



文献 CSTR:

32001.14.11-6035.csd.2023.0149.zh



文献 DOI:

10.11922/11-6035.csd.2023.0149.zh

数据 DOI:

10.57760/sciencedb.j00001.00920

文献分类: 信息科学

收稿日期: 2023-09-20

开放同评: 2023-09-25

录用日期: 2023-10-31

发表日期: 2024-03-28

面向对海监视的舰船目标跟踪与航迹融合数据集

刘丽华¹, 陈志豪¹, 杨皓宇¹, 肖开明^{1*}, 吴继冰¹, 陈海文¹, 黄宏斌¹

1. 国防科技大学, 大数据与决策实验室, 长沙 410073

摘要: 对海监视中航迹实时关联与轨迹融合任务是安全防控、区域态势监视、远程精确打击等军民应用领域的热点和难点问题, 高质量的数据集对推动目标跟踪与融合技术在该领域的研究有重要作用。本研究针对目标跟踪与融合领域技术研究的数据需求以及目前公开数据集所存在的数据缺乏、场景设计针对性差、数据格式单一、数据描述不全等问题, 通过仿真软件对复杂场景中多传感器多目标探测数据进行仿真, 提供了一套面向典型对海监视场景(以舰船为探测对象的2D雷达与侦察传感器(ESM))的目标跟踪与航迹融合数据集。其中仿真软件包括剧情发生器和传感器模拟器两部分, 是一套成熟的目标跟踪场景仿真环境, 提供逼真的探测数据模拟能力。本数据集涵盖的传感器对象包括2D雷达与侦察传感器, 目标包括典型的海上舰船类别, 并支持携带辐射源配置, 设计了高速运动、密集交通、多传感器数据融合、特定舰船侦测和交叉定位等多种典型场景。本数据集中共包含368155条目标点迹, 舰船数量为101条, 时间范围15000秒, 数据格式符合实际设备上报情景、探测误差模型符合实际。本数据集通过对数据误差进行正态性检验、对检测率、虚警率的场景检验以及实地调研, 完成了对数据的准确性评估与数据完备性控制, 可为舰船目标跟踪、轨迹融合等算法研究与验证提供基础数据。

关键词: 目标跟踪; 轨迹融合; 目标点迹; 雷达; 侦察; 数据集

数据库(集)基本信息简介

数据库(集)名称	面向对海监视的舰船目标跟踪与轨迹融合数据集
数据通信作者	肖开明(kmxiao@nudt.edu.cn)
数据作者	刘丽华、陈志豪、杨皓宇、肖开明、吴继冰、陈海文、黄宏斌
数据时间范围	15000秒
地理区域	地理范围为116.69751°E-121.23773°E, 22.57222°N-26.509774°N, 其中包括福建省、台湾省以及东海的部分区域
数据量	74.5 MB
数据格式	*.csv
数据服务系统网址	http://dx.doi.org/10.57760/sciencedb.j00001.00920
数据库(集)组成	数据集共包括3个数据文件, 其中: (1) radar-simulation.csv 是传感器探测数据, 数据量16.73 MB; (2) process.csv 是舰船轨迹数据, 数据量57.85 MB; (3) sensorinfo.csv 是传感器参数信息, 数据量1.11 KB。

* 论文通信作者

肖开明: kmxiao@nudt.edu.cn

引言

目标跟踪与轨迹合批技术在侦察监视、武器制导以及交通控制等领域有着至关重要的作用,通过高效准确实时的技术手段实现目标跟踪与轨迹合批是态势感知和有效决策的基础。基于雷达信号的航迹实时关联与轨迹融合是将数据融合技术运用于目标跟踪领域,能够解决军事中的诸多问题,具备广阔的应用前景和巨大的科研价值^[1]。由于多传感器系统具有较大的优越性,可以提高对目标的检测概率和跟踪精度,故而多传感器多目标跟踪问题越来越受到国内外专家学者的重视,与之相关的研究成果也日益增多^[2]。通过对来自多传感器的数据进行迅速而有效的处理,可以获取准确的目标状态和属性估计,从而进行完整且及时的态势评估和威胁估计,为火力控制、电子对抗、辅助决策提供可靠的依据。近年来,基于雷达信号的航迹实时关联与轨迹融合已经成为目标跟踪领域的重要研究方向。尤其是随着大数据技术的快速发展,基于大数据的融合算法设计以及算法参数生成问题,也逐渐成为该领域的研究热点。

目前雷达与侦察航迹关联是异类传感器航迹关联中的重要研究内容^[3]。2D 雷达可以获取探测范围内观测目标的距离、方位信息;而侦察设备探测范围通常大于 2D 雷达,能获取探测范围内电子设备的方位、载频、脉宽和重频信息,但无法获取距离信息,具有更强的目标判别能力。随着战场环境日益复杂,目标密集度越来越高,为提高态势的准确性,必须通过适当的目标跟踪和航迹融合方法来实现雷达和侦察情报信息充分利用。

尽管实际场景数据因其无可替代的真实性而被学界高度重视,然而获取这类数据所面临的高昂成本问题以及对舰船种群多样性与时空关系动态构造的局限性,使得单纯依赖实地观测数据推动算法研发与效能评估的研究路径面临严峻挑战^[4-5]。当前公开可获取的目标检测与跟踪数据集资源数量较少。针对现实中迫切的应用需求,以及领域内数据资源匮乏的现状,构建了本数据集。与现有的相关的三类数据集——真实数据集、非海上目标仿真数据集以及海上目标仿真数据集相比,本数据集展现出显著的优势和创新性特质。首先,与实际获取的真实数据集^[6-8]相比,本数据集运用先进的仿真技术精细复现了各类复杂的海洋环境背景和舰船运动状态,既克服了真实数据获取过程中的种种困难,又赋予了数据集更高的可控性和定制化潜能;其次,与一般的非海上目标仿真数据集(空中飞行目标仿真数据集^[9]和地上运动目标仿真数据集^[10])相比,本数据集以舰船目标为核心,系统地涵盖了多种舰船类型及多变的时空行为模式,从而更精准地服务于舰船识别与追踪等相关高级任务的研究需求;最后,相较于其他的海上目标仿真数据集^[11-13],本数据集在保证数据精确性的前提下,尤其注重对 ESM 交叉定位等专业应用场景的细致建模和深度挖掘,以此确保数据集在涵盖范围和内涵深度方面的完备性。当前已有的仿真数据集大多关注并集中收录的是雷达回波数据或舰船的图像数据^[11-13],对 ESM 交叉定位场景设计较少、数据描述单一,领域内的科研学术研究缺乏数据支撑,算法性能测试存在较大困难,因此急需一个质量较高,场景全面的数据集^[14]。

本数据集通过仿真模拟、数据处理标注、数据评测准则设计等步骤最大地还原了真实场景,设计了高速运动、密集交通、多传感器数据融合、特定舰船侦测和交叉定位等多种典型场景,构建了基于传感器信号的舰船目标跟踪与轨迹合批数据集。

1 数据采集和处理方法

仿真软件采用 MATLAB 内置的雷达与舰船仿真模型,通过对舰船行驶过程及传感器探测过程进

行仿真，模拟真实场景得到雷达数据与轨迹数据。具体过程如下：首先，通过软件设计 2D 雷达与 ESM 传感器相关参数和误差，并模拟多艘舰船的航行轨迹，保证舰船航行轨迹在传感器探测范围之内（非所有舰船每时每刻均在传感器探测范围内）；然后，仿真过程运行，同时打开雷达记录舰船行驶的整个过程所产生的数据信息；最后，将数据信息的初始文件进行提取与处理，分别得到雷达记录数据与舰船真实轨迹数据。数据仿真、采集和处理过程如图 1。

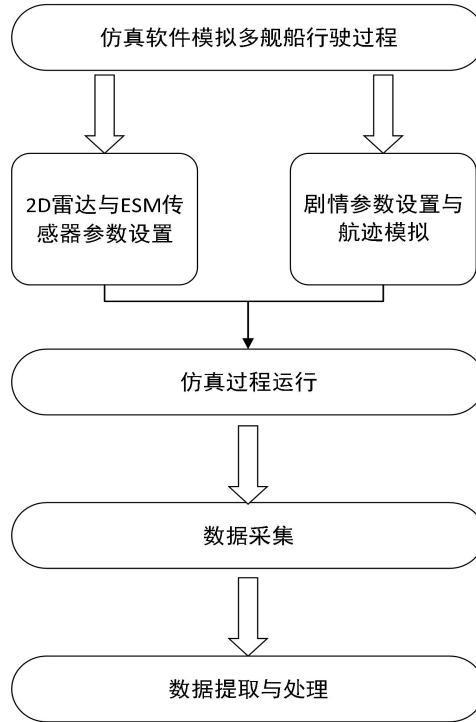


图 1 数据集采集和处理过程

Figure 1 Acquisition and processing of the dataset

1.1 数据仿真方法

首先，对 2D 雷达与 ESM 传感器进行创建和参数设置，充分考虑满足测试需求的仿真过程所需要各种参数和误差，统一设置其剧情参数和传感器参数，包括传感器类型、传感器位置、扫描周期、方位量程、距离量程、虚标率、检测概率、系统偏差（包括原点坐标偏差、时间同步偏差、测距偏差、方位偏差和俯仰偏差）和随机误差（包括类型、扫描周期、探测范围、方位范围、测距误差、方位误差、载频误差、脉宽误差、重频误差、位置）等，参数细节见表 1。表 1 当中所提到的偏差和误差均呈 3σ 处截断的正态分布，偏差（误差）值为其分布的 6σ 值，本数据集中我们设置 6 个 2D 雷达与 3 个 ESM 传感器。

表 1 参数及相应说明

Table 1 Parameters and corresponding descriptions

传感器类型	参数	说明
2D 雷达、ESM	传感器位置	表示传感器在大地坐标系下的经纬度
2D 雷达、ESM	扫描周期	表示传感器扫描一周所需要的时间
2D 雷达、ESM	方位量程	表示传感器可以探测的角度范围

传感器类型	参数	说明
2D 雷达	距离量程	表示传感器可以探测的距离范围
2D 雷达、ESM	高度	表示传感器在大地坐标系的高度
2D 雷达、ESM	检测概率	表示在雷达检测范围内目标的最低（最远处）被检测概率
2D 雷达、ESM	虚标率	表示在雷达检测范围内出现假目标的概率
2D 雷达、ESM	原点坐标偏差	表示雷达设置点在地心系当中的偏差
2D 雷达、ESM	时间同步偏差	表示不同传感器侦测信息的时间同步的偏差
2D 雷达	测距偏差	表示由于传感器本身因素导致的测量目标到传感器距离的偏差
2D 雷达、ESM	方位偏差	表示由于传感器本身因素导致的测量目标到传感器方位的偏差
2D 雷达、ESM	俯仰偏差	表示传感器本身俯仰角度的偏差
2D 雷达	测距误差	表示传感器由于多种因素影响，测得的目标到传感器距离的误差
2D 雷达、ESM	方位误差	表示传感器由于多种因素影响，测得的目标到传感器方位的误差
ESM	载频误差	表示传感器由于多种因素影响，测得的目标载频的误差
ESM	脉宽误差	表示传感器由于多种因素影响，测得的目标脉宽的误差
ESM	重频误差	表示传感器由于多种因素影响，测得的目标重频的误差

然后，对 2D 雷达与 ESM 传感器进行编组，其属于 3 个海边观测站，每个观测站当中分别包含一个 ESM 传感器和两个 2D 雷达，其中传感器 1、2、3 属于观测站 1，传感器 4、5、6 属于观测站 2，传感器 7、8、9 属于观测站 3。

最后，在雷达侦测范围内模拟 101 艘舰船的行驶轨迹，其中 50 艘舰船带有传感器，能够被 ESM 雷达检测到。考虑现实情形当中会发生的各种情况，人工设置舰船速度、航迹和舰船搭载的传感器参数等。

1.2 数据采集方法

在各个仿真参数设置达到要求后，运行仿真过程，进行数据采集，当剧情时间达到 15000 s 时，舰船行驶过程基本结束，结束仿真过程。仿真软件返回 3 个数据文件，分别对应着 101 艘舰船真实轨迹信息以及三个观测站中的传感器返回的侦测信息以及传感器参数。其中，传感器共收集到 368155 条数据，真实轨迹信息时间范围 15000 s。

2 数据样本描述

本数据集以 csv 格式进行存储，包含 radar-simulation.csv、process.csv 和 sensorinfo.csv 共 3 个文件，其中 radar-simulation.csv 内每一条记录是传感器返回的数据，process.csv 内每一条记录是某一舰船在某时刻的真实轨迹信息，sensorinfo.csv 则记录全部传感器的参数信息。

radar-simulation.csv 中包含：批号、时间、传感器编号、目标真实 ID、距离（km）、方位（度）、俯仰（度）、载频（GHz）、脉宽（us）、重频（Hz）。

process.csv 中包含：帧数、时间、目标真实 ID、经度、纬度。

sensorinfo.csv 中包含：传感器编号、传感器类型、隶属观测站编号、距离量程（公里）、方位量程（度）、测距误差（米）、方位误差（度）、载频误差（兆赫兹）、脉宽误差（微秒）、重频

误差（赫兹）、扫描周期（秒）、经度、纬度、高度（m）。其中：

- (1) 批号是传感器每个扫描周期所探测到的一条记录的编号。
- (2) 时间是剧情发展的时间。
- (3) 传感器编号是当前信息来源的传感器的编号。
- (4) 目标真实 ID 是这条信息返回的船的真实 ID。
- (5) 距离是在传感器的极坐标系下其探测到的目标与传感器的距离。
- (6) 方位是在传感器的极坐标系下其探测到的目标与传感器的方位。
- (7) 俯仰是被检测对象相对于传感器的角度。
- (8) 载频是 ESM 传感器所探测到的舰船传感器返回的载频。
- (9) 脉宽是 ESM 传感器所探测到的舰船传感器返回的脉宽。
- (10) 重频是 ESM 传感器所探测到的舰船传感器返回的重频。
- (11) 帧数是在剧情进行过程中每经过一个扫描周期后的时刻编号。
- (12) 经度和纬度（process 文件）是舰船在剧情当中的真实经纬度。具体样本数据展示如图 2、图 3、图 4 所示。

批号	时间	传感器编号	目标真实ID	距离	方位	俯仰	载频	脉宽	重频
1	12	3	0	-100	101.743	-100	10163.3	25.0606	2003.2
2	12	3	2	-100	93.664	-100	10073.2	24.8115	1991.14
3	12	3	4	-100	144.444	-100	9519.93	0.1904	1992.81
4	12	3	30	-100	93.6761	-100	9526.23	0.1938	737.991
5	12	3	33	-100	162.187	-100	9464.67	1.0436	771.282
6	12	3	36	-100	109.467	-100	9621.46	1.036	740.935
7	12	3	39	-100	113.933	-100	9370.71	0.1987	756.113
8	12	3	44	-100	36.1131	-100	10078.5	24.8252	2043.86

图 2 radar-simulation 数据集数据示例

Figure 2 Data sample of radar-simulation dataset

帧数	剧情当前时间	目标id	经度	纬度
0	6	0	119.69031	38.2126
0	6	1	123.94703	37.206525
0	6	2	123.31026	37.978529
0	6	3	118.96615	38.773319
0	6	4	121.18146	37.675098
0	6	5	120.57311	37.79158
0	6	6	124.59862	36.65736
0	6	7	121.60112	38.761117

图 3 process 数据集数据示例

Figure 3 Data sample of process dataset

3 数据质量控制和评估

3.1 数据准确性验证

本数据集选择在 MATLAB 环境下的雷达模型作为实验的基础框架，不仅基于科研领域对其广泛认可，也因为它能够对实际雷达信号进行精确模拟，能为我们的实验结果提供更强的可信度^[15]。通过对模型进行多次核验和微调，包括对参数设定的验证、对模型输出结果的反复测试等，使得实验数据更具准确性和可靠性。为确保模型参数设定符合实际，本研究线下调研部分应用单位，充分考虑多种海目标探测场景，以实现对各项参数的调整与优化；为确保模型输出结果的精准性，保证数

据质量, 设计场景进行模型输出结果测试, 对检测率、虚警率、测距偏差、方位偏差、载频误差、重频误差和脉宽误差进行分析和统计, 其结果均符合参数设定。

编号	传感器类型	隶属观测站	距离量程	方位量程	测距误差	方位误差	载频误差	脉宽误差	重频误差	扫描周期
0	ESM_1D	1	[0, 240]	[0, 360]	0	3	5	0.3	1.2	12
1	Radar_2D	1	[0, 120]	[0, 360]	20	1.2	0	0	0	6
2	Radar_2D	1	[0, 120]	[0, 360]	20	1.2	0	0	0	6
3	ESM_1D	2	[0, 220]	[0, 360]	0	3	8	0.2	1.2	12
4	Radar_2D	2	[0, 110]	[0, 360]	30	1.2	0	0	0	8
5	Radar_2D	2	[0, 110]	[0, 360]	30	1.2	0	0	0	8
6	ESM_1D	3	[0, 240]	[0, 360]	0	3	5	0.3	1.2	12
7	Radar_2D	3	[0, 120]	[0, 360]	20	1.2	0	0	0	6
8	Radar_2D	3	[0, 120]	[0, 360]	20	1.2	0	0	0	6
编号	经度	纬度	高度	虚标率	检测概率	原点坐标偏差	时间同步偏差	测距偏差	方位偏差	俯仰偏差
0	122.00841	39.158	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
1	122.2674	39.39308	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
2	121.60113	38.89039	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
3	121.26799	37.47584	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
4	121.87633	37.3356	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
5	120.8045	37.62781	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
6	120.81898	36.44707	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
7	121.03625	36.6575	0	0.000001	0.9	(0, 0, 0)	0	0	0	0
8	120.39894	36.13143	0	0.000001	0.9	(0, 0, 0)	0	0	0	0

图4 sensorinfo 数据集数据示例

Figure 4 Data sample of sensorinfo dataset

3.1.1 误差控制与评估

生成 2000 条点迹信号进行测试, 由于舰船行驶速度对误差生成的影响可忽略不计, 为方便统计误差, 我们选择同一艘静止舰船进行重复观测。场景具体设计如下: 设置 2D 雷达和 ESM 传感器各一个, 对其探测范围边缘位置同一艘舰船各观测 1000 次, 传感器参数除 2D 雷达探测距离属性外 (方便距离误差分布能够更清晰地被检测到) 与数据集保持一致, 统计误差情况并对误差进行正态分布检验, 误差分布直方图如图 5 所示。2D 雷达的距离、方位参数与 ESM 传感器的方位、载频、脉宽、重频参数是与检测信号质量紧密相关的参数, 清楚其分布对海目标侦测有重要意义。如图 5 所示, 其误差直方图分布与正态分布拟合曲线趋势基本一致, 且均呈现 3σ 处截断的正态分布, 与误差预期分布相符。

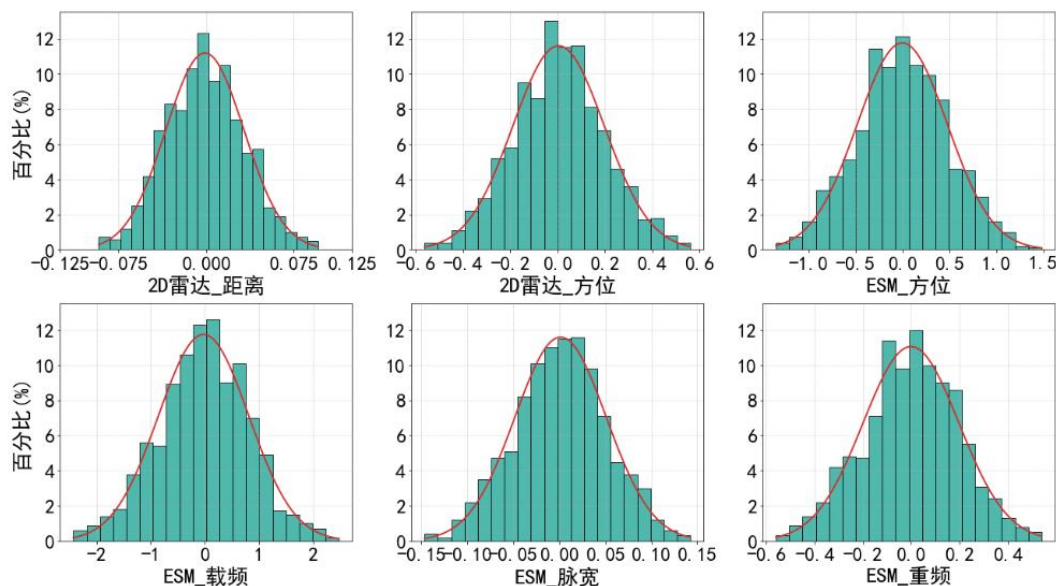


图5 数据集误差分布直方图

Figure 5 Histogram of the dataset error distribution

使用 Anderson-Darling 检验、Shapiro-Wilk 检验和 D'Agostino 检验三种方法验证其正态性，结果如表 2 所示。Anderson-Darling 检验是一种检验数据与特定分布（如正态分布）之间差异的方法，其为分布的各个部分赋予不同的权重，敏感地捕捉数据与理论分布之间的偏差^[16]。Anderson-Darling 统计量表达了其与正态分布的差异，统计量值越小，检测分布与正态分布的差异越小。在显著性 0.05 的情况下，Anderson-Darling 检验临界值为 0.784，6 个参数均小于临界值，因此可接受其分布接近正态分布。Shapiro-Wilk 检验方法基于数据样本的顺序统计量，与正态分布的理论顺序统计量进行比较，进而评估数据正态性，其统计量的值越接近 1，数据越可能符合正态分布^[17]。图中可见，6 个参数 Shapiro-Wilk 统计量均达到 0.99 以上，P 值超过 0.05，可接受其分布接近正态分布。D'Agostino 检验基于数据的偏度和峰度来判断数据是否遵循正态分布，偏度用于衡量数据分布的不对称性，峰度用于衡量数据分布尾部的厚度以及数据分布的中心峰值部分的锐利程度^[18-19]。6 个参数 D'Agostino 检验 P 值均超过 0.05，可接受其分布接近正态分布。三个检验方法从不同角度均证明数据满足正态性要求，进一步证明数据符合设定的质量要求。

表 2 正态分布检验统计量

Table 2 Test statistic for normal distribution

参数 检验方法	Anderson-Darling Statistic	Shapiro-Wilk Statistic	Shapiro-Wilk p-value	D'Agostino Statistic	D'Agostino p-value
2D 雷达_距离	0.40213	0.99782	0.21560	2.57526	0.27592
2D 雷达_方位	0.25582	0.99860	0.62108	0.33535	0.84563
ESM_方位	0.26091	0.99826	0.40797	2.58558	0.27450
ESM_载频	0.57307	0.99750	0.12925	1.76875	0.41297
ESM_脉宽	0.34356	0.99822	0.38610	1.95384	0.37647
ESM_重频	0.58011	0.99716	0.07403	3.78799	0.15047

3.1.2 检测率、虚警率控制与评估

由于检测率随着目标离雷达距离的增大而降低，我们设置数据集的检测率为 0.9。只需验证雷达检测范围边缘舰船的检测率即可。为此，我们设置 2D 雷达和 ESM 各重复观测检测范围边缘 1000 次。观测结果表明：2D 雷达观测到 905 次，ESM 观测到 891 次，检测概率接近 0.9，满足数据质量要求。同时，2D 雷达与 ESM 传感器在 1000 次观测中无虚假目标出现，数据满足虚警率 0.000001 的设置要求。

3.2 数据完备性控制

在保证数据准确性的基础上，充分考虑数据的完备性。对现实场景下雷达侦测的需求进行分析，设计包含高速运动和密集交通中航迹关联精准性问题、多源异构数据融合问题、航迹实时关联问题等关键方面的数据集，深入模拟了传感器的探测过程以及可能影响探测结果的因素。影响因素包括检测概率、虚警杂波、系统偏差和随机误差等多个方面，对模拟结果的真实性和完整性有着显著影响^[20]。此外，考虑影响因素的独立作用的同时，也充分考虑因素之间可能存在的交互效应，全方位地模拟雷达探测的复杂环境，使得数据集具有高度的完备性。

4 数据使用方法和建议

4.1 数据使用方法

本数据集由仿真生成，经过严格质量控制，可作为目标定位、航迹关联及轨迹合批等先进算法的评估与优化的基准。数据的推荐使用方法是：问题导向，针对海目标侦测场景中所出现的侦测需求和算法需求，将数据集中的雷达观测数据作为现实观测数据构建模型算法，将数据集中的舰船真实时空数据作为评估模型算法的标准。

数据集中文件的具体使用方法是：首先使用者使用 `radar-simulation` 文件当中的雷达侦测数据（数据集中除目标真实 id 以外的数据）和 `sensorinfo` 文件当中的参数信息建构合适算法，根据算法进行目标的检测跟踪和轨迹合批，得到检测结果；然后将 `radar-simulation` 文件当中的真实 id、时间数据与 `process` 文件的相同 id 舰船的经纬度进行匹配，得到真值数据；最后通过提供的评测准则将真值数据与检测结果进行比对计算得分，作为算法衡量的指标。

例如在第四届全国大数据与计算智能挑战赛（2023）中，根据实际需求提出三个问题，分别是：

（1）单观测站 2D 雷达观测范围内航行船只的实时目标跟踪任务，要求实时对单观测站 2D 雷达观测范围内的船只进行目标跟踪，以相同的 id 对同一艘船进行标记，并给出这些船的航行轨迹。

（2）搭载传感器的特殊船只的探测任务，多艘船只上搭载了无线电设备，该设备会被观测站中的 ESM 探测到，要求结合 2D 雷达与 ESM 探测信息给出这些船只在 2D 雷达探测范围内的航行轨迹。

（3）由于 ESM 和 2D 雷达探测范围不同，部分区域的船只只能被 ESM 探测到，要求根据 ESM 探测的方位信息，得到观测站 ESM 探测范围交叉重叠区域的船只轨迹。

三问分别考察了单观测站多源同构传感器航迹关联与目标跟踪问题，多观测站多源异构传感器目标检测与轨迹合批问题，多观测站 ESM 传感器远距离交叉定位问题。解决上述问题需要使用 `radar-simulation` 文件中的相关传感器的舰船距离、方位、载频等属性信息和 `sensorinfo` 文件当中的参数信息，首先构建算法进行点迹关联，将点迹信息转换为航迹信息，然后建构算法对多传感器信息进行合批，同时对雷达传感器航迹信息进行转换，将航迹信息转换为经纬度信息，将船的航迹表示出来。

然后需要对轨迹关联算法进行评估，通过一定的评估准则进行真实轨迹与检测轨迹相似度的评估。`process` 文件中为所有舰船的全部真实轨迹。所需范围真实轨迹的获得方法为：在 `radar-simulation` 文件中获得舰船真实 id、时间数据信息，对 `process` 文件相同真实 id 和时间的舰船进行关联，得到特定传感器范围内的真实轨迹。对真实轨迹与生成轨迹的相似度进行评估与比较，进而指导算法的改进。

4.2 算法评估准则

算法的评估准则需要考虑 5 个方面：准确性、实时性、鲁棒性、假警率和漏检率。准确性是目标检测和跟踪最基本的评价指标，其评估算法能否准确的检测和跟踪目标，以及能否正确的关联和合批航迹，描述了算法预测结果与实际情况之间的符合程度；实时性是算法响应和执行的关键特性，其评估算法能否即时地给出轨迹；鲁棒性是算法稳定性的重要衡量标准，其评估系统在面对异常输入或外部干扰时能否维持正常运行，以及能否抵御各种不确定因素，体现算法对各种挑战的适应和抵抗能力；假警率是指算法错误地将非目标识别为目标的频率，其是评价系统误报能力的关键指标；

漏检率描述了算法未能检测到真实目标的频率，其关联到系统的敏感性和可靠性。

基于上述算法评估准则，生成轨迹与数据集中真实轨迹的推荐评价方法为：GOSPA。GOSPA (Generalized optimal sub-pattern assignment metric) 是估计和评价生成轨迹集合与真实轨迹集合之间相似度的业界认可的通用领域方法，其充分考虑了准确性、假警率和漏检率^[21]。GOSPA 估计具体计算公式如下：

$$value = d_p^{(c)}(X, Y) = \left[\min_{\gamma \subseteq \Gamma} \left(\sum_{(i,j) \in \gamma} d(x_i, y_j)^p + \frac{c^p}{2} (|X| + |Y| - 2|\gamma|) \right) \right]^{\frac{1}{p}} \quad (1)$$

$$score = \frac{1}{1 + value} \quad (2)$$

其中 $c > 0, 1 \leq p \leq \infty$, X 表示融合出的轨迹的集合, Y 表示真实轨迹集合, x 和 y 表示某条具体轨迹, $d(x_i, y_i)$ 表示 x_i 与 y_i 之间的相似度, 集合: $\gamma \subseteq \Gamma, \Gamma = \{1, \dots, |X|\} \times \{1, \dots, |Y|\}$ 。

$d(x, y)$ 通过计算每个融合轨迹中的点与对应真实轨迹中同时刻的点的距离的均值求得。

如图 6 所示, 上下两条实线代表两个轨迹, 轨迹之间的虚线的起始端分别代表两个轨迹之间相似的点, 在本问题中即为时刻相同的点。使用所有这些相似点之间的距离的均值来衡量两个轨迹之间的相似性。要计算相似度的融合轨迹和对应的真实轨迹为 x 和 y 。

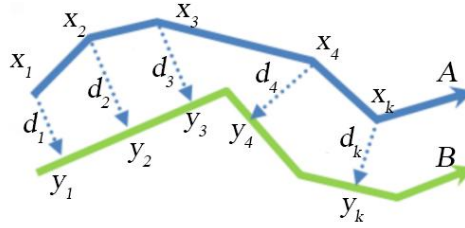


图 6 距离计算示意图

Figure 6 Distance calculation diagram

K_x 为融合轨迹的轨迹点数, K_y 为对应真实轨迹的轨迹点数。 $d = d_1, d_2, \dots, d_k$ 。 d 的含义为 $dis(x_k, y_k)$, $dis(x_k, y_k)$ 指 x 轨迹的第 k 个数据点和 y 轨迹的第 k 个数据点之间的欧几里得距离, 其中 x_k 表示 x 轨迹中的第 k 个轨迹点, y_k 表示 y 轨迹中的第 k 个轨迹点。两轨迹间的相似度 $d(x, y)$ 从 $d_1 = (x_1, y_1)$ 开始, 到 $d_k = (x_k, y_k)$ 结尾, $k = \max(K_x, K_y)$, 以保证融合轨迹 x 和 y 中的每个坐标都在 $d(x, y)$ 中出现。

最后要得到的融合轨迹与其对应的真实轨迹的相似度计算公式如下：

$$d(x, y) = \begin{cases} \frac{1}{K_x} \sum_{k=1}^{K_x} d_k + \frac{1}{K_y - K_x} \sum_{k=K_x+1}^{K_y} dis(x_{K_x}, y_k), & K_x < K_y \\ \frac{1}{K} \sum_{k=1}^K d_k, & K_x = K_y = K \\ \frac{1}{K_y} \sum_{k=1}^{K_y} d_k + \frac{1}{K_x - K_y} \sum_{k=K_y+1}^{K_x} dis(x_k, y_{K_y}), & K_x > K_y \end{cases} \quad (3)$$

算法的实时性需求可以通过要求算法对于每个时刻的输出只能使用当前时刻以及当前时刻之前的数据作为输入, 从而保证算法的实时性。算法的鲁棒性需求可以通过分割数据集交叉验证、人为加入噪声和异常值检测算法性能等方法进行检测。

5 数据价值

本数据集针对传感器信号的舰船目标跟踪与轨迹合批问题，设计了包含多种现实场景在内的时间范围为 15000 s 的仿真数据集，为当下更高效更适应现实要求的算法研究提供了基础。其不仅能满足常见传统的传感器目标跟踪问题的要求，还充分考虑了 ESM 传感器与 2D 雷达传感器的协同定位和多 ESM 传感器的交叉定位问题，设置了多种情况下的传感器交叉区域和舰船航行轨迹。在当代目标识别复杂化、数据集标准性差、数据质量低和开放性差的背景之下，对如何实现更高效更先进的算法应用研究具有重要意义。其中 ESM 交叉定位问题，目前国内尚缺乏权威高质且对外开放的应用的高效算法，数据集资源同样非常匮乏，因此构建一个能满足算法需求的高质数据集尤为重要^[22]。

本数据集中的数据已在 2023 全国大数据与计算智能挑战赛——基于传感器信号的舰船目标跟踪与轨迹合批挑战赛中发布供参赛选手测试及正式比赛使用。包含多所国内相关领域的知名高校及科研院所共计 421 人组成 260 支队伍报名参赛，提交有效作品超八百个。赛题三问选手总成绩 GOSPA 评分最高达到 0.89，关联正确率超过 90%，超过十五名选手总体成绩超过 0.5。其中，选手第二问和第三问的单问最高成绩均达到 0.95 以上，模型算法表现优异，同时也间接证明了本数据集的精准性、可靠性和完备性。此次大赛的成功举办也充分验证了本数据集中仿真数据和评分准则的科学性和可靠性，可作为广泛科研人员的数据素材。此外，在挑战赛结束后，众多参赛选手和国内同行对本数据集都提出了申请需求，这也印证了本数据集的科研价值。

数据作者分工职责

刘丽华（1981—），女，吉林省吉林市人，博士，副教授，研究方向为数据工程、数据融合、智能决策和雷达目标识别。主要承担工作：数据集仿真与构建、数据集核验、论文撰写。

陈志豪（1999—），男，山东省青岛市人，博士研究生，研究方向为数据工程。主要承担工作：数据集仿真与构建、数据集采集、论文撰写。

杨皓宇（1998—），男，山东省临沂市人，硕士研究生，研究方向为数据工程。主要承担工作：数据集仿真与构建、数据集预处理、论文修改。

肖开明（1990—），男，云南省腾冲市人，博士，讲师，研究方向为大数据与智能决策技术。主要承担工作：数据使用评价方法设计、数据集核验、论文撰写。

吴继冰（1987—），男，江苏省南通市人，博士，副研究员，研究方向为信息管理与智能决策。主要承担工作：数据使用评价方法设计、论文撰写指导。

陈海文（1988—），男，山东省临沂市人，博士，讲师，研究方向为数据工程、边缘数据组织。主要承担工作：数据使用评价方法设计、论文撰写指导。

黄宏斌（1975—），男，江苏省如皋市人，博士，研究员，研究方向为信息管理与智能决策、军事信息物理系统（MCPS）、大数据分析。主要承担工作：数据集构建方法指导、论文撰写指导，数据质量控制与综合管理。

参考文献

[1] MA K, ZHANG H G, WANG R T, et al. Target tracking system for multi-sensor data fusion[C]//2017

- IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Chengdu, China. IEEE, 2018: 1768–1772. DOI: 10.1109/ITNEC.2017.8285099.
- [2] 严怀成, 黄心汉, 王敏. 多传感器数据融合技术及其应用[J]. 传感器技术, 2005(10): 1–4. [YAN H C, HUANG X H, WANG M. Multi-sensor data fusion technique and its application[J]. Journal of Transducer Technology, 2005(10): 1–4.]
- [3] 王国宏, 何友. 雷达与 ESM 相关算法综述[J]. 上海航天, 1996, 13(2): 38–42. [WANG G H, HE Y. A summary of interrelating algorithm between radar and esm[J]. Aerospace Shanghai, 1996, 13(2): 38–42.]
- [4] BARTON D K. Radar System Analysis And Modeling [M]. Boston|London: Artech House Publishers, 2004.
- [5] 刘宁波, 董云龙, 王国庆, 等. X 波段雷达对海探测试验与数据获取[J]. 雷达学报, 2019, 8(5): 656–667. DOI: 10.12000/JR19089. [LIU N B, DONG Y L, WANG G Q, et al. Sea-detecting X-band radar and data acquisition program[J]. Journal of Radars, 2019, 8(5): 656–667. DOI: 10.12000/JR19089.]
- [6] WANG Y, WANG C, ZHANG H, et al. A SAR dataset of ship detection for deep learning under complex backgrounds[J]. Remote Sensing, 2019, 11(7): 765.
- [7] ZHANG T, ZHANG X, KE X, et al. LS-SSDD-v1. 0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images[J]. Remote Sensing, 2020, 12(18): 2997.
- [8] WEI S, ZENG X, QU Q, et al. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation[J]. Ieee Access, 2020, 8: 120234-120254.
- [9] 回丙伟, 宋志勇, 范红旗, 等. 地/空背景下红外图像弱小飞机目标检测跟踪数据集[J/OL]. 中国科学数据, 2020, 5(3). (2020-01-20). DOI: 10.11922/csdata.2019.0074.zh. [HUI B W, SONG Z Y, FAN H Q, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground / air background[J/OL]. China Scientific Data, 2020, 5(3). (2020-01-20). DOI: 10.11922/csdata.2019.0074.zh.]
- [10] LIU Z, LUO S, WANG Y. Mix mstar: A synthetic benchmark dataset for multi-class rotation vehicle detection in large-scale sar images[J]. Remote Sensing, 2023, 15(18): 4558.
- [11] HUANG L, LIU B, LI B, et al. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 11(1): 195-208.
- [12] MEYER F, WIN M Z. Joint navigation and multitarget tracking in networks[C]. 2018 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2018: 1-6.
- [13] KOSTA L, IRVINE J, SEAMAN L, et al. Many-target, Many-sensor Ship Tracking and Classification[C]. 2019 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, 2019: 1-7.
- [14] ZHAO W J, CHEN Z, JIN M L. Subband maximum eigenvalue detection for radar moving target in sea clutter[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(2): 281–285. DOI: 10.1109/LGRS.2020.2971589.

- [15] 窦林涛, 程健庆, 李素民. 基于 Matlab 的雷达信号处理系统仿真[J]. 指挥控制与仿真, 2006, 28(2): 78–82. [DOU L T, CHENG J Q, LI S. Simulation of radar signal processing based on Matlab[J]. Command Control & Simulation, 2006, 28(2): 78–82.]
- [16] ANDERSON T W, DARLING D A. Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes[J]. The Annals of Mathematical Statistics, 1952, 23(2): 193–212. DOI: 10.1214/aoms/1177729437.
- [17] SHAPIRO S S, WILK M B. An analysis of variance test for normality (complete samples)[J]. Biometrika, 1965, 52(3/4): 591–611. DOI: 10.1093/biomet/52.3-4.591.
- [18] D'AGOSTINO R B. Transformation to normality of the null distribution of g_1 [J]. Biometrika, 1970, 57(3): 679–681. DOI: 10.1093/biomet/57.3.679.
- [19] D'AGOSTINO R, PEARSON E S. Tests for departure from normality. empirical results for the distributions of b_2 and $\sqrt{b_1}$ [J]. Biometrika, 1973, 60(3): 613–622. DOI: 10.2307/2335012.
- [20] 李天成, 范红旗, 孙树栋. 粒子滤波理论、方法及其在多目标跟踪中的应用[J]. 自动化学报, 2015, 41(12): 1981–2002. [LI T C, FAN H Q, SUN S. Particle filtering: theory, approach, and application for multitarget tracking[J]. Acta Automatica Sinica, 2015, 41(12): 1981–2002.]
- [21] RAHMATHULLAH A S, GARCÍA-FERNÁNDEZ Á F, SVENSSON L. Generalized optimal sub-pattern assignment metric[C]//2017 20th International Conference on Information Fusion (Fusion). Xi'an, China. IEEE, 2017: 1–8.
- [22] 韩志钢, 卿利. 多节点传感器协同探测技术综述与展望[J]. 电讯技术, 2020, 60(3): 358–364. DOI: 10.3969/j.issn.1001-893x.2020.03.020. [HAN Z G, QING L. Overview and prospect of cooperative detection technology for multi-node's sensors[J]. Telecommunication Engineering, 2020, 60(3): 358–364. DOI: 10.3969/j.issn.1001-893x.2020.03.020.]

论文引用格式

刘丽华, 陈志豪, 杨皓宇, 等. 面向对海监视的舰船目标跟踪与航迹融合数据集[J/OL]. 中国科学数据, 2024, 9(1). (2024-03-28). DOI: 10.11922/11-6035.csd.2023.0149.zh.

数据引用格式

刘丽华, 陈志豪, 杨皓宇, 等. 面向对海监视的舰船目标跟踪与航迹融合数据集[DS/OL]. V2. Science Data Bank, 2023. (2024-03-28). DOI: 10.57760/sciencedb.j00001.00920.

A dataset of ship target tracking and trajectory fusion in maritime surveillance

LIU Lihua¹, CHEN Zhihao¹, YANG Haoyu¹, XIAO Kaiming^{1*}, WU Jibing¹,
CHEN Haiwen¹, HUANG Hongbin¹

1. Laboratory for Big Data and Decision, National University of Defense Technology, Changsha 410073, P.

R. China

*Email: kmxiao@nudt.edu.cn

Abstract: Real-time trajectory association and trajectory fusion in maritime surveillance pose great challenges and remain hot issues in security, regional situation monitoring, and long-range precision strikes for both military and civilian applications. High-quality datasets play a pivotal role in advancing research in target tracking and fusion technologies within this domain. This paper addresses the data requirements for technological research in target tracking and fusion, as well as the limitations of currently available datasets, including data scarcity, inadequate scene design specificity, uniform data formats, and incomplete data descriptions. We used simulation software to emulate multi-sensor multi-target detection data in complex scenarios, so as to present a dataset tailored for typical maritime surveillance scenarios—targeting ships using 2D radar and Electronic Support Measures (ESM) sensors. The simulation software comprises a scenario generator and a sensor simulator, creating a mature target tracking scenario simulation environment with realistic detection data modeling capabilities. The dataset includes data from 2D radar and ESM sensors, covering typical maritime ship categories, supporting configurations with radiation sources. It is designed for a variety of scenarios such as high-speed motion, dense traffic, multi-sensor data fusion, specific ship detection, and cross-positioning. The dataset contains a total of 368,155 target tracks from 101 ships, spanning a duration of 15,000 seconds. The data format conforms to actual equipment reporting scenarios, while the detection error model accurately reflects real-world conditions. Accuracy assessment and data completeness are ensured through various methods including normality testing of data error, scenario testing of detection rate and false alarm rate, as well as field research. This dataset can provide fundamental data for algorithmic research and validation in the ship's target tracking, trajectory fusion and other related areas.

Keywords: target tracking; trajectory fusion; target point track; radar; reconnaissance; dataset

Dataset Profile

Title	A dataset of ship target tracking and trajectory fusion in maritime surveillance
Data corresponding author	XIAO Kaiming(kmxiao@nudt.edu.cn)
Data authors	LIU Lihua, CHEN Zhihao, YANG Haoyu, XIAO Kaiming, WU Jibing, CHEN Haiwen, HUANG Hongbin
Time range	15000 s
Geographical scope	Fujian Province, Taiwan Province and part of the East China Sea (116.69751°E–121.23773°E and 22.57222 °N–26.509774 °N)
Data volume	74.5 MB
Data format	*.csv
Data service system	< http://dx.doi.org/10.57760/sciencedb.j00001.00920 >
Dataset composition	The dataset consists of a total of three data files: (1) radar-simulation.csv contains sensor detection data, with a size of 16.73 MB; (2) process.csv contains ship trajectory data, with a size of 57.85 MB; and (3) sensorinfo.csv contains sensor parameter information, with a size of 1.11 KB.