著优势, 在设备投入方面成本较低, 有效降低了技术应用的门槛; 实时性表现突出, 能够迅速反馈船只的姿态信息; 环境适应性良好, 即使在复杂光照等条件下仍能保持稳定运行。通过对算法流程的深入分析可知, 姿态估计精度主要受制于实例分割算法的精度。在实际应用场景中, 图像易受到光照、背景等多种因素的干扰, 进而导致实例分割精度降低。为推动该技术的持续发展, 后续需广泛采集多环境下的船只数据, 并通过模型训练提升实例分割算法的精度。

参考文献

- [1] 马燕妮. 三维姿态传感器的研制[D]. 西安: 西安工业大学, 2015.
- [2] 孙立刚. GPS姿态测量技术研究现状及发展趋势[J]. 自动化应用, 2017(4): 109-110.
SUN Ligang. Research status and development trend of GPS attitude measurement technology[J]. Automation Application, 2017(4): 109-110.
- [3] 李娜. 基于GPS/INS组合姿态方法研究[D]. 沈阳: 沈阳理工大学, 2015.
- [4] 郑义桀, 罗健欣, 陈卫卫, 等. 单目视觉SLAM综述[J]. 软件导刊, 2022(12): 242-251.
ZHENG Yijie, LUO Jianxin, CHEN Weiwei, et al. Review of monocular visual SLAM[J]. Software Guide, 2022(12): 242-251.
- [5] LINDBERGER T. Scale invariant feature transform[J]. Scholarpedia, 2012, 7(5): 10491.
- [6] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features(SURF)[J]. Computer Vision and Image Understanding, 2007, 110(3): 346-359.
- [7] RUBLEE E, RABAUDE V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 IEEE International Conference on Computer Vision(ICCV), 2011: 2564-2571.
- [8] VINCENT L, FRANCESC M-N, PASCAL F. EPnP: An accurate on solution to PnP problem[J]. International Journal of Computer Vision, 2008, 81(2): 155-166.
- [9] HINTERSTOISSER S, HOLZER S, CAGNIART V, et al. Multimodal templates for real - time detection of texture - less object in heavily cluttered scenes[C]//2011 IEEE International Conference on Computer Vision (ICCV), 2011: 858-865.
- [10] MELLADO N, AIGER D, MITRA N J. SUPER 4PCS fast global pointcloud registration via smart indexing[J]. Computer Graphics Forum: Journal of the European Association for Computer Graphics, 2014, 33(5): 205-215.
- [11] YANG J, LI H, CAMPBELL D, et al. GP-ICP: A globally optimal solution to 3D ICP point - set registration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2241-2254.
- [12] TEKIN B, SINHA S N, FUA P. Real - time seamless single shot 6D object pose prediction [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 292-301.
- [13] RAD M, LEPETIT V. A scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth [C]// 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 3848-3856.
- [14] PENG S, LIU Y, HUANG Q, et al. PVNet: Pixel - wise voting network for 6DoF pose estimation [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 4556-4565.
- [15] WANG C, XU D, ZHU Y, et al. Densefusion 6d object pose estimation by iterative dense fusion [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 3343-3352.
- [16] YU X, ZHUANG Z, KONIUSZ P, et al. 6DOF object pose estimation via differentiable proxy voting loss [EB/OL]. arXiv preprint arXiv: 200203923, 2020.
- [17] PARK K, PATTEN T, VINCZE M. Pix2pose: Pixel - wise coordinate regression of objects for 6d pose estimation [C]//2019 IEEE International Conference on Computer Vision(ICCV), 2019: 7668-7677.
- [18] KENDSLL A, GRIMES M, CIPOLLA R. PoseNet: A convolutional network for real time 6-DOF camera relocalization[C]// 2015 IEEE International Conference on Computer Vision(ICCV), 2015: 2938-2946.
- [19] KENDALL A, CIPOLLA R. Geometric loss functions for camera pose regression with deep learning [C]//2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 6555-6564.
- [20] XIANG Y, SCHMIDT T, NARAYANNAN V, et al. Posecnn: A convolutional neural network for 6D object pose estimation in cluttered scenes [EB/OL]. arXiv pre-

- print arXiv:171100199, 2017.
- [21] DO T-T, CAT M, PHAM T, et al. Deep-6DPose: Recovering 6D object pose from a single RGB image[EB/OL]. arXiv preprint arXiv:180210367, 2018.
- [22] HE Kaiming, ZHANG xiangyu, REN shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [23] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [24] GIRSHICK R. Fast R-CNN [C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.
- [25] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [26] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.
- [27] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [J]. CoRR, 2015, abs/1512.02325.
- [28] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [29] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2016: 779-788.
- [30] DUAN K, BAI S, XIE L X, et al. CenterNet: Keypoint triplets for object detection[C]//Proceedings of 2019 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6568-6577.
- [31] ZHOU X Y, ZHUO J C, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points [C]//Proceedings of 2019 IEEE Conference on Computer Vision And Pattern Recognition. Piscataway: IEEE Press, 2019: 850-859.
- [32] LAW H, DENG J. CornerNet: Detecting objects as paired keypoints [J]. International Journal of Computer Vision, 2020, 128(3): 642-656.
- [33] 张效娟, 赵元豪, 赵洋. 基于环形平滑 YOLOv5- Ghost 的唐卡元素自动检测算法[J]. 山西大学学报(自然科学版), 2023, 46(2): 342-351.
- ZHANG Xiaojuan, ZHAO Yuanhao, ZHAO Yang. An automatic detection algorithm for Thangka elements based on circular smoothing YOLOv5 - Ghost[J]. Journal of Shanxi University (Natural Science Edition), 2023, 46(2): 342-351.
- [34] 陈燕. 基于特征图像匹配的目标定位关键技术研究[D]. 南京: 南京航空航天大学, 2018.
- [35] 魏巍. 基于图像显著性的舰船目标检测与姿态识别[D]. 北京: 北京理工大学, 2016.
- [36] GAO Y C, WANG H T, LI M, et al. Automatic tandem dual BlendMask networks for everity assessment of wheat Fusarium head blight[J]. Agriculture, 2022, 12(9): 1493.
- [37] TIAN Z, SHEN C, CHEN H. Conditional convolutions for instance segmentation[C]//ECCV. Computer Vision-ECCV 2020: 16th European Conference. Glasgo: Springer International Publishing, 2020: 282-298.
- [38] 丁秀清, 周斌, 胡波. 基于改进 Yolo v8s-seg 的船舶旋转角度检测方法[J]. 中南民族大学学报(自然科学版) 2023, 43(2): 209-216.
- [39] LIU Z, YEOH J K W, GU X Y, et al. Automatic pixel-level detection of vertical cracks in asphalt pavement based on GPR investigation and improved mask R-CNN [J]. Automation in Construction, 2023, 146: 104689.
- [40] CHEN K, PANG J M, WANG J Q, et al. Hybrid task cascade for instance segmentation[C]//IEEE. Proceedings of the IEEE / CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4974-4983.
- [41] CAI Z W, VASCONCELOS N. Cascade R-CNN: High quality object detection and instance segmentation[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. New York: IEEE, 2021, 43(5): 1483-1498.
- [42] SUN X M, FANG W T, GAO C Q, et al. Remote estimation of grafted apple tree trunk diameter in modern orchard with RGB and point cloud based on SOLOv2[J]. Computers and Electronics in Agriculture, 2022, 199: 107209.

[作者简介]

- 闫 威 1974年生, 硕士, 高级工程师。
 郑智辉 1986年生, 博士, 高级工程师。
 郭宸瑞 1993年生, 硕士, 工程师。
 王 硕 1993年生, 硕士, 工程师。
 董昊天 1996年生, 硕士, 工程师。

(本文编辑: 杨秀丽)