ISSN 1001-5965
CODEN BHHDE8

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

2022-04
Vol.48 No.4

北航学报
赠 阅

北 京 航 空 航 天 大 学 学 报

第 48 卷 第 4 期 （总第 350 期） 2022 年 4 月

目 次

机场跑道水泥混凝土道面板尺寸分析…… 张献民，李梦晓，陈宇，李长辉，徐博庆（551）

基于自编码器和 HMM 的民机接地点远事件检测 …… 霍纬纲，李继龙，王慧芳（560）

含孔隙缺陷三维五向编织复合材料偏轴拉伸力学性能分析 …… 杨志贤，张明，李昂，张超（569）

颅缝早闭手术中颅骨切割方案生成方法…… 罗杨宇，贺佳宾，谢东升，陆珅宇，宫剑（578）

电力场景下基于无人机视觉的运动目标追踪方法…… 冯雪，杜猛俊，向新宇，钱锦，张敏（586）

复杂曲面上的自动铺放路径规划方法…… 赵安安，何大亮，王晗，郭俊刚，柯映林（595）

水下燃料组件转运系统耦合振动分析 …… 袁占航，李运华，刘昊东（602）

下肢康复机器人动力学建模约束违约抑制…… 徐亚茹，李克鸿，刘佳，刘荣，张建成（609）

纤维素类生物质水热法制备航油前驱物能耗分析 …… 董升飞，黄星华，杨晓奕（620）

基于 SINS/RFID 的隧道列车高精度定位方法 …… 张辰东，王兆瑞，金声震，艾国祥（632）

一种基于地转风模型的平流层风场插值方法 …… 胡康，周江华，张晓军（639）

基于双重属性信息的跨模态行人重识别算法…… 陈琳，高赞，宋雪萌，王英龙，聂礼强（647）

单元构架式抛物面天线张紧绳索多层设计方法…… 王辉，何天宇，都显琛，王春洁（657）

柔性翼气动力和变形特性的实验研究 …… 王涵斌，贺曦，王晋军（665）

基于联合分布适配的智能电表运行状态评价 …… 许丹，李游，李贺龙（674）

底凹结构减阻效应数值分析 …… 李彪，王良明，杨志伟（682）

基于 QZSS L1 信号的岸基 GNSS-R 码延迟海面测高 ……
…… 储倜，贺凯飞，高凡，贺匀峤，孟馨悦，徐天河（691）

复合材料层板冰雹高速冲击损伤预测及失效分析 …… 张超，方鑫，刘建春（698）

账户型区块链可扩展存储模型 …… 张栩豪，牛保宁，弓婷（708）

低复杂度自适应容积卡尔曼滤波算法 …… 李春辉，马健，杨永建，甘轶（716）

期刊基本参数： CN 11-2625/V * 1956 * m * A4 * 174 * zh * P * ￥50.00 * 900 * 20 * 2022-04

（**编 辑** 张 嵘 李艳霞 孙 芳 卞欢欢）

北京航空航天大学学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Vol. 48 No. 4 (Sum 350) April 2022

北航学报 赠阅

CONTENTS

Size of cement concrete pavement slab of airport runway
…… ZHANG Xianmin, LI Mengxiao, CHEN Yu, LI Changhui, XU Boqing (551)
Civil aircraft long touchdown exceedance detection based on autoencoder and HMM
…… HUO Weigang, LI Jilong, WANG Huifang (560)
Off-axis tensile mechanical properties of 3D five-directional braided composites with void defects
…… YANG Zhixian, ZHANG Ming, LI Ang, ZHANG Chao (569)
Skull cutting plan generation method in craniosynostosis surgery
…… LUO Yangyu, HE Jiabin, XIE Dongsheng, LU Shenyu, GONG Jian (578)
UAV-vision-based moving target tracking scheme in electric scenario
…… FENG Xue, DU Mengjun, XIANG Xinyu, QIAN Jin, ZHANG Min (586)
Automatic paving path planning method on complex surfaces
…… ZHAO An'an, HE Daliang, WANG Han, GUO Jungang, YE Yinglin (595)
Coupled vibration of underwater dual assembly transfer system
…… YUAN Zhanhang, LI Yunhua, LIU Haodong (602)
Constraint violation suppression for dynamics modeling of lower limb rehabilitation robot
…… XU Yaru, LI Kehong, LIU Jia, LIU Rong, ZHANG Jiancheng (609)
Energy consumption for production of jet fuel precursors from cellulosic biomass by hydrothermal method
…… DONG Shengfei, HUANG Xinghua, YANG Xiaoyi (620)
High-precision positioning method based on SINS/RFID for trains in tunnel
…… ZHANG Chendong, WANG Zhaorui, JIN Shengzhen, AI Guoxiang (632)
A stratospheric wind field interpolation method based on geostrophic wind model
…… HU Kang, ZHOU Jianghua, ZHANG Xiaojun (639)
A cross-modal pedestrian Re-ID algorithm based on dual attribute information
…… CHEN Lin, CAO Zan, SONG Xuemeng, WANG Yinglong, NIE Liqiang (647)
Multi-layer design method of tension rope for paraboloid antenna
…… WANG Hui, HE Tianyu, DU Xianchen, WANG Chunjie (657)
Experimental study on aerodynamic and deformation characteristics of flexible membrane wing
…… WANG Hanbin, HE Xi, WANG Jinjun (665)
Operating status evaluation of smart electricity meters based on joint distribution adaption
…… XU Dan, LI You, LI Helong (674)
Numerical analysis on drag reduction effect of base cavity
…… LI Biao, WANG Liangming, YANG Zhiwei (682)
Coastal GNSS-R ocean altimetry based on code delay of QZSS L1 signal
…… CHU Ti, HE Kaifei, GAO Fan, HE Yunqiao, MENG Xinyue, XU Tianhe (691)
Damage prediction and failure mechanism of composite laminates under high-velocity hailstone impact
…… ZHANG Chao, FANG Xin, LIU Jianchun (698)
Account-based blockchain scalable storage model
…… ZHANG Xuhao, NIU Baoning, GONG Ting (708)
Low-complexity adaptive cubature Kalman filter algorithm
…… LI Chunhui, MA Jian, YANG Yongjian, GAN Yi (716)

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0646

机场跑道水泥混凝土道面板尺寸分析

全下文载

张献民*，李梦晓，陈宇，李长辉，徐博庆
（中国民航大学 交通科学与工程学院，天津 300300）

摘 要：中国机场通常采用板宽不超过5 m的水泥混凝土道面板，国际上已有机场采用板宽6 m以上的水泥混凝土道面板。相比大尺寸水泥混凝土道面板，小尺寸道面板容易产生病害。为了研究道面板尺寸对道面性能的影响，基于弹性层状理论体系建立道面结构有限元模型，通过改变水泥混凝土道面板平面尺寸，研究了考虑温度梯度时飞机荷载作用下道面板内的应力变化。通过某机场实例分析了温度梯度和飞机荷载共同作用下道面板平面尺寸对道面使用寿命的影响，结果表明：道面板内应力受温度梯度影响较大，常见机型荷载作用下，夏季道面板内拉应力25%以上是由温度梯度荷载造成；不同尺寸道面板在相同荷载作用下，道面产生的损伤量不同，设计过程中加入道面板温度应力计算，选择合适的道面板尺寸，可以显著提高水泥混凝土道面的预期使用寿命。

关 键 词：道面板尺寸；温度梯度；临界荷载；累积损伤；疲劳寿命
中图分类号：V351.11
文献标志码：A **文章编号**：1001-5965(2022)04-0551-09

中国现行的民用机场水泥混凝土道面设计规范中没有专门针对道面板平面尺寸计算的部分，国内机场一般依经验选用尺寸为4.5 m×5 m或5 m×5 m宽的道面板。现行规范要求分块宜使主要飞机的主起落架从道面板中部通过，常见民航客机的主起落架间距为5～10 m，飞机滑行时会发生横向偏移现象，采用宽度为4.5 m或5 m的道面板会使飞机轮载更多作用在板缝处。对于常见的45 m和60 m宽水泥混凝土跑道，一般横向至少布置十块道面板，有较多板缝，增加了发生错台的概率，因此，板边和板角也是水泥混凝土道面板最易发生损坏的位置。减少板缝最直接的方式就是增大道面板尺寸，国际上已有机场采用宽度尺寸在6 m以上的水泥混凝土道面板。选用尺寸合适的道面板可提升飞机在道面上滑行的舒适性和安全性，因此，在设计过程中加入对水泥混凝土道面板尺寸的设计具有十分重要的意义。

翁兴中等[1]通过计算不同尺寸和厚度道面板的温度应力，认为同样环境和板厚下，温度应力决定了道面板的尺寸。孙道胜等[2]从混凝土材料入手，研究了在半刚性基层上扩大道面板尺寸的方法。傅明国等[3]结合工程实践，论述了大尺寸水泥混凝土道面板的质量控制措施。黄凯健等[4]在纤维混杂微膨胀混凝土大尺寸面板内部埋设应变计，总结了其应变变化规律。高志斌[5]通过等效荷载的方法，研究了水泥混凝土道面板温度应力与板块尺寸及板厚的关系，建议板长与板厚之比应大于13且小于20。

对水泥混凝土道面板尺寸的研究大多从道面板内部温度应力变化和水泥混凝土材料的角度出

收稿日期：2020-11-20；**录用日期**：2020-12-11；**网络出版时间**：2020-12-15 15:59
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20201215.1557.001.html
基金项目：中国工程院重点资询项目（2019C1-0002）；天津市教委科研计划（2019KJ122）；中国民航机场工程研究基地开放基金（205016010411，205016010413）；中国民航大学科研启动基金（2020KYQD37）
* **通信作者**. E-mail：cauczxm@126.com

引用格式：张献民，李梦晓，陈宇，等. 机场跑道水泥混凝土道面板尺寸分析[J]. 北京航空航天大学学报，2022，48(4)：551-559.
ZHANG X M，LI M X，CHEN Y，et al. Size of cement concrete pavement slab of airport runway[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：551-559 (in Chinese).

北京航空航天大学学报

北航学报 网阅

发,很少进行有关水泥混凝土道面板尺寸和道面疲劳寿命之间关系的研究。因此,本文使用数值模拟的方法,分析不同尺寸道面板在荷载和温度梯度共同作用下道面最大拉应力的变化情况,结合工程实际研究道面板尺寸对道面使用寿命的影响,探讨在水泥混凝土道面板设计过程中,加入温度梯度荷载计算和道面板平面尺寸设计的必要性。

1 有限元模型建立

1.1 道面参数

基于 ABAQU 有限元软件建立跑道全宽为 45 m的足尺实体模型,板块布置形式为 8 种,中线附近主要受荷板宽度分别为 4.5 m、5 m、5.5 m、6 m、6.5 m、7 m、7.5 m 和 8 m,跑道两侧位置道面板采用其他尺寸,以满足不同模型道面的宽度均为 45 m。8 种模型由跑道左侧至右侧横向布置形式分别为:10 块 4.5 m×5 m 板,1 块 7.5 m×5 m 板,6 块 5 m×5 m 板,1 块 7.5 m×5 m 板;1 块 6 m×5.5 m 板,6 块 5.5 m×5.5 m 板,1 块 6 m×5.5 m 板;1 块 4.5 m×6 m 板,6 块 6 m×6 m 板,1 块 4.5 m×6 m 板;2 块 4.75 m×6.5 m 板,4 块 6.5 m×6.5 m 板,2 块 4.75 m×6.5 m 板;1 块 8.5 m×7 m 板,4 块 7 m×7 m 板,1 块 8.5 m×7 m板;6 块 7.5 m×7.5 m 板;1 块 6.5 m×8 m 板,4 块 8 m×8 m 板,1 块 6.5 m×8 m 板。纵向布置三排道面板,三排道面板横向布置方式相同。边界条件为:面层、基层及土基侧面设置 x、y 两方向位移为 0;土基底面设置 x、y、z 三个方向位移为 0。道面板和基层层间法向无黏结,切向采用库伦摩擦方程表征,摩擦系数取值为 1.5,基层与土基完全黏结[6]。传力杆和纵缝拉杆按照现行规范布置,横缝传荷系数为 0.95,纵缝传荷系数为 0.75,其余道面结构参数如表 1 所示[7]。

表 1 道面结构参数[7]

Table 1 Parameters of pavement structure[7]

参数	水泥混凝土面层	水泥稳定碎石基层	土基
厚度/m	0.38	0.40	10
弹性模量/MPa	38 000	1 500	90
密度/(kg·m^{-3})	2 500	2 000	1 800
泊松比	0.15	0.25	0.35

1.2 飞机荷载

飞机荷载采用国内常见的 A320 和 B737-800 机型进行加载。飞机荷载通过机轮传递到道面上,绝大多数荷载由主起落架承担,一般通过主起落架荷载分配系数计算飞机单轮荷载,表达式为

$$P = \frac{\rho G}{m} \tag{1}$$

式中:P 为飞机主起落架单轮荷载;ρ 为主起落架荷载分配系数;G 为飞机荷载,通常采用飞机最大起飞重量;m 为主起落架机轮个数。

一般在机场道面设计中采用组合型轮迹,把轮印假定为 2 个半圆和矩形的组合,为了方便有限元模型的建立和模型计算精度,可根据面积等效原则,把轮印简化为矩形[8],如图 1 所示。轮印面积和尺寸按照式(2)和式(3)计算。A320 和 B737-800 的主起落架构型如图 2 所示,其他飞机计算参数如表 2 所示。

$$A = \frac{P}{1\,000q} \tag{2}$$

$$L = \sqrt{\frac{A}{0.522\,7}} \tag{3}$$

式中:A 为单个飞机轮印面积;q 为飞机轮胎胎压;L 为轮印长度。

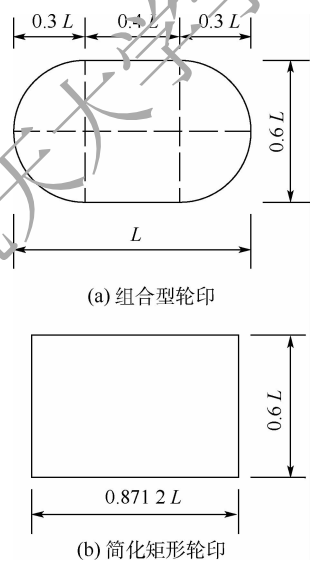


图 1 轮印形状

Fig. 1 Wheel tread profile

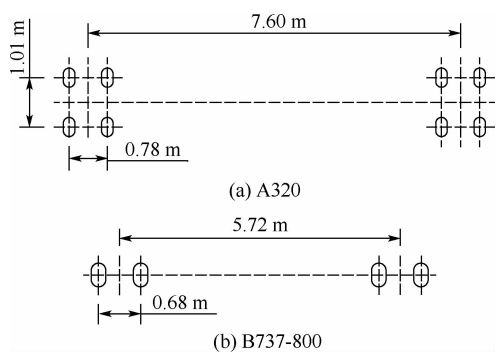


图 2 A320 和 B737-800 主起落架构型

Fig. 2 Main gear configuration of A320 and B737-800

北京航空航天大学学报

表 2　飞机参数

Table 2　Aircraft parameters

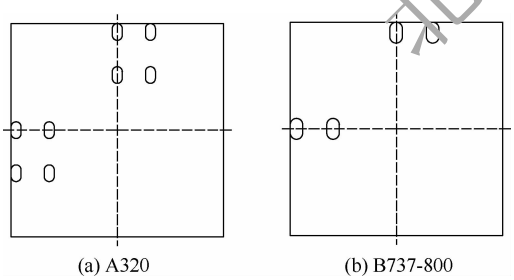
参数	A320	B737-800
最大起飞重量/kN	770.00	790.04
主起落架荷载分配系数	0.95	0.95
主起落架个数	2	2
主起落架间距/m	7.60	5.72
主起落架构型	双轴双轮	单轴双轮
主起落架轮距/m	0.78/1.01	0.86
主起落架胎压/MPa	1.14	1.47
主起落架单轮荷载/kN	91.44	187.63
主起落架轮印面积/m^2	0.080	0.128
主起落架轮印长度/m	0.392	0.494
0.871 2L/m	0.341	0.431
0.6L/m	0.235	0.296

2　道面板应力分析

水泥混凝土道面板内温度应力与板的尺寸相关，尺寸发生变化，温度应力也随之发生改变，采用有限元法计算道面温度应力，结果较为精确[1,9-10]。

温度梯度作用下，水泥混凝土道面板产生翘曲温度应力。中国不同地区自然条件差异较大，水泥混凝土道面温度梯度各不相同，可根据公路自然区划确定水泥混凝土道面温度梯度[11]。道面板产生最大荷载应力的临界荷位在板纵缝或横缝中部，如图 3 所示。

通过有限元软件计算不同公路自然区划 38 cm厚水泥混凝土道面在飞机临界荷载和温度梯度共同作用下，道面板产生的最大拉应力。不同公路自然区划 38 cm 厚水泥混凝土道面板最大温度梯度如表 3 所示，水泥混凝土线膨胀系数为 1×10^{-5}/℃。飞机荷载作用在各尺寸道面板横缝及纵缝临界荷位时的板边最大拉应力如图 4 所示。



(a) A320　(b) B737-800

图 3　临界荷位

Fig. 3　Critical load position

表 3　不同公路自然区划 38 cm 水泥混凝土面层温度梯度

Table 3　Temperature gradient of 38 cm cement concrete pavement with different natural zoning for highways

公路自然区划	Ⅱ、Ⅴ	Ⅲ	Ⅳ、Ⅵ	Ⅶ
温度梯度/(℃ · m^{-1})	52.3	56.7	54.2	58.0

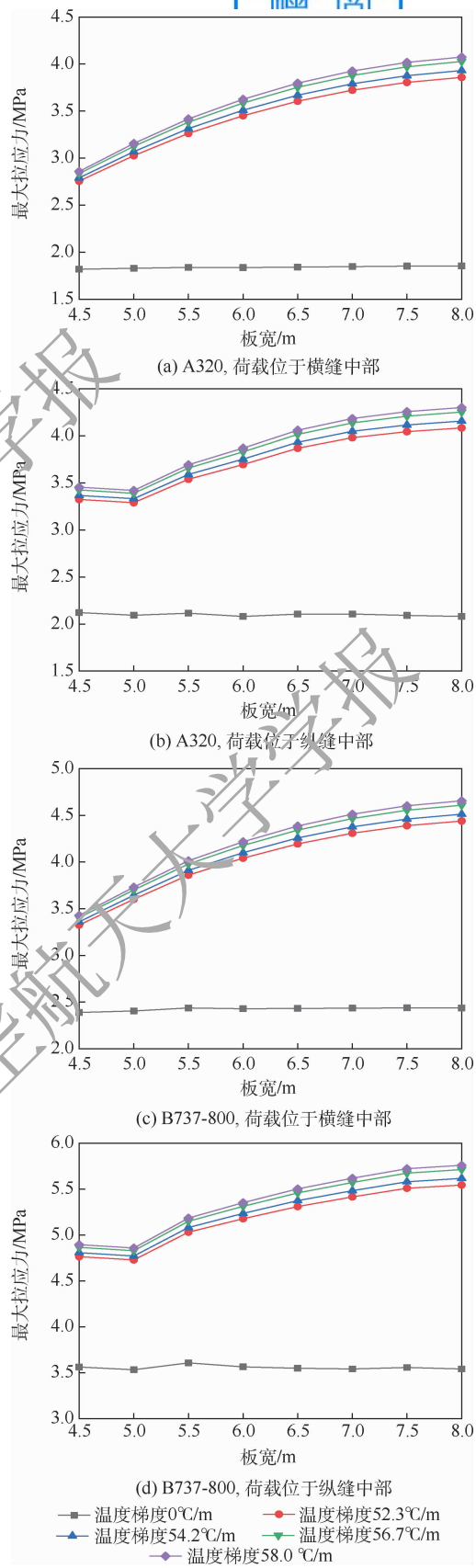


图 4　不同尺寸道面板板边应力曲线

Fig. 4　Edge stress curves of different sizes of pavement slab

无温度梯度时，相同飞机荷载作用下，各尺寸道面板产生的应力基本一致；存在温度梯度

北京航空航天大学学报

北航学报 网阅

时,随着尺寸增大,板内最大拉应力逐渐增加,因为在相同温度梯度作用下,无约束的道面板尺寸越大发生的翘曲变形越大,施加约束后变形大的道面板产生的温度应力更大,所以尺寸大的道面板产生的总应力更大;宽度为 4.5 m 和 5 m 道面板的长度都为 5 m,因此飞机荷载作用在纵缝临界荷位时,这 2 种道面板产生的最大拉应力基本保持一致。因为道面板纵缝传荷能力低于横缝,所以同一机型作用在道面板纵缝产生的最大拉应力大于作用在横缝产生的拉应力。B737-800 主起落架构型为单轴双轮,导致 B737-800 作用在纵缝产生的拉应力远大于横缝产生的拉应力[12]。

不同温度梯度和飞机荷载共同作用下,各尺寸道面板板边应力如表 4 和表 5 所示。4.5 m 宽道面板,在温度梯度和飞机荷载共同作用下,板底最大拉应力为 4.896 MPa,随着板尺寸增大,板底最大拉应力逐渐增大,当道面板宽度达到 8 m 时,板底最大拉应力达到 5.761 MPa。飞行区等级指标Ⅱ在 C 以上的机场,水泥混凝土设计弯拉强度为 5 MPa。在实际情况中,更大的飞机荷载,如 A380 和 B747 等机型作用在道面或更高的温度梯度产生时,道面板底最大拉应力将进一步增大,即使是尺寸较小的 4.5 m 道面板,飞机荷载和温度梯度共同作用时产生的板底最大拉应力将超过 5 MPa。因此,为防止道面板受拉开裂,可考虑在道面中线两侧的主要受荷板布置钢筋或钢丝网,分担道面板产生的拉应力,或增加道面板厚度,从而减小道面板产生的应力。如表 4 和表 5 所示,4.5 m 宽道面板考虑温度梯度和不考虑温度梯度时,飞机临界荷载作用下不同地区道面板产生的最大拉应力差值约在 1 MPa,占总应力的 25% ~ 35%。随道面板尺寸的增大,考虑温度梯度和不考虑温度梯度时产生的最大拉应力差值越来越大,道面板宽度为 8 m 时,差值最大达到 2.447 MPa,占总应力一半以上。因为随着道面板尺寸增加,板内温度应力逐渐增大,但由飞机荷载产生的板内应力并不随道面板尺寸变大而增加,所以随着板尺寸的增加,温度应力占板内应力的比例越来越高。

表 4 A320 飞机荷载作用下不同尺寸道面板板边应力

Table 4 Edge stresses of different sizes of pavement slabs under load of A320 aircraft

板宽/m	温度梯度 0℃/m		温度梯度 52.3℃/m		温度梯度 56.7℃/m		温度梯度 54.2℃/m		温度梯度 58.0℃/m	
	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa
4.5	1.819	2.121	2.756	3.324	2.835	3.425	2.790	3.367	2.858	3.455
5	1.829	2.093	3.026	3.290	3.126	3.390	3.069	3.333	3.156	3.420
5.5	1.839	2.115	3.263	3.539	3.381	3.658	3.314	3.590	3.417	3.693
6	1.837	2.081	3.451	3.695	3.586	3.830	3.509	3.753	3.626	3.870
6.5	1.842	2.105	3.605	3.868	3.753	4.016	3.669	3.932	3.797	4.060
7	1.847	2.106	3.722	3.981	3.879	4.138	3.79	4.049	3.926	4.185
7.5	1.852	2.092	3.805	4.065	3.97	4.21	3.876	4.116	4.018	4.258
8	1.854	2.081	3.858	4.085	4.027	4.254	3.931	4.158	4.074	4.301

表 5 B737-800 飞机荷载作用下不同尺寸道面板板边应力

Table 5 Edge stresses of different sizes of pavement slabs under load of B737-800 aircraft

板宽/m	温度梯度 0℃/m		温度梯度 52.3℃/m		温度梯度 56.7℃/m		温度梯度 54.2℃/m		温度梯度 58.0℃/m	
	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa
4.5	2.388	3.462	3.325	4.665	3.404	4.766	3.359	4.708	3.427	4.796
5	2.404	3.432	3.601	4.629	3.701	4.729	3.644	4.672	3.731	4.759
5.5	2.436	3.507	3.860	4.931	3.979	5.050	3.911	4.982	4.014	5.085
6	2.428	3.464	4.042	5.078	4.177	5.213	4.10	5.136	4.217	5.253
6.5	2.431	3.448	4.194	5.211	4.342	5.359	4.258	5.275	4.386	5.403
7	2.434	3.441	4.309	5.316	4.466	5.473	4.377	5.384	4.513	5.520
7.5	2.437	3.457	4.390	5.410	4.555	5.575	4.461	5.481	4.603	5.623
8	2.436	3.441	4.440	5.445	4.609	5.614	4.513	5.518	4.656	5.661

北京航空航天大学学报

北航学报 赠阅

3　道面寿命计算

机场道面在荷载长期作用下有一定的使用寿命,中国现已建成机场在设计时大多采用“设计飞机”法计算道面飞机荷载作用次数,计算时将各不同机型换算成一种“设计飞机”对道面重复作用,假定轮迹在通行宽度内均匀分布。然而,不同机型主起落架构型不同、轮迹横向分布不同,并且飞机轮迹横向分布近似服从正态分布,采用“设计飞机”近似换算的方法会导致计算结果不精确[13-15]。基于 Miner 损伤原理的累积损伤法,进行计算时叠加道面在不同机型作用下产生的损伤,充分考虑不同机型实际对道面造成的损伤情况,计算结果比“设计飞机”法更为准确[16-17]。

3.1　覆盖作用次数

飞机轮迹横向分布近似服从正态分布,国内学者将飞机主起落架轮迹横向分布叠加,得到不同情况下轮迹覆盖作用曲线[18]。

当飞机主起落架构型为单轴单轮时,若两侧主起落架距离大于通行宽度,则两轮迹不重叠,轮迹覆盖作用曲线函数表达式为

$$F_1(x)=\begin{cases}f_1(x)=\dfrac{1}{\sigma_0\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{x+S/2}{\sigma_0}\right)^2} & x<0\\ f_2(x)=\dfrac{1}{\sigma_0\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-S/2}{\sigma_0}\right)^2} & x\geqslant 0\end{cases} \tag{4}$$

式中:$F_1(x)$ 为单轴单轮飞机主起落架轮迹不重叠时的轮迹覆盖作用曲线;x 为道面某一点距跑道中线的垂直距离,沿飞机前进方向左侧为负、右侧为正;$f_1(x)$、$f_2(x)$ 分别为单轴单轮飞机左、右起落架轮迹横向分布曲线;σ_0 为飞机滑行时横向偏移标准差;S 为主起落架间距。

当飞机主起落架构型为单轴单轮时,若两侧主起落架距离小于通行宽度,则左、右主起落架轮迹重合,轮迹横向分布曲线为左、右轮迹分布曲线叠加,函数表达式为

$$F_2(x)=f_1(x)+f_2(x)=\frac{1}{\sigma_0\sqrt{2\pi}}\left[\mathrm{e}^{-\frac{1}{2}\left(\frac{x+S/2}{\sigma_0}\right)^2}+\mathrm{e}^{-\frac{1}{2}\left(\frac{x-S/2}{\sigma_0}\right)^2}\right] \tag{5}$$

式中:$F_2(x)$ 为单轴单轮飞机主起落架轮迹重叠时的轮迹覆盖作用曲线。

当飞机主起落架构型为单轴双轮时,主起落架距离一般大于通行宽度,两侧主起落架轮迹不重叠,计算轮迹覆盖作用曲线时,只需将单侧主起落架 2 个轮迹进行线性叠加。

$$f_{21}(x)=\frac{1}{\sigma_0\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left[\frac{x-(S-S_t)/2}{\sigma_0}\right]^2} \tag{6}$$

$$f_{22}(x)=\frac{1}{\sigma_0\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left[\frac{x-(S+S_t)/2}{\sigma_0}\right]^2} \tag{7}$$

$$F_3(x)=f_{21}(x)+f_{22}(x)=\frac{1}{\sigma_0\sqrt{2\pi}}\left[\mathrm{e}^{-\frac{1}{2}\left[\frac{x-(S+S_t)/2}{\sigma_0}\right]^2}+\mathrm{e}^{-\frac{1}{2}\left[\frac{x-(S-S_t)/2}{\sigma_0}\right]^2}\right] \tag{8}$$

式中:$F_3(x)$ 为单轴双轮飞机右侧轮迹覆盖作用曲线;$f_{21}(x)$、$f_{22}(x)$ 分别为右侧起落架内、外机轮的轮迹横向分布作用曲线;S_t 为轮距。

对于多轴单轮、多轴多轮或特殊构型起落架的覆盖作用曲线,可以根据单轮和双轮的覆盖作用曲线叠加确定。

飞机荷载对道面板边产生一次最大应力作用,称对该点进行了一次覆盖,道面覆盖作用次数按式(9)计算[13,19]:

$$N_i(x)=n_iF_i(x)W_i \tag{9}$$

式中:$N_i(x)$ 为第 i 种机型的覆盖作用次数;n_i 为第 i 种机型运行次数;$F_i(x)$ 为第 i 种机型的轮迹覆盖作用曲线;W_i 为第 i 种飞机的轮印宽度。

3.2　温度梯度确定

由第 2 节内容可知,道面板因温度梯度产生的温度应力不能忽视,因此在计算道面使用寿命时,应加入温度梯度荷载。因为道面板不是总处于温度梯度最大的极限状态,大多数机场会在夜间停航,所以在计算道面板使用寿命时,可采用道面板白天平均温度梯度进行应力计算。一天之内道面板温度梯度在午后达到正温差最大,温度梯度变化为正弦函数[10,20],白天温度梯度变化如式(10)所示,可根据式(11)计算出白天平均温度梯度。

$$T_g(t)=T_{g,\max}\sin(\pi t/12) \tag{10}$$

$$T_{ga}=\frac{2}{\pi}T_{g,\max} \tag{11}$$

式中:$T_g(t)$ 为白天温度梯度随时间变化函数;t 为时刻;$T_{g,\max}$ 为公路自然区划中规定的水泥混凝土道面板最大温度梯度;T_{ga} 为水泥混凝土道面板白天平均温度梯度。

3.3　允许作用次数确定

道面允许作用次数就是道面的疲劳寿命,根据道面板荷载应力和疲劳方程可求出道面允许作用次数。

对荷载作用下的道面板进行应力分析,将道面纵向各点的最大拉应力映射到横向边,板内横向最大应力分布可简化为分段函数,单块道面板 x 位置的应力为道面板宽度 B 的周期函数见

北京航空航天大学学报

式(12),应力分布情况如图 5 所示[18-19]。

$$\sigma(x) = \begin{cases} \sigma_Y - \dfrac{(\sigma_Y - \sigma_X)x}{0.15B} & 0 \leqslant x < 0.15B \\ \sigma_X & 0.15B \leqslant x \leqslant 0.85B \\ \sigma_X + \dfrac{(\sigma_Y - \sigma_X)(x - 0.85B)}{0.15B} & 0.85B < x \leqslant B \end{cases} \tag{12}$$

式中:$\sigma(x)$为距纵缝 x 处板边应力;σ_X、σ_Y 分别为温度梯度和飞机临界荷载共同作用下道面板横、纵向边最大拉应力,MPa;B 为道面板宽度。

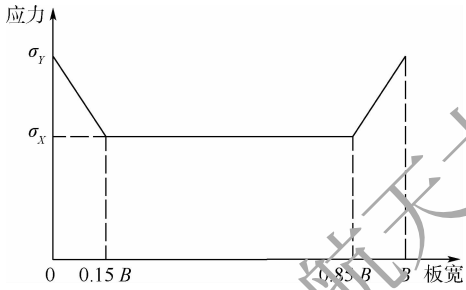


图 5 板边最大应力分布[18-19]

Fig. 5 Distribution of maximum stress at the edge of pavement slab[18-19]

美国 NCHRP1-26 疲劳方程采用有限元的计算方法,充分融合实验和理论研究成果,考虑了荷载和温度的作用,与实际状态接近。荷载应力通过有限元软件模拟得到,同时考虑板内温度应力,可采用 NCHRP1-26 疲劳方程,根据式(13)计算道面板在第 i 种机型作用下位置 x 处的疲劳寿命,即该位置的允许作用次数[21]。

$$N_{ri}(x) = \begin{cases} 10^{-1.713\,6u + 4.284} & u > 1.25 \\ 10^{2.812\,7u - 1.221\,4} & u \leqslant 1.25 \end{cases} \tag{13}$$

$$u = \frac{\sigma(x)}{f_r} \tag{14}$$

式中:$N_{ri}(x)$为道面板设计寿命期内第 i 种机型的允许作用次数;f_r 为道面设计弯拉强度;u 为应力比。

3.4 累积损伤量

根据 Miner 定理,机场水泥混凝土道面在不同机型荷载作用下,受到的累积损伤可以线性相加,得到累积损伤曲线。在第 i 种机型作用下,飞机荷载对水泥混凝土道面的疲劳消耗是该机型覆盖作用次数 N_i 与允许作用次数 N_{ri} 的比值,将不同机型对道面板的疲劳消耗进行叠加,得到道面横向各位置的累积损伤因子(cumulative damage factor, CDF)[19,21-22],表达式为

$$\mathrm{CDF}(x) = \sum_{i=1}^{k} \frac{N_i(x)}{N_{ri}(x)} \tag{15}$$

根据各位置 CDF 值大小可判断道面的损伤情况。如果 CDF < 1,说明道面未达到疲劳破坏极限状态;如果 CDF ≥ 1,说明道面达到疲劳破坏极限状态。

3.5 实例计算

为研究道面板尺寸对道面寿命的影响,分别计算某机场道面板宽度为 4.5 m、5 m、5.5 m、6 m、6.5 m、7 m、7.5 m 和 8 m 时的道面板累积损伤量。

某已建成机场飞行区等级 4F,道面结构参数为:水泥混凝土道面板厚度为 38 cm,尺寸为 4.5 m × 5 m,弹性模量为 38 000 MPa,泊松比为 0.15,设计弯拉强度为 5.5 MPa,水泥稳定碎石基层厚度为 40 cm,弹性模量为 1 500 MPa,土基回弹模量为 90 MPa。该机场的主要机型为 A320 和 B737-800,近 5 年的年均起降量分别为 10 454 架次和 22 871 架次,其对应轮迹横向分布函数分别为

$$F_{\mathrm{A320}}(x) = \frac{2}{1.2\sqrt{2\pi}}\left\{ \mathrm{e}^{-\frac{[(x-3.41)/1.2]^2}{2}} + \mathrm{e}^{-\frac{[(x-4.19)/1.2]^2}{2}} \right\} \tag{16}$$

$$F_{\mathrm{B737\text{-}800}}(x) = \frac{1}{1.2\sqrt{2\pi}}\left\{ \mathrm{e}^{-\frac{[(x-2.43)/1.2]^2}{2}} + \mathrm{e}^{-\frac{[(x-3.29)/1.2]^2}{2}} \right\} \tag{17}$$

该地区 38 cm 厚水泥混凝土道面板最大温度梯度为 52.3℃/m,根据式(13)可计算出白天平均温度梯度为

$$T_{\mathrm{ga}} = 2/\pi \times 52.29 = 33.29℃/\mathrm{m} \tag{18}$$

将飞机临界荷载和 33.29℃/m 温度梯度共同作用在道面板上,得到不同板块布置方式的道面板板边应力,如表 6 所示。道面温度梯度在全年随时间发生周期性变化,计算所采用的温度梯度为全年温度梯度最大值,一般出现在七月,假设一个月内温度梯度为定值,则道面板该月损伤量为全年最大值。

表 6 道面板板边最大拉应力

Table 6 Maximum tensile stress at the edge of pavement slab

板宽/m	A320		B737-800	
	横向边应力/MPa	纵向边应力/MPa	横向边应力/MPa	纵向边应力/MPa
4.5	2.581	3.106	3.150	4.447
5	2.807	3.071	3.382	4.410
5.5	3.008	3.284	3.605	4.676
6	3.169	3.413	3.760	4.796
6.5	3.304	3.567	3.893	4.910
7	3.408	3.667	3.995	5.002
7.5	3.483	3.723	4.068	5.088
8	3.531	3.758	4.113	5.118

假设该机场每月航班数量一致，则该机场单月主要机型起降架次为：A320 起降 968 次，B737-800 起降 1 906 次，根据式(9)、式(12)、式(13)、式(14)和式(15)计算得到该机场采用不同板块布置方式时温度梯度最大月份的道面单月损伤量，对道面板宽度进行归一化处理，得到中线右侧道面板夏季损伤曲线，如图 6 所示。道面板宽度为 4.5 m 时，道面板最大损伤位置为远离跑道中线的外侧纵向边，其次是靠近跑道中线的内侧纵向边，相比之下板块中部损伤不明显。随着板宽增加，外侧纵向边的损伤量逐渐减小，道面板横向边中部和内侧纵向边的损伤量逐渐增大。板宽达到 6 m 时，道面板横向边中部和内侧纵向边的损伤量超过外侧纵向边。当板宽继续增加到 6.5 m 后，横向边中部超过内侧纵向边的损伤量，成为道面板损伤最严重的部位。因为同一飞机荷载作用在道面板纵向边临界位置时产生的最大拉应力大于作用在横向边临界位置时产生的最大拉应力，且飞机在道面上滑行时，轮迹近似呈正态分布，A320 机型主起落架间距为7.6 m，B737-800 机型主起落架间距为 5.62 m，所以当道面板尺寸较小时，轮载更接近道面板荷载应力更大的外侧纵向边，造成外侧纵边损伤较大。随着道面板尺寸增大，板内温度应力逐渐增加，因此，道面板横向边中部和内侧纵向边的损伤量随着尺寸的增大而不断增加。但由于尺寸增大的过程中，轮载作用位置由外侧逐渐向板中移动，外侧纵向边覆盖作用次数减少，即使板内温度应力增加，外侧纵向边损伤量依旧逐渐减小。因为轮载覆盖作用位置主要在板中，所以当板内应力随尺寸增大而增加时，横向边中部比内侧纵向边的损伤量增长的更快，最终横向边中部的损伤量超过内侧纵向边中部，成为道面板损伤最严重的位置。

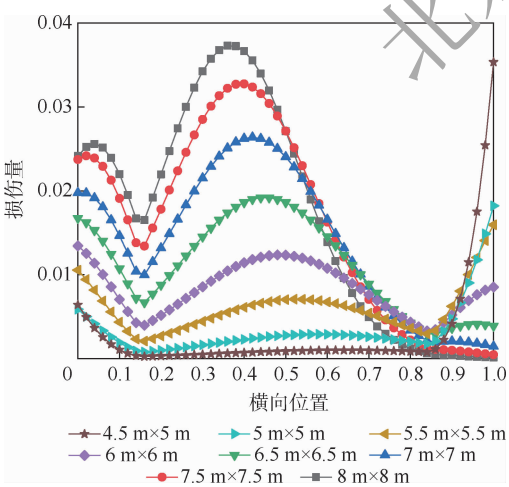


图 6　不同尺寸道面板单月最大损伤量

Fig. 6　Maximum damage of different sizes of pavement slabs in single month

以此可见，相同飞机荷载和温度梯度作用下，道面板尺寸不同的道面，产生的最大损伤量及其损伤位置都不相同。

各板型最大损伤量和损伤位置如表 7 所示。跑道中线两侧道面板宽度为 4.5 m、5 m 和 5.5 m 时，道面最大损伤位置为中线两侧板块远离中线的纵向边；中线两侧道面板宽度为 6 m 时，最大损伤位置为中线两侧板块靠近中线的纵向边；中线两侧道面板宽度为 6.5 m、7 m、7.5 m 和 8 m 时，最大损伤位置为中线两侧板块的横向边中部。该机场采用的 4.5 m×5 m 板块温度梯度最大月份的单月损伤量为 0.035，相同荷载作用下温度梯度最大月份产生损伤最小的板块是 6 m×6 m，单月损伤量为 0.013，比该机场原板块损伤降低了 62.9%，其他板块形式的损伤量相较于原板块均有不同程度的改变，说明选择合适的道面板尺寸有助于提升场道寿命。值得指出的是，提升道面寿命的方法有很多种，可以从增加道面配筋、提升道面板厚度、提高地基强度等多方面进行考虑，本文仅从改变道面板平面尺寸的角度分析提升道面寿命的方法。

表 7　不同尺寸道面板最大损伤量和损伤位置

Table 7　Maximum damage and its occurrence position of different sizes of pavement slabs

板宽/m	最大损伤位置/m	损伤量
4.5	4.5	0.035
5	5.0	0.018
5.5	5.5	0.015
6	0.0	0.013
6.5	2.9	0.019
7	2.9	0.026
7.5	2.9	0.033
8	2.9	0.037

4　结　论

本文采用 ABAQUS 有限元软件进行数值模拟，研究分析了不同温度梯度和飞机荷载下，不同尺寸道面板的应力分布、疲劳寿命和道面损伤量，得出如下结论：

1）随道面板尺寸增大而增加的板内应力，增加部分是因温度梯度而产生的温度应力，道面板尺寸越大，温度应力占板内总应力的比例越大。

2）道面板在夏季较大温度梯度和飞机荷载共同作用下，板底最大拉应力可能超过道面设计弯拉强度，建议在道面中线两侧的主要受荷板布

北航学报 瞰阅

置钢筋或钢丝网分担道面板产生的拉应力,或增加道面板厚度,从而减小道面板产生的应力。

3)相同飞机荷载和温度梯度作用下,不同尺寸道面板产生的最大损伤量及其损伤位置各不相同,这是不同尺寸道面板轮迹覆盖作用间的差异,以及各道面板产生的温度应力不同所导致的。

4)同一机场采用不同尺寸道面板时道面的使用寿命差异明显,建议在水泥混凝土道面设计过程中加入道面板温度应力计算部分,根据道面所在地区,结合不同月份水泥混凝土面板温度梯度和机场预期起降机型及其架次,选择合适的道面板尺寸,从而提升水泥混凝土道面的预期使用寿命。

参考文献(References)

[1] 翁兴中,谭麦秋,黄小明,等. 机场水泥混凝土道面板尺寸的确定方法[J]. 空军工程大学学报(自然科学版),2002,3(6):4-6.
WENG X Z, TAN M Q, HUANG X M, et al. A calculating method of size of large cement concrete slab[J]. Journal of Air Force Engineering University(Natural Science Edition),2002,3(6):4-6(in Chinese).

[2] 孙道胜,黄小明,邓敏,等. 道面水泥混凝土高性能化和放大道面板尺寸问题初探[J]. 混凝土,2004(5):21-25.
SUN D S,HUANG X M,DENG M,et al. The discuss of pursuing high performance of the pavement cement concrete and magnifying dimension of the pavement slab[J]. Concrete,2004(5):21-25(in Chinese).

[3] 傅明国,陈岩,黄业坤,等. 机场道面大尺寸板块混凝土配合比设计与施工技术[J]. 施工技术,2007,36(2):103-106.
FU M G,CHEN Y,HUANG Y K,et al. Mixture ratio design and construction technology of large size plate concrete at airport pavement[J]. Construction Technology,2007,36(2):103-106(in Chinese).

[4] 黄凯健,邓敏,何大贵,等. 高原地区纤维混凝土微膨胀混凝土大尺寸面板[J]. 建筑材料学报,2010,13(4):474-477.
HUANG K J,DENG M,HE D G,et al. Big dimensional pavement slabs with hybrid fiber reinforced micro expansion concrete in plateau[J]. Journal of Building Materials,2010,13(4):474-477(in Chinese).

[5] 高志斌. 温度应力对机场道面板厚度与分块设计影响研究[J]. 建筑结构,2011,41(1):1068-1072.
GAO Z B. Study on the impact of temperature stress on the airport pavement panel thickness and block design[J]. Building Structure,2011,41(1):1068-1072(in Chinese).

[6] 张献民,刘小兰,张子文. 基于振动特性的机场刚性道面传荷性能[J]. 北京航空航天大学学报,2018,44(9):1787-1796.
ZHANG X M,LIU X L,ZHANG Z W. Rigid pavement load transfer ability of airport based on vibration characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics,2018,44(9):1787-1796(in Chinese).

[7] 中国民用航空局. 民用机场水泥混凝土道面设计规范:MH/T5004—2010[S]. 北京:人民交通出版社,2010:23-24.
CACC. Specifications for airport cement concrete pavement design:MH/T5004—2010[S]. Beijing:China Communications Press,2010:23-24(in Chinese).

[8] 刘文,凌建明,赵鸿铎. 考虑接缝影响的机场水泥混凝土道面结构响应[J]. 公路交通科技,2007,24(12):15-18.
LIU W,LING J M,ZHAO H D. Analysis on mechanical responses of rigid airport pavement with load transfer of joints[J]. Journal of Highway and Transportation Research and Development,2007,24(12):15-18(in Chinese).

[9] 蒋应军,戴经梁. 刚性路面中关于温度应力计算问题[J]. 岩土工程学报,2007,29(6):837-842.
JIANG Y J,DAI J L. Computation of thermal stresses in rigid pavement[J]. Journal of Geotechnical Engineering,2007,29(6):837-842(in Chinese).

[10] 李新凯,侯相深,马松林. 轴载与温度作用下水泥路面板的变形与应力分析[J]. 重庆建筑大学学报,2008,30(5):53-57.
LI X K,HOU X S,MA S L. Deflection and stress analysis of a pavement slab under temperature gradient and axial loading coupling[J]. Journal of Chongqing Jianzhu University,2008,30(5):53-57(in Chinese).

[11] 翁兴中. 机场道面设计[M]. 3 版. 北京:人民交通出版社,2017:34-39.
WENG X Z. Design of airport pavement[M]. 3rd ed. Beijing:China Communications Press,2017:34-39(in Chinese).

[12] KANDEH D S,罗君豪. 典型飞机起落架荷载作用下机场水泥砼道面结构响应分析[J]. 公路与汽运,2019(6):55-57.
KANDEH D S,LUO J H. Analysis of airport cement concrete pavement structure response under the load of typical aircraft landing gear[J]. Highways and Automotive Applications,2019(6):55-57(in Chinese).

[13] 吴爱红,蔡良才,顾强康,等. 适应未来大型飞机的水泥混凝土道面设计方法[J]. 北京航空航天大学学报,2011,37(9):1169-1175.
WU A H,CAI L C,GU Q K,et al. Airport concrete pavement design for large aircraft in future[J]. Journal of Beijing University of Aeronautics and Astronautics,2011,37(9):1169-1175(in Chinese).

[14] 孙建斌,翁兴中. 军用机场沥青混凝土道面结构设计指标确定[J]. 空军工程大学学报(自然科学版),2005,6(1):8-10.
SUN J B,WENG X Z. Research on structure design criteria of asphalt concrete pavement for military airport[J]. Journal of Air Force Engineering University(Natural Science Edition),2005,6(1):8-10(in Chinese).

[15] BRILL D R,PARSONS I D. Three-dimensional finite element analysis in airport pavement design[J]. International Journal of Geomechanics,2001,1(3):273-290.

[16] 杨光. 寒区机场水泥混凝土道面剩余寿命评估预测研究[D]. 哈尔滨:哈尔滨工业大学,2017.
YANG G. Research on evaluation and prediction of residual life of cement concrete pavement of airport in cold region[D]. Harbin:Harbin Institute of Technology,2017(in Chinese).

北京航空航天大学学报

[17] TODINOV M T. Necessary and sufficient condition for additivity in the sense of the palmgren-miner rule[J]. Computational Materials Science,2001,21(1):101-110.

[18] 蔡良才,王海服,张罗利,等.基于累积损伤的机场道面剩余寿命预测模型[J].交通运输工程学报,2014,14(4):1-6.
CAI L C,WANG H F,ZHANG L L,et al. Prediction model of remaining life for airport pavement based on cumulative damage [J]. Journal of Traffic and Transportation Engineering,2014,14(4):1-6(in Chinese).

[19] 王振辉,蔡良才,顾强康,等.考虑荷载应力分布的机场刚性道面累积损伤优化模型[J].土木工程学报,2011,44(11):143-150.
WANG Z H,CAI L C,GU Q K,et al. Airport rigid pavement cumulative damage optimization model considering load stress distribution[J]. China Civil Engineering Journal,2011,44(11):143-150(in Chinese).

[20] 纪伦,张锐,徐慧宁,等.季冻区道路结构温度场分布特性研究[J].中外公路,2013,33(3):57-63.
JI L,ZHANG R,XU H N,et al. Research on temperature field distribution characteristics of road structure in seasonal frozen area[J]. Journal of China and Foreign Highway,2013,33(3):57-63(in Chinese).

[21] 李巧生,赵鸿铎,凌建明.适应大型特种飞机的机场水泥混凝土道面结构设计方法[J].土木工程学报,2011,44(1):121-126.
LI Q S,ZHAO H D,LING J M. Method of airport concrete pavement design for large military transport aircrafts[J]. China Civil Engineering Journal,2011,44(1):121-126(in Chinese).

[22] RISITANO A,RISITANO G. Cumulative damage evaluation in multiple cycle fatigue tests taking into account energy parameters[J]. International Journal of Fatigue,2013,48(48):214-222.

Size of cement concrete pavement slab of airport runway

ZHANG Xianmin*, LI Mengxiao, CHEN Yu, LI Changhui, XU Boqing

(School of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Chinese airports usually use cement concrete road slabs with a slab width of no more than 5 m, and existing airports in other countries use cement concrete slabs with a width more than 6 m. Small size cement concrete pavement slabs are more likely to suffer from disease than large size. In order to study the effect of pavement slab size on pavement performance, a finite element model of pavement structure was established based on the elastic layered theoretical system. By changing the size of the pavement slab, the stress changes of the pavement slab under the combined action of aircraft load and temperature gradient are studied. The internal stress change of the slab is used to analyze the influence of the plane size of the pavement slab on the service life of the pavement slab under the combined action of temperature gradient and aircraft load through an example of real airport. The results show that the internal stress of the pavement slab is greatly affected by the temperature gradient. Under the action of common aircraft load, more than 25% of the internal tensile stress of the pavement slab in summer is caused by the temperature gradient load; under the same load, the damage of different sizes of pavement slabs is different, and choosing the appropriate pavement size can significantly increase the expected service life of the cement concrete pavement.

Keywords: pavement slab size; temperature gradient; critical load; cumulative damage; fatigue life

Received: 2020-11-20; **Accepted**: 2020-12-11; **Published online**: 2020-12-15 15:59

URL: kns.cnki.net/kcms/detail/11.2625.V.20201215.1557.001.html

Foundation items: Key Consulting Research Project of China Academy of Engineering (2019C1-0002); Tianjin Municipal Education Commission Scientic Research Project (2019KJ122); Scientific Research Foundation of Civil Airport Engineering Research Base of China (205016010411, 205016010413); Scientific Research Foundation of Civil Aviation University of China (2020KYQD37)

* **Corresponding author**. E-mail: cauczxm@126.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0649

基于自编码器和 HMM 的民机接地点远事件检测

全下文载

霍纬纲*，李继龙，王慧芳
（中国民航大学 计算机科学与技术学院，天津 300300）

摘 要：已有的飞行品质监控(FOQA)标准仅由地速的积分距离判定接地点远事件(LTE)的发生，无法结合多个快速存取记录器(QAR)参数检测并解释发生该事件的原因。通过滑动窗口对包含多个参数的 QAR 样本进行分段，按照分段位置形成若干分段样本集。由长短时记忆网络(LSTM)自编码器得到 QAR 样本分段及分段内向量的特征表示，采用 K-means 分别对这些表示向量进行聚类处理，实现 QAR 样本及其分段的符号化。使用正常航班 QAR 样本集的符号序列建立隐马尔可夫模型(HMM)，以检测包含接地点远事件的航班。由发生及未发生接地点远事件的 QAR 样本各个分段的符号序列构建 HMM 后，采用 Viterbi 算法确定接地点远事件在 QAR 样本分段内的具体位置。在真实 QAR 数据集上的实验结果表明，与其他多维时间序列异常检测方法相比，所提方法不仅能有效检测接地点远事件，而且可以获得多个 QAR 参数的异常值，辅助领域专家分析发生该事件的原因。

关 键 词：异常检测；接地点远事件(LTE)；自编码器；隐马尔可夫模型(HMM)；快速存取记录器(QAR)
中图分类号：V328.1；TP183
文献标志码：A **文章编号**：1001-5965(2022)04-0560-09

飞机的着陆过程大致可分为下滑接地和接地后的减速滑跑 2 个阶段。下滑接地是飞机经跑道入口从离地垂直高度 50 ft 点(1 ft(英尺) = 0.304 8 m)，以进场速度开始进场，经过下滑拉平至主轮着地的阶段。飞机在下滑接地阶段经过的水平距离称为接地距离。接地点远事件(long touchdown exceedance，LTE)是指飞行着陆过程中的接地距离大于规定范围的超限事件。该事件是造成飞机冲出机场跑道的重要因素之一。目前，民航业内的飞行品质监控(flight operation quality assurance，FOQA)工作仅依据飞行下滑接地期间地速的积分距离判定接地点远事件的发生，无法结合多个快速存取记录器(quick access recorder，QAR)参数取值检测并分析发生接地点远事件的原因。以单个 QAR 参数进行超限事件检测效率较高，但是很容易出现由于相应参数记录值缺失而无法判断超限事件情形，或由于数据噪声而导致“假事件”现象。一方面，实际上在某一超限事件发生时，会有若干个相关联的 QAR 参数都有异常变化[1]，如果能综合运用多个 QAR 参数进行超限事件检测，便可以降低发生如上问题的机率。另一方面，现有的监控标准大多来自飞机制造商提供的各种手册或航空公司内部规定，这些标准大多只考虑了一般情况。然而飞机的飞行过程会受到运行环境、飞机自身机械状况、飞行员的驾驶水平等众多因素影响。在天气异常、或特殊机场等情形下，飞行员可能必须采取“大尺度”操纵，此类操纵也容易被误判为超限事件。QAR 记录了大量飞行参数的变化规律，反映了飞机运行环境及飞行员对各种事件的反应及处理过程。因此，若能从多个

收稿日期：2020-11-23；**录用日期**：2021-01-22；**网络出版时间**：2021-02-19 11:43
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210219.1138.001.html
基金项目：中央高校基本科研业务费专项资金（3122019190）
* **通信作者**. E-mail：wghuo@cauc.edu.cn

引用格式：霍纬纲，李继龙，王慧芳. 基于自编码器和HMM的民机接地点远事件检测[J]. 北京航空航天大学学报，2022，48(4)：560-568. HUO W G，LI J L，WANG H F. Civil aircraft long touchdown exceedance detection based on autoencoder and HMM [J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：560-568 (in Chinese).

北京航空航天大学学报

北航学报 增阅

QAR 参数的角度检测并解释超限事件,将有助于进一步提高 FOQA 的管理水平。

为提高航空安全管理水平,近年来研究人员围绕民航业内积累的大量 QAR 数据开展了许多研究工作。文献[2]通过对 QAR 数据聚类分析挖掘隐含的着陆阶段以油门和杆位表达的飞行操作模式,分析 FOQA 指标值与挖掘的操作模式之间的关联关系,量化飞行操作模式的风险水平。文献[3]分析了飞机着陆阶段拉平操作过程中 QAR 参数取值的方差特征,在此基础上采用回归模型分析了拉平操作对接地距离及重着陆事件的影响。文献[4]基于正态云理论建立了飞行员着陆操作风险评价模型。上述研究都是从安全风险评估的角度开展工作[2-4],这些模型涉及的 QAR 参数个数相对较少。另外机器学习算法已被应用于从海量多维 QAR 数据集中发现异常飞行事件[5-10]。文献[5-6]将多维时序 QAR 数据转化为高维向量,由基于密度的噪点空间聚类(density-based spatial clustering of applications with noise, DBSCAN)聚类算法检测异常航班。文献[7]应用聚类技术识别与飞行风险相关的 QAR 参数。文献[8]通过采样技术将单个 QAR 数据转化为多个时序向量,由基于高斯混合模型的聚类算法检测不安全事件。文献[9]采用向量自回归模型表示每个航班的 QAR 数据,基于回归模型参数计算航班之间的距离矩阵,由局部离群因子(local outlier factor, LOF)检测算法检测异常航班。文献[10]提出了融合半马尔可夫和向量自回归模型的飞行安全隐患检测方法。文献[5-8]中的模型训练前对 QAR 数据的特征提取方式有可能丢失检测飞行安全隐患所需的关键信息,不能较好地捕获 QAR 参数值的时序特征和参数之间的耦合关系。文献[9-10]的向量自回归模型仅能表达 QAR 参数之间的线性关系,且该类模型对噪声也比较敏感。

隐马尔可夫模型(hidden Markov model, HMM)为结构最简单的动态贝叶斯网,主要用于时序数据的建模。基于 HMM 的时间序列的异常检测方法一般主要包含 2 个重要步骤:①符号化时间序列;②参数学习与概率估计。符号化时,旨在以字符串序列表示原始时间序列,不仅可以达到数据降维的目的,而且符合 HMM 对观测序列的要求。文献[11]将 HMM 应用于飞机着陆操作的异常检测中,采用 K-means 聚类算法将原始时间序列转化成由 K 个簇标记表示的符号序列。文献[12]将 HMM 应用在多维时间序列上的异常检测中,分别采用模糊 C 均值(fuzzy C-means, FCM)聚类算法和模糊积分技术将多维时间序列转换成单维的符号序列,提高 HMM 的异常检测能力。文献[13]的研究表明如果在符号化过程之前不对原始序列进行特征提取,序列中噪声会对异常检测效果造成影响,提出了基于感知重要点技术的符号化方法,但其需要计算每个点对于时间序列的影响力,计算复杂度较高。文献[11-12]在符号化过程中均没有对原始序列进行特征提取。自编码器是一种由编码器和解码器构成的无监督学习算法,能从大量无标记的数据中学习数据的有效信息,实现对输入数据的非线性压缩和重构。文献[14]提出了基于自编码器和 HMM 的多维时序数据异常检测模型,由自编码器生成多维时间序列的低维特征表示,对这些特征表示聚类处理,实现多维时间序列的符号化,研究表明该模型能显著提高 HMM 在多维时序数据上的异常检测效果,但自编码器无法表达多维时序数据的时态依赖关系。

本文提出了一种基于长短时记忆网络自编码器(long short term memory networks auto encoder, LSTM-AE)[15]和 HMM 的接地点远事件检测方法(long touchdown exceedance detection method based on LSTM-AE and HMM, LTED-LSTM-HMM),该方法采用 LSTM-AE 学习多维 QAR 数据的特征表示,获取得的特征表示能更好地表达 QAR 数据的时序信息,并由此建立了基于 HMM 的接地点远事件检测模型,实验表明了本文方法的有效性。

1　LTED-LSTM-HMM 方法

LTED-LSTM-HMM 方法的具体流程如图 1 所示。从 QAR 译码文件中截取飞机着陆阶段相关 QAR 参数取值,将生成的数据集合划分为训练集、验证集和测试集,其中训练集只包含未发生接地点远事件的 QAR 样本,测试集和验证集包含发生和未发生接地点远事件的 QAR 样本。本文方法首先利用滑动窗口将所有 QAR 样本按固定分段数目进行分段,按分段位置形成若干 QAR 片段样本集。由训练集不同位置的 QAR 片段样本训练各个分段的 LSTM 自编码器网络,从而得到 QAR 样本各个片段的低维特征表示。采用 K-means 算法对这些表示向量集聚类处理,实现 QAR 样本的符号化。在训练集 QAR 样本的符号化序列上,采用 Baum-Welch 算法[16]构建检测接地点远事件的 HMM 模型 λ_1。在模型 λ_1 下,计算验证集中每个 QAR 样本符号序列出现的概率,并

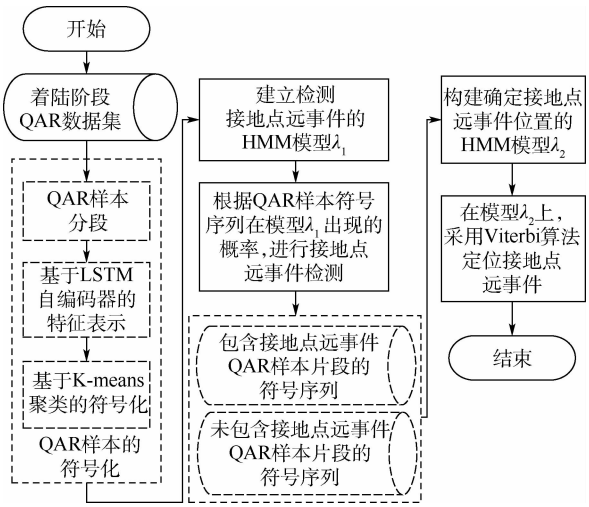


图 1 接地点远事件检测方法流程

Fig. 1 Flowchart of detection method of long touchdown exceedance

在出现概率的最大值和最小值之间均匀划分若干个值,根据 F_1 值最大原则来确定接地点远事件检测阈值。由包含发生及未发生接地点远事件 QAR 样本片段集训练 LSTM 自编码器,采用 K-means算法对 LSTM 编码器中每个 LSTM 单元隐藏层的输出向量进行聚类,实现 QAR 样本每个片段的符号化。由所有 QAR 样本分段的符号序列建立 HMM 模型 λ_1,采用 Viterbi 算法[16]确定接地点远事件在 QAR 样本片段内的具体位置。

1.1 QAR 样本分段及符号化

本文着陆阶段含有 N 个 QAR 样本的数据集可表示为 $X=(\boldsymbol{X}^1,\boldsymbol{X}^2,\cdots,\boldsymbol{X}^n,\cdots,\boldsymbol{X}^N)$,其中每个样本 $\boldsymbol{X}^n=(\boldsymbol{x}_0^n,\boldsymbol{x}_1^n,\cdots,\boldsymbol{x}_t^n,\cdots,\boldsymbol{x}_T^n)$ $(1\leqslant n\leqslant N)$ 表示长度为 T 的多维时间序列,$\boldsymbol{x}_t^n$ 表示第 n 个样本 t 时刻对应的 m 个 QAR 参数取值,每个样本 $\boldsymbol{X}^n$ 长度 T 的取值可以不相同。本文 QAR 样本的分段方式如图 2 所示。采用滑动窗口以固定分段数 d 和步长 s 对每个 QAR 样本 $\boldsymbol{X}^n$ $(1\leqslant i\leqslant N)$ 进行分段处理,每个样本的窗口大小为 $w=T-(d-1)s-1$,分段后的 QAR 样本可表示为 $\boldsymbol{X}^n=\{\boldsymbol{X}_1^n,\boldsymbol{X}_2^n,\cdots,\boldsymbol{X}_i^n,\cdots,\boldsymbol{X}_d^n\}$。

数据集 X 按上述方法分段后,形成 d 个 QAR 片段集,每个片段集记为 $X_i=\{\boldsymbol{X}_i^1,\cdots,\boldsymbol{X}_i^n,\cdots,\boldsymbol{X}_i^N\}$ $(1\leqslant i\leqslant d)$,其中 $\boldsymbol{X}_i^n$ 表示第 n 个 QAR 样本的第 i 个片段,由各个片段数据集训练 d 个 LSTM 自编码器,每个 LSTM 自编码器记为 L_i $(1\leqslant i\leqslant d)$。图 3 为 LSTM-AE 模型训练过程示意图,图 3(b)为 L_i 的网络结构示意图,由 2 个作为编码器和解码器的 LSTM 网络组成。L_i 的编码器输入是长度为 w 的 QAR 样本片段 $\boldsymbol{X}_i^n=\{\boldsymbol{x}_i^{(1)},\cdots,\boldsymbol{x}_i^{(t)},\cdots,\boldsymbol{x}_i^{(w)}\}$,$\boldsymbol{x}_i^{(t)}$ 表示片段 $\boldsymbol{X}_i^n$ 在 t 时刻的 m 维向量。$\boldsymbol{h}_{\mathrm{E}}^{(t)}$ 是 L_i 的编码器在 t 时刻由式(1)~式(5)计算得到的隐藏层输出向量[17],其中 $\boldsymbol{i}_t$、$\boldsymbol{f}_t$、$\boldsymbol{c}_t$、$\boldsymbol{o}_t$ 分别表示 t 时刻输入门、遗忘门、单元状态、输出门的输出,$\boldsymbol{W}_i$、$\boldsymbol{W}_f$、$\boldsymbol{W}_c$、$\boldsymbol{W}_o$ 分别表示输入门、遗忘门、更新单元状态、输出门的权重矩阵,$\boldsymbol{b}_i$、$\boldsymbol{b}_f$、$\boldsymbol{b}_c$、$\boldsymbol{b}_o$ 分别表示输入门、遗忘门、更新单元状态、输出门的偏置项,[ ]表示向量的串接操作,σ 表示 Logistic 函数。L_i 的解码器按逆序方式重构 $\boldsymbol{X}_i^n$。L_i 的编码器最后时刻隐藏层输出 $\boldsymbol{h}_{\mathrm{E}}^{(w)}$ 作为解码器初始网络单元隐藏的输出,即 $\boldsymbol{h}_{\mathrm{D}}^{(w)}=\boldsymbol{h}_{\mathrm{E}}^{(w)}$。$\boldsymbol{h}_{\mathrm{D}}^{(w)}$

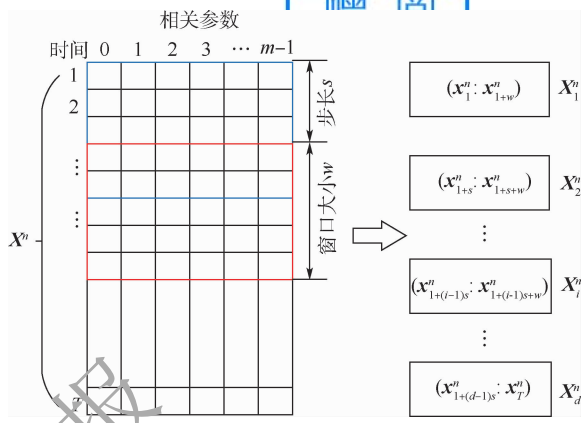
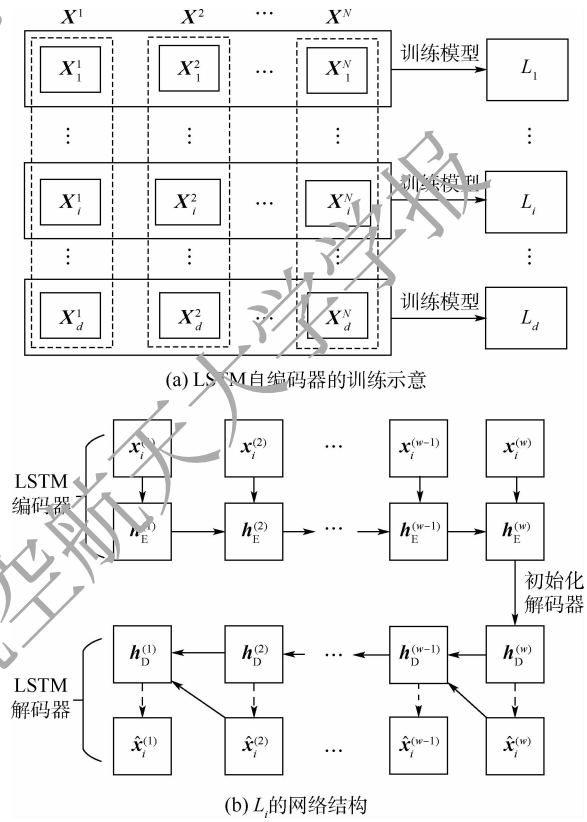


图 2 QAR 样本分段方法示意图

Fig. 2 Schematic diagram of QAR sample segmentation method



(a) LSTM自编码器的训练示意

(b) L_i的网络结构

图 3 LSTM-AE 训练过程示意图

Fig. 3 Schematic diagram of LSTM-AE training process

北京航空航天大学学报

北航学报 赠阅

由式(6)所示的线性变换后得到重构目标向量 $\hat{\boldsymbol{x}}_i^{(w)}$。由 $\hat{\boldsymbol{x}}_i^{(w)}$ 和 $\boldsymbol{h}_{\mathrm{D}}^{(w)}$ 计算得到解码器 $w-1$ 时刻的隐藏层输出 $\boldsymbol{h}_{\mathrm{D}}^{(w-1)}$,再经类似式(6)所示的线性变换后得到重构目标向量 $\hat{\boldsymbol{x}}_i^{(w-1)}$。$L_i$ 的训练目标为最小化式(7)所示的均方误差损失。

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{x}_i^{(t)}, \boldsymbol{h}_{\mathrm{E}}^{(t-1)}, \boldsymbol{c}_{t-1}] + \boldsymbol{b}_i) \tag{1}$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{x}_i^{(t)}, \boldsymbol{h}_{\mathrm{E}}^{(t-1)}, \boldsymbol{c}_{t-1}] + \boldsymbol{b}_f) \tag{2}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \tanh(\boldsymbol{W}_c \cdot [\boldsymbol{x}_i^{(t)}, \boldsymbol{h}_{\mathrm{E}}^{(t-1)}] + \boldsymbol{b}_c) \tag{3}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{x}_i^{(t)}, \boldsymbol{h}_{\mathrm{E}}^{(t-1)}, \boldsymbol{c}_t] + \boldsymbol{b}_o) \tag{4}$$

$$\boldsymbol{h}_{\mathrm{E}}^{(t)} = \boldsymbol{o}_t \tanh(\boldsymbol{c}_t) \tag{5}$$

$$\hat{\boldsymbol{x}}^{(w)} = \boldsymbol{W}\boldsymbol{h}_{\mathrm{D}}^{(w)} + \boldsymbol{b} \tag{6}$$

$$\sum_{X_i^n \in X_i} \sum_{t=1}^{w} \| \boldsymbol{x}_i^{(t)} - \hat{\boldsymbol{x}}_i^{(t)} \|^2 \tag{7}$$

QAR 样本片段 $\boldsymbol{X}_i^n$ 在训练后的 L_i 编码器最后时刻隐藏层输出记为 $\boldsymbol{h}_i^n$,将其作为 $\boldsymbol{X}_i^n$ 的特征表示。划分为 d 个分段的 QAR 样本 $\boldsymbol{X}^n$ 可表示为 $H^n = \{\boldsymbol{h}_1^n, \cdots, \boldsymbol{h}_i^n, \cdots, \boldsymbol{h}_d^n\}$。QAR 样本集 X 的表示向量集为 $H = \{H^1, \cdots, H^n, \cdots, H^N\}$。对 H 中的向量集进行 K-means 聚类处理,将每个向量距离最近的聚类中心所在簇标记作为其符号表示。符号化后的样本集 X 表示为 $O = \{O^1, \cdots, O^n, \cdots, O^N\}$,其中每个 QAR 样本 $\boldsymbol{X}^n$ 的符号序列为 $O^n = (O_1^n, \cdots, O_i^n, \cdots, O_d^n)$。具体符号化过程见算法 1 描述。本文将 $\boldsymbol{X}_i^n$ 在训练后的 L_i 编码器每个时刻隐藏层的输出 $\boldsymbol{p}_t$ 作为片段内每个向量 $\boldsymbol{x}_i^{(t)}$ 的特征表示。片段 $\boldsymbol{X}_i^n$ 可表示为 $P_i^n = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_t, \cdots, \boldsymbol{p}_w)$,基于 K-means 算法对样本集 X 所有的 $P_i^n (1 \leqslant i \leqslant d, 1 \leqslant n \leqslant N)$ 聚类处理,实现片段 $\boldsymbol{X}_i^n$ 的符号化。符号化后的 $\boldsymbol{X}_i^n$ 记为 $S_i^n = (s_1, \cdots, s_t, \cdots, s_w)$。$\boldsymbol{X}_i^n$ 的具体符号化过程与 QAR 样本 $\boldsymbol{X}^n$ 的符号化原理相同。

算法 1　QAR 样本集 X 符号化算法。

输入:样本集 X 的表示向量集 H。

输出:样本集 X 的符号表示 O。

Begin

由肘部方法确定 K-means 聚类个数 K

采用 K-means 算法对 H 聚类处理,生成 K 个聚类中心 $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_K\}$

For 每个 $H^n (1 \leqslant n \leqslant N)$

For 每个片段表示 $\boldsymbol{h}_i^n (1 \leqslant i \leqslant d)$

计算 $\boldsymbol{h}_i^n$ 与 K 个聚类中心的距离,将距离最近的聚类中心所在类簇标记作为 $\boldsymbol{h}_i^n$ 的符号表示,即

$$\theta_j = \arg \min_{j \in \{1, \cdots, K\}} \| \boldsymbol{h}_i^n - \boldsymbol{\mu}_j \|^2$$

$O_i^n = \theta_j$

$O^n = O^n \cup O_i^n$

End for

$O = O \cup O^n$

End for

End

1.2　基于 HMM 的接地点远事件检测

HMM 是由隐藏状态序列和观测序列构成的双重随机过程,序列的每一位置对应一个时刻的隐藏状态和观测值。检测接地点远事件的 HMM 记为 λ_1。λ_1 由 Q_1、V_1、$\boldsymbol{\Pi}_1$、$\boldsymbol{A}_1$、$\boldsymbol{B}_1$ 描述,其中 Q_1 表示隐藏状态集合,本文每个状态表达 QAR 样本内数据变化趋势,共设有平稳、轻微上升、轻微下降、突然上升和突然下降 5 种隐状态,V_1 表示所有可观测的表达 QAR 样本的符号集,其取值由符号化过程中 K-means 算法确定,$\boldsymbol{\Pi}_1$ 为初始状态概率向量,由各个隐藏状态的初始概率组成,$\boldsymbol{A}_1$ 为状态转移概率矩阵,由隐藏状态之间的转换概率组成,$\boldsymbol{B}_1$ 为输出概率矩阵,由隐藏状态下输出观测值的概率组成。

将 QAR 样本的符号序列集 O 划分为训练集 O^{T}、验证集 O^{V} 和测试集 O_{T}。O^{T} 只包含未发生接地点远事件样本的符号序列,O^{V} 和 O_{T} 包含发生及未发生接地点远事件样本的符号序列,样本数目分别记为 V_{N} 和 T_{N}。因为无法得到与观测序列对应的隐状态序列,所以在训练集 O^{T} 上采用无监督的 Baum-Welch 算法[16] 学习模型 λ_1 的参数 $\boldsymbol{\Pi}_1$、$\boldsymbol{A}_1$、$\boldsymbol{B}_1$。计算 O^{V} 中每个符号序列 O_n^{V} 在模型 λ_1 下的出现概率 $P(O_n^{\mathrm{V}} | \lambda_1), 1 \leqslant n \leqslant V_n$。在验证集 O^{V} 样本出现概率的最小值与最大值之间均匀划分 1 000 个值作为接地点远事件的检测候选阈值,计算每个候选阈值在 O^{V} 上的 F_1 得分,F_1 取值最高的候选阈值作为在测试集 O_{T} 上接地点远事件的最终阈值 τ。因为模型 λ_1 是由未发生接地点远事件的样本集训练生成,所以在模型 λ_1 下测试样本出现概率越小,其发生接地点远事件的可能性越大。检测接地点远事件的具体算法描述如下。

算法 2　检测接地点远事件算法。

输入:在 O^{T} 上训练的 HMM 模型 λ_1;验证集 O^{V};测试集 O_{T}。

输出:O_{T} 中每个符号序列 $O_{\mathrm{T}}^n (1 \leqslant n \leqslant T_{\mathrm{N}})$ 是否含接地点远事件。

Begin

For O^{V} 中每个符号序列 $O_n^{\mathrm{V}} (1 \leqslant n \leqslant V_{\mathrm{N}})$

计算 O_n^{V} 在模型 λ_1 下的出现概率 $P(O_n^{\mathrm{V}} | \lambda_1)$,并存入集合 P^{V}

End for

在集合 P^{V} 中的最小值和最大值间均匀划分 1 000 个值,计算每个候选阈值在 O^{V} 上接地点远

北京航空航天大学学报

北航学报 瞰阅

事件的 F_1 值。F_1 取值最高的候选阈值作为 O_T 上的检测阈值 τ

For O_T 中每个符号序列 $O_T^n(1\leqslant n\leqslant T_N)$
计算 O_T^n 在模型 λ_1 下的出现概率 $P(O_n^V\mid\lambda_1)$
If $P(O_n^V\mid\lambda_1)\leqslant\tau$ then
判定 O_T^n 中含有接地远点事件
End if
End for
End

1.3 基于 HMM 的接地点远事件位置检测

本文检测接地点远事件位置的 HMM 模型 λ_2 的结构如图 4 所示。模型 λ_2 由 Q_2、V_2、$\boldsymbol{\Pi}_2$、$\boldsymbol{A}_2$、$\boldsymbol{B}_2$ 刻画。Q_2 表示隐状态集合，包含正常、异常 2 种状态，正常表示无接地点事件发生，用 0 表示，异常表示发生接地点远事件，用 1 表示，V_2 表示 QAR 样本片段的符号集合，$\boldsymbol{\Pi}_2$、$\boldsymbol{A}_2$、$\boldsymbol{B}_2$ 的含义与 1.2 节中模型 λ_1 的 $\boldsymbol{\Pi}_1$、$\boldsymbol{A}_1$、$\boldsymbol{B}_1$ 含义类似。将发生及未发生接地点远事件的 QAR 样本集 X 划分为训练集和测试集。按 1.1 节描述的分段符号化原理对 X 中的样本片段符号化。训练集 QAR 样本分段的符号序列集记为 S^T，S^T 中每个样本为长度 e 的符号序列。测试集 QAR 样本分段的符号序列集记为 S_T，S_T 中包含 eT'_N 个片段符号序列，T'_N 为测试集样本数目。由 S^T 采用无监督 Baum-welch 算法[16]训练 HMM 模型 λ_2 的参数。基于 Viterbi 算法[16]计算 S_T 中每个片段符号序列在模型 λ_2 下的最大概率状态转换路径，由该路径确定接地点远事件在片段内的具体位置。具体算法描述如下。

算法 3 检测接地点远事件位置算法。

输入：在 S^T 上训练的 HMM 模型 λ_2；测试集分段符号序列集 S_T。

输出：S_T 中每个片段符号序列中接地点远事件位置信息。

Begin
For S_T 中的每个符号序列 $S_i^n=(s_1,\cdots,s_t,\cdots,s_w)(1\leqslant i\leqslant e,1\leqslant n\leqslant T'_N)$

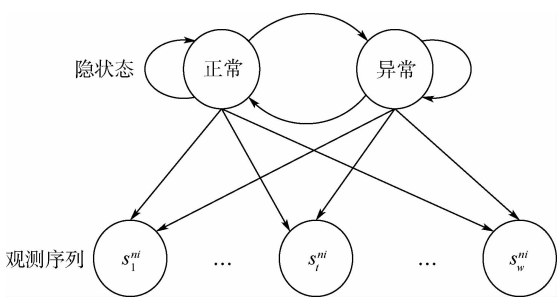


图 4 检测接地点远事件位置的 HMM 模型结构示意

Fig. 4 Structure of HMM model for detecting the position of long touchdown exceedance

采用 Viterbi 算法计算模型 λ_2 下最大概率 $P(\boldsymbol{I}\mid S_i^n)$ 的状态转换路径 $\boldsymbol{I}=(i_1,\cdots,i_w)$
For 每个状态 $i_m(1\leqslant m\leqslant w)$
If $i_m=1$ then
S_i^n 的 m 时刻发生接地点远事件
End if
End for
End for
End

2 实验与分析

2.1 实验数据与环境

本文实验数据为某航空公司同一架机型 737-800 的飞机 242 个航班 QAR 数据，截取了飞机着陆阶段相关 QAR 参数取值。其中 180 个航班未发生接地点远事件，62 个航班包含接地点远事件。接地点远事件检测实验中，训练集由 144 个正常 QAR 样本组成，测试集和验证集都由 22 个正常 QAR 样本和 31 个发生接地点远事件的 QAR 样本构成。接地点远事件位置检测实验中，共使用了 115 个 QAR 样本，训练集由随机抽取的 31 个正常 QAR 样本和 31 个接地点远事件的 QAR 样本组成，剩余 QAR 样本作为测试集。实验的软件环境为：Tensorflow1.10.0，Python3.6.2，Windows 10 64 位操作系统；硬件环境为：Intel(R) Core(TM) i7-3770 处理器，4 GB 内存。

根据飞机着陆过程的动力学模型[18]，与接地点远事件直接相关的因素包括经跑道入口时的高度、下滑角、拉平阶段圆弧运动的半径。当飞机高于 50 ft 的垂直距离进入跑道时，飞机的接地时间会变慢，从而导致接地距离变长。此外进近速度偏大会导致飞机下降率大，飞行员有可能拉杆防止重着陆，这也可能导致飞机接地距离变长。根据以上分析，选取了某航空公司在飞行品质日常业务工作中常用的 20 个 QAR 参数。这些参数既包含接地点远事件直接机理因素数，也含有飞机姿态和运行状态的相关参数。表 1 给出了 20 个参数名称及意义说明。

在训练检测接地点远事件的 HMM 过程中有 2 个重要参数：一个是 QAR 样本的片段个数，另一个是 QAR 样本符号化过程中 K-means 算法的聚类数目 K。本文中采用步长为 2 的滑动时间窗口将每个 QAR 样本划分为 6 个不等长的多维时间序列片段。K-means 算法的 K 值由肘部法确定。图 5 为肘部法确定符号个数过程中，聚类数目 K 与平均畸变程度的变化趋势。其中图 5(a)为 QAR 样本符号化的实验结果，当 $K=8$ 时，平均畸

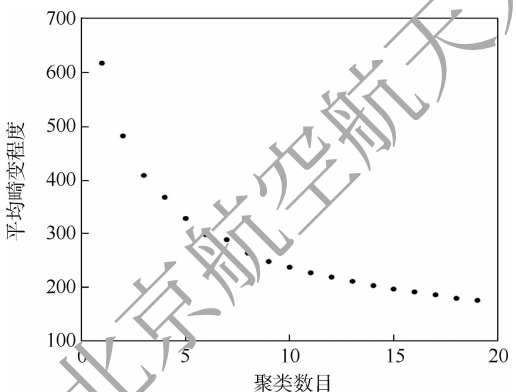
北京航空航天大学学报

北航学报

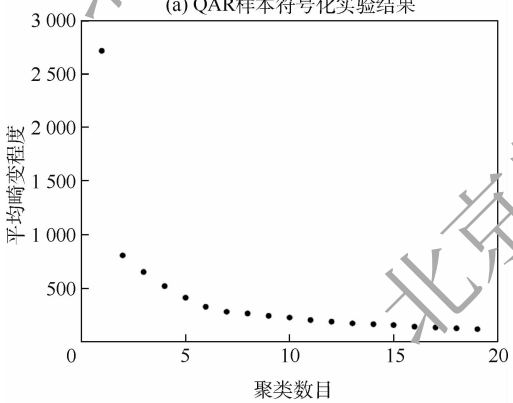
表 1　参数说明

Table 1　Parameter description

参数	意义
标准气压高度	飞机与标准气压平面之间的距离
无线电高度	飞机距离地面的真实高度
N11、N21	1、2 号发动机低压压气机转子负荷百分比
N12、N22	1、2 号发动机高压压气机转子负荷百分比
EGT1、EGT2	1、2 号发动机排气温度
左、右迎角	左、右机翼弦线与空气流速的夹角
俯仰角	机身纵轴线与水平面的夹角
倾侧角	飞机在横滚轴上的倾斜角度
方向舵偏角	方向舵左右偏转角度
空速	根据标准大气压下空速与动压关系测定的空速
垂直加速度	飞机在垂线上的加速度
纵向加速度	飞机在纵轴上的加速度
横向加速度	飞机在横轴上的加速度
下降率	飞机的垂直速度
FLAP1、FLAP2	飞机左右两边襟翼角度



(a) QAR样本符号化实验结果



(b) QAR样本片段符号化实验结果

图 5　聚类数目 K 与类簇平均畸变程度的变化趋势

Fig. 5　Changing trend between clusters number K and average distortion degree of cluster

变程度变化较高,因此符号个数取为 8;图 5(b)为 QAR 样本片段符号化的实验结果,当 $K=5$ 时,平均畸变程度变化较高,故符号个数取为 5。

2.2　接地点远事件检测实验结果

为了验证 LTED-LSTM-HMM 方法的有效性,分别采用了文献[12]和文献[14]提出的多维时间序列异常检测方法做对比。文献[12]分别采用主成分分析(principal component analysis,PCA)和 FCM 将多维时间序列转化成单维时间序列,再基于正常单维时间序列进行 HMM 建模,将训练后的 HMM 用于异常检测;文献[14]提出了基于自编码器和 HMM 的多维时间序列异常检测方法(autoencoder and HMM-based anomaly detection,AHMM-AD),该方法的实现原理与 LTED-LSTM-HMM 方法相同,主要区别为:AHMM-AD 方法采用具有3 层神经网络的自编器学习多维时间序列片段的低维特征表示。实验结果如表 2 所示。

表 2　异常航班数据检测的对比实验

Table 2　Comparison experiment of abnormal flight data detection

检测方法	准确率	召回率	F_1
PCA + HMM	0.43	0.89	0.609
FCM + HMM	0.432	0.708	0.611
AHMM-AD	0.493	0.682	0.570
LTED-LSTM-HMM	0.600	0.968	0.741

从表 2 中的准确率、召回率、F_1 值可知,基于 PCA 和 HMM 的方法异常检测效果较差,可能的原因是:PCA 对 QAR 样本降维时,只取了线性变换后方差最大方向上的样本信息,忽略了其他投影方向上的信息。基于 FCM 和 HMM 的方法引入隶属度的量化方法描述了 QAR 样本的各参数取向量与符号的关系,在符号化转换过程中使用了所有 QAR 参数信息,故异常检测效果有所提高。AHMM-AD 方法的实验结果显示准确率比 PCA + HMM、FCM + HMM 方法提高了约 0.06,但 F_1 值分别减少了 0.039 和 0.041。这主要是因为:AHMM-AD 方法在符号序列之前对 QAR 样本片段进行了非线性的特征表示,但该方法只能处理等长多维时间序列数据,表示学习过程中将 QAR 样本划分为了窗口大小为 28 的等长时序片段,在划分不等长的 QAR 样本时去除了部分数据,这导致了召回率和 F_1 值降低。LTED-LSTM-HMM 方法通过固定时间序列分段数解决了不等长 QAR 样本的特征表示问题,而且基于 LSTM 自编器的表示考虑了 QAR 样本片段内的时序信息。本文的 LTED-LSTM-HMM 方法与 AHMM-AD 方法相比,准确率提高了约 0.11,召回率提高了 0.286,F_1 值提高了 0.171,实验结果表明了本文方法的有效性。

2.3　接地点远事件位置检测实验结果

为便于分析发生接地点远事件的原因,缩小查找异常 QAR 参数的范围,依据领域专家经验,选取表 1 中的左右迎角、俯仰角、倾侧角、方向舵偏角、空速、垂直加速度、纵向加速度、下降率 8 个 QAR 参数取值训练用于检测接地点远事件位置的 HMM 模型 λ_2。使用 22 个正常 QAR 样本和

北京航空航天大学学报

31 个接地点远事件 QAR 样本的 318 个数据片段作为测试集。QAR 片段的每个位置都有表征是否接地点远事件的标签,该标签由 QAR 译码软件标注。模型 λ_2 的接地点远事件位置检测准确率为 78.2%,位置检测准确率由检测出的真实接地点远事件时刻点总数除以检测出的接地点远事件时刻点总数计算。图 6 为某 QAR 样本片段实际状态值序列与异常检测结果状态值序列的对比,状态值 1 表示该位置异常,状态值 0 表示该位置正常。图 6 的实线中取值为 1 的部分表示发生接地点远事件的真实位置,虚线中取值为 1 的部分表示本文模型检测出的发生接地点远事件位置。虚线取值为 1 的片段可以覆盖实线取值为 1 的部分,因此本文方法能在 QAR 片段内正确标识出发生接地点远事件的位置。从实验结果中还发现,接地点远事件位置大多出现在 QAR 样本片段的开始部分。

选取 2.2 节实验过程中离检测阈值较近的观测概率较高的 2 个发生接地点远事件的 QAR 样本和观测概率最低的正常 QAR 样本的原始参数取值进行对比。图 7 分别展示了 4 个样本在发生

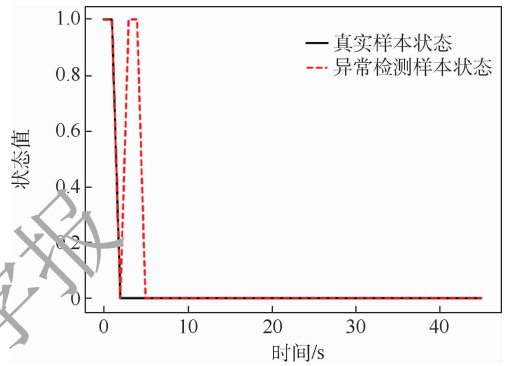


图 6 真实与异常检测结果状态值序列对比

Fig. 6 Comparison of state value sequence between real and anomaly detection results

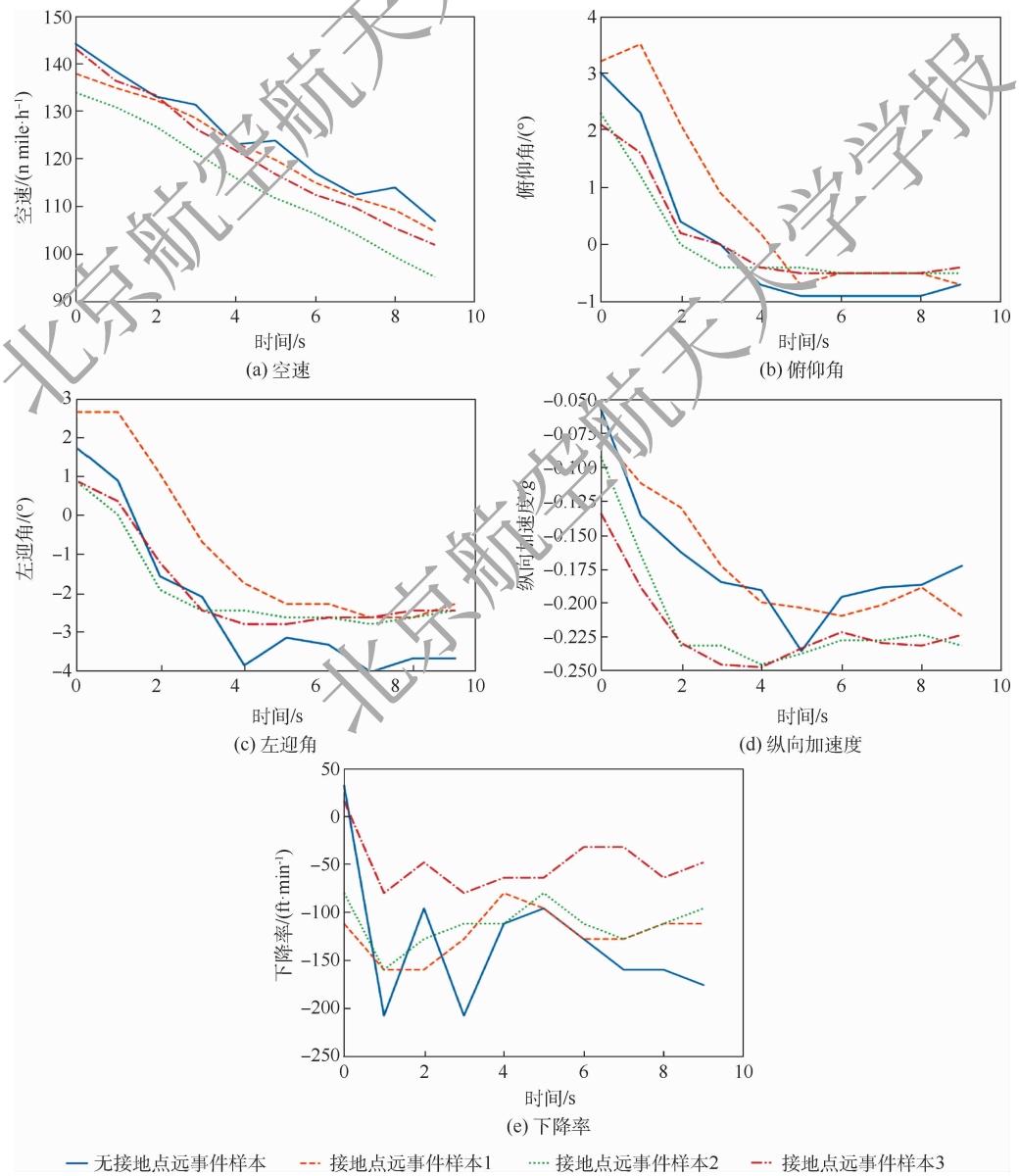


图 7 有无接地点远事件 QAR 样本典型参数取值对比

Fig. 7 Comparison of typical parameter values between the normal and the long touchdown exceedance QAR sample

北航学报 网阅

接地点远事件的片段 10 s 内空速、俯仰角、左迎角、纵向加速度、下降率 5 个 QAR 参数取值变化趋势。从图 7(a)可知,发生接地点远事件的 3 个样本空速都低于正常样本的取值。但是根据领域专家经验,着陆过程中空速越大越容易发生接地点远事件。本文中发现的 3 个样本虽然空速较低,但却出现了地速偏大造成接地点远事件的情况。结合其他5个参数的取值原因分析如下:①从图 7(b)和图 7(c)可以看出,发生接地点远事件的 3 个样本的俯仰角、左迎角片段取值较大,为了获得合适的升力,飞行员有可能收油门降低速度。另外从图 7(e)中无接地点远事件样本的下降率变化趋势整体高于其他 3 个异常样本。俯仰角大,下降率低都有可能延长飞机下滑拉平距离,导致飞机超出接地区接地。②从图 7(d)可知,接地点远事件样本 2 和样本 3 的减速方向上的纵向加速度大于正常样本,这使得接地点远事件样本 2 和样本 3 的空速低于正常样本。但从图 7(e)发现,样本 2 和样本 3 的下降率比正常样本低很多,这有可能是飞行员为了防止大下降率,造成重着陆事件,从而延迟了飞机接地时间。

3　结　论

1) LSTM 自编码器能较好地学习反映不等长 QAR 样本的特征表示,因此与已有的基于 HMM 的多维时序异常检测方法相比, LTED-LSTM-HMM 方法能更好地检测包含接地点远事件的航班。

2) 通过 QAR 样本片段建立的 HMM 和 Viterbi 算法,本文方法能以较高的准确率确定接地点远事件在 QAR 片段内的发生位置,由此得到除地速之外的其他 QAR 参数异常取值,辅助领域专家分析发生接地点远的原因。

下一步可把本文方法推广应用于其他典型超限事件检测及分析,以期提高航空安全管理水平。

参考文献 (References)

[1] 孙瑞山,韩文律. 基于差异检验的飞行超限事件参数特征分析[J]. 中国安全生产科学技术,2011,7(2):22-27.
SUN R S,HAN W L. Anaylysis on parameters characteristics of flight exceedance events based on distinction test[J]. Journal of Safety Science and Technology,2011,7(2):22-27(in Chinese).

[2] 郑磊,池宏,邵雪焱. 基于 QAR 数据的飞行操作模式及其风险分析[J]. 中国管理科学,2017,25(10):109-118.
ZHENG L,CHI H,SHAO X Y. Pattern recognition and risk analysis for flight operations[J]. Chinese Journal of Management Science,2017,25(10):109-118(in Chinese).

[3] WANG L,REN Y,WU C X. Effects of flare operation on landing safety:A study based on ANOVA of real flight data[J]. Safety Science,2018,102:14-25.

[4] 汪磊,郭世广,任勇. 基于飞行数据正态云的着陆操作风险评价方法[J]. 安全与环境学报,2019,19(5):1555-1561.
WANG L,GUO S G,REN Y. Landing operation risk evaluation based on the normal cloud of the flight data[J]. Journal of Safety and Environment,2019,19(5):1555-1561(in Chinese).

[5] LI L,DAS S,HANSMAN R J,et al. Analysis of flight data using clustering techniques for detecting abnormal operations[J]. Journal of Aerospace Information Systems,2015,12(9):587-598.

[6] SHERIDAN K,PURANIK T G,MANGORTEY E,et al. An application of DBSCAN clustering for flight anomaly detection during the approach phase[C]// AIAA SciTech Forum. Reston:AIAA,2020,1851:1-20.

[7] EUGENE M,DYLAN M,JAMEY A,et al. Application of machine learning techniques to parameter selection for flight risk identification[C]// AIAA SciTech Forum. Reston:AIAA,2020,1850:1-39.

[8] LI L,HANSMAN R J,PALACIOS R,et al. Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring[J]. Transportation Research Part C:Emerging Technologies,2016,64:45-57.

[9] MELNYK I,MATTHEWS B,VALIZADEGAN H,et al. Vector autoregressive model-based anomaly detection in aviation systems[J]. Journal of Aerospace Information Systems,2016,13(4):161-173.

[10] MELNYK I,BANERJEE A,MATTHEWS B,et al. Semi-Markov switching vector autoregressive model-based anomaly detection in aviation systems[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York:ACM,2016:1065-1074.

[11] KEYGHOBADI H,SEYEDIN A. Abnormality detection in a landing operation using hidden Markov model[J]. Journal of Computer & Robotics,2017,10(1):31-37.

[12] LI J,PEDRYCZ W,JAMAL I. Multivariate time series anomaly detection:A framework of hidden Markov models[J]. Applied Soft Computing,2017,60:229-240.

[13] JIA Y Z,XU M Q,WANG R X. Symbolic important point perceptually and hidden Markov model based hydraulic pump fault diagnosis method[J]. Sensors,2018,18(12):4460.

[14] 霍纬纲,王慧芳. 基于自编码器和隐马尔可夫模型的时间序列异常检测方法[J]. 计算机应用,2020,40(5):1329-1334.
HUO W G,WANG H F. Time series anomaly detection method based on autoencoder and HMM[J]. Journal of Computer Applications,2020,40(5):1329-1334(in Chinese).

[15] MALHOTRA P,RAMAKRISHNAN A,ANAND G,et al. LSTM-based encoder-decoder for multi-sensor anomaly detection[C]// Anomaly Detection Workshop at 33rd International Conference on Machine Learning,2016.

[16] 李航. 统计学习方法[M]. 2 版. 北京:清华大学出版社,2019:193-201.
LI H. Statistical learning method [M]. 2nd ed. Beijing:Tsing-

北京航空航天大学学报

hua University Press,2019:193-201(in Chinese).

[17] SRIVASTAVA N,MANSIMOV E,SALAKHUDINOV R. Unsupervised learning of video representations using LSTMs[C]// Proceedings of the 32nd International Conference on Machine Learning,2015,37:843-852.

[18] 占欣. 基于 QAR 数据的冲/偏出跑道风险评估研究[D]. 天津:中国民航大学,2019:15-20.
ZHAN X. Research on risk evaluation of runway excursion based on QAR data [D]. Tianjin:Civil Aviation University of China,2019:15-20(in Chinese).

Civil aircraft long touchdown exceedance detection based on autoencoder and HMM

HUO Weigang*, LI Jilong, WANG Huifang

(School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The existing flight operation quality assurance (FOQA) standard only uses the integral distance of the ground speed to define the long touchdown exceedance (LTE), which cannot detect and explain the exceedance using multiple quick access recorder (QAR) parameters. The QAR samples with multiple parameters were segmented by the sliding window, and the segmented sample sets were generated according to the segmentation position. The representation of the QAR sample segmentation and the vector within each QAR segment was obtained by the long short-term memory (LSTM) networks autoencoder, and the representation vectors were clustered by K-means to realize the symbolization of the QAR samples and the QAR segments. The hidden Markov model (HMM) model was built by using the symbolic sequence of the QAR sample set of the normal flights, which was used to detect the flights with the LTE. The second HMM model was constructed from the symbolic sequences of the segment of the normal QAR samples and the QAR samples including the LTE. Then the Viterbi algorithm was used to determine the specific positon of the LTE in the QAR sample segment. Experimental results on real QAR data sets show that, compared with other multi-dimensional time series anomaly detection methods, the proposed method can not only effectively detect the LTE, but also obtain the outliers of multiple QAR parameters, which can assist domain experts to analyze the cause of the exceedance.

Keywords: anomaly detection; long touchdown exceedance (LTE); autoencoder; hidden Markov model (HMM); quick access recorder (QAR)

Received: 2020-11-23; **Accepted**: 2021-01-22; **Published online**: 2021-02-19 11:43

URL: kns.cnki.net/kcms/detail/11.2625.V.20210219.1138.001.html

Foundation item: the Fundamental Research Funds for the Central Universities (3122019190)

* **Corresponding author**. E-mail: wghuo@cauc.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0615

含孔隙缺陷三维五向编织复合材料偏轴拉伸力学性能分析

全下文载

杨志贤，张明，李昂，张超*
（江苏大学 机械工程学院，镇江 212013）

摘 要：针对孔隙缺陷对材料力学性能的重要影响，在材料细观模型中引入随机分布的孔隙缺陷，研究了含孔隙缺陷三维五向编织复合材料的偏轴拉伸力学性能。基于2种典型编织角试件，讨论了孔隙率对材料正轴力学性能的影响，通过与实验数据对比，确定了材料的孔隙率。结合周期性边界条件施加偏轴拉伸载荷，获取了不同偏轴角度下材料的应力-应变曲线，并预测了材料的强度性能。模拟了典型偏轴角度下材料的细观损伤起始、演化过程，分析了材料的失效机理，为其他复合材料结构孔隙缺陷问题及偏轴载荷问题数值分析提供了一定的参考。

关 键 词：三维五向；编织复合材料；孔隙缺陷；偏轴拉伸；有限元分析
中图分类号：TB332
文献标志码：A **文章编号**：1001-5965(2022)04-0569-09

三维编织复合材料是以编织技术织造的预制件为增强体，与树脂等基体复合固化而形成的新型纺织结构复合材料。其增强体中纤维束相互交织形成立体网状结构，各个方向均有纤维增强，很好地满足了航空航天领域对材料结构减重和高承载性能的要求，已成为航空航天领域重要的结构材料[1]。

三维五向编织预制件是在三维四向编织工艺基础上，在编织成型方向加入一组不参与编织的轴向纱而形成的。由于在轴向引入了一组纱线，材料的轴向力学性能得到提高，为编织结构材料作为主承力构件应用提供了基础[1]。目前，关于三维五向编织复合材料力学性能的研究已经取得了较好的研究成果，包括几何模型、刚度性能预测、强度及损伤性能分析[2-5]等方面。

孔隙是三维五向编织复合材料客观存在的缺陷，对材料力学性能具有重要影响。但在上述研究工作中，为降低建模难度，并未考虑材料中孔隙的存在。目前，一些学者在三维机织[6-8]、三维四向编织[9-11]复合材料细观模型中引入孔隙，研究了孔隙缺陷对材料刚度、强度性能的影响。由于孔隙不可避免地存在于三维编织复合材料的制备过程中，在细观模拟中考虑孔隙缺陷，可以更加真实地模拟材料的损伤过程。然而目前，关于三维五向编织复合材料中孔隙缺陷对其力学性能影响的研究还极其有限[12]。

偏轴载荷是最典型的复杂载荷形式之一，偏轴载荷下材料力学行为研究是其他复杂载荷下相关问题研究的基础。然而目前，三维编织复合材料力学性能研究工作均局限在正轴载荷情况下，现有复合材料偏轴力学性能研究工作主要集中在单向复合材料和机织复合材料[13-16]。牛序铭[13]研究了单向陶瓷基复合材料偏轴拉伸变形与失效行为。杨昌[14]提出了复合材料单向板偏轴疲劳寿命的预测方法。Lu 等[15]对2.5维机织复合材料偏轴拉伸载荷下的应力-应变行为和损伤过程开展了多尺度渐进损伤分析。Wang 等[16]采用细

收稿日期：2020-11-04；**录用日期**：2021-04-02；**网络出版时间**：2021-04-28 08:54
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210427.1655.002.html
基金项目：江苏省自然科学基金（BK20180855）；机械结构力学及控制国家重点实验室开放课题（MCMS-E-0219Y01）
* **通信作者**. E-mail：zhangchao@ujs.edu.cn

引用格式：*杨志贤，张明，李昂，等. 含孔隙缺陷三维五向编织复合材料偏轴拉伸力学性能分析[J]. 北京航空航天大学学报，2022，48(4)：569-577. YANG Z X，ZHANG M，LI A，et al. Off-axis tensile mechanical properties of 3D five-directional braided composites with void defects[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：569-577 (in Chinese).*

北京航空航天大学学报

观有限元法,模拟了斜纹机织复合材料偏轴载荷下的损伤演化过程。

三维编织复合材料的偏轴力学性能指标与正轴加载相距甚远,材料承载性能无法由正轴载荷测试直接表征,材料的损伤失效机理也更为复杂,此方面的研究工作亟需开展。本文将孔隙缺陷引入到材料细观模拟中,讨论不同孔隙率对材料正轴力学性能的影响,确定与实验数据较吻合的孔隙率。在引入孔隙的基础上,基于周期性边界条件,施加偏轴拉伸载荷,讨论偏轴角度对三维五向编织复合材料力学性能的影响,分析典型偏轴角度下材料的损伤扩展过程及失效机理。

1 细观有限元模型

1.1 单胞结构模型

Xu 等[2]考虑了纤维束的相互挤压,假设编织纱截面为外切于椭圆的六边形,轴向纱截面为正方形,系统建立了三维五向编织材料单胞模型及其与工艺参数的关系,本文采用此单胞结构模型。

如图 1 所示,编织纱截面为内切一个长、短半轴分别为 a 和 b 椭圆的六边形,有

$$a = b\sqrt{r^2 + 4r + 3}\cos\gamma \tag{1}$$

式中:γ 为内部编织角;r 为轴向纱截面尺寸因子。

轴向纱截面是正方形,边长为 rb。单胞结构宽度为 W,厚度为 T,高度为 h,分别为

$$W = T = 2(2 + r)\sqrt{2}b \tag{2}$$

$$h = 4(2 + r)b/\tan\gamma \tag{3}$$

编织纱截面积为 S_b,轴向纱截面积为 S_a,分别为

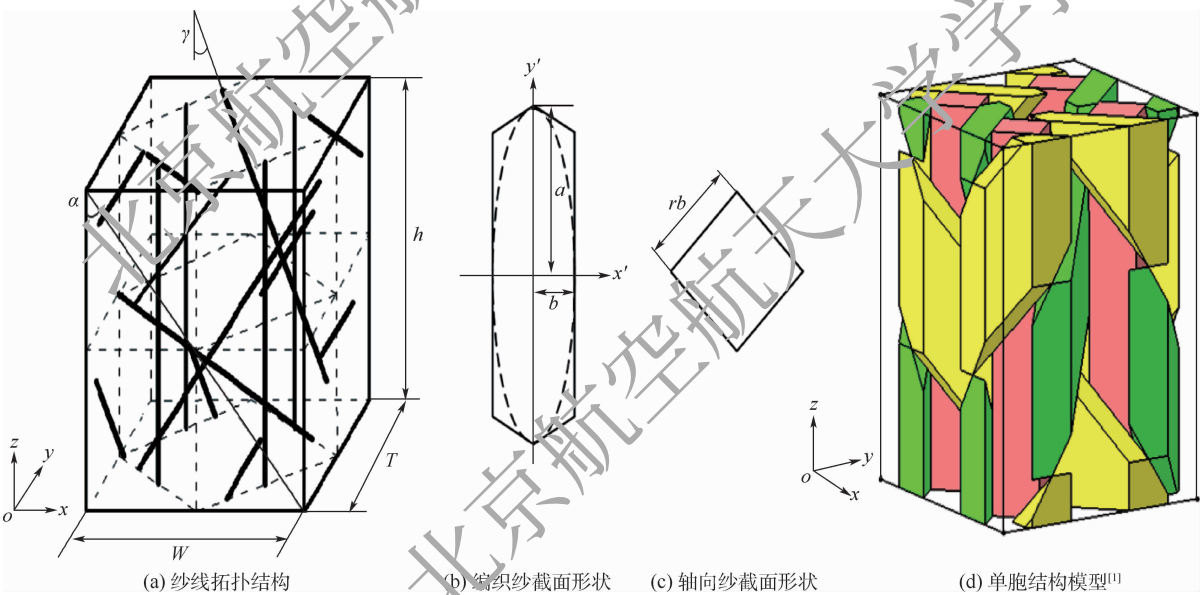
$$S_b = 2b^2(2r + 3)\cos\gamma \tag{4}$$

$$S_a = r^2b^2 \tag{5}$$

单胞模型中纤维体积含量 V_f 为

$$V_f = \frac{V_b\kappa_b + V_a\kappa_a}{V} = \frac{2(2r + 3)\kappa_b + r^2\kappa_a}{2(2 + r)^2} \tag{6}$$

式中:V_b、V_a 和 V 分别为编织纱、轴向纱和单胞模型体积;κ_b、κ_a 分别为编织纱、轴向纱填充系数。



(a) 纱线拓扑结构 (b) 编织纱截面形状 (c) 轴向纱截面形状 (d) 单胞结构模型[1]

图 1 三维五向编织复合材料单胞结构模型

Fig. 1 Unit-cell structure model of 3D five-directional braided composites

1.2 周期性边界条件

细观有限元分析需要施加周期性边界条件来体现细观结构的周期性特征。含理想周期性单胞结构的材料,其周期性位移场可以表示为

$$u_i = \bar{\varepsilon}_{ik}x_k + u_i^* \qquad i,k = 1,2,3 \tag{7}$$

式中:$\bar{\varepsilon}_{ik}$ 为单胞平均应变;x_k 为单胞内任意点的坐标;u_i^* 为周期性位移修正量。

显然,式(7)满足变形协调条件,但由于 u_i^* 通常为未知量且依赖于所施加的全局载荷,该周期性位移场很难被直接应用于基于单胞的分析中。

对于具有平行相对面的单胞,在一对平行边界面上,周期性位移场可以表示为[17]

$$u_i^{j^+} = \bar{\varepsilon}_{ik}x_k^{j^+} + u_i^* \qquad j = 1,2,3 \tag{8}$$

$$u_i^{j^-} = \bar{\varepsilon}_{ik}x_k^{j^-} + u_i^* \qquad j = 1,2,3 \tag{9}$$

式中:上标 j^+ 和 j^- 分别为沿 x_j 轴的正方向和负方向。

u_i^* 在周期性单胞的平行相对面上是相同的,式(8)和式(9)相减可得

$$u_i^{j^+} - u_i^{j^-} = \bar{\varepsilon}_{ik}(x_k^{j^+} - x_k^{j^-}) = \bar{\varepsilon}_{ik}\Delta x_k^j \tag{10}$$

对于立体单胞模型,Δx_k^j 为常数。一旦给定 $\bar{\varepsilon}_{ik}$,式(10)右侧位移差则变为常值。通过施加多点约束

北京航空航天大学学报

方程(MPC)来实现周期性边界条件的施加[1]。

1.3　有限元模型及相关参数

采用文献[18]所提供的三维五向编织复合材料拉伸试验结果作为对比验证数据。试件的工艺参数及单胞模型的结构参数如表 1 所示,组分材料力学性能参数如表 2 所示。纤维束的刚度、强度性能参数根据表 2 组分材料参数,采用 Chamis细观力学公式[19]计算获得。

表 1　试件工艺参数[18]及单胞模型结构参数

Table 1　Process parameters of specimen[18] and structure parameters of unit-cell model

试件	纤维单丝数量		试件尺寸/(mm × mm × mm)	编织角 α/(°)	W/mm	h/mm	V_f
	编织纱	轴向纱					
5DS1	9 000	6 000	20 × 4 × 150	20	3.052	8.415	45
5DS2	9 000	6 000	20 × 4 × 150	40	3.357	4.012	50

表 2　组分材料力学性能参数

Table 2　Mechanical properties of component materials

组分材料	E_{f1}/GPa	E_{f2}/GPa	G_{f12}/GPa	G_{f23}/GPa	μ_{f12}	E_m/GPa	μ_m	X_T/MPa	X_C/MPa	S/MPa
T300	220	13.89	9	4.8	0.2			3 000	2 070	
酚醛						3.2	0.35	75	180	60

试件 5DS1 和 5DS2 的编织角分别为 20° 和 40°,为典型的小编织角和大编织角试件。试件 5DS1 细观有限元模型由 39 501 个节点、152 694 个四面体单元(C3D4)组成;试件 5DS2 细观有限元模型由 20 385 个节点、77 965 个四面体单元组成。经数值计算,结果表明此细观有限元模型的网格尺寸满足数值计算收敛性的要求。

2　渐进损伤模型

基于单胞的渐进损伤模型包括应力分析、失效分析和材料性能退化方案等。应力分析涉及单胞模型中各组分的本构关系;失效分析需要选取合适的损伤起始准则来识别复合材料的主要失效模式;另外,在损伤起始后,需要建立合适的损伤演化模型来退化材料的刚度性能。

2.1　损伤起始准则

本文三维编织复合材料单胞模型由纤维束和基体构成。纤维束的损伤起始采用三维 Hashin 失效准则[20],基体损伤起始采用最大应力准则。三维 Hashin 失效准则区分了纤维束的 4 种不同损伤模式。

纤维束 L 向拉伸损伤 ($\sigma_L \geqslant 0$):

$$\sigma_{Lt} = \left(\frac{\sigma_L}{F_L^t}\right)^2 + \beta\left(\frac{\sigma_{LT}}{S_{LT}}\right)^2 + \beta\left(\frac{\sigma_{LZ}}{S_{LZ}}\right)^2 \geqslant 1 \tag{11}$$

纤维束 L 向压缩损伤 ($\sigma_L < 0$):

$$\varphi_{Lc} = \left(\frac{\sigma_L}{F_L^c}\right)^2 \geqslant 1 \tag{12}$$

纤维束 T 和 Z 向拉剪损伤 ($\sigma_T + \sigma_Z \geqslant 0$):

$$\varphi_{T(Z)t} = \left(\frac{\sigma_T + \sigma_Z}{F_T^t}\right)^2 + \frac{1}{S_{TZ}^2}(\sigma_{TZ}^2 - \sigma_T\sigma_Z) + \left(\frac{\sigma_{LT}}{S_{LT}}\right)^2 + \left(\frac{\sigma_{LZ}}{S_{LZ}}\right)^2 \geqslant 1 \tag{13}$$

纤维束 T 和 Z 向压剪损伤 ($\sigma_T + \sigma_Z < 0$):

$$\varphi_{T(Z)c} = \left(\frac{\sigma_T + \sigma_Z}{2S_{TZ}}\right)^2 + \left(\frac{\sigma_T + \sigma_Z}{F_T^c}\right)\left[\left(\frac{F_T^c}{2S_{TZ}}\right)^2 - 1\right] + \frac{1}{S_{TZ}^2}(\sigma_{TZ}^2 - \sigma_T\sigma_Z) + \left(\frac{\sigma_{LT}}{S_{LT}}\right)^2 + \left(\frac{\sigma_{LZ}}{S_{LZ}}\right)^2 \geqslant 1 \tag{14}$$

式中:F_L^t 和 F_L^c 分别为纤维束轴向拉伸和压缩强度;F_T^t 和 F_T^c 分别为纤维束横向拉伸和压缩强度;S_{LT}、S_{LZ}和 S_{TZ}分别为纤维束 LT、LZ 和 TZ 剪切强度;β 为剪切应力对纤维束 L 拉伸失效的贡献因子,β 取 0.5。L-T-Z为纤维束的局部直角坐标系,L、T 和 Z 轴分别为纤维束轴向、横向和法向。

基体损伤起始采用最大应力准则:

$$\varphi_{Mt} = \left|\frac{\sigma_1^t}{F_m^t}\right| \geqslant 1 \tag{15}$$

$$\varphi_{Mc} = \left|\frac{\sigma_3^c}{F_m^c}\right| \geqslant 1 \tag{16}$$

式中:F_m^t 和 F_m^c 分别为基体的拉伸和压缩强度;σ_1^t 和 σ_3^c 分别为基体中最大拉应力和压应力。

2.2　损伤演化模型

考虑到损伤的不可逆性,不同损伤模式下损伤变量的演化规律为

$$d_I = \max\left\{d_I, \frac{\delta_{I,eq}^f(\delta_{I,eq} - \delta_{I,eq}^0)}{\delta_{I,eq}(\delta_{I,eq}^f - \delta_{I,eq}^0)}\right\} \tag{17}$$

式中:$I = Lt, Lc, Tt, Tc, Zt, Zc, Mt, Mc$ 表示纤维束和基体的损伤模式;$\delta_{I,eq}^0$和 $\delta_{I,eq}^f$分别为某一失效模式的初始损伤等效位移和完全失效等效位移,可以由式(18)、式(19)计算:

$$\delta_{I,eq}^0 = \frac{\delta_{I,eq}}{\sqrt{\varphi_I}} \tag{18}$$

$$\delta_{I,eq}^f = \frac{2G_I}{\sigma_{I,eq}/\sqrt{\varphi_I}} \tag{19}$$

北京航空航天大学学报

北航学报 赠阅

式中:φ_1 为损伤起始准则的值;G_1 为断裂能密度;$\delta_{I,\mathrm{eq}}$和 $\sigma_{I,\mathrm{eq}}$分别为某一损伤模式的等效位移和等效应力,表达式为

$$\delta_{I,\mathrm{eq}} = l\sqrt{\varepsilon_{ii}^2 + \alpha_i \sum_{j=1,j\neq i}^{3} \varepsilon_{ij}^2} \tag{20}$$

$$\sigma_{I,\mathrm{eq}} = l\left(\sigma_{ii}\varepsilon_{ii} + \sum_{j=1,j\neq i}^{3} \sigma_{ij}\varepsilon_{ij}\right)/\delta_{I,\mathrm{eq}} \tag{21}$$

式中:$i=1,2,3$,表示 3 个坐标轴方向;l 为单元特征长度。

2.3 损伤本构模型

采用 Murakami 损伤张量[21]来表示材料各向异性损伤状态,即

$$\boldsymbol{D} = \sum_i \boldsymbol{D}_i \boldsymbol{n}_i \otimes \boldsymbol{n}_i \qquad i = L,T,Z \tag{22}$$

式中:$\boldsymbol{D}_i$和 $\boldsymbol{n}_i$分别为 3 个方向损伤张量主值和主方向单位矢量。

考虑对称性,有效应力 $\boldsymbol{\sigma}^*$ 与名义应力 $\boldsymbol{\sigma}$ 间有

$$\boldsymbol{\sigma}^* = \frac{1}{2}[(\boldsymbol{I}-\boldsymbol{D})^{-1}\boldsymbol{\sigma} + \boldsymbol{\sigma}(\boldsymbol{I}-\boldsymbol{D})^{-1}] = \boldsymbol{M}(\boldsymbol{D})\boldsymbol{\sigma} \tag{23}$$

采用应变能等效假设,得

$$\boldsymbol{C}(\boldsymbol{D}) = \boldsymbol{M}^{-1}(\boldsymbol{D}):\boldsymbol{C}_0:\boldsymbol{M}^{T,-1}(\boldsymbol{D}) \tag{24}$$

式中:$\boldsymbol{C}_0$ 为未损伤材料的刚度矩阵。

因此,含损伤材料的刚度矩阵为

$$\boldsymbol{C}(\boldsymbol{D}) = \begin{bmatrix} d_L^2 C_{11} & & & & \\ d_L d_T C_{12} & d_T^2 C_{22} & & \mathrm{sym} & \\ d_L d_Z C_{13} & d_T d_Z C_{23} & d_Z^2 C_{33} & & \\ 0 & 0 & 0 & d_{TZ} C_{44} & \\ 0 & 0 & 0 & 0 & d_{ZL} C_{55} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad d_{LT} C_{66} \tag{25}$$

式中:sym 表示对称;$d_L = 1 - D_L$;$d_T = 1 - D_T$;$d_Z = 1 - D_Z$;$d_{TZ} = \left(\frac{2d_T d_Z}{d_T + d_Z}\right)^2$;$d_{ZL} = \left(\frac{2d_Z d_L}{d_Z + d_L}\right)^2$;$d_{LT} = \left(\frac{2d_L d_T}{d_L + d_T}\right)^2$。

对于纤维束,L、T、Z 方向的损伤变量主值由式(26)~式(28)确定:

$$D_L = \max(d_{Lt}, d_{Lc}) \tag{26}$$

$$D_T = \max(d_{Tt}, d_{Tc}) \tag{27}$$

$$D_Z = \max(d_{Zt}, d_{Zc}) \tag{28}$$

式中:$d_I(I = Lt, Lc, Tt, Tc, Zt, Zc)$为纤维束失效模式相关的损伤变量。

对于基体,损伤变量主值为

$$D_M = D_L = D_T = D_Z = \max(d_{Mt}, d_{Mc}) \tag{29}$$

式中:d_{Mt}和 d_{Mc}分别为与基体拉伸和压缩失效模式相关的损伤变量。

3 数值模型验证

3.1 孔隙缺陷的引入

在三维五向编织复合材料细观模型中引入孔隙缺陷,评价孔隙对材料力学性能的影响。考虑到细观模型中孔隙的分布是随机的,基于 random 函数,编写筛选程序,对有限元模型中各组分的单元进行随机选取,将选中的单元统一赋予极小的刚度参数,达到赋予孔隙材料属性的目的,以模拟材料中的孔隙缺陷。对于三维五向编织复合材料 2 种典型编织角试件的细观有限元模型,对纤维束和基体单元分别进行处理,随机选取 0.1%、0.5%、1%、1.5%和 2%的体积含量单元作为孔隙缺陷引入。

3.2 正轴拉伸应力-应变曲线

对不同孔隙率单胞模型进行正轴拉伸数值模拟,得到 2 种典型编织角试件的应力-应变曲线,如图 2 所示。可以看出,引入不同孔隙率后,小编织角试件的拉伸强度和断裂应变随着孔隙率的增大逐步下降,且下降幅度依次减小;大编织角试件的拉伸强度在引入孔隙之后下降幅度较大,且不同孔隙率材料拉伸强度之间的差别较小。

对于三维五向编织复合材料,小编织角试件强度主要取决于轴向纱的强度,而大编织角试件的强度由轴向纱强度和基体强度共同决定。小编织角试件在引入孔隙后,轴向纱断裂将引起材料快速失去承载能力;随着孔隙率的不断增加,轴向纱的强度逐渐降低,小编织角试件的强度也不断下降。大编织角试件在轴向纱断裂后,由基体和编织纤维束共同承载,强度由基体强度决定;引入孔隙后,孔隙对轴向纱强度影响较大,对基体强度影响较小,故不同孔隙率对大编织角试件的拉伸强度影响较小。

表 3 和表 4 分别为 2 种编织角试件在不同孔

北京航空航天大学学报

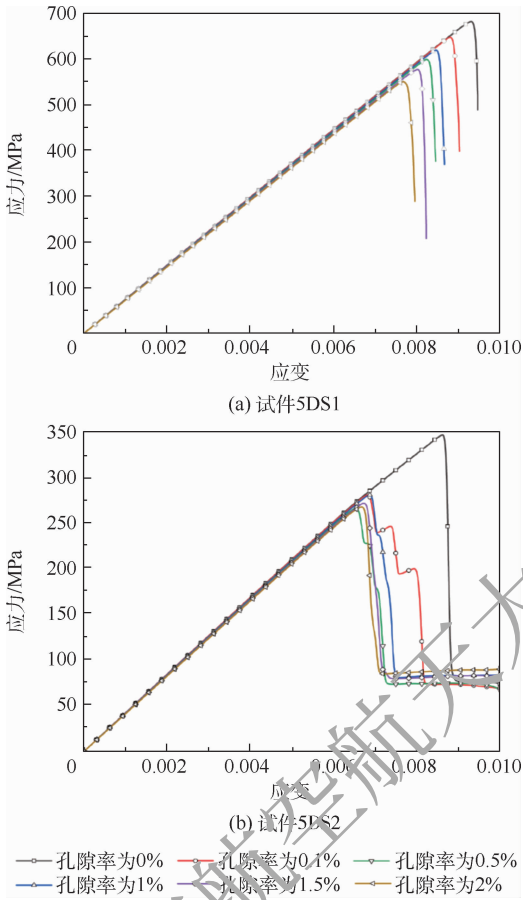


图 2　不同孔隙率试件的应力-应变曲线

Fig. 2　Stress-strain curves of specimen with different void contents

隙率下拉伸强度预测值与实验结果的对比。可以发现，小编织角试件在孔隙率为 0.5% 时的强度预测值与实验数据较为吻合；大编织角试件在孔隙率为 1%、1.5% 和 2% 时的拉伸强度预测值之间相差较小，3 种不同孔隙率下拉伸强度的预测结果都与实验数据较为相近。

$$\boldsymbol{A}=\begin{bmatrix} l_1^2 & m_1^2 & n_1^2 & 2l_1m_1 & 2m_1n_1 & 2n_1l_1 \\ l_2^2 & m_2^2 & n_2^2 & 2l_2m_2 & 2m_2n_2 & 2n_2l_2 \\ l_3^2 & m_3^2 & n_3^2 & 2l_3m_3 & 2m_3n_3 & 2n_3l_3 \\ l_1l_2 & m_1m_2 & n_1n_2 & l_1m_2+l_2m_1 & m_1n_2+m_2n_1 & n_1l_2+n_2l_1 \\ l_1l_3 & m_1m_3 & n_1n_3 & l_3m_1+l_1m_3 & m_3n_1+m_1n_3 & n_3l_1+n_1l_3 \\ l_2l_3 & m_2m_3 & n_2n_3 & l_2m_3+l_3m_2 & m_2n_3+m_3n_2 & n_2l_3+n_3l_2 \end{bmatrix} \tag{30}$$

式中：l_i、m_i、n_i 分别为局部坐标系和全局坐标系各对应坐标轴方向之间夹角的余弦值，$l_i=\cos(i,x)$，$m_i=\cos(i,y)$，$n_i=\cos(i,z)$。

通过位移加载的方式来实现在有限元模型上正轴载荷施加，偏轴载荷的施加同样通过施加各个方向的正轴位移来实现。

通过空间应变转换矩阵可以将偏轴坐标系下的偏轴应变转换为正轴坐标系下各个方向的应变分量，即

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{23} \end{bmatrix}=\boldsymbol{A}^{\mathrm{T}}\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{xy} \\ \gamma_{xz} \\ \gamma_{yz} \end{bmatrix} \tag{31}$$

表 3　试件 5DS1 拉伸强度预测值与实验结果对比

Table 3　Comparison between predicted and experimental tensile strengths of specimen 5DS1

试件	孔隙率/%	拉伸强度/MPa	
		预测值	实验结果
5DS1	0	681	632
	0.1	647	
	0.5	618	
	1	598	
	1.5	576	
	2	549	

表 4　试件 5DS2 拉伸强度预测值与实验结果对比

Table 4　Comparison between predicted and experimental tensile strengths of specimen 5DS2

试件	孔隙率/%	拉伸强度/MPa	
		预测值	实验结果
5DS1	0	346	259
	0.1	280	
	0.5	280	
	1	266	
	1.5	271	
	2	267	

4　偏轴拉伸力学性能分析

4.1　偏轴加载及应力状态分析

三维五向编织复合材料在承载偏轴载荷时，不再处于单一应力状态，而是处于复杂载荷状态。随着偏轴角度的不同，在材料的各个主方向会有不同的主应力和剪应力与之对应。在偏轴加载时，通常将局部坐标系中偏轴载荷分解到材料的全局坐标系中，便于分析求解。在局部坐标系中，偏轴载荷作用下的应力，可以通过空间应力转轴公式分解为全局坐标下的 3 个主应力及 3 个剪应力，具体空间应力转换矩阵为

北京航空航天大学学报

式中：$\boldsymbol{A}^{\mathrm{T}}$ 为空间应变转换矩阵，由空间应力转换矩阵转置得到；ε_1、ε_2、ε_3、γ_{12}、γ_{13}、γ_{23}为三维编织复合材料正轴坐标系下的应变分量；ε_x、ε_y、ε_z、γ_{xy}、γ_{xz}、γ_{yz}为偏轴坐标系下的应变分量。通常，对于三维五向编织复合材料偏轴拉伸来说，$\varepsilon_x \neq 0$，$\varepsilon_y = \varepsilon_z = \gamma_{xy} = \gamma_{xz} = \gamma_{yz} = 0$。

在将偏轴应变转化为相应的正轴坐标系下的应变分量后，根据周期性边界条件得到相应的位移施加到细观有限元模型中。三维五向编织复合材料正轴坐标系下的各个主方向的主应力及剪应力可以由式(32)求得

$$\overline{\sigma_{ij}} = \frac{(P_i)_j}{S_j} \tag{32}$$

式中：$(P_i)_j$ 为第 j 面上节点 i 方向的约束反力之和；S_j 为单胞的第 j 面；指标 j 不求和。

根据空间应力转轴公式，可以将正轴坐标系下的应力分量转换为偏轴坐标系下的应力分量，即

$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{xz} \\ \tau_{yz} \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \tau_{12} \\ \tau_{13} \\ \tau_{23} \end{bmatrix} \tag{33}$$

4.2 偏轴角度对材料力学性能的影响

选取孔隙率为 0.5% 的试件 5DS1 和孔隙率为 1.5% 的试件 5DS2 施加偏轴载荷。在 x-z 平面内，z 轴分别偏轴 0°、30°、45°、60°，得到 2 个编织角试件不同偏轴角度下的应力-应变曲线。

图 3 为孔隙率为 0.5% 的试件 5DS1 和孔隙率为 1.5% 的试件 5DS2 在不同偏轴角度下的应力-应变曲线。可以看出，2 个编织角试件的拉伸强度随偏轴角度的增大而不断减小，其中小编织角试件的拉伸强度随偏轴角度的增加下降较大；大编织角试件的拉伸强度随偏轴角度的增加下降幅度较小，虽然随着偏轴角度的增大拉伸强度不断下降，但是断裂应变在不断增大。对于小编织角试件，当承担偏轴拉伸载荷时，试件中的轴向纱会同时承受轴向拉伸、横向拉伸和轴向剪切载荷。轴向剪切使得轴向纱更易于出现纤维断裂失效，导致小编织角试件的偏轴拉伸强度降低。对于大编织角试件，在轴向纱断裂之后，基体和编织纤维束仍可以继续承载；且偏轴载荷的施加，使轴向纱受到的轴向拉伸载荷降低，只有施加更大的位移载荷，材料才会达到最大应力而失效，因此其断裂

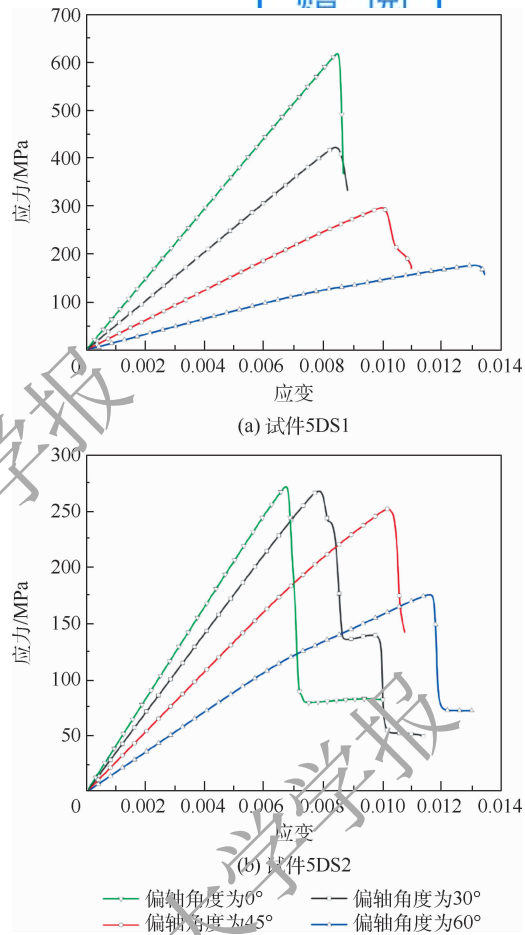


图 3 不同偏轴角度下试件的应力-应变曲线

Fig. 3 Stress-strain curves of specimen with different off-axis angles

应变也就相应增大。

表 5 和表 6 分别为三维五向编织复合材料各个角度偏轴拉伸下的拉伸强度和断裂应变。可以看出，小编织角试件的拉伸强度对偏轴角度较为敏感。当偏轴角度为 45°和 60°时，小编织角试件随着偏轴角度的增大，其拉伸强度不断减小，断裂应变不断增大；大编织角试件在偏轴角度较小时，其拉伸强度变化不大，断裂应变明显增加，随着偏轴角度的不断增加，大编织角试件的拉伸强度和断裂应变的变化与小编织角试件呈现相似的特征，其拉伸强度不断下降，断裂应变不断增加。

综合以上数据可以看出，偏轴角度对三维五向编织复合材料拉伸强度和断裂应变等性能的影响较大，对不同编织角材料的影响也不同。在偏轴角度较小时，对小编织角的试件的拉伸强度影响较大，对大编织角试件的断裂应变影响较大；随着偏轴角度的不断增大，2 种编织角的拉伸强度都不断下降，断裂应变不断增加。

北京航空航天大学学报

北航学报 网阅

表 5　试件 5DS1 偏轴拉伸力学性能

Table 5　Mechanical properties of specimen 5DS1 under off-axial tension

试件	偏轴角度/(°)	拉伸强度/MPa	断裂应变/%
5DS1	0	618.9	0.847
	30	423.1	0.842
	45	297.3	0.998
	60	177.1	1.313

表 6　试件 5DS2 偏轴拉伸力学性能

Table 6　Mechanical properties of specimen 5DS2 under off-axial tension

试件	偏轴角度/(°)	拉伸强度/MPa	断裂应变/%
5DS2	0	271.5	0.673
	30	267.8	0.787
	45	251.9	1.021
	60	175.6	1.157

4.3　偏轴拉伸损伤分析

图 4 给出了孔隙率为 0.5% 的试件 5DS1 在 45°偏轴拉伸时轴向纱 T 拉剪损伤、编织纱 Z 拉剪损伤和基体损伤的扩展过程。试件 5DS1 为小编织角试件，在承载偏轴拉伸载荷时其主要的失效模式为轴向纱 T 拉剪损伤、编织纱 Z 拉剪损伤，同时还有轴向纱 Z 压剪损伤、少量编织纱 Z 压剪和 T 拉剪损伤。

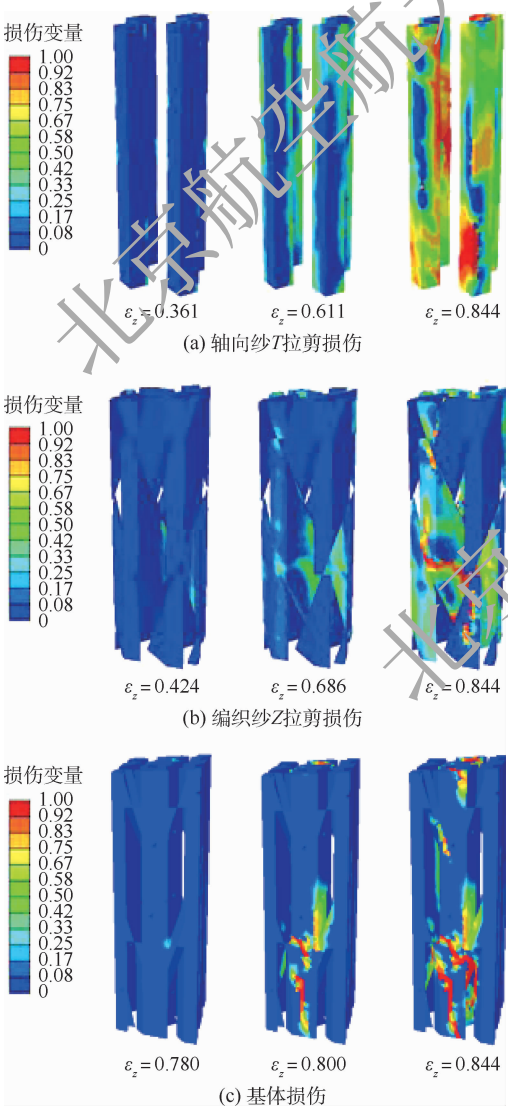


图 4　0.5% 孔隙率试件 5DS1 45°偏轴拉伸损伤扩展过程

Fig. 4　Tensile damage propagation process of specimen 5DS1 with 0.5% void content under 45° off-axial loading

图 4(a)为试件 5DS1 中轴向纱 T 拉剪损伤分布及扩展过程。在偏轴拉伸时，轴向纱中先出现损伤，且出现在全局坐标系下试件的棱边处，随着偏轴载荷的不断增加，损伤由棱边沿纵向和横向扩展，扩展速度逐渐加快。图 4(b)为编织纱 Z 拉剪损伤分布和扩展，其损伤最早出现在与轴向纱交界处，随着载荷增加，损伤由内向外不断扩展延伸，速度相对稳定。在图 4(c)中可以看到，基体中损伤先出现在与轴向纱和编织纱交错的区域，随着载荷的不断增加，沿着与编织纱和轴向纱的交界处不断延伸到外表面处，最终主要集中在基体的外表面；由于载荷重分配，基体中损伤扩展速度在试件应力-应变曲线达到最大值之前较慢，在达到最大值后损伤迅速增加。

图 5 为孔隙率 1.5% 的试件 5DS2 在 45°偏轴拉伸时轴向纱 T 拉剪损伤、编织纱 Z 拉剪损伤及基体损伤的分布和扩展过程。试件 5DS2 为大编织角试件，偏轴拉伸下其主要失效模式为轴向纱

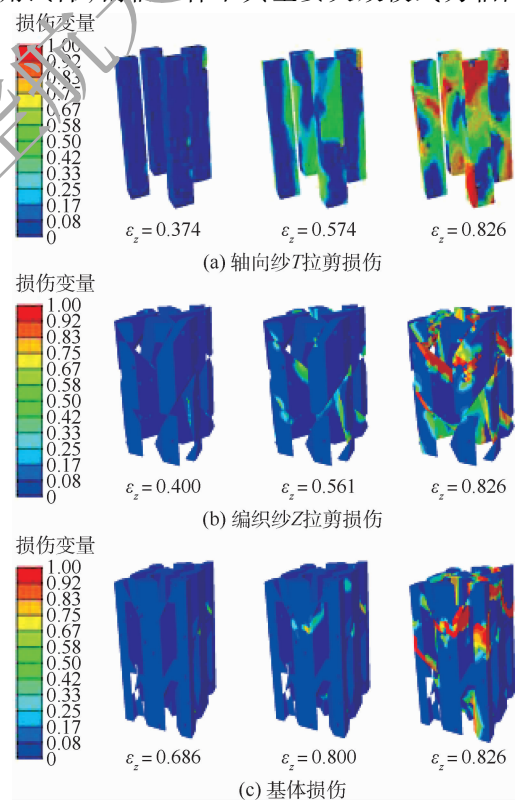


图 5　1.5% 孔隙率试件 5DS2 45°偏轴拉伸损伤扩展过程

Fig. 5　Tensile damage propagation process of specimen 5DS2 with 1.5% void content under 45° off-axial loading

北京航空航天大学学报

北航学报 赠阅

T 拉剪损伤、编织纱 Z 拉剪损伤和基体损伤，轴向纱中还有 Z 压剪损伤和少量 L 拉伸损伤，编织纱还有 L 拉伸损伤和少量 T 拉剪和 Z 压剪损伤。通过对比图 5 和图 4 可以发现，试件 5DS2 和试件 5DS1 相似，损伤最早出现在轴向纱中，位于全局坐标系下试件的棱边上；随着载荷的增加，不断向相近的棱边扩展，扩展速度持续加快。编织纱中的损伤出现在编织纱与轴向纱棱边交界处，沿着横向朝外部不断扩展，速度较为稳定。基体中损伤出现在轴向纱和编织纱棱边交错处，不断向外扩展延伸；相对于轴向纱和编织纱，基体中的损伤出现较晚，且在试件应力-应变曲线达到最大值之前，损伤扩展速度较慢；但在达到最大值之后，载荷重新分配导致损伤迅速增加，应力-应变曲线骤降，材料失去承载能力。

5 结 论

1）通过在 2 种典型编织角三维五向编织复合材料单胞模型中引入孔隙缺陷，发现小编织角试件的拉伸强度随着孔隙率的增加逐步下降，且下降幅度逐渐减小；大编织角试件的拉伸强度在引入孔隙后下降很大，但不同孔隙率对试件拉伸强度的影响较小。

2）通过含孔隙缺陷三维五向编织复合材料偏轴拉伸数值模拟发现，当偏轴角度较小时，对小编织角试件的拉伸强度影响较大，对大编织角试件的断裂应变影响较大，但这 2 种编织角试件的断裂应变和拉伸强度都随偏轴角度的增大而减小。

3）通过对 2 种试件的损伤分布和扩展过程分析发现，应力-应变曲线达到最大值时的断裂应变随偏轴角度的增加而增加。同一偏轴角度下，轴向纱损伤在应力-应变曲线达到最大值之前，损伤扩展速度逐渐加快；编织纱中损伤累计的速度相近；基体损伤在应力-应变曲线达到最大值之前变化很小，在达到最大值后，损伤急速增加。

4）引入孔隙缺陷的三维五向编织复合材料细观有限元模型，能够获得与实验数据更加吻合的预测结果，且所提出的偏轴加载方式和损伤模型具有良好的通用性，研究结果可以为其他复合材料结构孔隙缺陷问题及偏轴载荷问题数值分析提供一定的参考。

参考文献（References）

[1] 张超. 三维多向编织复合材料宏细观力学性能及高速冲击损伤研究[D]. 南京：南京航空航天大学，2013.
ZHANG C. Research on macro-meso-mechanical properties and high velocity impact damage of 3D multi-directional braided composites[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2013(in Chinese).

[2] XU K, XU X W. Finite element analysis of mechanical properties of 3D five-directional braided composites[J]. Materials Science and Engineering: A, 2008, 487(1-2): 499-509.

[3] LI D S, LU Z X, CHEN L, et al. Microstructure and mechanical properties of three-dimensional five-directional braided composites[J]. International Journal of Solids and Structures, 2009, 46(18-19): 3422-3432.

[4] 韩小进，孙慧玉，闫光，等. 三维五向编织复合材料渐进损伤分析的数值方法[J]. 复合材料学报，2012，29(6)：219-224.
HAN X J, SUN H Y, YAN G, et al. Numerical method for progressive damage analysis of 3D five-directional braided composites[J]. Acta Materiae Compositae Sinica, 2012, 29(6): 219-224(in Chinese).

[5] ZHANG D T, SUN Y, WANG X M, et al. Meso-scale finite element analyses of three-dimensional five-directional braided composites subjected to uniaxial and biaxial loading[J]. Journal of Reinforced Plastics and Composites, 2015, 34(24): 1989-2005.

[6] HUANG T, GONG Y H. A multi-scale analysis for predicting the elastic properties of 3D woven composites containing void defects[J]. Composite Structures, 2018, 185: 401-410.

[7] MEHDIKHANI M, PETROV Y A, STRAUMIT I, et al. The effect of voids on matrix cracking in composite laminates as revealed by combined computations at the micro- and meso-scales [J]. Composites Part A: Applied Science and Manufacturing, 2019, 117: 180-192.

[8] LIU T, FAN W, WU X Y. Comparisons of influence of random defects on the impact compressive behavior of three different textile structural composites[J]. Materials & Design, 2019, 181: 108073.

[9] XU K, QIAN X M. An FEM analysis with consideration of random void defects for predicting the mechanical properties of 3D braided composites[J]. Advances in Materials Science and Engineering, 2014, 2014: 1-12.

[10] DONG J W, HUO N F. A two-scale method for predicting the mechanical properties of 3D braided composites with internal defects[J]. Composite Structures, 2016, 152: 1-10.

[11] GAO X H, YUAN L, FU Y T, et al. Prediction of mechanical properties on 3D braided composites with void defects[J]. Composites Part B: Engineering, 2020, 197: 108164.

[12] GE L, LI H M, LIU B S, et al. Multi-scale elastic property prediction of 3D five-directional braided composites considering pore defects[J]. Composite Structures, 2020, 244: 112287.

[13] 牛序铭. 偏轴载荷下单向陶瓷基复合材料拉伸行为数值模拟及应用[D]. 南京：南京航空航天大学，2018：83-98.
NIU X M. Numerical simulation of tensile behavior of unidirectional ceramic matrix composite under off-axial loading and its application[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2018: 83-98(in Chinese).

[14] 杨昌. 复合材料单向板偏轴疲劳寿命预测研究[D]. 南京：南京航空航天大学，2014：36-41.

北京航空航天大学学报

YANG C. Research on fatigue life prediction of unidirectional composite[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2014: 36-41(in Chinese).

[15] LU Z X, ZHOU Y, YANG Z Y, et al. Multi-scale finite element analysis of 2.5D woven fabric composites under on-axis and off-axis tension[J]. Computational Materials Science, 2013, 79: 485-494.

[16] WANG L, ZHAO B, WU J Y, et al. Experimental and numerical investigation on mechanical behaviors of woven fabric composites under off-axial loading[J]. International Journal of Mechanical Sciences, 2018, 141: 157-167.

[17] XIA Z H, ZHANG Y F, ELLYIN F. A unified periodical boundary conditions for representative volume elements of composites and applications[J]. International Journal of Solids and Structures, 2003, 40(8): 1907-1921.

[18] 李仲平，卢子兴，冯志海，等. 三维五向碳/酚醛编织复合材料的拉伸性能及破坏机理[J]. 航空学报，2007，28(4)：869-873.
LI Z P, LU Z X, FENG Z H, et al. Investigation of the tensile properties and failure mechanism of integrally-braided 5D carbon/phenolic composites[J]. Acta Aeronautica et Astronautica Sinica, 2007, 28(4): 869-873(in Chinese).

[19] CHAMIS C C. Mechanics of composites materials: Past, present and future[J]. Journal of Composites Technology and Research, 1989, 11(1): 3-14.

[20] HASHIN Z. Failure criteria for unidirectional fiber composites[J]. Journal of Applied Mechanics, 1980, 47(2): 329-334.

[21] MURAKAMI S. Mechanical modeling of material damage[J]. Journal of Applied Mechanics, 1988, 55(2): 280-286.

Off-axis tensile mechanical properties of 3D five-directional braided composites with void defects

YANG Zhixian, ZHANG Ming, LI Ang, ZHANG Chao*

(School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: In view of the important influence of void defects on the mechanical properties of materials, the off-axis tensile mechanical properties of 3D five-directional braided composites with voids are studied by introducing randomly distributed void defects into the meso-scale model of material. For two specimens with typical braided angles, the effect of void contents on the on-axis mechanical properties was discussed and the appropriate void contents were determined by comparing with the available experimental data. Based on the periodic boundary conditions, the off-axis load was applied to obtain the stress-strain curves of the material under different off-axis angles and the strength properties were thus predicted. The meso-scale damage initiation and evolution processes of the composites under typical off-axis angles were simulated and the failure mechanism was analyzed in detail, which provides a proper reference for the numerical analysis of void defects and off-axis loading problems of other composite structures.

Keywords: 3D five-directional; braided composites; void defect; off-axial tensile; finite element analysis

Received: 2020-11-04; **Accepted**: 2021-04-02; **Published online**: 2021-04-28 08:54

URL: kns.cnki.net/kcms/detail/11.2625.V.20210427.1655.002.html

Foundation items: Natural Science Foundation of Jiangsu Province (BK20180855); Research Fund of State Key Laboratory of Mechanics and Control of Mechanical Structures (MCMS-E-0219Y01)

* **Corresponding author**. E-mail: zhangchao@ujs.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0631

颅缝早闭手术中颅骨切割方案生成方法

全下文载

罗杨宇[1,*]，贺佳宾[1]，谢东升[1]，陆珅宇[1]，宫剑[2]
（1. 中国科学院自动化研究所，北京 100190； 2. 首都医科大学附属北京天坛医院，北京 100070）

摘 要：为实现颅缝早闭手术方案制定的规范化，提出一种结合深度学习、立体视觉和点云处理技术的颅骨切割轨迹生成方法，用于建立切割方案模板库和生成新的病例切割方案。该方法将深度学习应用于颅骨外表面的实例分割中，利用 Mask R-CNN 对手术区域进行检测和分割，利用简化轮廓提取算法提取切割轨迹，并结合点云处理技术将切割轨迹坐标进行 2D-3D 映射，实现切割轨迹自动化提取；在此基础上建立典型病例模板库，通过模板匹配方法自动生成新病例的切割方案。实验证明：该轨迹提取方法可以准确高效地检测出颅骨切割轨迹，并将轨迹坐标进行 3D 映射，点云深度测量误差小于 3 mm，达到临床可用标准；模板匹配方法也有效生成新病例的切割轨迹，符合资深医生的手术方案。

关 键 词：颅缝早闭；轨迹提取；深度学习；深度相机；点云处理
中图分类号：T24
文献标志码：A **文章编号**：1001-5965(2022)04-0578-08

颅缝早闭是婴幼儿常见疾病，发病率约为 1/2 000，患者通常表现为颅面畸形，严重影响其外貌和脑部正常发育。当前，通常采用外科手术对颅骨进行切割和拼接来达到扩容和美观的目的。由于手术对象为婴幼儿软脆的颅骨，对术者的临床经验和操作水平都提出了很大的挑战，其中的难点之一在于如何针对不同患者制定切割方案，以最少量的切割骨片达到理想的治疗效果。目前针对颅缝早闭手术方案设计，南京市儿童医院沈卫民等[1]使用梅花颅骨瓣头颅盖成型法对 3 个月以上的舟状头患儿实施手术，取得良好效果；冯胜之和张涤生[2]针对不同颅面畸形探讨了不同手术方法及其效果。当前手术方案的制定严重依赖医生经验和个人判断，由于在术前没有模拟演示而手术机会近乎只有 1 次，外科医师在术前对于患者个体化的术后效果并没有明确概念，这使得术后效果有时并不尽如人意。

当前人工智能方法在颅脑相关领域有广泛的应用，尤其在颅内医学影像识别和分割领域，Kleesiek 等[3]将深度学习应用于磁共振成像（magnetic resonance imaging，MRI）的颅内肿瘤识别，该方法在颅内肿瘤识别分割测试集上取得了较好的效果；Chilamkurthy 等[4]使用经过训练的深度学习方法来识别各种颅内出血、颅骨骨折、中线转移及质量效应等 CT 异常，有效提高了影像科医生的工作效率。上述方法主要针对颅腔内容物的识别和分割，都取得了较好的效果。但目前鲜有将人工智能方法应用于颅腔表面区域的检测和分割的研究，对畸形颅骨的区域分割是人工智能应用的有益探索。

本文将人工智能方法引入颅缝早闭手术方案的制定中，将资深医生的手术方案设计作为输入，利用深度学习方法将手术方案设计规范化，提高一致性。

收稿日期：2020-11-12；**录用日期**：2021-02-10；**网络出版时间**：2021-03-30 11:09
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210329.1516.001.html
基金项目：国家重点研发计划（2017YFE0121200）
*通信作者. E-mail：yangyu.luo@ia.ac.cn

引用格式：罗杨宇，贺佳宾，谢东升，等. 颅缝早闭手术中颅骨切割方案生成方法[J]. 北京航空航天大学学报，2022，48(4)：578-585. LUO Y Y，HE J B，XIE D S，et al. Skull cutting plan generation method in craniosynostosis surgery[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：578-585 (in Chinese).

北京航空航天大学学报

北航学报 网阅

1　基于 Mask R-CNN 的颅骨手术区域分割

1.1　Mask R-CNN 实例分割算法

Mask R-CNN[5]是当前用于像素级别实例分割的经典框架,且目前尚未有在颅骨外表面进行实例分割的应用,而资深医生在颅骨 3D 打印模型上标记的切割方案将颅骨表面划分为若干鲜明分区,使得选用 Mask R-CNN 进行颅骨表面手术区域的检测和分割具有先天优势。

Mask R-CNN 整体分为 2 个阶段:第 1 阶段利用卷积神经网络(CNN)对全图进行卷积操作,生成候选区域;第 2 阶段则针对前一阶段生成的候选区域进行物体的分类、包围框和 mask 的生成。

第 1 阶段使用 ResNet-101[6]作为基础网络,通过串联特征金字塔网络(FPN)[7]融合图像的低层特征和高层语义特征,提高第 1 阶段网络提取特征图的表达能力,之后利用区域候选网络(RPN)[8]生成用于进一步分类回归的 ROI[9]。第 2 阶段网络主要对第 1 阶段提取到的 ROI 进行分类、包围框偏移量回归和 mask 像素级别分类,最终输出目标物体的类别、包围框坐标和 mask 像素类别。

Mask R-CNN 使用的损失函数为

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}} \tag{1}$$

上述损失函数由 Mask R-CNN 的 3 个分支组成,其中一个分支针对每个 ROI 输出 $K+1$ 类离散的概率 $p = \{p_0, p_1, \cdots, p_K\}$,对应的损失函数 L_{cls} 为交叉熵损失(cross-entropy loss),表达式为

$$L_{\text{cls}}(p,u) = -\ln p_u \tag{2}$$

式中:L_{cls} 为每个 ROI 的分类损失;p 使用 softmax 计算得到;u 为每一个 ROI 类别的 ground-truth 标签。

另一分支输出预测包围框的偏移量回归值,对应的损失函数 L_{box} 为 smooth L_1 损失(见式(3)和式(4)),表示每个 ROI 预测框的坐标偏移量的回归损失,$t^k = (t_x^k, t_y^k, t_w^k, t_h^k)$ 表示 K 类中类别 k 对应的目标包围框的偏移量回归值,包围框回归值的 ground-truth 记为 v。

$$L_{\text{box}}(t^u, v) = \sum_{i \in \{x,y,w,h\}} \text{smooth}_{L_1}(t_i^u - v_i) \tag{3}$$

式中:

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & \text{其他} \end{cases} \tag{4}$$

最后一个分支输出每个 ROI 内逐像素的分类概率,L_{mask} 为平均二值交叉熵损失(average binary cross-entropy loss),表达式为

$$L_{\text{mask}} = -\frac{1}{m^2} \sum_{1 \leqslant i,j \leqslant m} [y_{ij} \ln \hat{y}_{ij}^k + (1 - y_{ij}) \ln(1 - \hat{y}_{ij}^k)] \tag{5}$$

式中:L_{mask} 为每个 ROI 内所有像素点的平均分类损失;y_{ij} 为像素 (i,j) 在对应的 ground-truth mask 中的标签;$\hat{y}_{ij}^k$ 为像素 (i,j) 在被预测为真实类别 k 时对应生成的 mask 中的标签,当前像素属于背景记为 0,属于类别 k 记为 1。

选择召回率(recall)、精度(precision)和所有类别精度的平均值(mean average precision,mAP)作为模型评价指标。

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

式中:TP、FN 和 FP 分别为真阳、假阴和假阳的像素点数量。

1.2　患者样本数据采集

人类颅骨存在天然缝隙,称为颅缝,如图 1 所示,包括冠状缝、矢状缝、人字缝,这些颅缝将颅骨分为若干边界分明的区域,其中任一颅缝提早闭合就会引起婴幼儿颅面畸形发育,这其中又以矢状缝早闭和冠状缝早闭最为常见。矢状缝早闭患者的常见颅骨形态呈现前后狭长的舟状,冠状缝早闭患者的常见颅骨形态表现为颅骨单侧畸形,对患儿外貌和颅脑发育影响极大。

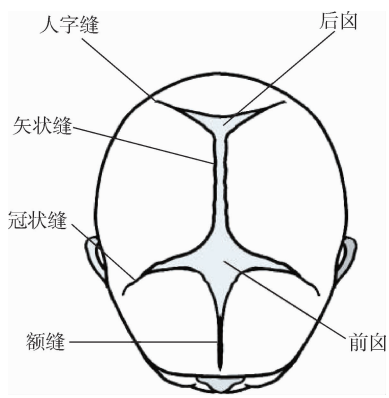


图 1　婴儿颅缝示意图

Fig. 1　Schematic diagram of infant cranial suture

资深医生从过往患者数据库中选择 20 例典型病例作为手术方案模板,由于患者术后颅骨形态发生改变,使得无法复现切割方案,需要先使用 3D-Slicer 软件重建患者原始颅骨模型,为符合医生操作习惯、降低标注难度并提高标注精度,未直接使用软件在颅骨模型曲面上进行切割轨迹标

北京航空航天大学学报

注,而是在打印颅骨模型后由资深医生在 3D 模型上标注切割轨迹,将标记了切割方案的 3D 颅骨模型作为本文数字化提取的对象。标记效果如图 2 所示。



图 2 医生标注方案

Fig. 2 Cutting plan marked by doctor

使用同一套深度相机用于采集颅骨 3D 打印模型的 2D 图像和 3D 点云,其中 2D 图像均选自左相机输出,3D 点云用于后期切割方案的 2D-3D 映射。使用该套深度相机采集 20 例颅骨 3D 打印模型在不同视角下的 500 张图片作为训练集,另外拍摄 50 张图片作为验证集(均采集左相机图像作为实验数据),图片尺寸均为 640 × 480。

样本图片采集完成后使用 Labelme 软件进行人工标注,获得每幅图片标注出的多边形区域对应的灰度图及包含多边形区域边界顶点 2D 坐标的 json 文件用于之后的训练。

1.3 模型训练

针对训练过程中训练样本数目不足的问题,使用经过 COCO 数据集预训练的模型参数而非随机值作为训练的初始参数,同时加入了旋转、裁剪等数据增强手段,有效降低了样本数量对模型训练的影响。

设置训练参数为:输入图片尺寸为 640 × 480;总训练轮数为 200;每轮训练的步数为 500;批尺寸为 1;学习率为 0.000 1;训练监测指标为验证集上的损失函数 L_{mask};总训练步数为 100 000。

图 3 和图 4 分别展示了训练过程中训练集和验证集上损失函数 L 和 L_{mask} 的变化曲线。

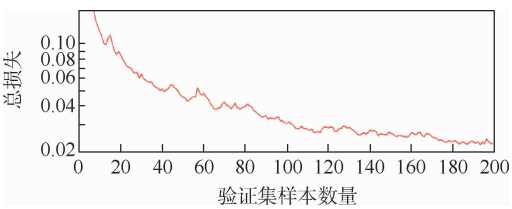


图 3 训练过程中训练集上的 L 曲线

Fig. 3 L curve of training dataset during training

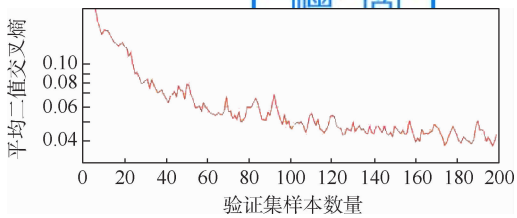


图 4 训练过程中验证集上的 L_{mask} 曲线

Fig. 4 L_{mask} curve of validation dataset during training

1.4 实验结果

颅骨 3D 模型上的切割轨迹标注均由资深医生完成,且所有实验在同一实验环境下完成。

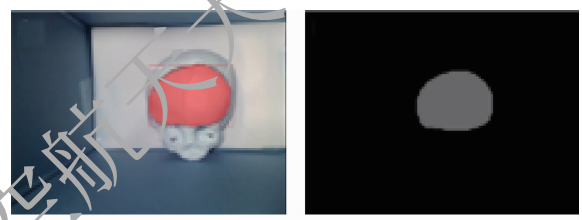
表 1 为以 ResNet-101 作为 backbone 的模型性能。可以看出,以 ResNet-101 作为 backbone 的模型在各个指标上均表现优异。

图 5 和图 6 展示了模型在不同视角和光照条件下对 3D 打印颅骨进行实例分割的结果。尤其是针对一些困难场景下的目标区域检测和分割,如图 7 所示场景,光照不足且目标位置发生较大旋转,模型依然能够精准检测并分割。

表 1 模型精度

Table 1 Model precision

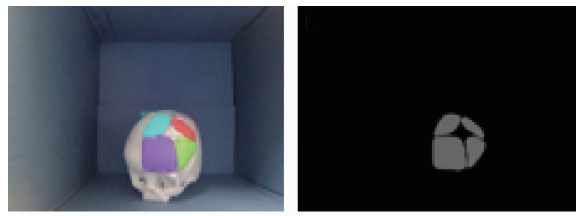
模型	backbone	精度/%	召回率/%	mAP/%
Mask R-CNN	ResNet-101	84.44	58.25	99.68



(a) 分割的目标mask(红色区域) (b) mask灰度图

图 5 冠状缝早闭颅骨分割结果

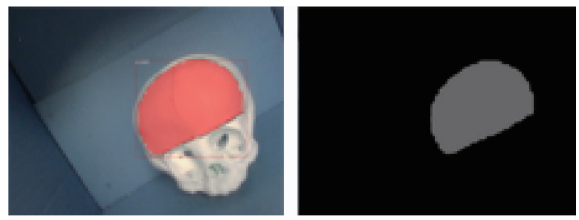
Fig. 5 Skull segmentation result for coronary craniosynostosis



(a) 分割的目标mask(红色区域) (b) mask灰度图

图 6 矢状缝早闭颅骨分割结果

Fig. 6 Skull segmentation result for sagittal craniosynostosis



(a) 分割的目标mask(红色区域) (b) mask灰度图

图 7 变形情况下颅骨分割结果

Fig. 7 Skull segmentation result under deformity circumstance

北京航空航天大学学报

之后随着新病例的增加,可以加入更多新的训练数据继续优化模型性能。

2　轮廓提取

2.1　轮廓提取算法

在图像处理领域,要提取图像中包含的目标轮廓[10],需要先将图像进行二值化处理,生成的二值图像中背景点和孔洞点的灰度值置 0,而前景目标点的灰度值置 1,利用灰度值的差异将全图分为不同的分区,不同分区的分界线即为轮廓。

图像的轮廓在数学上定义如下:对于灰度值为 1 的像素点(i,j),若在其 8 -(4 -)邻域内存在像素点(x,y)的灰度值为 0,则认为点(i,j)是轮廓点[11]。其中 i 表示像素点所在行号,j 表示列号。同一行上轮廓的左侧起点和右侧终点满足以下性质[12]。

性质 1　若像素点$(i,j-1)$的灰度值为 0,而像素点(i,j)的灰度值为 1,则像素点(i,j)为轮廓的左侧起点。

性质 2　若像素点(i,j)的灰度值为 1,而像素点$(i,j+1)$的灰度值为 0,则像素点(i,j)为轮廓的右侧终点。

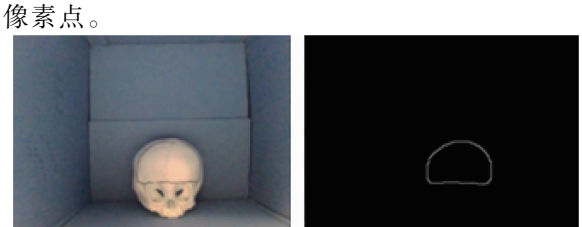
颅骨只存在唯一最外轮廓,此外,图 8 所示方法可以得到目标区域包围框,因此可以对上述算法进行简化,简化后的算法步骤如下:

步骤 1　从包围框的左上角像素按照从左到右的顺序搜索,找到满足性质 1 的像素点记为该行轮廓起点 S,并置灰度值为 255。

步骤 2　从 S 开始向右继续搜索,若当前像素点满足轮廓点定义,则将当前点标记为轮廓点 C,灰度值置为 255。

步骤 3　继续向右搜索直到当前像素点满足性质 2,将当前像素点记为该行轮廓终点 E,并置灰度值为 255。

步骤 4　从下一行包围框起点开始向右搜索,重复步骤 1 ~ 步骤 3 直至达到包围框右下角像素点。



(a) 目标mask(红色区域)　　(b) 目标mask灰度图

图 8　冠状缝早闭简化轮廓提取算法提取结果

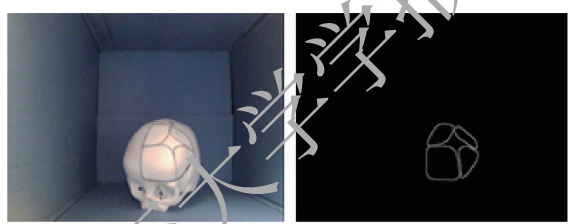
Fig. 8　Extraction result of simplified contour extraction algorithm for coronary craniosynostosis

2.2　轮廓提取结果

首先对图 5 和图 6 所示的目标 mask 灰度图进行二值化处理,然后使用上述简化轮廓提取算法进行目标区域轮廓提取。

图 8 和图 9 为简化轮廓提取算法的结果,图 10为使用 OpenCV 轮廓提取算法对矢状缝早闭手术切割方案提取的结果。可以看出,在同样提取完整切割方案的情况下,通过 Mask R-CNN 结合简化轮廓提取算法可以有效避免噪声的干扰,算法更加鲁棒。

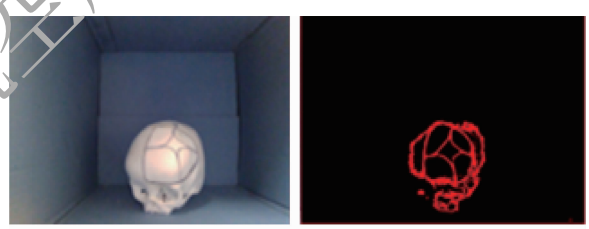
为验证上述简化轮廓提取算法的效率,以 VS Code 作为开发平台,在配置为 Intel i7 9700K 3.6 GHz 8 核 CPU,8 GB 内存,12 GB 独显的机器上对尺寸为 640 × 480 的图像使用上述简化轮廓提取算法。如图 11 所示,简化后的轮廓提取算法与未简化算法相比,总用时大幅缩减 50% 以上,在包围框占比小于 8% 的情况下,算法效率提升 80% 以上。



(a) 目标mask(彩色区域)　　(b) 目标mask灰度图

图 9　矢状缝早闭简化轮廓提取算法提取结果

Fig. 9　Extraction result of simplified contour extraction algorithm for sagittal craniosynostosis



(a) 目标mask(红色区域)　　(b) 目标mask灰度图

图 10　矢状缝早闭 OpenCV 轮廓提取算法提取结果

Fig. 10　Extraction result of OpenCV contour extraction algorithm for sagittal craniosynostosis

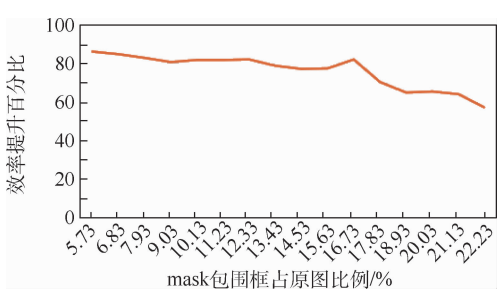


图 11　轮廓提取算法简化前后效率对比

Fig. 11　Efficiency comparison between original and simplified contour extraction algorithm

北京航空航天大学学报

北航学报 瞰阅

3 切割轨迹 3D 映射

3.1 颅骨点云获取

使用第 1 节深度学习方法和第 2 节简化轮廓提取算法可以获得 2D 图像上的切割方案像素坐标，为了实现切割方案的 3D 数字化，还需获取切割方案在颅骨点云上对应的 3D 坐标。在前期利用深度相机采集训练样本 2D 图像时，已经采集了颅骨 3D 打印模型的点云信息。获取点云信息的步骤如下：

步骤 1 特征描述子选择。选择尺度不变特征变换(scale-invariant feature transform，SIFT)[13]作为特征描述子来进行左右相机图像对应点的匹配，由于颅骨 3D 打印模型表面较为光滑，缺乏纹理，不利于特征点的检测和匹配，在深度相机系统中使用主动式投射结构光的方法增加场景纹理，进一步改善左右视图特征点匹配结果。

步骤 2 视差计算。深度相机视觉系统基于小孔成像原理，利用左右视图像素视差并基于三角形相似原理计算场景深度[14-15]，原理如图 12 所示。

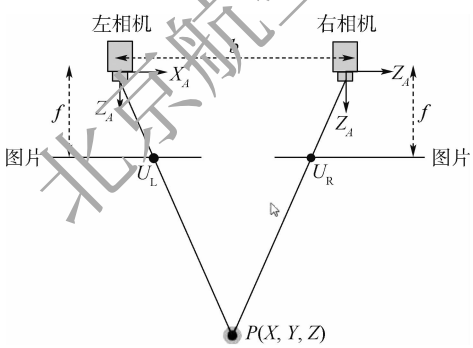


图 12 深度相机简单模型

Fig. 12 Simplified model of depth camera vision

点云深度计算公式为

$$X = b(U_L - U_0)/d \tag{8}$$

$$Y = b(V_L - V_0)/d \tag{9}$$

$$Z = bf/d \tag{10}$$

式中：X 为世界坐标点 P 在左相机坐标系下的横坐标值；Y 为世界坐标点 P 在左相机坐标系下的纵坐标值；Z 为世界坐标点 P 在左相机坐标系下的深度值；(U_L, V_L) 表示点 P 在左相机坐标系下成像点的坐标，(U_R, V_R) 表示点 P 在右相机坐标系下成像点的坐标，二者为对应点；(U_0, V_0) 表示左图像中心点在左相机坐标系下的坐标；d 为点 P 在左右相机坐标系下对应点的像素差异 $(U_L - U_R)$ 的绝对值；b 为基线长度；f 为像素化后的相机焦距值。

步骤 3 生成点云。遍历左相机采集图像，计算 2D 图像上所有像素点对应的点云体素 3D 坐标，至此便可以获得颅骨 3D 打印模型点云。

3.2 2D-3D 坐标映射

获取颅骨点云后，通过 2D-3D 的坐标映射便可以将上述 2D 图像上切割轨迹信息映射到颅骨点云上，利用深度相机组成的立体视觉系统同时获取 2D 图像和场景 3D 信息，通过格式转换，可以获得所需颅骨点云。

2D-3D 的坐标转换主要分为以下步骤：

步骤 1 对左相机采集到的 2D 图像使用深度学习方法和简化轮廓提取算法获得如图 8 和图 9 所示的轮廓提取结果。

步骤 2 对包围框内的每一像素 (U_i, V_i) 利用 3.1 节的方法获得其对应的 3D 坐标 (X_i, Y_i, Z_i)，并将对应点云体素的 RGB 标记为白色(255，255，255)，同时对切割轨迹像素坐标 (U_i, V_i) 对应的点云体素 (X_i, Y_i, Z_i) 的 RGB 标记为红色(255，0，0)，至此便可以实现坐标的 2D-3D 映射。

冠状缝早闭颅骨切割方案坐标 2D-3D 映射结果如图 13 所示，矢状缝早闭颅骨切割方案坐标 2D-3D 映射结果如图 14 所示。

为验证深度相机获取点云的精度，将轨迹点 Z 方向的定位误差作为评价指标，通过设置固定距离标准平面对该点云获取方法精度进行检测，

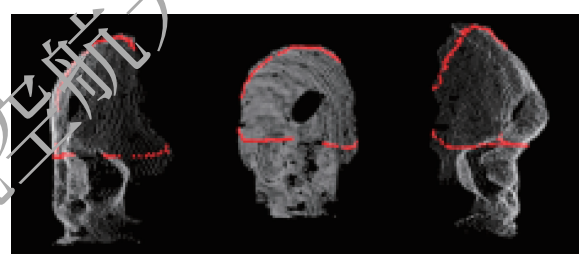


图 13 冠状缝早闭手术方案的 2D-3D 映射效果

Fig. 13 2D-3D mapping result of coronary craniosynostosis surgery plan

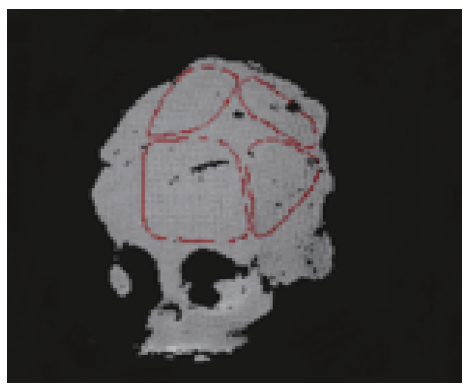


图 14 矢状缝早闭手术方案的 2D-3D 映射效果

Fig. 14 2D-3D mapping result of sagittal craniosynostosis surgery plan

北京航空航天大学学报

北航学报 赠阅

实验表明在 1 m 范围内轨迹点 Z 方向定位误差不超过 3 mm,符合实际操作要求。

4　测试病例颅骨上切割轨迹生成

基于以上轨迹提取方法构建矢状缝早闭手术方案模板库和冠状缝早闭手术方案模板库,模板库中病例均为首都医科大学附属北京天坛医院资深医生从现有病例中挑选的具有代表性的典型案例。其中矢状缝早闭模板病例包含 8 位患者数据(均为男性),月龄 6 月到 30 月均匀分布,相应病例的颅骨模型如图 15 所示;冠状缝早闭模板病例包含 8 位患者数据(2 名男性,6 名女性),月龄7 月到 14 月均匀分布,相应病例的颅骨模型如图 16 所示。

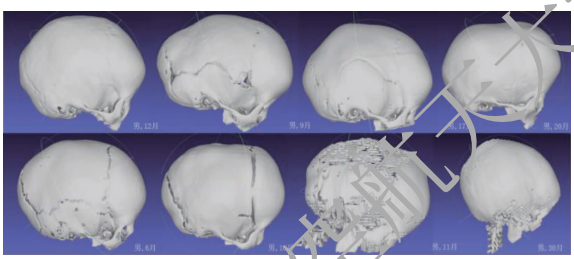


图 15　矢状缝早闭模板病例颅骨外观

Fig. 15　Skull appearance of sagittal craniosynostosis template case

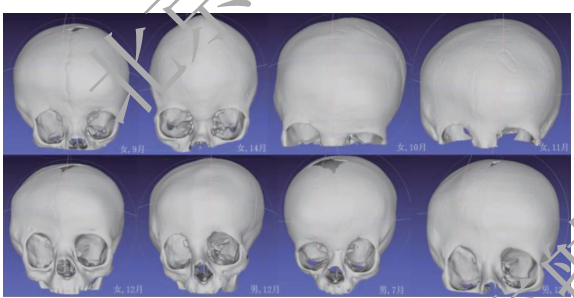


图 16　冠状缝早闭模板病例颅骨外观

Fig. 16　Skull appearance of coronary craniosynostosis template case

依托上述构建的手术方案模板库,在对测试病例进行手术方案设计时,先采用基于点云配准的方法依次计算测试病例与模板库病例间的相似度,二者点云配准误差越小,相似度就越高。将配准误差最小的模板病例对应的手术方案设计作为测试病例的手术方案计算初值,并采用模板匹配的方法快速生成测试病例的手术切割方案。

针对测试病例的具体操作如下:

1)利用测试病例原始颅骨 CT 数据重建 3D 模型,获得测试病例颅骨模型 P_t,使用点云配准方法依次计算测试病例颅骨模型 P_t 与模板库中任一模板病例颅骨模型 P_i 的配准误差 E_i,并得到测试病例模型与模板病例模型之间的旋转矩阵 $\boldsymbol{R}_i$ 和平移矩阵 $\boldsymbol{T}_i$。

2)将所有的配准误差进行降序排列,选择配准误差最小的模板点云 P_m 作为测试病例手术方案设计的参考模板。

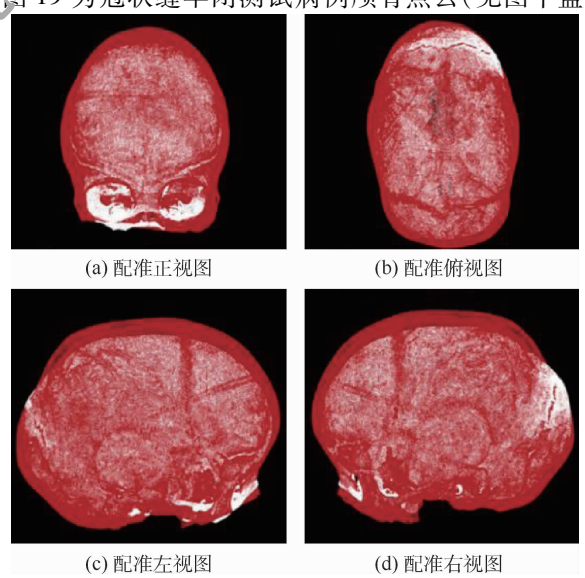
3)利用 P_m 与测试病例颅骨模型 P_t 之间的旋转矩阵 $\boldsymbol{R}_m$ 和平移矩阵 $\boldsymbol{T}_m$,将模板点云 P_m 上提取并数字化的颅骨切割方案 3D 坐标 $\boldsymbol{C}_i$ 进行变换,得到测试病例颅骨模型 P_t 对应的切割方案坐标 $\boldsymbol{C}''_i = \boldsymbol{R}_m \boldsymbol{C}_i + \boldsymbol{T}_m$,依次遍历模板病例上所有切割方案坐标,最终在测试病例颅骨模型 P_t 上生成对应的颅骨切割方案。

由于测试病例颅骨 P_t 与最终选定的模板病例颅骨 P_m 之间仍存在细微差别,采用上述方法直接生成的测试病例颅骨切割方案坐标 $\boldsymbol{C}''_i$ 可能偏离测试病例颅骨点云。为解决这一问题,对上述生成的测试病例切割方案坐标是否落在测试病例颅骨表面进行判断并更新,具体操作如下:

1)对任一测试病例切割方案坐标 $\boldsymbol{C}''_i$,遍历测试病例颅骨点云 P_t 的所有点,并计算两空间点之间的欧氏距离。

2)选择测试病例颅骨点云 P_t 上与 $\boldsymbol{C}''_i$ 欧氏距离最小的点作为实际切割方案坐标,并更新 $\boldsymbol{C}''_i$。

图 17 为矢状缝早闭测试病例颅骨点云(见图中白色区域)与匹配程度最高的模板颅骨点云(见图中红色区域)之间的配准结果,对应点集中所有对应点对的平均空间距离误差小于 1.8 mm。图 18 为矢状缝早闭测试病例基于模板匹配方法生成的切割方案,其中红色点为切割方案轨迹点。图 19 为冠状缝早闭测试病例颅骨点云(见图中蓝



(a) 配准正视图　(b) 配准俯视图

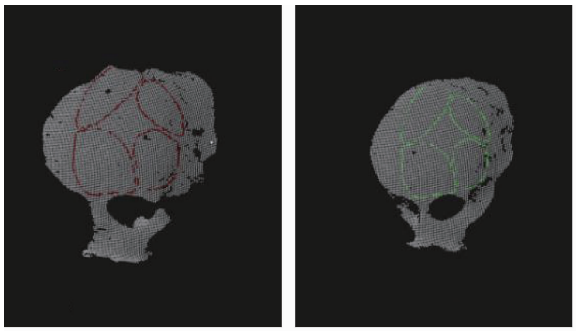
(c) 配准左视图　(d) 配准右视图

图 17　矢状缝早闭测试病例颅骨点云与模板颅骨点云配准结果

Fig. 17　Registration result of skull point cloud of sagittal craniosynostosis testing case with that of templates

北京航空航天大学学报

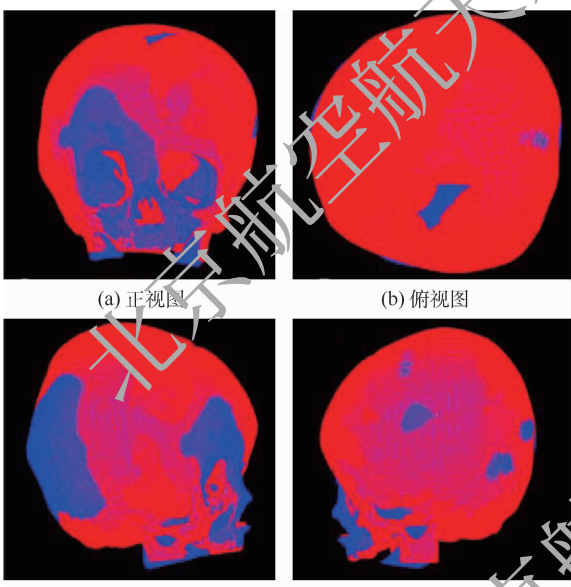
色区域)与匹配程度最高的模板颅骨点云(见图中红色区域)之间的配准结果,对应点集中所有对应点对的平均空间距离误差小于 1.3 mm。图 20



(a) 模板颅骨切割方案 (b) 矢状缝早闭测试病例切割方案

图 18 矢状缝早闭测试病例基于模板匹配生成的初步切割方案

Fig. 18 Preliminary cutting plan of sagittal craniosynostosis testing case based on template matching

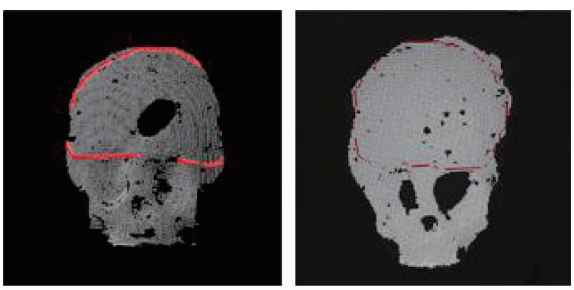


(a) 正视图 (b) 俯视图

(c) 左视图 (d) 右视图

图 19 冠状缝早闭测试病例颅骨点云与模板颅骨点云配准结果

Fig. 19 Registration result of skull point cloud of coronary craniosynostosis testing case with that of templates



(a) 模板颅骨切割方案 (b) 冠状缝早闭测试病例切割方案

图 20 冠状缝早闭测试病例基于模板匹配生成的初步切割方案

Fig. 20 Preliminary cutting plan of coronary craniosynostosis testing case based on template matching

为冠状缝早闭测试病例基于模板匹配方法生成的切割方案,其中红色点为轨迹点。

5 结 论

本文提出了一种深度学习结合点云处理的方法,实现了手术方案设计中的切割轨迹自动提取。通过该方法提取样本切割轨迹,分别构建了针对矢状缝早闭和冠状缝早闭的颅骨切割方案数字化模板库。通过点云配准技术实现模板手术方案的迁移,为新病例快速生成个性化手术方案。

本文方法实现了颅缝早闭手术切割方案的规范化和标准化,同时缩短了术前方案设计时间和术中操作时间。

后续将研究如何提升模板库的数量和质量,优化模板匹配的切割轨迹,推动该方法逐步应用于临床。

参考文献(References)

[1] 沈卫民,王刚,吴玉新,等. 先天性颅缝早闭的颅骨成形——附 6 例报道[J]. 实用美容整形外科杂志,2002,13(4):180-183.
SHEN W M,WANG G,WU Y X,et al. Plastic skull for craniosynostosis: A report of 6 cases[J]. Journal of Practical Aesthetic and Plastic Surgery,2002,13(4):180-183(in Chinese).

[2] 冯胜之,张涤生. 先天性颅缝早闭症的治疗[J]. 中华整形烧伤外科杂志,1995,11(6):406-411.
FENG S Z,ZHANG D S. Surgical treatment of craniosynostosis [J]. Chinese Journal of Plastic Surgery,1995,11(6):406-411 (in Chinese).

[3] KLEESIEK J,URBAN G,HUBERT A,et al. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping[J]. NeuroImage,2016,129:460-469.

[4] CHILAMKURTHY S,GHOSH R,TANAMALA S,et al. Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study[J]. Lancet,2018,392(10162):2388-2396.

[5] HE K,GKIOXARI G,DOLLAR P,et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE Press,2017:2961-2969.

[6] HE K,ZHANG X Y,REN S Q,et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press,2016:770-778.

[7] LIN T Y,DOLLAR P,GIRSHICK R,et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2017:936-944.

[8] REN S,HE K M,GIRSHICKR R,et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(6):1137-1149.

北航学报 赠阅

[9] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.

[10] ROSENFELD A, KAK A C. Digital picture processing[M]. New York: Academic Press, 1976: 113-116.

[11] ROSENFELD A. Connectivity in digital pictures[J]. Journal of the ACM, 1970, 17(1): 146-160.

[12] SUZUKI S, BE K. Topological structural analysis of digitized binary images by border following[J]. Computer Vision Graphics and Image Processing, 1985, 30(1): 32-46.

[13] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[14] MARR D, POGGIO T. A computational theory of human stereo vision[J]. Proceedings of the Royal Society of London, 1991, 204(1156): 301-328.

[15] HEIKO H. Stereo processing by semi-global matching and mutual information[J]. IEEE Transactions on Pattern Analysis and Machine Intelligenc, 2008, 30(2): 328-341.

Skull cutting plan generation method in craniosynostosis surgery

LUO Yangyu[1,*], HE Jiabin[1], XIE Dongsheng[1], LU Shenyu[1], GONG Jian[2]

(1. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;
2. Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China)

Abstract: To standardize the plan generation of craniosynostosis surgery, we propose a skull cutting trajectory generation method which combines deep learning, stereo vision and point cloud processing technology to establish cutting plan template library and generate cutting trajectory for new cases. The proposed method, for the first time, uses the deep learning to segment the external surface of skull. First, it takes the advantage of Mask R-CNN to detect and segment the surgical cutting area. Then, a simplified contour extraction algorithm is explored to extract the surgical cutting trajectory. After that, the point cloud processing technology is used to map the surgical cutting trajectory to three-dimensional coordinates and realize the automatic extraction of cutting trajectory. Finally, a template library with typical cases is established and the cutting plans of new cases are automatically generated by template matching method. The experiment shows that the proposed trajectory extraction method can detect the skull surgical cutting trajectory accurately and efficiently and conduct three-dimensional mapping of trajectory coordinate. The depth measurement error of point cloud is less than 3 mm, which meets the clinical available standard. Using the template matching method, the new case's cutting trajectory can be generated effectively, which conforms to surgical plan of senior doctors.

Keywords: craniosynostosis; trajectory extraction; deep learning; depth camera; point cloud processing

北京航空航天大学学报

Received: 2020-11-12; **Accepted**: 2021-02-10; **Published online**: 2021-03-30 11:09

URL: kns.cnki.net/kcms/detail/11.2625.V.20210329.1516.001.html

Foundation item: National Key R & D Program of China (2017YFE0121200)

* **Corresponding author**. E-mail: yangyu.luo@ia.ac.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0613

电力场景下基于无人机视觉的运动目标追踪方法

全下文载

冯雪，杜猛俊*，向新宇，钱锦，张敏
（国网浙江省电力有限公司杭州供电公司，杭州 310000）

摘 要：随着人工智能技术的发展，面向电力系统的运动目标追踪技术逐渐得到关注，现有方法虽有一定成效，但是大多基于固定摄像头的监控视频录制，不能灵活追踪运动目标，当运动目标离开摄像头视野时，存在运动目标丢失问题。为此，利用无人机设备，并基于深度学习和核相关滤波技术，提出了一个电力场景下基于无人机视觉的运动目标追踪方法（MTTS_UAV）。所提方法采用改进的目标追踪方法与目标检测方法相结合的方式来追踪运动目标隐患，并引入2种无人机飞行控制模块：启发式和数据驱动式，使得无人机的飞行速度和方向可以根据目标移动情况自适应地调节。在真实变电站的安全帽人员数据集上进行了大量实验，对所提方法的追踪效果进行评估，结果表明：所提方法在真实数据集上的平均像素误差（APE）和平均重叠率（AOR）分别可达到2.37和0.67，验证了方法的有效性。

关 键 词：目标追踪；目标检测；深度学习；无人机视觉；电力场景
中图分类号：TP37
文献标志码：A **文章编号**：1001-5965(2022)04-0586-09

近年来，随着中国经济的飞速发展，电力建设的规模也随之逐年扩大。据国家电网统计，截至2018年底，110 kV及以上输电线路长度已达103.34×10^4 km。这对电网运行的稳定性和安全性带来了更高挑战，对变电、输电、配电等电力系统各环节的巡检提出了更高要求。传统的人工巡检方式由于巡检周期长、效率低、无法覆盖复杂地形和恶劣环境等问题，已无法满足当前电网的巡检要求。随着人工智能和计算机视觉领域的蓬勃发展，智能电力巡检方式逐渐成为当前电力领域的研究热点。

事实上，由于输电线路和杆塔等电力设备长期暴露在户外，常常受到各种鸟类、风筝、气球或者空中漂浮塑料袋等移动物体的干扰和威胁，以及工作人员进行维修攀爬等作业时带来的安全隐患。因此，对运动目标隐患的检测和追踪是电力传输系统亟待解决的重要问题。现有针对电力传输系统监控视频流中运动物体的检测方法主要是基于对固定摄像头拍摄的视频帧进行目标检测，虽然该方法可以取得不错的准确率，但是存在两方面局限：①由于需要对各视频帧进行目标检测，检测的实时性无法得到保证。②由于固定摄像头的视野有限，对于运动物体常常存在追踪丢失问题，即运动物体移出视野外。因此，为提高目标追踪的准确率和实时性，本文旨在基于无人机视觉，并结合目标检测和目标追踪技术对电力场景中运动目标进行实时追踪。该方法具有操作灵活、效率高、对环境要求低等优势，有利于解决电力领域的实际应用问题。

然而，在电力场景下，基于无人机视觉的运动目标追踪技术的研究存在诸多挑战。①为了保证方法的实时性，现有基于目标检测和目标追踪相结合的方法，通常假设运动目标的移动速度较慢，即目标的前后帧位移较小。因此，如何有效追踪各

收稿日期：2020-11-03；**录用日期**：2020-11-12；**网络出版时间**：2021-03-02 12:10
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210301.1821.004.html
基金项目：国网浙江省电力有限公司科技项目(5211HZ19014U)
* **通信作者**. E-mail：du_mengjun@zj.sgcc.com.cn

引用格式：*冯雪，杜猛俊，向新宇，等. 电力场景下基于无人机视觉的运动目标追踪方法[J]. 北京航空航天大学学报，2022，48(4)：586-594. FENG X，DU M J，XIANG X Y，et al. UAV-vision-based moving target tracking scheme in electric scenario[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：586-594 (in Chinese).*

北京航空航天大学学报

北航学报 网阅

种运动速度的物体,降低追踪丢失的问题是本文工作的首要挑战。②为提高方法的准确性,保证运动目标始终保持在无人机视野的中心,如何有效控制无人机的飞行速度和方向是另一个重要挑战。

为了解决以上挑战,本文设计了一个电力场景下基于无人机视觉的运动目标追踪方法(moving target tracking scheme based on UAV vision for electric scenario, MTTS_UAV),包括追踪模块、飞行控制模块两部分。对于追踪模块,本文基于目标检测方法 YOLOv3[1] 和目标追踪方法核相关滤波(kernel correlation filter, KCF)[2] 对隐患目标进行实时低功耗定位与追踪;对于飞行控制模块,本文提出了 2 种解决办法:启发式飞行控制模块和数据驱动的飞行控制模块。其中,启发式飞行策略旨在针对不同速度的移动物体自适应调节无人机的飞行速度和方向,进而使隐患目标一直处在画面中心;数据驱动的飞行策略根据预先提取的无人机追踪目标位移序列,基于门控循环神经网络(gated recurrent neural network, GRU)[3] 预测每一帧无人机的位移。

本文提出了一个电力场景下基于无人机视觉的运动目标追踪方法。所提方法通过结合目标检测和目标追踪方法,使用无人机可以对电力场景下的运动目标隐患进行准确地实时追踪。特别地,本文提出了 2 种飞行控制模块:启发式飞行控制模块和数据驱动的飞行控制模块。这些模块可以使无人机自适应地调整飞行速度和方向。此外,本文基于电力场景中的真实数据集,开展了大量实验。相关结果证实了本文所提模型的有效性。

1　相关工作

1.1　目标检测

目标检测任务旨在找出图像中所有感兴趣的目标(物体),确定它们的类别和位置,这是计算机视觉领域的核心问题之一,也是电力场景中智能巡检的重要组成部分。传统的目标检测方法大多基于手工特征所构建。例如,2001 年 Viola 和 Jones[4] 建立了一种称为积分图像的特征,在极为有限的计算资源下第 1 次实现了人脸的实时检测,即 Viola-Jones 目标检测器,速度是同期检测方法的几十甚至上百倍;之后,文献[5]提出了方向梯度直方图特征的 HOG 行人检测器,为了兼顾速度和性能,采用的分类器通常为线性分类器[6] 或级联决策分类器[7] 等。由于深度学习的快速发展和广泛运用,许多检测能力和实时性都很优秀的基于深度学习的算法被提出,如 Fast R-CNN[8]、Faster R-CNN[9]、YOLOv3[1]、SSD[10] 等。面向智能电网领域,这些基于深度学习的目标检测方法也取得了一些成绩。例如,吴冬梅等[11] 针对作业人员不佩戴安全帽、越界操作等违规行为,提出了一种基于改进的 Faster R-CNN 的安全帽佩戴检测及身份识别方法。董召杰[12] 针对自动巡检的精度和速度需求,提出了基于 YOLOv3 的电力线多种关键部件的实时检测方法,构建了包含 5 种电力线关键部件的数据集,并在该数据集上应用 YOLOv3 实现了高精度实时检测。虽然现有方法在面向电力领域的目标检测任务中取得了一定成功,但是其在电力场景中运动目标追踪任务中的应用尚未被探索。本文利用目前先进的目标检测方法纠正目标追踪过程中的目标丢失或追错问题,提高了电力系统中运动目标追踪的准确性和实时性。

1.2　目标追踪

目标追踪任务旨在给定目标的初始位置,计算出后续每帧图像中目标的位置信息。国内外的科研工作者对该问题进行了广泛研究。例如,Xiang[13] 提出了一种单 PTZ 摄像头下的快速运动物体的实时跟踪方法,引入了卡尔曼滤波算法和感兴趣区域(region of interest, ROI)来达到跟踪快速物体的效果。Zhang 等[14] 提出了一个基于单 PTZ 摄像头的视频监控系统,利用背景差分的方法检测出运动目标。然而这些传统的静态摄像头下的目标追踪,受限于摄像头视角的有限性,难以对感兴趣目标进行大范围和长时间的跟踪。近年来,随着无人机技术的发展,基于无人机进行目标追踪被广泛应用于航拍、电力巡检、物流、军事侦察、目标打击等领域。例如,Dong 和 Zou[15] 介绍了基于视觉的无人机目标追踪方法,采用数传将摄像机采集到的图像数据发送到远端进行处理后再发送指令给无人机。林勇等[16] 介绍了一种基于合作目标的无人机目标追踪方法,能够实现准确跟踪,主要用于危险物排除。

尽管上述方法取得了突破性进展,但大多只考虑了改进追踪方法,而没有采用一定的控制策略以控制无人机飞行,使其达到变速飞行的效果,无法在复杂电力场景中对运动隐患目标进行有效追踪。因此,本文提出了一种电力场景下基于视觉的运动目标追踪方法,通过目标检测和目标追踪相结合的方式,再引入启发式和数据驱动的飞行控制模块,采用一定的控制策略控制无人机飞行,对电力系统中的运动隐患实时追踪,保障电力系统安全运行。

北京航空航天大学学报

北航学报 网络阅

2 模 型

2.1 问题定义

假设无人机拍摄的实时视频流表示为 $\boldsymbol{v}=[I_1,I_2,\cdots,I_N]$，$I_j$ 表示视频流中的第 j 帧图像，N 为视频的帧数。本文使用 $\boldsymbol{t}_0=(x_0,y_0,w_0,h_0)$ 表示待追踪的运动目标的初始位置，其中，(x_0,y_0) 表示图像中初始目标框的左上角坐标，w_0 和 h_0 分别为目标框的宽度和高度。本文所提出的电力场景下基于无人机视觉的运动目标追踪方法，旨在基于给定的目标初始位置 $\boldsymbol{t}_0$，通过实时调整无人机的飞行速度和方向，也就是无人机的帧间位置偏移量，追踪目标在视频流中各帧的位置，即 $\boldsymbol{t}_j=(x_j,y_j,w_j,h_j)$，$j=1,2,\cdots,N$，其中，$(x_j,y_j)$ 表示运动目标框在第 j 帧图像中的左上角坐标，w_j 和 h_j 分别为目标框的宽度和高度。

2.2 追踪模块

为了准确实时地对电力场景下运动的隐患目标进行追踪，本文采用目标追踪和目标检测相结合的方式，在目标追踪方法追错或追丢的情况下，能够通过目标检测方法及时纠正或找回，进而保障追踪的准确性。具体而言，本文有机结合了简洁高效的目标追踪方法 KCF 和目标检测方法 YOLOv3，其具体模型结构如图 1 所示。值得一提的是，本文所提框架也适用于其他的目标检测和目标追踪方法。

2.2.1 KCF 目标追踪

经典的 KCF 追踪算法，通过使用循环移位得到的循环矩阵来采集正负样本，利用循环矩阵在傅里叶空间可对角化的性质，将矩阵的运算转化为元素的点乘，从而降低了运算量，使得算法满足实时性要求，在众多目标追踪相关任务中取得了出色的效果[17-20]。为了保证追踪效果，本文选择 KCF 常见的提高算法判别能力的高斯核函数[2]，其表达式为

$$k^{xx'}=\exp\left[-\frac{1}{\sigma^2}\left(\|\boldsymbol{x}\|^2+\|\boldsymbol{x}'\|^2-2\mathcal{F}^{-1}\hat{\boldsymbol{x}}^*\odot\hat{\boldsymbol{x}}'\right)\right] \tag{1}$$

式中，$\boldsymbol{x}$ 和 $\boldsymbol{x}'$ 分别为 2 个任意样本；σ 为标准差；$\mathcal{F}^{-1}$ 代表傅里叶逆变换；“$\odot$”表示内积。

根据 KCF 算法，可以得到待追踪运动目标在每一帧的坐标位置 $\boldsymbol{t}_j=(x_j,y_j,w_j,h_j)$。

为了降低追踪目标丢失的可能性，本文提出采用 YOLOv3 来重新捕获可能丢失的运动目标，以提高模型效果。如果基于 KCF 算法预测的相邻 2 帧的目标区域 $\boldsymbol{t}_j$ 和 $\boldsymbol{t}_{j-1}$ 之间的颜色直方图相似性小于一定阈值 γ_1，或者当前帧未检测到运动目标，则触发目标检测方法 YOLOv3 来重新选定目标区域，以开展后续的目标追踪。本文相似性采用最直观的相邻帧直方图相似性。令 $\boldsymbol{H}_{j-1}$ 和 $\boldsymbol{H}_j$ 分别表示第 $j-1$ 和第 j 帧的目标区域 $\boldsymbol{t}_{j-1}$ 和 $\boldsymbol{t}_j$ 的颜色直方图向量。当前帧 I_j 和前一帧 I_{j-1} 的目标区域的直方图相似性 S 的计算如下：

$$S=\frac{\sum_\alpha(\boldsymbol{H}_j(\alpha)-\overline{\boldsymbol{H}}_j)(\boldsymbol{H}_{j-1}(\alpha)-\overline{\boldsymbol{H}}_{j-1})}{\sqrt{\sum_\alpha(\boldsymbol{H}_j(\alpha)-\overline{\boldsymbol{H}}_j)^2\sum_\alpha(\boldsymbol{H}_{j-1}(\alpha)-\overline{\boldsymbol{H}}_{j-1})^2}} \tag{2}$$

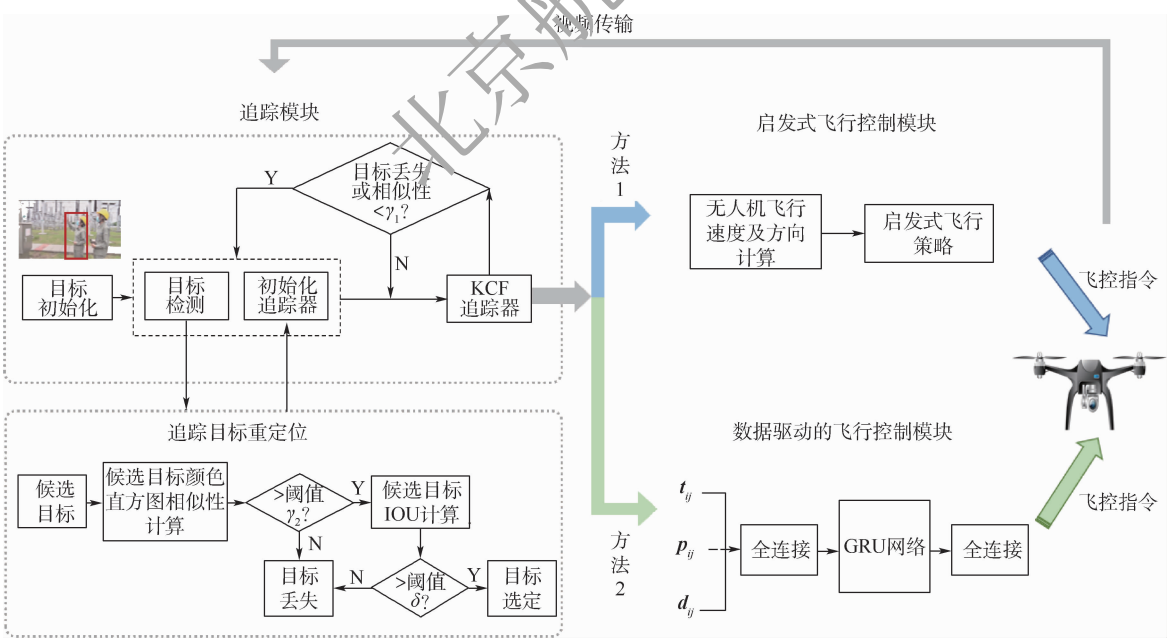


图 1 电力场景下基于无人机视觉的运动目标追踪方法结构

Fig. 1 Framework of UAV-vision-based moving target tracking scheme in electric scenario

北航学报 赠阅

式中：$\overline{\boldsymbol{H}}_j$ 和 $\overline{\boldsymbol{H}}_{j-1}$ 为直方图均值；α 为直方图的区间个数。

2.2.2　追踪目标重定位

由于 YOLOv3 借鉴残差网络结构，形成了更深的 Darknet-53 网络结构，以及采用多尺度检测，提升了 mAP 及小物体检测效果，其在目标检测的诸多相关任务中得到了广泛应用并取得了优异的检测效果[21-22]。因此，本文采用 YOLOv3 网络结构 $\mathcal{Y}$ 来对运动目标进行目标检测，以弥补追踪过程中目标丢失的问题。YOLOv3 算法可以描述为

$$\hat{\boldsymbol{t}}_j = \mathcal{Y}(\boldsymbol{I} \mid \boldsymbol{\theta}_{\mathcal{Y}}) \tag{3}$$

式中：$\boldsymbol{\theta}_{\mathcal{Y}}$ 为 YOLOv3 网络 $\mathcal{Y}$ 的参数；$\hat{\boldsymbol{t}}_j = (\hat{x}_j, \hat{y}_j, \hat{w}_j, \hat{h}_j)$ 作为算法输出表示基于 YOLOv3 检测到的运动目标在第 j 帧图像中的坐标位置。

此时，根据 YOLOv3 预测的当前帧目标区域 $\hat{\boldsymbol{t}}_j$ 与 KCF 追踪算法预测的前一帧目标区域 $\boldsymbol{t}_{j-1}$ 的相似性是否大于阈值 γ_2，判定当前 YOLOv3 检测到的目标区域 $\hat{\boldsymbol{t}}_j$ 是否应当被选作新的目标，供后续 KCF 追踪。为了增强模型的准确性，不仅考虑了基于直方图的相似度，也考虑了当前帧和前一帧目标框的空间相似度，即交并比 IOU，设其阈值为 δ，计算公式如下：

$$\mathrm{IOU} = \frac{\boldsymbol{t}_{j-1} \cap \hat{\boldsymbol{t}}_j}{\boldsymbol{t}_{j-1} \cup \hat{\boldsymbol{t}}_j} \tag{4}$$

具体的目标重定位流程如图 1 中的追踪目标重定位模块所示。本文所提的追踪模块的具体算法如算法 1 所示。

算法 1　追踪模块。

输入：待追踪目标的初始坐标位置 $\boldsymbol{t}_0$ 及无人机传输的视频流 $\boldsymbol{v} = (I_1, I_2, \cdots, I_N)$。

输出：每帧图像中运动目标的坐标位置 $\boldsymbol{t}_j$ 及前后 2 帧图像中目标中心点的坐标差值 $\boldsymbol{d}_t = (d_x, d_y)$。

1.　初始化 KCF 算法和 YOLOv3 网络的参数 $\boldsymbol{\theta}_{\mathcal{Y}}$、$\boldsymbol{\theta}_K$

2.　REPEAT：

3.　　初始化 KCF 追踪器

4.　　FOR from I_1 to I_N

5.　　IF KCF 追踪器对当前帧 I_j 输出是目标丢失或者颜色直方图相似性 S 小于 γ_1

6.　　根据式(1)得到当前帧 I_j 的候选目标的坐标 $\boldsymbol{t}_j$

7.　　根据式(2)和式(3)得到当前帧与前一帧图像 I_{j-1} 的目标颜色直方图相似性 S

8.　　IF $S > \gamma_2$

9.　　根据式(4)，计算当前帧与前一帧的目标框 IOU

10.　　IF IOU $> \delta$

11.　　目标选定

12.　　ELSE

13.　　目标丢失

14.　　ELSE

15.　　追踪器得到的隐患目标的坐标 $\boldsymbol{t}$ 输入到飞行控制模块

16.　　END FOR

17. UNTIL：代价函数收敛

2.3　飞行控制模块

基于 2.2 节所述追踪模块，可以获得当前帧运动目标的位置。为了增强追踪效果，使待追踪目标保持在无人机的视野中心，本文引入了 2 种飞行控制模块，以引导无人机以合适的速度和方向追踪目标，具体为启发式飞行控制模块和数据驱动的飞行控制模块。

2.3.1　启发式飞行控制模块

很显然，在控制无人机追踪时，不能单纯地依靠单一速度来追踪不同的物体。因此，本文提出无人机应能够自适应地追踪具有不同速度的运动目标。直观上说，无人机的速度应当和运动物体的速度保持一致，以防止无人机速度过快带来的摇摆情况和速度过慢引起的目标丢失等问题。为此，本文首先按照式(5)计算无人机速度 v_j，即

$$\begin{cases} v_j = \psi(d(c_{j-1}, c_j)) \\ \psi(d) = wd + b \\ d(c_{j-1}, c_j) = \sqrt{(\bar{x}_{j-1} - \bar{x}_j)^2 + (\bar{y}_{j-1} - \bar{y}_j)^2} \end{cases} \tag{5}$$

式中：w、b 分别为权重偏差；$c_{j-1} = \{\bar{x}_{j-1}, \bar{y}_{j-1}\}$ 和 $c_j = \{\bar{x}_j, \bar{y}_j\}$ 分别表示前后 2 帧图像 I_{j-1} 和 I_j 中目标框的中心点坐标；$d(c_{j-1}, c_j)$ 为 2 个中心点之间的距离，即运动目标在当前帧的位置相对于前一帧偏移的距离，该距离越大，说明物体移动的速度越快，无人机的速度配置也应当加快。为了保证无人机和运动目标保持在同一视野之中，以方便仿真实验，本文采用了线性函数 $\psi(d)$ 对前后帧目标框中心点之间的距离进行了缩放。考虑到真实场景中无人机大多不能同时在多个方向上移动，本文设计了如下启发式飞行策略：如果 $\bar{x}_j - \bar{x}_{j-1} > 0$，则使无人机沿其 x 轴的正半轴方向以 v_j 的速度移动；反之，则沿 x 轴的负半轴方向以 v_j 的速度移动。倘若一次单方向移动，无法使运动目标在无人机视野中的 x 轴上保持居中，则继续计算 v_{j+1}，并进行下一次的单方向移动。该策略在无人机视野中的 y 轴方向上同样适用。

2.3.2　数据驱动的飞行控制模块

2.3.1 节所述启发式飞行控制模块受人为经

北京航空航天大学学报

北航学报 赠阅

验影响较大,且需要多次无人机移动。因此,本文设计了数据驱动的飞行控制模块。首先,使用预训练的追踪方法在各视频上进行追踪。在追踪过程中,令用来模拟无人机的仿真窗口在视频的每一帧上进行随机移动,直到无法追踪到运动目标为止。通过多次随机移动,最终选择长度最长的若干序列所对应的位移序列$\{\boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_M\}$作为真实标签来训练该飞行控制模块。其中,$\boldsymbol{Q}_i=\{\boldsymbol{q}_{i1},\boldsymbol{q}_{i2},\cdots,\boldsymbol{q}_{i(Q_i-1)}\}$,$\boldsymbol{q}_{ij}=[\boldsymbol{q}_{ij}^x,\boldsymbol{q}_{ij}^y]$表示无人机在该序列的第$j$帧到第$j+1$帧间发生的位移,$Q_i$表示第$i$个序列的长度。针对第$i$个序列,可以通过训练好的追踪模块,得到其运动目标在第j帧的位置信息$\boldsymbol{t}_{ij}$。然后,利用 GRU 网络[3],根据运动目标的位置信息$\boldsymbol{t}_{ij}$,对应无人机仿真窗口位置信息$\boldsymbol{p}_{ij}$,以及前后相邻帧目标中心点的位移$\boldsymbol{d}_{ij}=(\boldsymbol{d}_{ij}^x,\boldsymbol{d}_{ij}^y)$,$\boldsymbol{d}_{ij}^x=\bar{x}_{ij}-\bar{x}_{ij-1}$,$d_{ij}^y=\bar{y}_{ij}-\bar{y}_{ij-1}$,来预测无人机的当前帧飞行位移$\tilde{\boldsymbol{d}}_{ij}=(\tilde{\boldsymbol{d}}_{ij}^x,\tilde{\boldsymbol{d}}_{ij}^y)$。

具体公式如下:

$$\begin{cases}\boldsymbol{z}_{ij}=\boldsymbol{t}_{ij}\|\boldsymbol{p}_{ij}\|\boldsymbol{d}_{ij}\\ \boldsymbol{z}'_{ij}=\boldsymbol{W}_1\boldsymbol{z}_{ij}+\boldsymbol{b}_1\\ \boldsymbol{h}_{ij}=\mathcal{G}(\boldsymbol{z}'_{ij},\boldsymbol{h}_{(j-1)})\\ \tilde{\boldsymbol{d}}_{ij}=\boldsymbol{W}_2\boldsymbol{h}_{ij}+\boldsymbol{b}_2\end{cases}\tag{6}$$

式中:"‖"表示向量的串联操作;$\mathcal{G}$表示 GRU 网络;$\boldsymbol{W}_1$、$\boldsymbol{W}_2$和$\boldsymbol{b}_1$、$\boldsymbol{b}_2$分别为全连接网络的权重和偏置参数;$\boldsymbol{h}_{ij}$为 GRU 网络中隐藏层向量表征。

采用均方误差(mean square error, MSE)损失,代价函数可以表示为

$$L=\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{Q_i}\|\boldsymbol{q}_{ij}-\tilde{\boldsymbol{d}}_{ij}\|^2\tag{7}$$

式中:M为位移序列总数。

3 实验结果与分析

3.1 数据集

为全面验证本文提出的面向电力场景中基于无人机视觉的运动目标追踪方法的效果和实时性,采用真实变电站的数据集进行了充分实验,其中,真实变电站数据集共计 24 个电力场景佩戴安全帽人员追踪视频。

为训练数据驱动的飞行控制模块,对每个视频进行 100 次随机控制追踪。每个视频都可以获得 100 个长短不一的目标序列和对应的无人机位置序列。随后,对所有获得的视频序列根据序列长度进行排序,挑选出长度最长的 1 000 个序列作为训练样本。由于视频序列长度较长,以及考虑无人机控制的稳定性,进一步对这些序列进行了截取。具体地,将序列截取成长度为10 s的序列。最终,获得了 1 354 个序列数据。

3.2 实验设置

使用的视频流中的所有图像帧经过预处理后,大小均为 960×540。对于真实变电站安全帽人员数据集进行划分,其中 80% 作为训练集,10% 作为验证集,10% 作为测试集。采用网格搜索策略来决定网络中超参数的最优值,同时根据已有经验,分别设置了追踪模块中的颜色直方图相似性、追踪目标重定位模块中的颜色直方图相似性、追踪模块中的空间相似度阈值,即$\gamma_1=0.7$,$\gamma_2=0.1$,$\delta=0.01$。通过多次迭代训练优化模型,得到最佳效果的本文方法模型。对于数据驱动的飞行控制模块,使用 PyTorch 构建模型并完成训练。其中,GRU 与全连接层的隐含状态个数均设置为 10,学习效率设置为 0.001,优化方法使用随机梯度下降法,通过多次迭代训练优化模型。

3.3 评价指标

为进一步评估本文方法在电力场景中对运动隐患目标的追踪效果,采用目标追踪任务中常用的评价指标——平均像素误差(average pixel error, APE)、平均重叠率(average overlap rate, AOR)、平均重启次数及平均连续追踪帧数。

根据 OTB 标准[23],平均像素误差指的是预测目标中心位置与真实目标中心位置的平均欧氏距离,平均重叠率指的是预测目标框和真实目标框面积的平均交并比,其具体计算公式如下:

$$\mathrm{APE}=\sqrt{(a_j-\hat{a}_j)^2+(b_j-\hat{b}_j)^2}\tag{8}$$

式中:(a_j,b_j)和$(\hat{a}_j,\hat{b}_j)$分别为第j帧真实目标和预测目标中心点的坐标位置。

$$\mathrm{AOR}=\frac{A_j\cap\hat{A}_j}{A_j\cup\hat{A}_j}\tag{9}$$

式中:A_j和$\hat{A}_j$分别为第j帧真实目标框和预测目标框的面积。

另外,在视频跟踪处理的一段序列中,当目标丢失或者颜色直方图相似性低于阈值γ_1(本文设置的γ_1为 0.7),则判定为跟踪失败,并重新初始化目标框。因此,根据 VOT2013[24]标准,本文采用平均重启次数来评价追踪器跟踪目标的稳定性。追踪器的鲁棒性定义如下:

$$\rho=\frac{1}{N_{\mathrm{rep}}}\sum_{k=1}^{N_{\mathrm{rep}}}F(k)\tag{10}$$

式中:$F(k)$为追踪器在第k次重复追踪中失败的

北京航空航天大学学报

次数;N_{rep}为重复追踪次数。

同时,根据 VOT2015[25] 标准,本文定义了评估指标平均连续追踪帧数,即上一次跟踪失败后能够持续跟踪的视频帧的平均数量。

3.4　结果分析

由于控制无人机进行实验不具备可重复性,本文在视频追踪上设置了与真实情况下无人机相对应的感知窗口,对无人机的移动进行模拟实验。在安全帽数据集上,分别对本文提出的启发式飞行控制模块和数据驱动的飞行控制模块进行了评估。共有 2 组对照实验:实验 1 比较了 3 个不同策略下平均像素误差和平均重叠率,实验 2 比较了启发式飞行控制模块和数据驱动的飞行控制模块的追踪效果;实验 3 对模型的参数敏感性进行了分析。

3.4.1　实 验 1

在本节实验中,分别设置了 3 组对照实验:目标检测 + 启发式飞行控制模块,目标检测 + KCF + 启发式飞行控制模块和目标检测 + KCF + 数据驱动的飞行控制模块。

1)目标检测 + 启发式飞行控制模块(方法 1)。目标检测采用 YOLOv3 算法,飞行控制模块通过对目标检测中目标移动距离的预测值进行加权,得到窗口移动的位移值。该位移模拟了无人机追踪场景下无人机的位移。其中无人机的飞行速度由超参 w 控制。

2)目标检测 + KCF + 启发式飞行控制模块(方法 2)。在方法 1 的基础上,添加 KCF 算法,旨在当目标丢失时,再启动目标检测方法。

3)目标检测 + KCF + 数据驱动的飞行控制模块(本文方法)。在方法 2 的基础上,替换启发式飞行控制模块为基于深度学习的数据驱动的飞行控制模块。

3.4.2　实 验 2

本节实验中在启用 KCF 的平等环境下,对比启发式飞行控制模块和数据驱动的飞行控制模块,比较其在追踪目标时的平均重启次数与平均连续追踪帧数。重启目标检测的次数越少,意味着追踪越不容易丢失目标,越流畅。连续追踪帧数越大,代表着追踪时长越长,追踪能力越强。

3.4.3　实 验 3

为了全面验证本文模型,对模型的 3 个重要参数:追踪中的当前帧和前一帧目标框颜色直方图相似性阈值 γ_1、目标重定位中的直方图相似性阈值 γ_2、空间相似度阈值 δ,分别进行了参数敏感性实验。将追踪中的颜色直方图相似性阈值 γ_1 设置在[0.1,0.9]范围内,并以 0.2 为步长进行调节;将目标重定位中的直方图相似性阈值 γ_2 设置在[0.1,0.5]范围内,并以 0.1 为步长进行调节;将空间相似度阈值 δ 设置在[0.001,0.003,0.01,0.03,0.1]中取值。

3.4.4　实验结果分析

从实验 1 模块的对比中可以看出,启发式飞行控制模块中超参 w 的设置对检测的效果有明显的影响。随着启发式飞行控制模块参数 w 从初始值 1 开始增加,检测的效果开始逐渐提升,并在 $w=2$ 后这种趋势逐渐放缓。同时,从表 1 和表 2的对比结果可以看出,尽管本文提出的数据驱动的飞行控制方法效果只是接近方法 1 的最优值($w=5$),但是由于方法 1 需要对每一帧进行目标检测,其时间代价远远大于目标检测与目标追踪结合的方法 2 和本文方法。

实验 2 对比了启发式飞行控制模块和数据驱动的飞行控制模块的平均重启次数和平均连续追踪帧数。如表 3 和表 4 所示,数据驱动的飞行控制模块平均重启次数 36.70 明显小于启发式飞行控制模块,即在相同的帧数内数据驱动的飞行控制模块的追踪效果更好。对比连续追踪帧数可以发现,数据驱动的飞行控制模块可以连续追踪到 41 帧,也高于启发式飞行控制模块,说明其追踪效果较强。

实验 3 对模型进行了参数敏感度分析,分别探索了超参(γ_1、γ_2 和 δ)对于目标追踪效果的影响。如图 2 所示,本文模型对于各参数均表现出了良好的鲁棒性。当 γ_1 在 0.3 ~ 0.9 之间,γ_2 在 0.1 ~ 0.5 之间,δ 在 0.003 ~ 0.1 之间,模型取得的平均像素误差和平均重叠率均变化不大。但是当 $\gamma_1=0.1$ 或者 $\delta=0.001$ 时,模型在平均像素误差这一评价指标上变化较大。说明当颜色直方图相似性阈值或者空间相似度阈值过低,可能会造成目标追丢或者追错问题。因此,合理设置超参对目标追踪的效果能起到一定的促进作用。

表 1　平均像素误差结果统计

Table 1　Statistical results of average pixel error

w	平均像素误差		
	方法 1	方法 2	本文方法
1	20.03	13.61	2.37
2	2.64	2.66	2.37
3	2.50	2.50	2.37
4	2.41	2.31	2.37
5	2.35	2.56	2.37

北航学报 赠阅

表 2 平均重叠率结果统计

Table 2 Statistical results of average overlap rate

w	平均重叠率		
	方法 1	方法 2	本文方法
1	0.39	0.50	0.67
2	0.65	0.65	0.67
3	0.66	0.64	0.67
4	0.67	0.67	0.67
5	0.68	0.64	0.67

表 3 平均重启次数统计

Table 3 Statistics of average restart times

w	平均重启次数	
	方法 2	本文方法
1	55.00	36.70
2	48.70	36.70
3	50.10	36.70
4	48.20	36.70
5	46.20	36.70

表 4 平均连续追踪帧数统计

Table 4 Statistics of the number of average consecutive tracking frames

w	平均连续追踪帧数	
	方法 2	本文方法
1	38.20	41.00
2	35.50	41.00
3	33.90	41.00
4	35.30	41.00
5	36.90	41.00

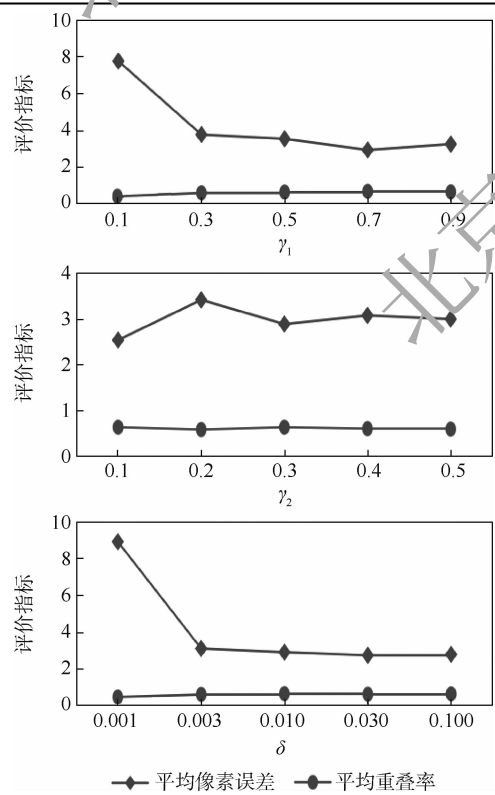


图 2 参数敏感性实验结果

Fig. 2 Sensitivity analysis w. r. t. different γ_1, γ_2 and δ

4 结 论

1）本文方法可实现较为优异的电力场景下的运动目标追踪性能，基于真实数据集，平均像素误差和平均重叠率分别可达到 2.37 和 0.67，平均重启次数和平均连续追踪帧数分别可达到 36.70 和 41.00，验证了方法的有效性。

2）通过采用目标检测和目标追踪相结合的方式，能够弥补追丢和追错目标的不足。

3）设计了 2 种飞行控制模块，使无人机可以自适应地调节飞行速度和方向。

本文方法目前着眼于考虑图像特征。在未来工作中，将引入光流、红外、声音等信息，进一步提高电力场景中对运动隐患目标的追踪效果。

参考文献（References）

[1] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2021-11-01]. https://arxiv.org/abs/1804.02767.

[2] HENRIQUES J F, RUI C, 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.

[3] CHO K, MERRIENBOER B V, GÜLÇEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1724-1734.

[4] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2001, 1: 511.

[5] SUGIMURA D, FUJIMURA T, HAMAMOTO T. Enhanced cascading classifier using multi-scale HOG for pedestrian detection from aerial images[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2016, 30(3): 1655009.

[6] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2005: 886-893.

[7] ZHU Q, AVIDAN S, YEH M C, et al. Fast human detection using a cascade of histograms of oriented gradients[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2006: 1491-1498.

[8] GIRSHICK R. Fast R-CNN[C]// International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.

[9] 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 and Machine Intelligence, 2017, 39(6): 1137-1149.

[10] LIU W, ANGUELOV D, ERHAN D. SSD: Single shot multibox

北京航空航天大学学报

北航学报 赠阅

detector[C]// European Conference on Computer Vision. Berlin:Springer,2016,9905:21-37.

[11] 吴冬梅,王慧,李佳. 基于改进 Faster RCNN 的安全帽检测及身份识别[J]. 信息技术与信息化,2020(1):17-20.
WU D M, WANG H, LI J. Helmet detection and identification based on improved Faster RCNN[J]. Journal of Information Technology and Informatization,2020(1):17-20(in Chiense).

[12] 董召杰. 基于 YOLOv3 的电力线关键部件实时检测[J]. 电子测量技术,2019,42(43):173-178.
DONG Z J. Key components of power line based on YOLOv3 real-time detection[J]. Journal of Electronic Measurement Technique,2019,42(43):173-178(in Chiense).

[13] XIANG G. Real-time follow up tracking fast moving object with an active camera[C]//2nd Internationl Congress on Image and Signal Processing. Piscataway:IEEE Press,2009:2075-2078.

[14] ZHANG J E, WANG Y T, CHEN J, et al. A framework of surveillance system using a PTZ camera[C]// IEEE International Conference on Computer Science and Information Technology (ICCSIT). Piscataway:IEEE Press,2010:658-662.

[15] DONG Q, ZOU Q H. Visual UAV detection method with online feature classification[C]// Proceedings of 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference. Piscataway:IEEE Press,2017:464-467.

[16] 林勇,秦文静,戚国庆. 基于机载视觉引导的无人机自主循迹研究[J]. 电子设计工程,2018,26(8):21-24.
LIN Y, QING W J, QI G Q. Autonomous tracking research of UAV based on airborne visual guidance[J]. Journal of Electronic Design Engineering,2018,26(8):21-24(in Chiense).

[17] XU Z H, ZHANG Y, LI H B, et al. A new shadow tracking method to locate the moving target in SAR imagery based on KCF[C]// Proceedings of the 2017 International Conference on Communications, Signal Processing, and Systems. Berlin:Springer, 2017:2661-2669.

[18] FAN J X, LI D W, WANG H L, et al. UAV low altitude flight threat perception based on improved SSD and KCF[C]// International Conference on Control and Automation. Piscataway: IEEE Press,2019:356-361.

[19] 徐小超,严华. 引入目标分块模型的核相关滤波目标追踪算法[J]. 计算机应用,2020,40(3):683-688.
XU X C, YAN H. A target tracking algorithm based on kernel correlation filter is introduced[J]. Computer Application, 2020,40(3):683-688(in Chinese).

[20] HAN S B, XIE Z J, LI H J, et al. Scale estimation for KCF tracker based on feature fusion[C]// The 4th International Conference on Innovation in Artificial Intelligence. New York: ACM,2020:32-36.

[21] SIMON M, AMENDE K, KRAUS A, et al. Complexer-YOLO: Real-time 3D object detection and tracking on semantic point clouds[C]// Conference on Computer Vision and Pattern Recognition Workshops. Piscataway:IEEE Press,2019:1190-1199.

[22] ZHANG H B, QIN L F, LI J, et al. Real-time detection method for small traffic signs based on Yolov3[J]. IEEE Access,2020, PP(99):1.

[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, MATAS J, LEONARDIS A, et al. The visual object tracking VOT2013 challenge results[C]// IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press,2013:98-111.

[25] KRISTAN M, JIRI M, LEONARDIS A. The visual object tracking VOT2015 challenge results[C]// IEEE International Conference on Computer Vision Workshop. Piscataway: IEEE Press,2015:564-586.

北京航空航天大学学报

北航学报 赠阅

UAV-vision-based moving target tracking scheme in electric scenario

FENG Xue, DU Mengjun*, XIANG Xinyu, QIAN Jin, ZHANG Min

(Hangzhou Power Supply Company of Zhejiang Power Co. Ltd. of State Grid Corporation of China, Hangzhou 310000, China)

Abstract: With the rapid development of artificial intelligence technology, the moving target tracking technology for the power system has gradually attracted researchers' attention. Although the existing methods have achieved great success, most of them are based on the fixed camera surveillance video recording, which cannot track the moving target flexibly. When the moving object leaves the camera's field of view, there is a problem of losing the moving object. In light of this, we propose a moving target tracking scheme based on UAV vision (MTTS_UAV) in electric scenario. In particular, to ensure the real-time feature, we combine the improved target tracking algorithm and target detection algorithm to track the hidden dangers. To ensure the accuracy, we introduce two UAV flight control modules: heuristic mode and data-driven mode, so that the UAV's flight speed and direction can be adjusted adaptively according to the target movement. Extensive experiments have been conducted on a real-world dataset of hardhat personnel in real substations, which demonstrate that the average pixel error (APE) and average overlap rate (AOR) are 2.37 and 0.67 respectively and verify the effectiveness of the proposed method.

Keywords: target tracking; target detection; deep learning; UAV vision; electric scenario

北京航空航天大学学报 北京航空航天大学学报

Received: 2020-11-03; **Accepted**: 2020-11-12; **Published online**: 2021-03-02 12:10

URL: kns.cnki.net/kcms/detail/11.2625.V.20210301.1821.004.html

Foundation item: Science and Technology Project of Zhejiang Power Co. Ltd. of State Grid Corporation of China (5211HZ19014U)

* **Corresponding author**. E-mail: du_mengjun@zj.sgcc.com.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0619

复杂曲面上的自动铺放路径规划方法

全下文载

赵安安[1,2]，何大亮[2]，王晗[3]，郭俊刚[2,*]，柯映林[3]

（1. 浙江大学 工程师学院，杭州 310058； 2. 中航飞机股份有限公司，西安 710089；
3. 浙江大学 机械工程学院，杭州 310058）

摘 要：自动铺丝是一种先进的复合材料自动制造技术，广泛应用于制造形面复杂的复合材料构件。大曲率区域丝束的压实质量较差、铺放路径的方向偏差和转弯半径难以同时满足等问题，导致了各种铺放缺陷。研究了曲面曲率对路径性能的影响，建立了分区域路径规划机制。采用压辊到曲面的距离来表征丝束压实情况，并根据曲面曲率估算可压实丝束的数量。建立了局部方向偏差和转弯半径计算方法，并在路径密化过程中实时评价，在路径性能超出约束时进行分区。考虑了丝束压实、方向偏差及转弯半径对铺放路径的约束，保证了整张曲面的铺放质量和铺放效率。在翼梢小翼曲面上完成了路径规划方法的仿真和实验验证，结果表明：所提方法能够在复杂曲面上获得满足性能要求的铺放路径。

关 键 词：复杂曲面；自动铺丝；路径规划；铺放分区；性能约束
中图分类号：V261.94；TH164
文献标志码：A **文章编号**：1001-5965(2022)04-0595-07

碳纤维复合材料具有优异的刚度、强度性能，广泛应用于航空领域[1]。自动铺丝是一种高效的复合材料自动化制造技术，广泛应用于高精度复合材料构件的制造。在自动铺丝过程中，多根丝束同时铺放到模具曲面上，每根丝束能够独立地重送和剪断，这使得自动铺丝能够铺放大曲率的复杂构件[2-3]。

路径规划是自动铺丝中的重要工序，铺放路径的质量直接决定了构件的铺放质量。丝束的压实质量、方向偏差和转弯半径是评价路径质量的3个主要指标。最理想的铺放路径必须保证丝束压实充分、方向偏差小、转弯半径大。在3项指标中，压实质量相对独立，不受方向偏差和转弯半径的影响，只与曲面曲率、压辊尺寸和同时铺放的丝束数量有关。但方向偏差和转弯半径互相影响，减小方向偏差的代价往往是同时减小了路径的转弯半径，导致不可接受的铺放褶皱。因此，路径规划需要在保证丝束压实质量的基础上兼顾方向偏差和转弯半径约束，从而获得性能优异的铺放路径。

路径规划过程大致分为2步：初始路径生成和路径密化。目前，主流的初始路径生成方法有定角度法、测地线法和变刚度法。文献[4-6]采用解析法在参数曲面上计算了初始路径，而文献[7]以铺层角度信息为依据生成了网格曲面上的铺放路径。文献[8]对比了曲面上的定角度路径和测地路径，分析结果表明，定角度路径的实际铺放方向与理论方向重合，但转弯半径往往偏小，容易导致铺放褶皱，而测地线路径恰好相反，测地线路径的转弯半径很大，不会出现褶皱缺陷，但实际铺放方向与理论铺放方向差距较大，导致构件性能偏离设计值。与定角度路径和测地线路径相

收稿日期：2020-11-06；**录用日期**：2021-03-19；**网络出版时间**：2021-05-19 11:53
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210518.1754.004.html
基金项目：国家自然科学基金重大研究计划（91948301）；浙江省重点研发计划（2020C01039）
* **通信作者**. E-mail：guojg@avic.com

引用格式：*赵安安，何大亮，王晗，等. 复杂曲面上的自动铺放路径规划方法[J]. 北京航空航天大学学报，2022，48(4)：595-601. ZHAO A A，HE D L，WANG H，et al. Automatic paving path planning method on complex surfaces[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：595-601 (in Chinese).*

北京航空航天大学学报

北航学报 赠阅

比,变刚度路径的设计性更强,能够根据不同的设计原则获得不同的铺放路径。文献[9-10]的铺放角度沿路径呈线性变化,而文献[11]的铺放路径是按照主应力方向设计的,使构件的力学性能得到显著提升。

当然,初始路径满足性能要求并不代表所有路径都满足性能约束,还需要配合优秀的密化方法。Rousseau 等[12]总结了 3 种常用的密化方法,即平行法、平移法和生成法。平行法是最常用的密化方法,文献[13]提出了一种近似计算曲面上平行线的方法,并在文献[14]中做了进一步改进,有效提高了计算精度。平移法通常与变刚度路径搭配使用,因为平移法可以有效继承初始路径的优良性能[15]。生成法是指所有铺放路径都按照初始路径的生成方法进行生成,会导致大量的间隙重叠,因此较少使用。

本文针对大曲率曲面铺放路径质量较差难以满足铺放要求的问题,建立了一种适用的铺放规划流程,实现了复杂曲面铺放路径自动生成。该方法综合考虑了铺放路径的方向偏差、转弯半径和丝束压实质量,确保了复杂曲面上的铺放质量和铺放效率。基于 CATIA 平台,采用二次开发技术编写了相应的软件模块,并以翼梢小翼为例验证了方法的有效性。

1 丝束数量的确定

丝束数量是路径规划过程中的重要参数,其不仅受铺放设备能力的约束,也与曲面曲率有很大关系。因此,本文首先分析了压辊在曲面上的压实特征,并根据曲面不同区域的丝束数量进行了初始分区。

1.1 铺丝过程的压实分析

自动铺放设备的压辊往往采用硅橡胶材料,在一定范围内,硅橡胶的应力和应变成线性关系,因此,可以用压辊变形量近似衡量该位置丝束的压实情况。变形量越大,该位置的铺放压力越大,丝束压实情况越好。

压辊的允许变形和曲面曲率是影响可铺放丝束数量的重要因素。压辊和模具之间的几何关系如图 1 所示。其中,P 为压辊与曲面的接触点。在 P 点处建立局部标架,$\boldsymbol{e}_1$ 为理想铺放方向,$\boldsymbol{e}_2$ 为压辊轴线方向,$\boldsymbol{e}_3$ 为曲面法矢。以压辊轴线上的点到曲面的距离作为判定丝束压实质量的条件。根据丝束的可压实区域宽度可以计算该位置的适宜丝束数量。采用压实点处曲面沿压辊轴向方向的法曲率 k 来近似计算丝束的可铺放数量 n:

$$n \leqslant \left| 2\sqrt{\frac{2D_{\mathrm{A}} - D_{\mathrm{A}}^2 k}{k W_{\mathrm{f}}^2}} \right| \tag{1}$$

式中:D_{A} 为压辊的可变形量;W_{f} 为单根丝束宽度;法曲率 k 根据欧拉公式获得,即

$$k = k_1 \cos^2\varphi + k_2 \sin^2\varphi \tag{2}$$

$$\cos\varphi = \frac{\boldsymbol{v}_1 \cdot \boldsymbol{e}_2}{|\boldsymbol{v}_1||\boldsymbol{e}_2|} \tag{3}$$

式中:k_1、k_2 分别为点处的最大、最小主曲率;$\boldsymbol{v}_1$ 为对应 k_1 的主方向。

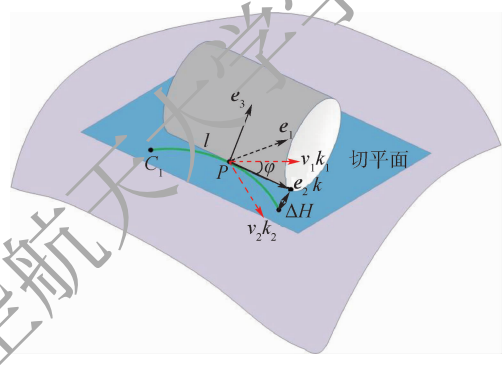


图 1 压辊与模具间的几何关系

Fig. 1 Geometric relationship between pressure roller and mold

1.2 压辊变形量确定

压辊的可变形量主要取决于压辊的材料、规格和铺放压力的大小。本文分别采用仿真和实验的方法获得不同铺放压力下的压辊可变形量。压辊的材料参数和尺寸如表 1 所示。

表 1 压辊模型的主要参数

Table 1 Main parameters of pressure roller model

参数	压辊内径/mm	压辊外径/mm	压辊长度/mm	压辊轴弹性模量/MPa	压辊轴泊松比	模具弹性模量/MPa	模具泊松比
数值	40	85	130	210 000	0.3	210 000	0.3

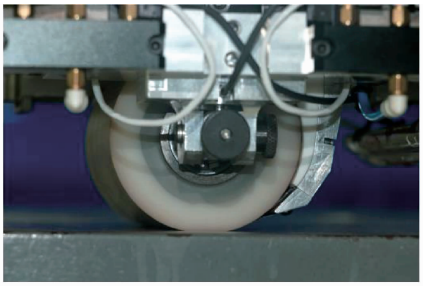
在 ABAQUS 中建立有限元模型,分析不同曲面上的压辊压实状态。采用 Mooney-Rivlin 超弹本构来描述压辊的变形状态。

$$W = C_{10}(I_1 - 3) + C_{01}(I_2 - 3) \tag{4}$$

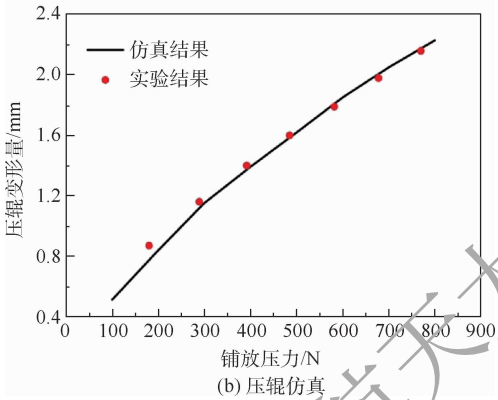
式中:W 为应变能密度;I_1、I_2 分别为第一、第二应变不变量;实验中所用压辊材料为硅橡胶,硬度为 40 HA,其对应的本构常数为 $C_{10} = 0.226$,$C_{01} = 0.056\,5$[16]。

首先通过仿真和实验确定不同压力下压辊在平面上的变形量。由图 2 可以看出,有限元仿真

北京航空航天大学学报



(a) 平面铺放实验



(b) 压辊仿真

图 2　压辊仿真与平面铺放实验

Fig. 2　Pressure roller simulation and flat paving experiment

结果与实验结果十分吻合，因此该有限元模型能够有效预测压辊的变形情况。当压辊压实到曲面上时，曲率情况较为多变，难以获得较为一致的变形结果。但可以肯定的是，由于压力集中效应，曲面上的压辊最大变形量要大于同压力下压辊在平面上的变形量，以压辊在平面上的变形量作为压辊可变形量的保守估计。在后续的应用过程中，面对不同材料不同尺寸的压辊只需进行仿真即可得到变形量的估计值。

2　铺放路径规划

2.1　初始路径的生成

由于复杂曲面的参数表达难以获得，建立了一种基于离散点的固定角度铺放路径生成方法。如图 3所示，对于给定铺放曲面，P_i 为路径点，TP 为 P_i 处的切平面，$\boldsymbol{V}$ 为路径的全局参考方向，θ 为铺层角度，λ 为步长值。所有路径点按照以下步骤获得：

步骤 1　选择模具曲面上的一点 P_i 作为起始点。

步骤 2　计算曲面在点 $P_i(x_i, y_i, z_i)$ 处的法矢 $\boldsymbol{n}_i$，并将参考方向 $\boldsymbol{V}$ 投影到切平面上得到投影矢量 $\boldsymbol{v}_i$。

$$\boldsymbol{v}_i = \boldsymbol{V} - \boldsymbol{n}_i \cdot \boldsymbol{V} \cdot \boldsymbol{n}_i \tag{5}$$

步骤 3　根据铺层设计角度计算点 P_i 处的铺放方向矢量。

$$\boldsymbol{t}_i = \boldsymbol{v}_i \cdot \cos\theta + (\boldsymbol{v}_i \times \boldsymbol{n}_i) \cdot \sin\theta \tag{6}$$

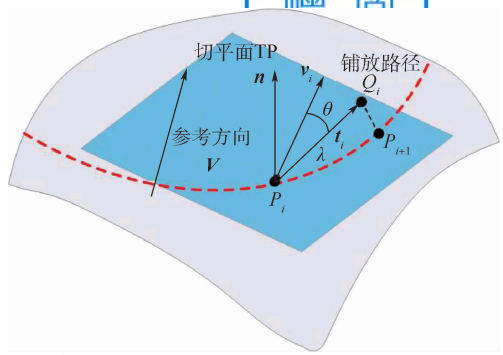


图 3　初始路径生成方法示意图

Fig. 3　Schematic diagram of initial path generation method

步骤 4　在切平面上做出特征点 Q_i，使得 $\overrightarrow{P_iQ_i} = \lambda \cdot \boldsymbol{t}_i$，然后将 Q_i 投影到模具曲面上得到路径点 O_{i+1}。

步骤 5　重复步骤 2 ~ 步骤 4，直到路径点到达曲面的边界。将所有路径点拟合为一条样条曲线作为自动铺丝的定角度初始路径。

对路径生成过程中每个路径点处的转弯半径进行近似估算。如图 4 所示，将上一路径点 P_{i-1} 向切平面 TP 上做投影，得到 Q'_{i-1}。用过点 Q'_{i-1}、P_i 和 Q_i 的圆弧半径 R_i 来近似计算铺放路径在点 P_i 处的局部转弯半径。则 R_i 可用式(7) ~ 式(9)进行近似计算。

$$R_i = \frac{\sqrt{\lambda_i^2 + \lambda^2 - 2\lambda_i\lambda\cos\beta_i}}{2\sin\beta_i} \tag{7}$$

式中：λ_i 为点 P_i 和 Q'_{i-1}之间的距离；β_i 为 $\overrightarrow{P_iQ'_{i-1}}$ 与铺放方向 $\boldsymbol{t}_i$ 之间的夹角。

$$\lambda_i = |\overrightarrow{P_iQ'_{i-1}}| \tag{8}$$

$$\cos\beta_i = \overrightarrow{P_iQ'_{i-1}} \cdot \boldsymbol{t}_i / \lambda_i \tag{9}$$

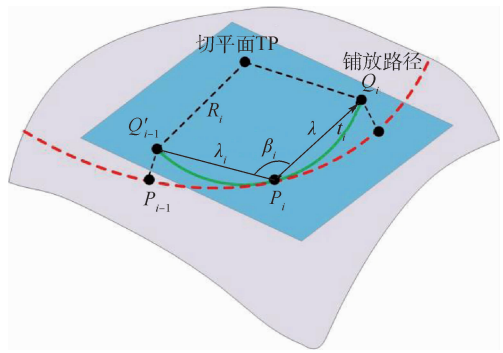


图 4　铺放路径局部转弯半径估算方法

Fig. 4　Estimation method of local turning radius of paving path

为进一步在允许的方向偏差范围内增大路径的转弯半径，在路径点的生成过程中采用黄金分割法对局部铺放方向进行优化。具体优化方法如下：

步骤 1　在路径点处根据上文所述方法获得定角度铺放方向 $\boldsymbol{t}_i$。

北京航空航天大学学报

步骤 2 根据最大允许的方向偏差 Δ_{max} 在铺放方向 $\boldsymbol{t}_i$ 的左右两侧做出矢量 $\boldsymbol{s}_1$、$\boldsymbol{s}_2$，且 $\boldsymbol{s}_1$、$\boldsymbol{s}_2$ 与 $\boldsymbol{t}_i$ 之间的夹角均为 Δ_{max}。

步骤 3 在 $\boldsymbol{s}_1$、$\boldsymbol{s}_2$ 所围成的区域内采用黄金分割法获得使路径转弯半径取得最大值的铺放方向 $\boldsymbol{t}_i$。黄金分割法的详细流程如图 5 所示。

步骤 4 根据优化后的铺放方向 $\boldsymbol{t}_i$ 生成下一个路径点。

步骤 5 重复步骤 1 ~ 步骤 4，直到路径点到达曲面的边界。将所有路径点拟合为一条样条曲线作为优化后的铺放初始路径。

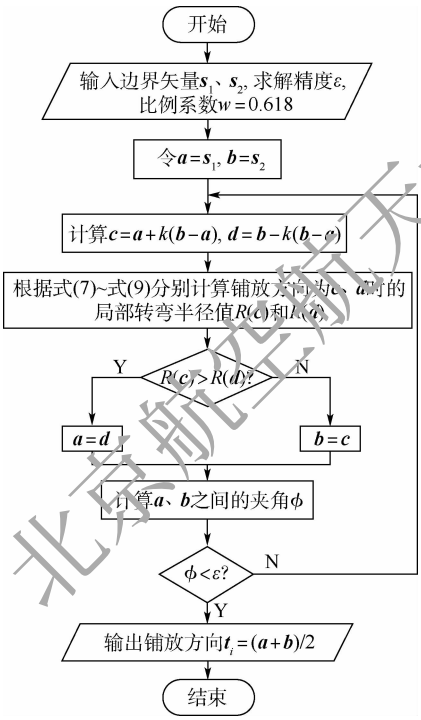


图 5 黄金分割法优化局部铺放方向

Fig. 5 Optimization of local paving direction using golden section method

2.2 铺放路径密化方法

在生成初始路径之后，需要将初始路径按照一定的规则进行密化，使得铺放路径铺满整个曲面。为避免丝束之间的间隙和重叠，采用平行等距方法对初始路径进行密化，平行等距原理如图 6所示。同时，为保证方向偏差和转弯半径约束，在密化过程中对每个偏移点处的方向偏差和转弯半径进行检查，如果偏移路径的转弯半径和方向偏差不满足要求，则偏移结束。平行等距路径的生成过程如下：

步骤 1 将初始路径按照一定的规则进行离散，获得每个离散点处路径的副法矢 $\boldsymbol{e}_2$。

步骤 2 计算当前区域的可压实丝束数量，并根据单丝宽度获得每个离散点处的平行等距距离 L_i，取 L_i 的最小值作为路径偏移距离 L。

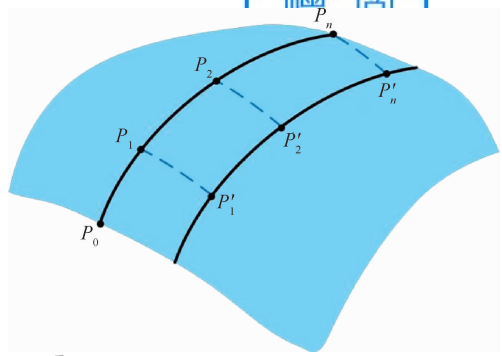


图 6 平行等距密化方法示意图

Fig. 6 Schematic of parallel isometric densification method

步骤 3 根据距离 L 和副法矢方向 $\boldsymbol{e}_2$ 生成指定起始点、方向和长度的测地线段。

步骤 4 如果测地线段的长度小于 L，则放弃该位置，在下一个离散点处执行步骤 2 ~ 步骤 4；如果测地线段长度等于 L，则测地线末端点为所求平行等距路径的路径点。

步骤 5 将所有平行等距路径的路径点拟合成一条样条曲线，并判断该曲线端点与曲面边界的位置关系。如果端点不在曲面边界上，则将样条曲线按 G2 连续进行外插延伸直到曲面边界，得到的曲线为所求铺放路径。

步骤 6 分析所得铺放路径的方向偏差和转弯半径。

步骤 7 将所得铺放路径作为起始路径，重复步骤 1 ~ 步骤 6，直到铺放路径布满整个曲面，或路径的方向偏差、转弯半径超差。

如果偏移路径并未铺满整张曲面，但路径的方向偏差和转弯半径已经超过限制值，此时不能再对路径进行平行等距计算，而应在路径未覆盖区域重新按 2.1 节所述方法生成初始路径，并重新进行平行等距，直到铺放路径完全覆盖曲面。

3 仿真与实验结果

根据第 1、2 节所述方法，在翼梢小翼曲面上完成路径生成和仿真分析，以及对应的实验验证。实验所用自动铺丝设备所能铺放的最大丝束数量为 16 丝束，单根丝束的宽度为 6.35 mm，实验中设定铺放压力为 500 N，根据 2.2 节的实验确定压辊可变形量为 1.6 mm，由此可得到曲面不同区域的可铺放丝束数量，如图 7 所示，16 丝束区域内的可铺放丝束数量为 16，而 8 丝束区域内的可铺放丝束数量为 8。可以看出，整个 0°铺层均可采用 16 丝束进行铺放，整个 −45°铺层均采用 8 丝束进行铺放。而 45°铺层和 90°铺层均有部分区域可铺放 16 丝束，部分区域仅能铺放 8 丝束。在这些区域内进行铺放路径生成，并按照对应的

北京航空航天大学学报

丝束数量进行路径密化。设定路径的最大方向偏差为 10°，最小转弯半径为 1 500 mm。在路径密化过程中，检查每条铺放路径的方向偏差和转弯半径性能。如果密化路径的性能不能满足上述要求，则采用 2.1 节所述的方法重新生成铺放路径。最终得到如图 8 所示的铺放路径规划结果。分别对路径的方向偏差、转弯半径和压实情况进行分析和评价，结果如图 9 所示。方向偏差的最大值、转弯半径的最小值和压实距离的最大值列在表 2 中。可以看出，所有铺放路径的方向偏差、转弯半径和压实情况均能满足预设的指标值。按上述路径在翼梢小翼曲面上进行铺放实验，最终铺放效果如图 10 所示。可以看出，丝束与模具曲面贴合

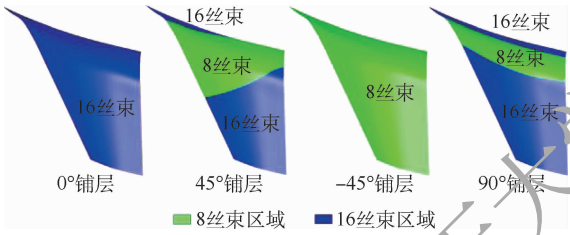


图 7　翼梢小翼曲面上的可铺放丝束数量分布

Fig. 7　Distribution of the number of tows that can be paved on winglet surface

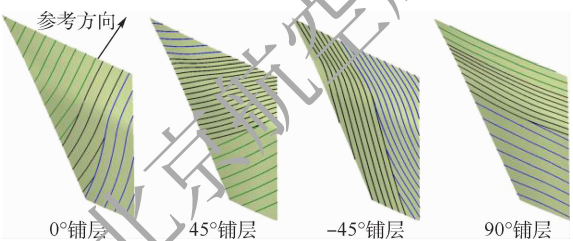


图 8　翼梢小翼曲面上的铺层路径

Fig. 8　Pavement path on winglet surface

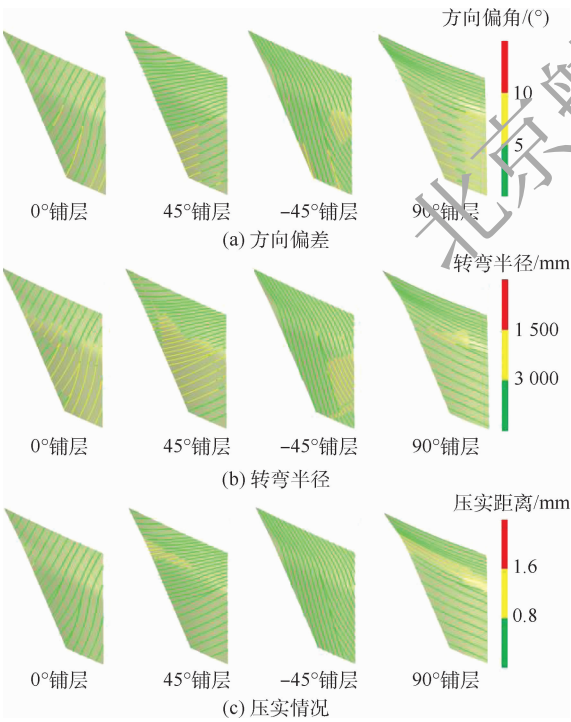


图 9　铺放路径性能仿真分析结果

Fig. 9　Simulation analysis results of paving path performance

表 2　方向偏差、转弯半径和压实情况

Table 2　Direction deviation, turning radius and compaction

铺层角度/(°)	方向偏差/(°)	转弯半径/mm	压实距离/mm
0	9.8	1 557.4	1.2
45	9.6	1 530.9	1.3
-45	8.5	1 975.0	0.9
90	9.7	1 552.6	1.5

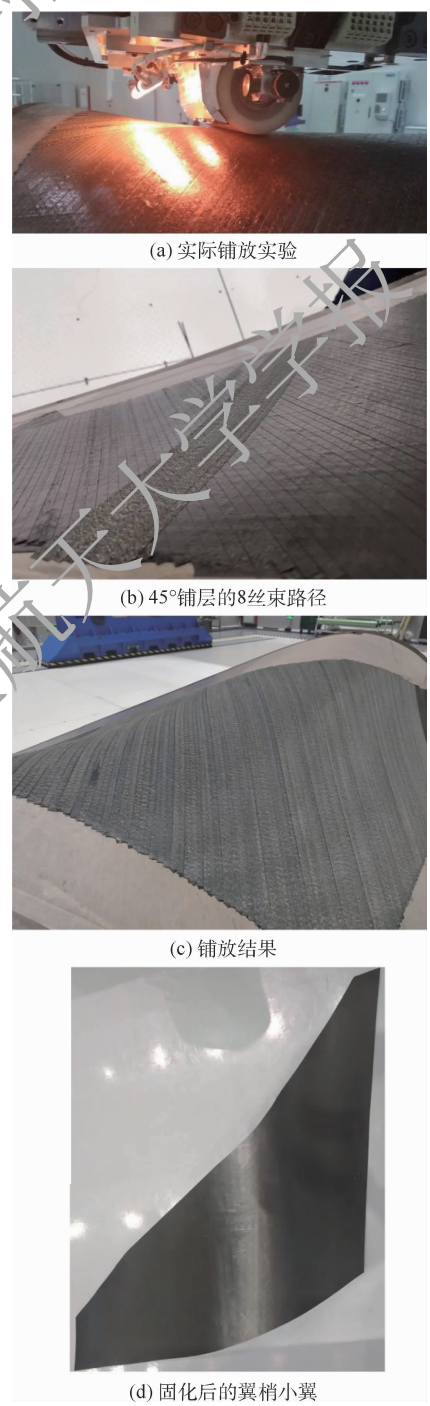


图 10　实际铺放实验和铺放结果

Fig. 10　Actual placement experiment and placement results

北京航空航天大学学报

良好,既不存在明显的架桥缺陷,也没有丝束褶皱,这能够充分说明该路径生成方法具有很强的工程实用性,能够很好地保证丝束铺放质量。

4 结 论

1)建立了曲面曲率与可铺放丝束数量之间的关系。通过压辊变形量和曲面沿压辊轴向的法曲率来近似计算可铺放丝束数量。针对复杂曲面曲率情况多变的问题,提出以可铺放丝束数量为依据对铺放曲面进行分区,有效保证了铺放效率。

2)建立了同时考虑铺放路径方向偏差和转弯半径的局部铺放方向优化方法,能够有效解决铺放路径方向偏差和转弯半径难以同时满足的问题,获得同时满足方向偏差和转弯半径约束的铺放路径。

3)基于 CATIA 平台开发了路径规划软件模块,能够根据性能约束自动生成曲面上的铺放路径,可用于各类飞机复合材料零件的自动铺放路径规划。

参考文献(References)

[1] BEAKOU A, CANO M, LE J B, et al. Modelling slit tape buckling during automated prepreg manufacturing: A local approach [J]. Composite Structures, 2011, 93(10): 2628-2635.

[2] DEBOUT F, CHANAL H, DUC E. Tool path smoothing of a redundant machine: Application to automated fiber placement [J]. Computer Aided Design, 2011, 43: 122-132.

[3] DIRK H J A L, WARD C, POTTER K D. The engineering aspects of automated prepreg layup: History, present and future [J]. Composites Part B: Engineering, 2012, 43(3): 997-1009.

[4] LAND I B. Design and manufacture of advanced composite aircraft structures using automated tow placement [D]. Cambridge: Massachusetts Institute of Technology, 1996.

[5] WANG X, ZHANG W, ZHANG L. Intersection of a ruled surface with a free-form surface [J]. Numerical Algorithms, 2007, 46(1): 85-100.

[6] HAN Z, FU Q, FAN Y, et al. A path planning algorithm of closed surface for fiber placement [C]// Proceedings of the 1st International Conference on Mechanical Engineering and Material Science. Paris: Atlantis Press, 2012: 475-479.

[7] PEI J, WANG X, PEI J, et al. Path planning based on ply orientation information for automatic fiber placement on mesh surface [J]. Applied Composite Materials, 2018, 25(6): 1477-1490.

[8] BLOM A W, SETOODEH S, HOL J M A M, et al. Design of variable-stiffness conical shells for maximum fundamental eigenfrequency [J]. Computers & Structures, 2008, 86(9): 870-878.

[9] BLOM A W, TATTING B F, HOL J M A M, et al. Fiber path definitions for elastically tailored conical shells [J]. Composites Part B: Engineering, 2009, 40(1): 77-84.

[10] GURDAL Z, TATTING B F, WU C K. Variable stiffness composite panels: Effects of stiffness variation on the in-plane and buckling response [J]. Composites Part A: Applied Science and Manufacturing, 2008, 39(5): 911-922.

[11] HYER M W, CHARETTE R F. Use of curvilinear fiber format in composite structure design [J]. AIAA Journal, 1991, 29(6): 1011-1015.

[12] ROUSSEAU G, WEHBE R, HALBRITTER J, et al. Automated fiber placement path planning: A state-of-the-art review [J]. Computer-Aided Design & Applications, 2019, 16(2): 172-203.

[13] SCHUELER K, MILLER I, HALE R. Approximate geometric methods in application to the modeling of fiber placed composite structures [J]. Journal of Computing and Information Science in Engineering, 2004, 4(3): 251-256.

[14] SHIRINZADEH B, CASSIDY G, OETOMO D, et al. Trajectory generation for open-contoured structures in robotic fibre placement [J]. Robotics and Computer-Integrated Manufacturing, 2007, 23(4): 380-394.

[15] LOPES C S, GURDAL Z, CAMANHO P P. Variable-stiffness composite panels: Buckling and first-ply failure improvements over straight-fibre laminates [J]. Computers & Structures, 2008, 86(9): 897-907.

[16] 王伟,邓涛,赵树高. 橡胶 Mooney-Rivlin 模型中材料常数的确定[J]. 特种橡胶制品,2004(4):8-10.
WANG W, DENG T, ZHAO S G. Determination for material constants of rubber Mooney-Rivlin model [J]. Special Purpose Rubber Products, 2004(4): 8-10(in Chinese).

北京航空航天大学学报

北航学报 赠阅

Automatic paving path planning method on complex surfaces

ZHAO An'an[1,2], HE Daliang[2], WANG Han[3], GUO Jungang[2,*], KE Yinglin[3]

(1. Polytechnic Institute, Zhejiang University, Hangzhou 310058, China; 2. Avic Aircraft Co., Ltd., Xi'an 710089, China; 3. School of Mechanical Engineering, Zhejiang University, Hangzhou 310058, China)

Abstract: Automated fiber placement is an advanced automatic manufacturing technology for composite materials, which is widely used in manufacturing composite components with complex shapes. However, the compaction quality of the fiber in the area of large curvature is poor. In addition, it is difficult to satisfy the directional deviation and the turning radius of the placement path at the same time, resulting in various placement defects. In this paper, the influence of surface curvature on path performance is studied and a new path planning method for complex surfaces with area partition is established. The distance between roller and the mold surface is used to estimate the fiber compaction quality and the fiber number placed simultaneously can be calculated by the surface curvature. The local direction deviation and steering radius are evaluated during the path generation process, and a new sector is partitioned when the evaluation results are out of constraints. The method fully considers the influence of compaction quality, direction deviation and turning radius on paving path, so that the placement quality and efficiency of the entire curved surface can be ensured. Simulations and experiments are completed on the winglet surface to verify the path planning method, and the results show that, by the proposed method, satisfactory component placement quality can be obtained on complex surface.

Keywords: complex surfaces; automated fiber placement; path planning; sector partition; performance constraints

北京航空航天大学学报

Received: 2020-11-06; **Accepted**: 2021-03-19; **Published online**: 2021-05-19 11:53

URL: kns.cnki.net/kcms/detail/11.2625.V.20210518.1754.004.html

Foundation items: Major Research Plan of National Natural Science Foundation of China (91948301); Key Research and Development Program of Zhejiang Province (2020C01039)

* **Corresponding author**. E-mail: guojg@avic.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0640

全下文载

水下燃料组件转运系统耦合振动分析

袁占航[1,2], 李运华[1,*], 刘昊东[1]
(1. 北京航空航天大学 自动化科学与电气工程学院, 北京 100083; 2. 中国核电工程有限公司, 北京 100840)

摘 要: 核动力厂燃料组件转运系统运行的平稳性和系统结构的可靠性对确保燃料组件安全至关重要。为了解运行过程中系统的振动情况,建立了考虑水影响的梁-小车+刚性杆-质量+弹簧+阻尼的耦合系统动力学模型,推导了系统的非线性运动微分方程。在此基础上,采用数值算法计算了小车在不同行走速度下系统的响应。研究结果表明:运行过程中,小车行走速度越小系统振动越弱,相对越平稳;由于静变形,即使小车行走速度很低,吊篮仍会承受一定的横向力作用,需注意吊篮薄弱部位的强度;导轨支撑附近位置为吊篮的相对危险位置,必要时可以在此设置监测点以保证燃料组件安全。

关 键 词: 燃料组件转运系统; 水; 耦合系统动力学; 响应; 安全
中图分类号: O326
文献标志码: A **文章编号**: 1001-5965(2022)04-0602-07

核动力厂燃料组件转运系统载荷为核燃料组件,系统运行过程的平稳与组件的安全至关重要。如果导轨变形过大,则可能会发生机械故障;如果转运过程中吊篮晃动过大,有可能使得吊篮根部破坏,且其所装载的燃料组件与吊兰磕碰而损坏,故有必要在设计阶段对系统进行分析,确定合适的方案、工艺参数,以保证燃料组件的安全。

燃料组件转运系统为一个转运小车装载燃料组件而行走在金属导轨上,与常见的车辆与桥梁耦合情况类似。对于车-桥耦合问题,工程技术人员及学者做了大量研究,主要有 3 种典型的分析模型[1-6],分别为移动力模型、移动质量模型和移动振动模型;典型的外载荷主要考虑风载的影响[7];朱素红和徐斌[8]研究了梁的几何非线性及桥面不平顺对桥梁响应的影响;Michaltsos[9]分析了载荷变速运动对桥梁响应的影响;Sun 等[10]分析了吊重质量及阻尼比等因素对运行的平稳性和加速度的影响;对于多跨连续梁,Wang 等[11]研究了其在多种载荷下的振动情况;Chen 等[12]对大跨度梁在多重载荷下进行了疲劳分析。

由于核动力厂燃料组件转运系统中装载燃料组件的吊篮为细长结构,高度比较大,有必要在计算模型中考虑该尺寸,且吊篮与转运小车为刚性连接,通常的车-桥耦合模型已不完全适用;另外,该系统在水下工作,需要综合考虑运行过程中水的影响。本文在已有典型车-桥耦合振动模型基础上,考虑了吊篮的高度与横向刚度,将导轨-小车-吊篮+燃料组件简化为梁-小车+刚性杆-质量+弹簧+阻尼模型,并考虑水的综合作用,导出了系统的运动微分方程,分析了不同行走速度下系统的响应。本文所提方法对系统结构部件设计和选择运行参数具有指导意义。

1 燃料组件转运系统动力学模型

燃料组件转运系统示意如图 1 所示。图中,吊篮为细长的金属框架结构,横截面尺寸为 0.29 m×0.29 m,高度为 4 m,通过销子固定在小车上,其内装载有燃料组件,小车前后端由钢丝绳

收稿日期: 2020-11-16; **录用日期**: 2021-02-05; **网络出版时间**: 2021-02-24 10:33
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210223.1734.001.html
* **通信作者**. E-mail: yhli@buaa.edu.cn

引用格式: 袁占航, 李运华, 刘昊东. 水下燃料组件转运系统耦合振动分析[J]. 北京航空航天大学学报, 2022, 48(4): 602-608.
YUAN Z H, LI Y H, LIU H D. Coupled vibration of underwater fuel assembly transfer system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 602-608 (in Chinese).

北京航空航天大学学报

北航学报 赠阅

牵引行走在金属导轨上，导轨支撑在伸出墙面的支撑上，整个系统浸没在水中。

根据燃料组件转运系统的结构特点，假设车轮与导轨始终接触，忽略次要因素，将其简化为图 2所示的梁-小车 + 刚性杆-质量 + 弹簧 + 阻尼耦合振动模型。

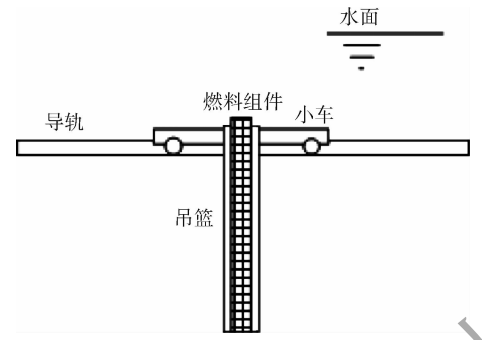


图 1　燃料组件转运系统示意

Fig. 1　Sketch of fuel assembly transfer system

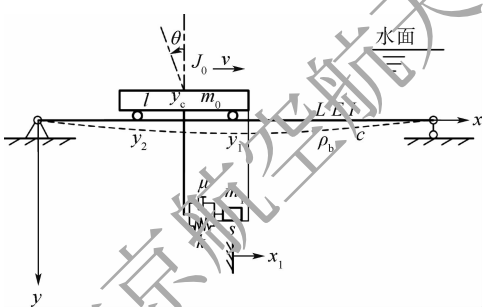


图 2　系统耦合振动模型

Fig. 2　Coupled vibration model of system

导轨简化为简支的 Euler-Bernoulli 梁，长度为支撑跨距 L，截面惯性矩为 I，其材料弹性模量为 E；小车体结构简单，在此简化为长度为 l 的刚体，轮距为 d，绕中心轴转动惯量为 J_0。吊篮横向尺寸远小于竖向尺寸，为弹性梁结构，由于关注的是吊篮底部的响应，在此将吊篮视做弹簧，简化为刚性杆 + 弹簧阻尼模型，弹簧阻尼与刚性杆下端固联，弹簧刚度 k 和阻尼 μ 为吊篮下端的横向等效刚度和等效阻尼系数，刚性杆长度为吊篮高度 h。将吊篮和燃料组件的质量简化为集中在刚性杆底部与弹簧阻尼相联的质量块 m_1（不考虑尺寸），m_1 可在与小车固联的 s 面上光滑运动。

1.1　小车、吊篮、燃料组件动力学方程

建立坐标系如图 2 所示，x 向右为正，y 向下为正；x_1 坐标固连于 s（英小）平面，水平向右为正；θ 为小车绕中心轴转角，逆时针为正。根据牛顿第二定律，小车和燃料组件的运动微分方程分别为

$$m_0\ddot{y}_c = F_1 + F_2 + m_0 g - F_f - F_{m0} + F_{z0} \tag{1}$$

$$J_0\ddot{\theta} = -(F_1 - F_2)\frac{d}{2} + k(x_1 - h\theta)h + \mu(\dot{x}_1 - h\dot{\theta})h - M_m + M_z \tag{2}$$

$$m_1\ddot{x}_1 = -k(x_1 - h\theta) - \mu(\dot{x}_1 - h\dot{\theta}) - F_{m1} + F_{z1} \tag{3}$$

$$y_c = \frac{y_1 + y_2}{2} \tag{4}$$

$$\theta = -\frac{y_1 - y_2}{d} \tag{5}$$

式中：g 为重力加速度；m_0 为小车、燃料组件和吊篮质量之和；y_c 为小车质心处竖向位移；y_1、y_2 分别为小车右、左两轮处的竖向位移；F_1、F_2 分别为小车右、左两轮处的受力，假设其向下（为正）；因导轨截面尺寸较小，故忽略水对导轨的作用，F_f、F_{m0}、F_{z0} 分别为系统除导轨外所受浮力、竖向水的附加质量力、阻力；F_{m1}、F_{z1} 分别为燃料组件所受水的横向附加质量力、阻力；M_m、M_z 分别为小车绕中心轴转动的水的附加质量力矩、阻力矩。

小车及燃料组件模型受力分析如图 3 所示。

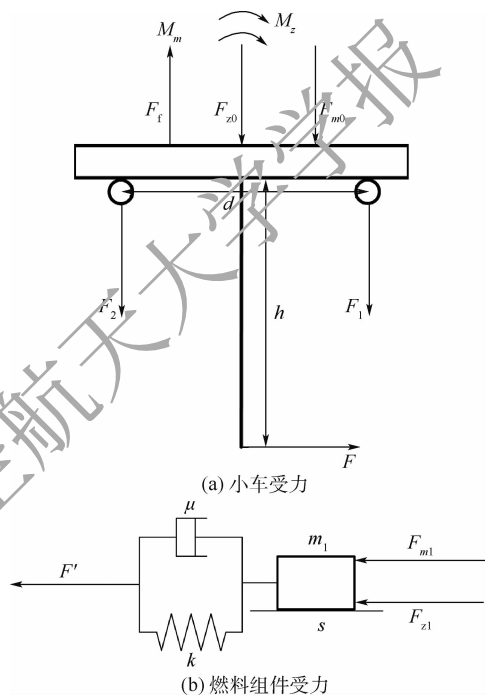


图 3　小车及燃料组件模型受力分析

Fig. 3　Force analysis of truck and fuel assembly model

根据文献[13-14]，可得

$$F_f = \rho_w V g \tag{6}$$

$$F_{m0} = C_m \rho_w V \dot{v}_y \tag{7}$$

$$F_{z0} = -\frac{1}{2} C_d \rho_w A_y \dot{v}_y^2 \operatorname{sign}(v_y) \tag{8}$$

$$F_{m1} = C_m \rho_w V \frac{\mathrm{d}\dot{x}}{\mathrm{d}t} \tag{9}$$

$$F_{z1} = -\frac{1}{2} C_d \rho_w A_{zj} (\dot{x}_1 + v)^2 \operatorname{sign}(\dot{x}_1 + v) \tag{10}$$

$$M_m = 2\int_0^{l/2} C_m \rho_w A \frac{\mathrm{d}\dot{\theta}}{\mathrm{d}t} x^2 \mathrm{d}x = \frac{1}{12} C_m \rho_w V \frac{\mathrm{d}\dot{\theta}}{\mathrm{d}t} l^2 \tag{11}$$

式中：ρ_w 为水的密度；V 为水中构件排开水的体

北航学报 赠阅

积；C_m、C_d 分别为水的附加质量力、阻力系数，C_m、C_d 均取 1.0[15]；v 为小车行走速度；v_y 为车体竖向振动速度；A_y 为小车竖向的投影面积；A_{zj} 为燃料组件横向投影面积；A 为小车体在垂直于 x 轴平面上的截面积。

水的附加转动惯量力系数 C_J 为

$$C_J = \frac{1}{12}C_m l^2$$

则

$$M_m = C_J \rho_w V \frac{\mathrm{d}\dot{\theta}}{\mathrm{d}t} \tag{12}$$

小车在行走过程中，其位移 x 始终为正，所以有

$$M_z = 2\int_0^{l/2} -\frac{1}{2}C_d \rho_w b \mathrm{d}x(\dot{\theta}x)^2 x \cdot \mathrm{sign}(\dot{\theta}x) - \frac{1}{2}C_d \rho_w A_y \dot{\theta}^2 \frac{l^3}{32}\mathrm{sign}(\dot{\theta})$$

式中：b 为小车体在垂直于 x 轴平面上的截面宽度尺寸。

小车转动时流体阻力系数 C_{dJ} 为

$$C_{dJ} = \frac{l^2}{32}C_d$$

则

$$M_z = -\frac{1}{2}C_d \rho_w A_y \dot{\theta}^2 \mathrm{sign}(\dot{\theta}) \tag{13}$$

小车在行走过程中还受到水的水平方向阻力和附加质量力，由于小车是由动力系统通过钢丝绳牵引运行的，认为牵引力始终与水平向载荷平衡，小车保持匀速运动。

1.2 导轨动力学方程

导轨为等截面匀质量，根据导轨受力和 Euler-Bernoulli 梁理论[16]，可以得到导轨的振动偏微分方程为

$$EI\frac{\partial^4 y(x,t)}{\partial x^4} + \rho_b \frac{\partial^2 y(x,t)}{\partial t^2} + c\frac{\partial y(x,t)}{\partial t} = -\delta(x - vt)F_1(x,t)\delta_1 - \delta(x - vt + d)F_2(x,t)\delta_2 \tag{14}$$

式中：E 为导轨材料弹性模量；I 为导轨截面惯性矩；ρ_b 为导轨材料密度；c 为导轨材料阻尼系数；δ 为Dirac 函数；δ_1 和 δ_2 的函数表达式分别为

$$\delta_1(t) = \begin{cases} 1 & 0 \leqslant t \leqslant \dfrac{L}{v} \\ 0 & \text{其他} \end{cases} \tag{15}$$

$$\delta_2(t) = \begin{cases} 1 & \dfrac{d}{v} \leqslant t \leqslant \dfrac{L+d}{v} \\ 0 & \text{其他} \end{cases} \tag{16}$$

2 振型与方程解耦

梁的响应用振型叠加表示为

$$y(x,t) = \sum_{j=1}^{\infty} \phi_j(x) q_j(t) \tag{17}$$

式中：$q(t)$ 为广义坐标。

简支梁的振型函数为

$$\phi_j(x) = \sin\left(\frac{j\pi x}{l}\right) \qquad j = 1,2,\cdots$$

将式(17)代入梁的微分方程(14)，对所得方程式左右同乘以 $\phi_j(x)$，并从 0 到 L 对坐标 x 积分，利用梁振动本征函数的正交性条件得到微分方程组为

$$\omega_j^2 q_j(t) + \frac{\mathrm{d}^2 q_j(t)}{\mathrm{d}t^2} + \frac{c}{2\rho_b}2\frac{\mathrm{d}q_j(t)}{\mathrm{d}t} = -\frac{2}{\rho_b L}\int_0^L \phi_j(x)\delta(x - vt)F_1\delta_1 \mathrm{d}x - \frac{2}{\rho_b L}\int_0^L \phi_j(x)\delta(x - vt + d)F_2\delta_2 \mathrm{d}x \tag{18}$$

式中：$\omega_j = \left(\dfrac{j\pi}{L}\right)^2\sqrt{\dfrac{EI}{\rho_b}}$。

将式(17)代入式(3)，可得

$$m_1\ddot{x}_1 = k\left\{\frac{h}{d}\left[-\sum_{i=1}^{\infty}\phi_i(x)_{x=vt}q_i(t) + \sum_{i=1}^{\infty}\phi_i(x)_{x=vt+d}q_i(t)\right] - x_1\right\} - \mu(\dot{x}_1 - h\dot{\theta}) - F_{m1} + F_{z1} \tag{19}$$

由式(1)～式(5)可以导出 F_1、F_2，并将式(6)～式(13)代入，整理后可以得到系统运动的动力方程组。取梁的前 n 阶振型，方程组写成矩阵形式为

$$\boldsymbol{M}\ddot{\boldsymbol{X}} + \boldsymbol{C}\dot{\boldsymbol{X}} + \boldsymbol{K}\boldsymbol{X} = \boldsymbol{F} \tag{20}$$

式中：$\boldsymbol{X}$ 为广义位移向量，$\boldsymbol{X} = [q_1 \quad q_2 \quad \cdots \quad q_n \quad x_1]^{\mathrm{T}}$；$\boldsymbol{C}$ 为广义阻尼矩阵，$\boldsymbol{C} = [C_{j,i}]_{(n+1)\times(n+1)}$；$\boldsymbol{K}$ 为广义刚度矩阵，$\boldsymbol{K} = [K_{j,i}]_{(n+1)\times(n+1)}$；$\boldsymbol{F}$ 为广义力向量，$\boldsymbol{F} = [F_1 \quad F_2 \quad \cdots \quad F_n \quad F_{n+1}]^{\mathrm{T}}_{1\times(n+1)}$；$\boldsymbol{M}$ 为广义质量矩阵，表达式为

$$\boldsymbol{M} = \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,n-1} & M_{1,n} & 0 \\ M_{2,1} & M_{2,2} & \cdots & M_{2,n-1} & M_{2,n} & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ M_{n-1,1} & M_{n-1,2} & \cdots & M_{n-1,n-1} & M_{n-1,n} & 0 \\ M_{n,1} & M_{n,2} & \cdots & M_{n,n-1} & M_{n,n} & 0 \\ 0 & 0 & \cdots & 0 & 0 & m_{1e} \end{bmatrix}$$

其中：

北京航空航天大学学报

北航学报 赠阅

$$M_{j,i}=\frac{2}{\rho_b L}\phi_j(vt)\delta_1\left\{\frac{1}{2}\cdot\frac{m_{0e}}{2}(\phi_i(vt)+\phi_i(vt-d))-\frac{1}{2}\left[\frac{2}{d}J_{0e}(\phi_i(vt-d)-\phi_i(vt))/d\right]\right\}+\frac{2}{\rho_b L}\phi_j(vt-d)\delta_2\left\{\frac{1}{2}\cdot\frac{m_{0e}}{2}(\phi_i(vt)+\phi_i(vt-d))+\frac{1}{2}\left[\frac{2}{d}J_{0e}(\phi_i(vt-d)-\phi_i(vt))/d\right]\right\}$$

$$j=1,2,\cdots,n;\ i=1,2,\cdots,n\ 且\ j\neq i$$

$$M_{j,j}=1+\frac{2}{\rho_b L}\phi_j(vt)\delta_1\left\{\frac{1}{2}\cdot\frac{m_{0e}}{2}(\phi_j(vt)+\phi_j(vt-d))-\frac{1}{2}\left[\frac{2}{d}J_{0e}(\phi_j(vt-d)-\phi_j(vt))/d\right]\right\}+\frac{2}{\rho_b L}\phi_j(vt)\delta_1\left[\frac{1}{2}\cdot\frac{m_{0e}}{2}(\phi_j(vt)+\phi_j(vt-d))+\frac{1}{2}\left[\frac{2}{d}J_{0e}(\phi_j(vt-d)-\phi_j(vt))/d\right]\right\}$$

$$j=1,2,\cdots,n$$

式中：$m_{0e}=m_0+C_m\rho_w V=k_e m_0=1.13m_0$；$J_{0e}$为合并水的附加质量力之后的惯量，$J_{0e}=k_e J_0=1.13J_0$；$m_{1e}$为合并水的附加质量力之后的等效质量，$m_{1e}=k_e m_1=1.13m_1$。

$$C_{j,i}=\frac{2}{\rho_b L}\phi_j(vt)\frac{h\mu}{d}\left[\frac{-h}{d}(\phi_i(vt-d)-\phi_i(vt))\right]+\frac{2}{\rho_b L}\phi_j(vt-d)\frac{-h\mu}{d}\left[\frac{-h}{d}(\phi_i(vt-d)-\phi_i(vt))\right]$$

$$j=1,2,\cdots,n;\ i=1,2,\cdots,n\ 且\ j\neq i$$

令 $\omega_b=\frac{c}{2\rho_b}$，则

$$C_{j,i}=2\omega_b+\frac{2}{\rho_b L}\phi_j(vt)\frac{h\mu}{d}\left[\frac{-h}{d}[\phi_i(vt-d)-\phi_i(vt))\right]+\frac{2}{\rho_b L}\phi_j(vt-d)\frac{-h\mu}{d}\left[\frac{-h}{d}(\phi_i(vt-d)-\phi_i(vt))\right]$$

$$j=1,2,\cdots,n;\ i=1,2,\cdots,n\ 且\ j=i$$

$$C_{j,n+1}=\frac{2}{\rho_b L}\phi_j(vt)\frac{h\mu}{d}+\frac{2}{\rho_b L}\phi_j(vt-d)\frac{-h\mu}{d}$$

$$j=1,2,\cdots,n$$

$$C_{n+1,i}=\frac{-\mu h}{d}(-\phi_i(vt)+\phi_i(vt-d))$$

$$i\neq n+1$$

$$C_{n+1,n+1}=\mu$$

$$K_{j,i}=\omega_j^2-\frac{2}{\rho_b L}\phi_j(vt)\delta_1\frac{kh^2}{d^2}(\phi_i(vt-d)-\phi_i(vt))+\frac{2}{\rho_b L}\phi_j(vt-d)\delta_2\frac{kh^2}{d^2}(\phi_i(vt-d)-\phi_i(vt))$$

$$j=1,2,\cdots,n;\ i=1,2,\cdots,n$$

$$K_{j,n+1}=\frac{2}{\rho_b L}\phi_j(vt)\delta_1\frac{kh}{d}-\frac{2}{\rho_b L}\phi_j(vt-d)\delta_2\frac{kh}{d}$$

$$K_{n+1,i}=\frac{h}{d}(\phi_i(vt)-\phi_i(vt-d))k$$

$$i=1,2,\cdots,n$$

$$K_{n+1,n+1}=\frac{kh}{d}$$

$$F_j=-\frac{2}{\rho_b L}\varphi_j(vt)\delta_1\frac{-1}{2}\left[m_0 g-\rho_w V g-\frac{1}{2}C_d\rho_w A_y\left(\frac{\dot{y}_1+\dot{y}_2}{2}\right)^2\mathrm{sign}\left(\frac{\dot{y}_1+\dot{y}_2}{2}\right)-\frac{2}{d}M_z\right]-\frac{2}{\rho_b L}\phi_j(vt-d)\delta_2\frac{-1}{2}\left[m_0 g-\rho_w V g-\frac{1}{2}C_d\rho_w A_y\left(\frac{\dot{y}_1+\dot{y}_2}{2}\right)^2\mathrm{sign}\left(\frac{\dot{y}_1+\dot{y}_2}{2}\right)+\frac{2}{d}M_z\right]$$

$$j=1,2,\cdots,n$$

$$\dot{y}_1=\sum_{i=1}^{n}\phi_i(vt)\dot{q}_i$$

$$\dot{y}_2=\sum_{i=1}^{n}\phi_i(vt-d)\dot{q}_i$$

$$F_{n+1}=-\frac{1}{2}C_c\rho_w A_{zj}(\dot{x}_1+v)^2\mathrm{sign}(\dot{x}_1+v)$$

3　算例与结果分析

式(20)为非线性微分方程组，用 Runge-Kutta 法求解。取某工程中实际系统，各参数如下：$L=2$ m，$m_0=950$ kg，$J_0=3\ 945$ kg·m^2，$m_1=864$ kg，$d=0.44$ m，$h=4$ m，$k=137\ 584$ N/m，$E=2\times10^{11}$ Pa，$I=1.834\ 461\ 3\times10^{-6}$ m^4，$\rho_w=1\ 000$ kg/m^3，$\rho_b=7\ 850$ kg/m^3；$A_y=0.46$ m^2，$A_{zj}=0.467$ m^2，$g=9.8$ m/s^2。计算中忽略结构的阻尼，$c=0$，$\mu=0$。

目前，工程应用中小车设计行走速度为 6～12 m/min，在此基础上，出于提高效率(行走速度)的考虑，分别计算小车行走速度 $v=6$、12、24、36、48 m/min时系统的响应，初始条件均取为零。不同行走速度下，计算导轨中间位置位移 y_c，右、左轮的位移 y_1、y_2，小车体转角 θ、m_1 位移 x_1、加速度 a_1。

不同行走速度下导轨中间位置的位移响应如图 4 所示。可见，小车在行走过程中，导轨在小车作用下会发生变形，其变形趋势随小车的行走而先增大后减小且有一定的波动，导轨中间位置的位移响应曲线与之相符。从图 4 中可以看出，不

北京航空航天大学学报

北航学报 赠阅

同小车行走速度下导轨中间处的位移响应总体趋势是一致的,但具体的波动幅值则随行走速度越大而越大,说明控制小车速度对控制系统的振动有效,但对控制导轨最大变形的作用有限。

小车右、左轮的位移响应如图 5 所示。可见,导轨受外力后变形,小车右、左轮的竖向位移会随驶入导轨的距离而呈先增大后减小的趋势,而在车轮驶入导轨前后,其位移为零,右、左轮的曲线与之相符。

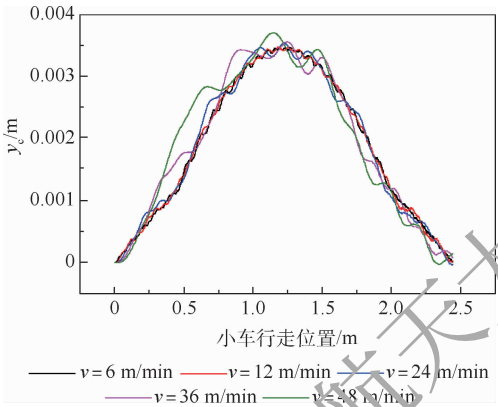


图 4 不同行走速度下导轨中间位置的位移响应

Fig. 4 Displacement of mid-span guide rail at different traveling speeds

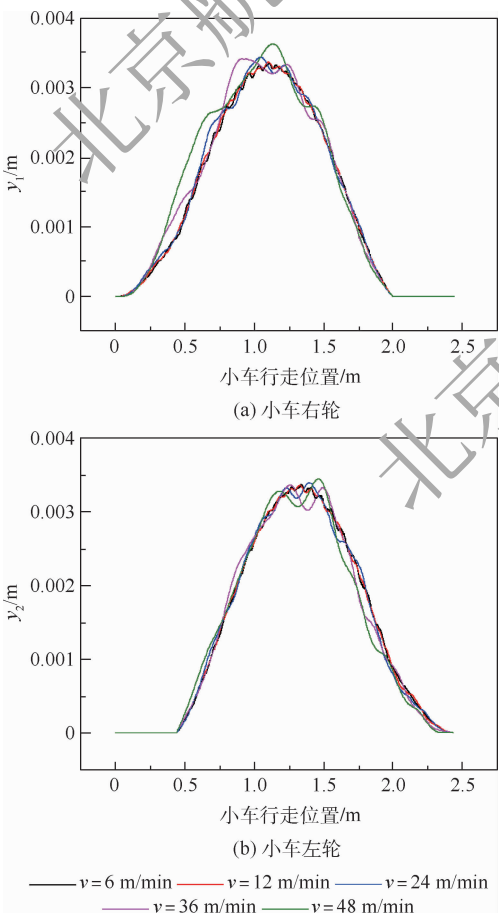


图 5 不同行走速度下小车右、左轮的位移响应

Fig. 5 Displacement of right and left wheel of truck at different traveling speeds

不同行走速度下小车体转角如图 6 所示,m_1 的位移响应如图 7 所示。

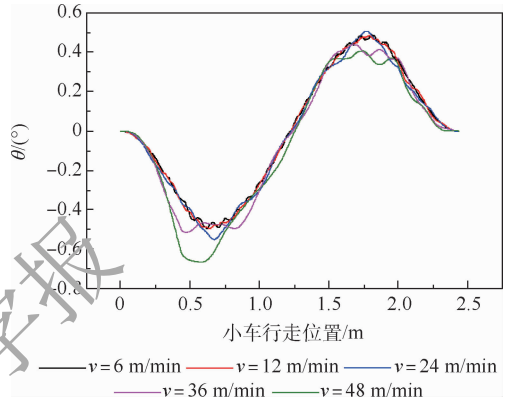


图 6 不同行走速度下小车体转角

Fig. 6 Rotation angle of truck at different traveling speeds

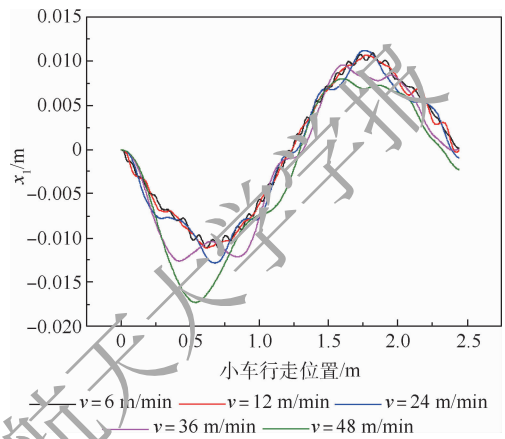


图 7 不同行走速度下 m_1 的位移响应

Fig. 7 Displacement of m_1 at different traveling speeds

由图 6 和图 7 可见:

1)相应的车体的转角和 m_1 的位移响应与右、左轮的位移相关,为先负后正的趋势,曲线与之相符。说明所建立的系统模型能正确反映系统的动力学行为。

2)同导轨情况类似,不同小车行走速度下小车体转角 θ、m_1 的位移响应 x_1 总体趋势基本相同,主要是由于小车两轮先后驶入驶出导轨。同样,小车行走速度越高,叠加在总体趋势上的波动次数越少,波动幅值越大;行走速度越低,波动次数越多,波动幅值越小,但在行走速度为 6、12 m/min 时,m_1 的位移响应和加速度响应的差别已经不明显。

3)图 6、图 7 中响应曲线并不是关于零位置对称,而是整体向纵轴下部偏移,并且行走速度越大,向下偏移的趋势越明显,主要是因为行走过程中燃料组件受到水的阻力作用。

不同行走速度下 m_1 的横向加速度如图 8 所示,吊篮的横向力如图 9 所示。

北京航空航天大学学报

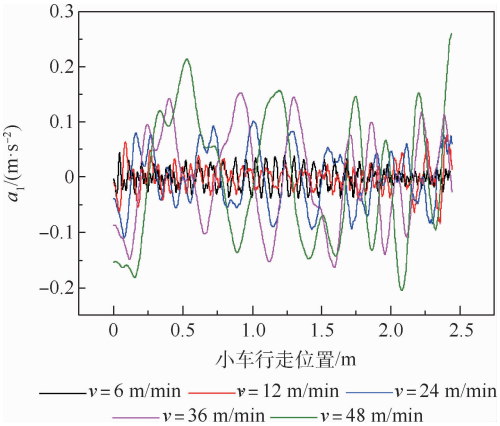


图 8　不同行走速度下 m_1 的横向加速度

Fig. 8　Horizontal acceleration of m_1 at different traveling speeds

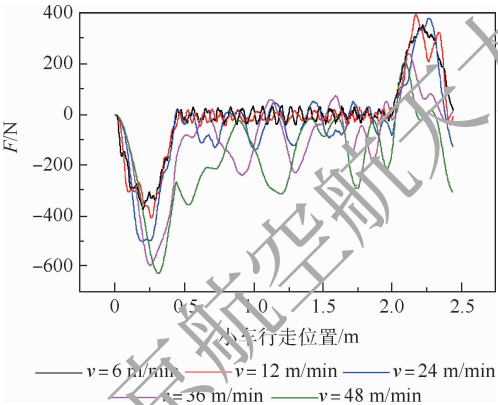


图 9　不同行走速度下吊篮的横向力

Fig. 9　Horizontal force of basket at different traveling speeds

由图 8、图 9 可以看到，小车行走速度越大，m_1 的横向加速度曲线及吊篮的横向力曲线波动幅值越大，由于水的阻力作用，吊篮的横向力曲线向下偏移量也越大。吊篮所受的最大横向力出现在小车刚驶入、驶出一跨导轨的位置，之后则大幅减小，但需注意到，行走速度为 24 m/min 及以上时，即使是小车处于导轨中间附近位置，横向力依旧有一定的波动幅值，此时，横向力曲线相较于 12 m/min时的情况变化比较大，但在行走速度为 6、12 m/min 时，横向力曲线差别不明显，结合图 6 和图 7，说明当小车行走速度小到一定程度时，再继续减小其速度对于控制某些指标的作用已经不大。

图 9 还可以看出，即使小车行走速度很小，在小车刚驶入、驶出一跨导轨的一段位置，吊篮依旧要承受一定的横向力，主要是由于导轨变形使车体倾斜引起。吊篮高度约为 4 m，吊篮与小车相连部位的力矩会比较大。因此，在控制小车行走速度的同时，也应根据情况对吊篮根部结构进行适当加强。

4　结　论

根据核动力厂水下燃料组件转运系统的结构形式和系统性能，建立了转运小车行走过程的梁-小车 + 刚性杆-质量 + 弹簧 + 阻尼的耦合动力学模型，并考虑了水的影响，推导了小车行走过程的微分方程，分析了不同小车行走速度时系统的响应，可得以下结论：

1）小车行走速度越低，整个系统相对越平稳，但由于存在静变形，通过控制小车行走速度对控制导轨最大变形和吊篮组件的变形及横向力等的作用有限。

2）小车行走过程中吊篮组件会不可避免地承受一定的横向力。由于水阻力作用，速度越大，吊篮的横向力越大，且力的方向越趋于同小车运行方向反向。由于吊篮高度尺寸较大，应注意吊篮与小车连接部位的强度。

3）在小车驶入、驶出导轨的一段距离内，对于吊篮变形、吊篮所受横向力及组件加速度来说都是最大的，需要重点关注，必要时可在导轨支撑位置附近设置监测点，以保证系统和燃料组件的安全。

参考文献（References）

[1] FRYBA L. Vibration of solids and structures under moving load [M]. Groninger: Noordhoff International Publishing, 1973: 3-243.

[2] CAI Y, CHEN S S, ROTE D M, et al. Vehicle/guideway interaction for high speed vehicles on a flexible guideway[J]. Journal of Sound and Vibration, 1994, 175(5): 625-646.

[3] NIELSEN J C O, IGELAND A. Vertical dynamic interaction between train and track-influence of wheel and track imperfections [J]. Journal of Sound and Vibration, 1995, 187(5): 825-839.

[4] DIANA G, CHELI F. Dynamic interaction of railway systems with large bridge[J]. Vehicle System Dynamics, 1989, 18(1/2/3): 71-106.

[5] 肖新标，沈火明. 3 种车桥耦合振动分析模型的比较研究[J]. 西南交通大学学报，2004，39(2)：172-175.
XIAO X B, SHEN H M. Comparison of three models for vehicle-bridge coupled vibration analysis[J]. Journal of Southwest Jiaotong University, 2004, 39(2): 172-175(in Chinese).

[6] MICHALTSOS G, SOPHIANOPOULOS D, KOUNADIS A N. The effect of a moving mass and other parameters on the dynamic response of a simply supported beam[J]. Journal of Sound and Vibration, 1996, 191(3): 357-362.

[7] LI Y L, QIANG S Z, LIAO H L, et al. Dynamics of wind-rail vehicle-bridge systems[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2005, 93(6): 483-507.

[8] 朱素红，徐斌. 考虑桥面不平顺的车桥耦合系统非线性振动[J]. 公路工程，2016，41(6)：66-69.

北京航空航天大学学报

ZHU S H, XU B. Nonlinear dynamic of vehicle-bridge coupled interaction system considering bridge surface roughness[J]. Highway Engineering, 2016, 41(6): 66-69(in Chinese).

[9] MICHALTSOS G T. Dynamic behavior of a single-span beam subjected to loads moving with variable speeds[J]. Journal of Sound and Vibration, 2002, 258(2): 359-372.

[10] SUN W J, GONG D, ZHOU J S, et al. Influences of suspended equipment under car body on high speed train ride quality[J]. Procedia Engineering, 2011, 16(1): 812-817.

[11] WANG L B, KANG X, JIANG P W. Vibration analysis of a multi-span continuous bridge subject to complex traffic loading and vehicle dynamic interaction[J]. KSCE Journal of Civil Engineering, 2016, 20(1): 323-332.

[12] CHEN Z W, XU Y L, XIA Y, et al. Fatigue analysis of long-span suspension bridges under multiple loading: case study[J]. Engineering Structures, 2011, 33(12): 3246-3256.

[13] 沈勇,王昌明,狄长安,等. 水中弹丸运动状态分析及试验研究[J]. 弹箭与制导学报, 2002, 22(1): 45-48.
SHEN Y, WANG C M, DI C A, et al. The analysis of underwater projector's motion[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2002, 22(1): 45-48(in Chinese).

[14] DONG R G. Effective mass and damping of submerged structures[R]. California: University of California Livermore, 1978: 10.

[15] 高涵,张明路,张小俊,等. 水下机械手动力学模型及力矩影响研究[J]. 机械设计与制造, 2017(3): 68-71.
GAO H, ZHANG M L, ZHANG X J, et al. Research on underwater manipulator dynamics model and torque influence[J]. Machinery Design & Manufacture, 2017(3): 68-71(in Chinese).

[16] 倪振华. 振动力学[M]. 西安: 西安交通大学出版社, 1988: 367-389.
NI Z H. Vibration mechanics[M]. Xi'an: Xi'an Jiao Tong University Press, 1988: 367-389(in Chinese).

Coupled vibration of underwater fuel assembly transfer system

YUAN Zhanhang[1,2], LI Yunhua[1,*], LIU Haodong[1]

(1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China;
2. China Nuclear Power Engineering Co., Ltd., Beijing 100840, China)

Abstract: The stability of fuel assembly transfer system and the reliability of system structure are very important to ensure the safety of fuel assembly in nuclear power plant. In order to obtain the vibration of system during the running process, the dynamic model of coupling system of the beam—truck + link—mass + spring + damping was established. Considering the effect of water, the nonlinear differential equation of the system is derived. The response of the system under different traveling speeds of truck was calculated by numerical algorithm. After comparison and analysis, it is considered that the smaller the truck speed is, the weaker the system vibration is and the more stable it is; due to the static deformation, even when the speed is very low, the basket will still bear a certain lateral force, so it is necessary to pay attention to the strength of the weak part of the basket; the position near the guide rail support is the dangerous position for the basket, and monitoring points can be set here if necessary to ensure the safety of the fuel assembly.

Keywords: fuel assembly transfer system; water; coupled system dynamics; response; safety

Received: 2020-11-16; **Accepted**: 2021-02-05; **Published online**: 2021-02-24 10:33

URL: kns.cnki.net/kcms/detail/11.2625.V.20210223.1734.001.html

* **Corresponding author**. E-mail: yhli@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0628

下肢康复机器人动力学建模约束违约抑制

全下文载

徐亚茹[1]，李克鸿[1]，刘佳[2]，刘荣[3]，张建成[1,*]

（1. 北京联合大学 机器人学院，北京 100020； 2. 石家庄学院 机电学院，石家庄 050035；
3. 北京航空航天大学 机器人研究所，北京 100083）

摘 要：U-K 理论为获得约束多体系统的解析动力学方程提供了新的理念，但由于数值近似和截断误差等因素的影响，动力学方程在位置和速度层面上存在约束违约。Baumgarte 约束违约稳定法（BSM）通过约束修正得到稳定的动力学方程。然而，Baumgarte 参数的选择通常涉及一个试错过程，可能会出现失效的仿真结果。为此，利用经典的四阶 Runge-Kutta 法研究了 Baumgarte 参数选取问题，创建了基于 BSM 修正后的 U-K 理论的机器人系统解析动力学方程。以下肢康复机器人为研究对象仿真分析，结果表明：利用所提方法可以有效抑制约束违约，关节角度误差控制在 -3×10^{-3}（°）～5×10^{-3}（°）范围内；关节角速度误差控制在 -2×10^{-4}～2×10^{-4} rad/s 范围内；机器人末端执行器运行轨迹能够很好地贴近系统预定的目标。

关 键 词：U-K 理论；约束违约；Baumgarte 约束违约稳定法（BSM）；Baumgarte 参数；下肢康复机器人

中图分类号：TH113

文献标志码：A **文章编号**：1001-5965(2022)04-0609-11

多体系统动力学建模方法有 Newton-Euler 方法、Lagrange 乘子法、Gibbs-Appell 方法、Kane 方法和 U-K 理论等。其中，U-K 理论[1]不需要辅助变量，可描述具有完整或非完整、独立或非独立约束关系的约束机械系统的解析动力学方程，给出解析形式的约束力；另外，由于 U-K 理论借助于广义逆这一数学工具，当系统约束方程的个数少于系统的广义变量个数时，即约束方程个数不限定出现约束矩阵不满秩时，仍可对系统动力学方程进行求解。因此，U-K 理论得到了国内外学者的重点关注[2-7]。约束多体系统动力学方程是将约束方程对时间求二阶导数获得加速度约束方程，并与动力学方程联立形成的方程组，因此没有违反加速度约束条件。然而，由于数值近似和截断误差等因素的影响，在位置和速度层面上的约束是违约的。因此，必须利用特殊的方法，以消除或最大限度地减小对约束的违约。通常，提高多体系统数值仿真精度最简单的方法是减小积分时间步长。然而，使用非常小的步长可能会使数值分析效率变得很低。应用较大的步长可以提高数值积分的效率，但同时也会使数值误差增大。对于约束系统，数值积分不精确的一个重要度量是约束违约。不受控制的约束违约会影响数值仿真的可靠性[8]。

约束违约稳定法通过在动力学方程中引入校正项[9]，间接地校正系统的广义坐标和广义速

收稿日期：2020-11-11；**录用日期**：2021-02-05；**网络出版时间**：2021-02-24 12:06
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210224.1111.001.html
基金项目：国家重点研发计划（2018YFB1307001）；北京市教委科研计划（KM202111417006）；河北省高等学校科学技术研究项目（QN2020510）；北京联合大学科研项目（ZK80202004）

* **通信作者**. E-mail：robotbuu@126.com

引用格式：*徐亚茹，李克鸿，刘佳，等. 下肢康复机器人动力学建模约束违约抑制[J]. 北京航空航天大学学报，2022，48(4)：609-619. XU Y R，LI K H，LIU J，et al. Constraint violation suppression for dynamics modeling of lower limb rehabilitation robot[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：609-619 (in Chinese).*

北京航空航天大学学报

北航学报 赠阅

度,以达到抑制约束违约的目的。Baumgarte 约束违约稳定法(Baumgarte's constraint violation stabilization methods, BSM)[10]因其简单性和计算实现的简便性而成功地应用于许多场合[11-15]。然而,BSM 在选择稳定参数时具有不确定性,可能会导致失效的仿真结果,即使对于具有有效解的系统也是如此[16]。事实上,正如 Baumgarte[10]所指出的,稳定参数的选择通常涉及一个试错过程。因此,Baumgarte 参数的选择问题是近几十年来许多研究工作的热点。

因此,本文研究了基于 BSM 修正后的 U-K 理论中的 Baumgarte 参数选取问题,建立了初始条件满足或不满足约束方程时的机器人系统的解析动力学方程。以下肢康复机器人为仿真对象,对比了利用本文方法在不同步长和 Baumgarte 参数取值下修正后的关节角度误差和关节角速度误差,验证了该方法可有效抑制约束违约。

1 基于 BSM 修正后的 U-K 理论

基于 U-K 理论的建模过程需要以下 3 步[1]:

步骤 1 未受约束多体系统动力学方程。

无约束时,机械系统的运动方程可以表述为

$$\begin{cases} \boldsymbol{M}(\boldsymbol{q},t)\ddot{\boldsymbol{q}} = \boldsymbol{Q}(\boldsymbol{q},\dot{\boldsymbol{q}},t) \\ \boldsymbol{q}(0) = \boldsymbol{q}_0 \\ \dot{\boldsymbol{q}}(0) = \dot{\boldsymbol{q}}_0 \end{cases} \tag{1}$$

式中:$\boldsymbol{M}(\boldsymbol{q},t) \in \mathbf{R}^{n\times n}$ 为 $n \times n$ 维广义质量矩阵;$\boldsymbol{Q}(\boldsymbol{q},\dot{\boldsymbol{q}},t) \in \mathbf{R}^n$ 为 n 维广义主动力矩阵,包括重力、外力和离心力/科氏力;$\dot{\boldsymbol{q}} \in \mathbf{R}^n$ 和 $\ddot{\boldsymbol{q}} \in \mathbf{R}^n$ 分别为广义速度向量和广义加速度向量;$\boldsymbol{q}_0$ 和 $\dot{\boldsymbol{q}}_0$ 分别为初始时刻的广义坐标和广义速度。

步骤 2 约束描述。

约束方程可表示为

$$\boldsymbol{\Phi} = [\phi_1 \quad \phi_2 \quad \cdots \quad \phi_p]^{\mathrm{T}} = 0 \tag{2}$$

式中:ϕ_i 为第 i 个约束。

假设约束(2)具有充足的光滑度,不可积分约束对时间 t 求一次导数,可积分约束对时间 t 求两阶导数,则约束方程表示为

$$\boldsymbol{A}(\boldsymbol{q},\dot{\boldsymbol{q}},t)\ddot{\boldsymbol{q}} = \boldsymbol{b}_v(\boldsymbol{q},\dot{\boldsymbol{q}},t) \tag{3}$$

式中:$\boldsymbol{A}(\boldsymbol{q},\dot{\boldsymbol{q}},t) \in \mathbf{R}^{p\times n}$ 为约束矩阵;$\boldsymbol{b}_v(\boldsymbol{q},\dot{\boldsymbol{q}},t) \in \mathbf{R}^p$ 为 p 维列向量。

步骤 3 约束多体系统动力学方程。

由于上述约束的存在,产生了约束力矩,使得系统的动力学方程变为

$$\boldsymbol{M}(\boldsymbol{q},t)\ddot{\boldsymbol{q}} = \boldsymbol{Q}(\boldsymbol{q},\dot{\boldsymbol{q}},t) + \boldsymbol{Q}^{\mathrm{c}}(\boldsymbol{q},\dot{\boldsymbol{q}},t) \tag{4}$$

式中:$\boldsymbol{Q}^{\mathrm{c}}(\boldsymbol{q},\dot{\boldsymbol{q}},t) \in \mathbf{R}^n$ 为系统满足约束条件所需的约束力。

Udwadia 和 Kalaba[1]给出了约束多体系统加速度的解析表达式为

$$\ddot{\boldsymbol{q}} = \boldsymbol{M}^{-1}\boldsymbol{Q} + \boldsymbol{M}^{-1/2}(\boldsymbol{A}\boldsymbol{M}^{-1/2})^{+}(\boldsymbol{b}_v - \boldsymbol{A}\boldsymbol{M}^{-1}\boldsymbol{Q}) \tag{5}$$

约束多体系统的解析动力学方程,即 U-K 理论,具有如下形式:

$$\boldsymbol{M}\ddot{\boldsymbol{q}} = \boldsymbol{Q} + \boldsymbol{M}^{1/2}(\boldsymbol{A}\boldsymbol{M}^{-1/2})^{+}(\boldsymbol{b}_v - \boldsymbol{A}\boldsymbol{M}^{-1}\boldsymbol{Q}) \tag{6}$$

数值积分法是对方程采用逐步数值积分求解而不作任何变换或修正的方法。约束违约稳定法通过引入包含约束方程的修正项以达到抑制约束违约的目的。针对初始条件与约束方程不相容的情况,为了减少约束违约对系统的影响,BSM 将反馈控制的思想引入微分方程数值积分方法中,达到对直接积分法改进的目的[10]。该方法基于反馈控制原理,将位置约束和速度约束引入加速度约束方程中,通过约束修正得到稳定化的动力学方程。在开环系统中,外界如果有任何扰动,$\boldsymbol{\Phi}$ 就不会收敛到零,导致系统失稳。对于闭环系统,只要 $\alpha_i,\beta_i > 0$,则修正的加速度约束方程[10]

$$\ddot{\boldsymbol{\Phi}}(\boldsymbol{q},t) + \boldsymbol{\alpha}\dot{\boldsymbol{\Phi}}(\boldsymbol{q},t) + \boldsymbol{\beta}\boldsymbol{\Phi}(\boldsymbol{q},t) = \boldsymbol{0} \tag{7}$$

是稳定的,即 $\boldsymbol{\Phi}$ 和 $\dot{\boldsymbol{\Phi}}$ 将收敛到零。式中:$\boldsymbol{\alpha} = \mathrm{diag}\{\alpha_1,\alpha_2,\cdots,\alpha_p\}$,$\boldsymbol{\beta} = \mathrm{diag}\{\beta_1,\beta_2,\cdots,\beta_p\}$ 为使该微分方程的解获得稳定的反馈控制参数;$\boldsymbol{\alpha}\dot{\boldsymbol{\Phi}}$ 和 $\boldsymbol{\beta}\boldsymbol{\Phi}$ 可看作控制项。

初始条件与约束方程不相容会导致约束违约,通常会使得仿真结果与实际情况相差较大,因此,需要修正动力学方程。根据上述分析,基于 BSM 修正后的 U-K 理论可得到稳定化的约束多体系统动力学方程,并表示为

$$\boldsymbol{M}\ddot{\boldsymbol{q}} = \boldsymbol{Q} + \boldsymbol{M}^{1/2}(\boldsymbol{A}\boldsymbol{M}^{-1/2})^{+}(\boldsymbol{b}_v - \boldsymbol{\alpha}\dot{\boldsymbol{\Phi}} - \boldsymbol{\beta}\boldsymbol{\Phi} - \boldsymbol{A}\boldsymbol{M}^{-1}\boldsymbol{Q}) \tag{8}$$

2 Baumgarte 参数选择

在初始条件不满足约束方程时,BSM 在某些应用中给出了准确的结果,但是没有可靠的方法来选择控制参数,不恰当的选择可能导致错误的结果。多体系统动力学方程为一系列二阶微分方程,一般的求解方法是将其初值问题转化为一阶微分方程的初值问题。

北京航空航天大学学报

北航学报 赠阅

考虑一阶微分方程的初值问题为[17]

$$\begin{cases} u' = f(x,u) \\ u(x_0) = u_0 \end{cases} \tag{9}$$

目标是寻求函数 $u(x)$ 在节点 x_k 上的近似值 u_k。选取等距节点 $x_k = x_0 + kh(k = 0,1,2,\cdots)$，$h$ 为步长。

求解初值问题(式(9))的近似计算方法

$$u_{k+1} = u_k + h\varphi(x_k,u_k,h) \quad k = 0,1,2,\cdots \tag{10}$$

为单步法。式中：$\varphi(x,u,h)$ 由函数 $f(x,u)$ 的某种形式确定。

为了构造函数 $\varphi(x_k,u(x_k),h)$，令

$$\begin{cases} K_1 = f(x_k,u_k) \\ K_2 = f(x_k + s_2h, u_k + c_{21}hK_1) \\ K_3 = f(x_k + s_3h, u_k + c_{31}hK_1 + c_{32}hK_2) \\ \vdots \\ K_m = f(x_k + s_mh, u_k + h\sum_{i=1}^{m-1} c_{mi}K_i) \end{cases} \tag{11}$$

得到近似计算表达式[17]

$$u_{k+1} = u_k + h\sum_{i=1}^{m} \lambda_i K_i \quad k = 0,1,2,\cdots \tag{12}$$

式中：$K_1 = f(x_k,y_k)$，$K_2 = f\left(x_k + \frac{h}{2}, u_k + \frac{h}{2}K_1\right)$，$K_3 = f\left(x_k + \frac{h}{2}, u_k + \frac{h}{2}K_2\right)$，$K_4 = f(x_k + h, u_k + hK_3)$。

获得具有收敛解的 Baumgarte 参数依赖于求解该一阶微分方程所采用的数值解法。基于经典的四阶 Runge-Kutta 法来讨论如何选取 Baumgarte 参数。首先，重新改写方程(7)为下列矩阵形式[18]：

$$\dot{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{X} \tag{13}$$

式中：$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{\Phi} \\ \dot{\boldsymbol{\Phi}} \end{bmatrix}$；$\boldsymbol{P} = \begin{bmatrix} 0 & 1 \\ -\boldsymbol{\beta} & -\boldsymbol{\alpha} \end{bmatrix}$。

四阶 Runge-Kutta 法表示为

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n + \frac{1}{6}(\boldsymbol{K}_1 + 2\boldsymbol{K}_2 + 2\boldsymbol{K}_3 + \boldsymbol{K}_4) \tag{14}$$

式中：

$$\boldsymbol{K}_1 = h[\boldsymbol{P}(\boldsymbol{X})]$$

$$\boldsymbol{K}_2 = h\left[\boldsymbol{P}\left(\boldsymbol{X} + \frac{1}{2}\boldsymbol{K}_1\right)\right] = h\boldsymbol{P}\left[\boldsymbol{X} + \frac{1}{2}(h\boldsymbol{P}\boldsymbol{X})\right] = \left(h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2\right)\boldsymbol{X}$$

$$\boldsymbol{K}_3 = h\left[\boldsymbol{P}\left(\boldsymbol{X} + \frac{1}{2}\boldsymbol{K}_2\right)\right] = h\boldsymbol{P}\left[\boldsymbol{X} + \frac{1}{2}\left(h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2\right)\boldsymbol{X}\right] = \left(h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2 + \frac{1}{4}h^3\boldsymbol{P}^3\right)\boldsymbol{X}$$

$$\boldsymbol{K}_4 = h[\boldsymbol{P}(\boldsymbol{X} + \boldsymbol{K}_3)] = h\boldsymbol{P}\left[\boldsymbol{X} + \left(h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2 + \frac{1}{4}h^3\boldsymbol{P}^3\right)\boldsymbol{X}\right] = \left(h\boldsymbol{P} + h^2\boldsymbol{P}^2 + \frac{1}{2}h^3\boldsymbol{P}^3 + \frac{1}{4}h^4\boldsymbol{P}^4\right)\boldsymbol{X}$$

将 $\boldsymbol{K}_1$、$\boldsymbol{K}_2$、$\boldsymbol{K}_3$ 和 $\boldsymbol{K}_4$ 代入式(14)得

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n + \frac{1}{6}(\boldsymbol{K}_1 + 2\boldsymbol{K}_2 + 2\boldsymbol{K}_3 + \boldsymbol{K}_4) = \boldsymbol{X}_n + \frac{1}{6}\left[h\boldsymbol{P} + 2\left(h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2\right) + 2\left(h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2 + \frac{1}{4}h^3\boldsymbol{P}^3\right) + \left(h\boldsymbol{P} + h^2\boldsymbol{P}^2 + \frac{1}{2}h^3\boldsymbol{P}^3 + \frac{1}{4}h^4\boldsymbol{P}^4\right)\right] \cdot \boldsymbol{X}_n$$

$$\boldsymbol{X}_n = \boldsymbol{X}_n + \frac{1}{6}\left(6h\boldsymbol{P} + 3h^2\boldsymbol{P}^2 + h^3\boldsymbol{P}^3 + \frac{1}{4}h^4\boldsymbol{P}^4\right) \cdot \boldsymbol{X}_n$$

$$\boldsymbol{X}_n = \left(\boldsymbol{I} + h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2 + \frac{1}{6}h^3\boldsymbol{P}^3 + \frac{1}{24}h^4\boldsymbol{P}^4\right)\boldsymbol{X}_n \tag{15}$$

对式(15)进行 Z 变换得[19]

$$Z\boldsymbol{X}_n(Z) = \left(\boldsymbol{I} + h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2 + \frac{1}{6}h^3\boldsymbol{P}^3 + \frac{1}{24}h^4\boldsymbol{P}^4\right)\boldsymbol{X}_n(Z) \tag{16}$$

系统的特征方程为

$$\left|Z\boldsymbol{I} - \left(\boldsymbol{I} + h\boldsymbol{P} + \frac{1}{2}h^2\boldsymbol{P}^2 + \frac{1}{6}h^3\boldsymbol{P}^3 + \frac{1}{24}h^4\boldsymbol{P}^4\right)\right| = \left|Z\boldsymbol{I} - \left(\boldsymbol{I} + \hat{\boldsymbol{P}} + \frac{1}{2}\hat{\boldsymbol{P}}^2 + \frac{1}{6}\hat{\boldsymbol{P}}^3 + \frac{1}{24}\hat{\boldsymbol{P}}^4\right)\right| = 0 \tag{17}$$

式中：$h\boldsymbol{P} = \begin{bmatrix} 0 & h \\ \boldsymbol{\beta}h & -\boldsymbol{\alpha}h \end{bmatrix}$；$\hat{\boldsymbol{P}} = \begin{bmatrix} 0 & 1 \\ -\hat{\boldsymbol{\beta}} & -\hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\boldsymbol{\beta}h^2 & -\boldsymbol{\alpha}h \end{bmatrix}$，$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}h$，$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}h^2$。用 $\hat{\boldsymbol{P}}$ 代替 $h\boldsymbol{P}$ 不会改变系统的特征方程[18]。

Baumgarte 参数选取步骤为：依据控制理论相关知识，稳定边界是单位圆 $|Z| = 1$，当 $|Z| < 1$ 时，$\dot{\boldsymbol{\Phi}}$ 和 $\boldsymbol{\Phi}$ 将会收敛到零，而当 $|Z| > 1$ 时，$\dot{\boldsymbol{\Phi}}$ 和 $\boldsymbol{\Phi}$ 将会发散。图 1 描述了当 $|Z| \leqslant 1$ 时 $\hat{\boldsymbol{\alpha}}$、$\hat{\boldsymbol{\beta}}$ 的选取点。根据图 1，在分布区域可选择合适的 $\hat{\boldsymbol{\alpha}}$、$\hat{\boldsymbol{\beta}}$；再结合 $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}/h$，$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}/h^2$，当步长 h 给定时，Baumgarte 参数则得以确定。

北京航空航天大学学报

北航学报 瞰阅

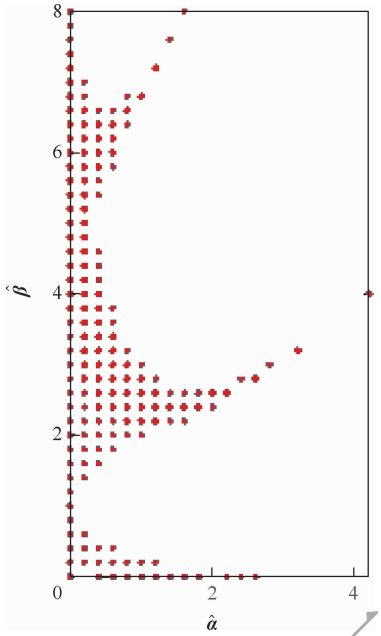


图 1 $|Z|\leqslant 1$ 时 $\hat{\boldsymbol{\alpha}}$、$\hat{\boldsymbol{\beta}}$ 的分布情况示意图

Fig. 1 Schematic diagram of distribution of $\hat{\boldsymbol{\alpha}}$、$\hat{\boldsymbol{\beta}}$($|Z|\leqslant 1$)

3 仿真结果及分析

下肢康复机器人是外骨骼机器人技术与康复训练相结合的产物，本质上属于约束多体系统。在康复训练过程中，根据预计的下肢末端轨迹，求解各关节所需的驱动力是控制策略制定和康复效果评估的依据。根据人体生物力学中的人体多刚体假设基本理论，可将人体各部分假设为刚体，并通过转动副连接。若恢复患者的髋关节、膝关节和踝关节在矢状面的协调转动，即可实现康复训练的目的，从而帮助患者恢复行走功能[20]。连续被动的运动可促使患者进行长时间的康复训练，以恢复运动灵活性和被动活动范围。如图 2 所示，该机器人末端执行器作重复性的圆周运动。

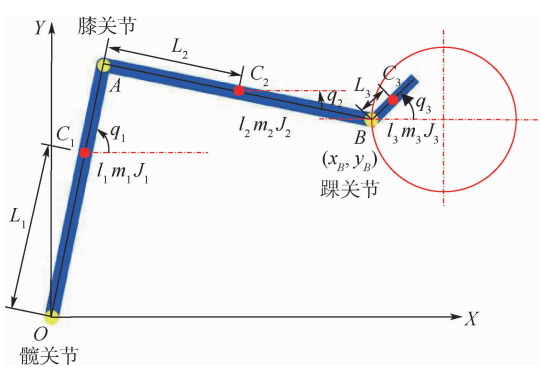


图 2 下肢康复机器人系统示意图

Fig. 2 Schematic diagram of lower limb rehabilitation robot system

将人体下肢各段与康复机器人各段对应，形成下肢康复机器人系统，简化为平面三连杆机构。其中，$i=1,2,3$ 分别表示大腿段、小腿段和足段。坐标原点 O 为髋关节，m_i 为系统第 i 连杆的质量，$m_i=m_i'+m_i''$，m_i'为下肢康复机器人第 i 体段的质量，m_i''为人体下肢第 i 体段的质量。l_i 为系统第 i 连杆的长度，人体下肢与下肢康复机器人各体段长度相等。L_i 为系统第 i 连杆质心 C_i 到其关节 i 的距离，$L_i=(m_i'd_i'+m_i''d_i'')/(m_i'+m_i'')$，$d_i'$为下肢康复机器人第 i 体段质心到关节 i 的距离，d_i''为人体下肢第 i 体段质心到关节 i 的距离。J_i 为系统第 i 连杆对质心的转动惯量，$J_i=m_iL_i^2/3$。q_i 为系统第 i 连杆和水平轴之间的夹角，逆时针为正；$(x_B,\ y_B)$为系统的踝关节 B 在基坐标系中的坐标。

对于未受轨迹约束的下肢康复机器人，其动力学方程的表现形式为

$$\boldsymbol{\tau}=\boldsymbol{M}(\boldsymbol{q})\cdot\ddot{\boldsymbol{q}}+\tilde{\boldsymbol{C}}(\boldsymbol{q})\cdot\dot{\boldsymbol{q}}^2+\boldsymbol{G}(\boldsymbol{q}) \tag{18}$$

式中：$\boldsymbol{M}(\boldsymbol{q})\in\mathbf{R}^{3\times 3}$为对称正定质量（转动惯量）矩阵；$\tilde{\boldsymbol{C}}\cdot\dot{\boldsymbol{q}}\in\mathbf{R}^{3\times 3}$为 3×3 离心力和科氏力相关矩阵；$\boldsymbol{G}(\boldsymbol{q})\in\mathbf{R}^{3\times 1}$为重力列矩阵；$\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}}\in\mathbf{R}^{3\times 1}$分别代表广义的位置、速度、加速度；$\boldsymbol{\tau}\in\mathbf{R}^{3\times 1}$为 3×1 的关节力矩，且

$$\begin{aligned}
M_{11}&=[m_1L_1^2+(m_2+m_3)l_1^2]+J_1\\
M_{12}&=-(m_2L_2+m_3l_2)l_1\cos(q_1-q_2)\\
M_{13}&=m_3l_1L_3\cos(q_1-q_3)\\
M_{21}&=(m_2L_2+m_3l_2)l_1\cos(q_1-q_2)\\
M_{22}&=(m_2L_2^2+m_3l_2^2)+J_2\\
M_{23}&=m_3l_2L_3\cos(q_2-q_3)\\
M_{31}&=m_3l_1L_3\cos(q_1-q_3)\\
M_{32}&=m_3l_2L_3\cos(q_2-q_3)\\
M_{33}&=m_3L_3^2+J_3\\
\tilde{C}_{12}&=(m_2L_2+m_3l_2)l_1\sin(q_1-q_2)\\
\tilde{C}_{13}&=m_3l_1L_3\sin(q_1-q_3)\\
\tilde{C}_{21}&=-(m_2L_2+m_3l_2)l_1\sin(q_1-q_2)\\
\tilde{C}_{23}&=m_3l_2L_3\sin(q_2-q_3)\\
\tilde{C}_{31}&=-m_3l_1L_3\sin(q_1-q_3)\\
\tilde{C}_{32}&=-m_3l_2L_3\sin(q_2-q_3)\\
\tilde{C}_{11}&=\tilde{C}_{22}=\tilde{C}_{33}=0\\
G_1&=[m_1L_1+(m_2+m_3)l_1]g\cos q_1\\
G_2&=(m_2L_2+m_3l_2)g\cos q_2\\
G_3&=m_3gL_3\cos q_3
\end{aligned}$$

假设末端执行器轨迹约束描述为

北京航空航天大学学报

北航学报 赠阅

$$\boldsymbol{\Phi}=\begin{bmatrix} l_1\cos q_1+l_2\cos q_2-0.58-0.1\cos(0.4\pi t+\pi) \\ l_1\sin q_1+l_2\sin q_2-0.3-0.1\sin(0.4\pi t) \\ q_3-0.25\pi-0.006\pi t \end{bmatrix}=\boldsymbol{0} \tag{19}$$

将式(19)对时间 t 求两阶导数，得到如式(20)所示表达式：

$$\begin{bmatrix} -l_1\sin q_1 & -l_2\sin q_2 & 0 \\ l_1\cos q_1 & l_2\cos q_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \\ \ddot{q}_3 \end{bmatrix}=\begin{bmatrix} l_1\dot{q}_1^2\cos q_1+l_2\dot{q}_2^2\cos q_2-0.1(0.4\pi)^2\cos(0.4\pi t+\pi) \\ l_1\dot{q}_1^2\sin q_1+l_2\dot{q}_2^2\sin q_2-0.1(0.4\pi)^2\sin(0.4\pi t) \\ 0 \end{bmatrix} \tag{20}$$

利用基于 BSM 修正后的 U-K 理论，式(20)可转化为

$$\begin{bmatrix} -l_1\sin q_1 & -l_2\sin q_2 & 0 \\ l_1\cos q_1 & l_2\cos q_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \\ \ddot{q}_3 \end{bmatrix}=\begin{bmatrix} l_1\dot{q}_1^2\cos q_1+l_2\dot{q}_2^2\cos q_2-0.1(0.4\pi)^2\cos(0.4\pi t+\pi)- \\ \alpha_1(-\dot{q}_1l_1\sin q_1-\dot{q}_2l_2\sin q_2+0.1(0.4\pi)\sin(0.4\pi t+\pi))- \\ \beta_1(l_1\cos q_1+l_2\cos q_2-0.58-0.1\cos(0.4\pi t+\pi)) \\ \hline l_1\dot{q}_1^2\sin q_1+l_2\dot{q}_2^2\sin q_2-0.1(0.4\pi)^2\sin(0.4\pi t)- \\ \alpha_2(\dot{q}_1l_1\cos q_1+\dot{q}_2l_2\cos q_2-0.1(0.4\pi)\cos(0.4\pi t))- \\ \beta_2(l_1\sin q_1+l_2\sin q_2-0.3-0.1\sin(0.4\pi t)) \\ \hline -\alpha_3(\dot{q}_3-0.006\pi)-\beta_3(q_3-0.25\pi-0.006\pi t) \end{bmatrix} \tag{21}$$

取 $\alpha=\alpha_1=\alpha_2=\alpha_3$，$\beta=\beta_1=\beta_2=\beta_3$。结合图 1，以及 $\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}/h$，$\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}/h^2$，根据所取不同步长求解出 Baumgarte 参数值，如表 1 所示。

理论上，在任意时刻，零阶和一阶的广义变量的初始值必须满足约束方程及其一阶导数。然而，初始条件与约束方程不相容的情况更为常见。因此，假设初始条件为：$q_1(0)=78°$，$q_2(0)=-11°$，$q_3(0)=44.8°$；$\dot{q}_1(0)=0.061$ rad/s，$\dot{q}_2(0)=0.298$ rad/s，$\dot{q}_3(0)=0.021$ rad/s。仿真参数为：$m_1=13.3$ kg，$m_2=5.5$ kg，$m_3=3$ kg；$l_1=0.39$ m，$l_2=0.41$ m，$l_3=0.08$ m，$L_1=0.287$ m，$L_2=0.234$ m，$L_3=0.08$ m。

表 1　不同 Baumgarte 参数取值

Table 1　Different Baumgarte parameter values

步长	Baumgarte 参数		
$h=0.01$	$\hat{\alpha}=0.002$ $\hat{\beta}=0.000\ 02$ $\alpha=0.2$ $\beta=0.2$	$\hat{\alpha}=0.015$ $\hat{\beta}=0.000\ 2$ $\alpha=1.5$ $\beta=2$	$\hat{\alpha}=0.01$ $\hat{\beta}=0.003$ $\alpha=1$ $\beta=3$
$h=0.01$	$\hat{\alpha}=0.02$ $\hat{\beta}=0.002$ $\alpha=2$ $\beta=20$	$\hat{\alpha}=0.15$ $\hat{\beta}=0.02$ $\alpha=15$ $\beta=200$	$\hat{\alpha}=0.1$ $\hat{\beta}=0.03$ $\alpha=10$ $\beta=300$
$h=0.001$	$\hat{\alpha}=0.002$ $\hat{\beta}=0.000\ 02$ $\alpha=2$ $\beta=20$	$\hat{\alpha}=0.015$ $\hat{\beta}=0.000\ 2$ $\alpha=15$ $\beta=200$	$\hat{\alpha}=0.01$ $\hat{\beta}=0.003$ $\alpha=10$ $\beta=300$

北京航空航天大学学报

北航学报 赠阅

仿真结果如图 3 ~ 图 6 所示。当 $h=0.01$，Baumgarte 参数取 $\alpha=0.2,\beta=0.2,\alpha=1.5,\beta=2,\alpha=1,\beta=3$ 时，仿真结果如图 3 和图 4 所示；当 $h=0.01$，Baumgarte 参数取 $\alpha=2,\beta=20,\alpha=15,\beta=200,\alpha=10,\beta=300$ 时，仿真结果如图 3 和图 4 所示；当 $h=0.001$，Baumgarte 参数取 $\alpha=2,\beta=20,\alpha=15,\beta=200,\alpha=10,\beta=300$ 时，仿真结果如图 3 和图 4 所示。

1）图 3 为基于 BSM 修正后的 U-K 理论在不同步长和参数取值下所得的关节角度误差。对比可知，当 $h=0.001,\alpha=10,\beta=300$ 时，关节角度误差在零值上下浮动，明显优于其他情况。为便于观察仿真结果，在图 3 设置了放大窗口，可明显观察到实线所示的关节角度误差最小。

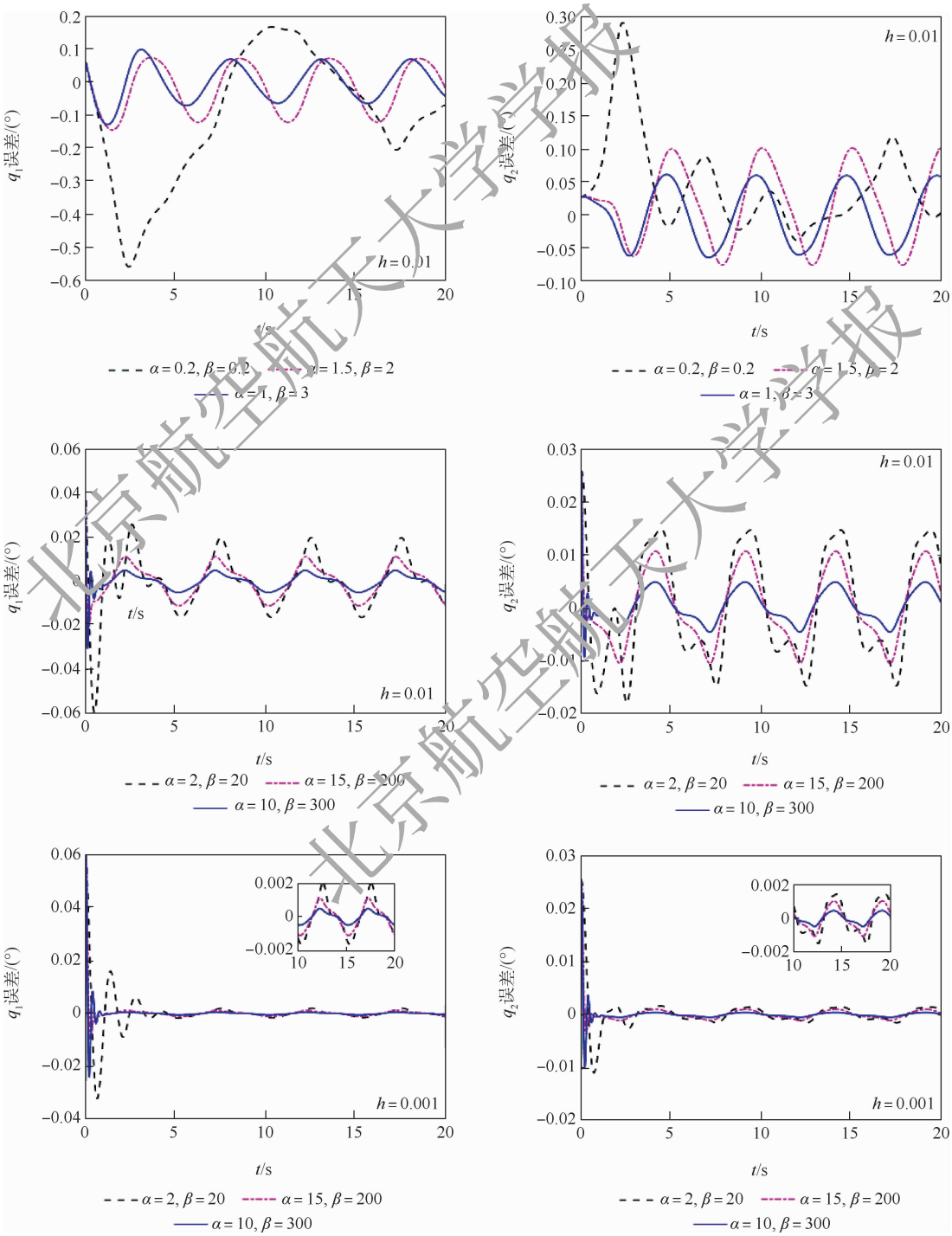


图 3 修正后的关节角度误差对比(不同步长和 Baumgarte 参数)

Fig. 3 Comparison of corrected joint angle errors (different step lengths and Baumgarte parameters)

北航学报 赠阅

2）图 4 为基于 BSM 修正后的 U-K 理论在不同步长和参数取值下所得的关节角速度误差。仿真结果表明，修正后的数值解均优于修正前的数值解。为便于观察 $h=0.001$，$\alpha=10$，$\beta=300$ 的仿真结果，在图 4 设置了放大窗口。与其他 Baumgarte 参数取值情况对比，明显地观察到实线所示的关节角速度误差更小。

3）当 $h=0.01$，$\alpha=10$，$\beta=300$ 时，仿真结果如图 5 和图 6 所示。与图 5（a）和图 5（c）相比，图 5（b）所示的关节角度和图 5（d）所示的关节角速度的修正数值解与理论值的曲线贴合良好。与图 5（e）相比，图 5（f）所示为修正后的运行轨迹的

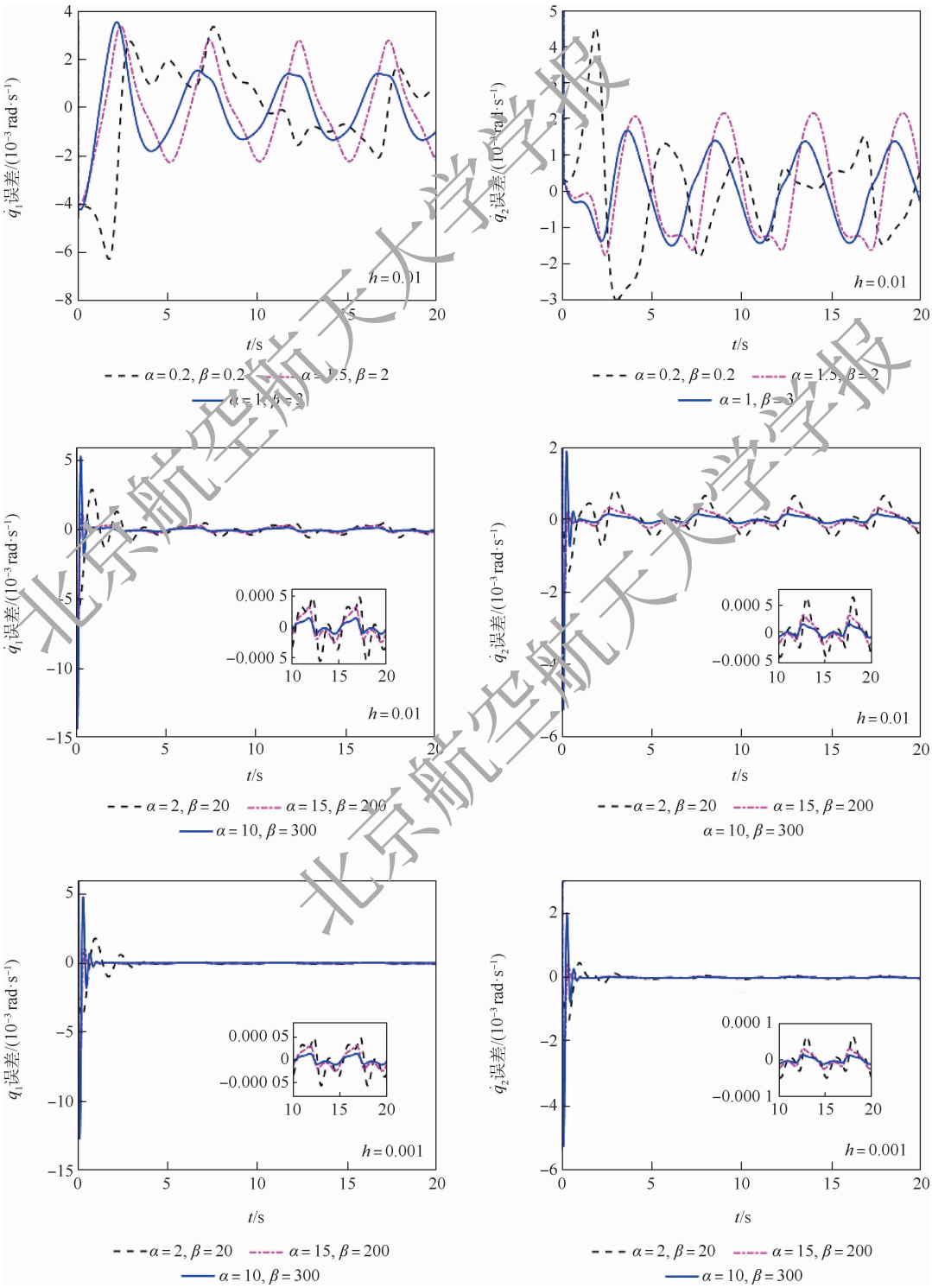


图 4　修正后的关节角速度误差对比（不同步长和 Baumgarte 参数）

Fig. 4　Comparison of corrected joint angular velocity errors (different step lengths and Baumgarte parameters)

北京航空航天大学学报

北航学报 赠阅

整体效果得到了很大的改善。具体地，与图 6(a)相比，图 6(c)、6(e)所示的关节角度误差大幅减小，均控制在 -5×10^{-3}(°) ~ 5×10^{-3}(°)范围内；与图 6(b)相比，图 6(d)、6(f)所示的关节角速度误差回归至零值附近波动，均控制在 -2×10^{-4} ~ 2×10^{-4} rad/s 范围内。

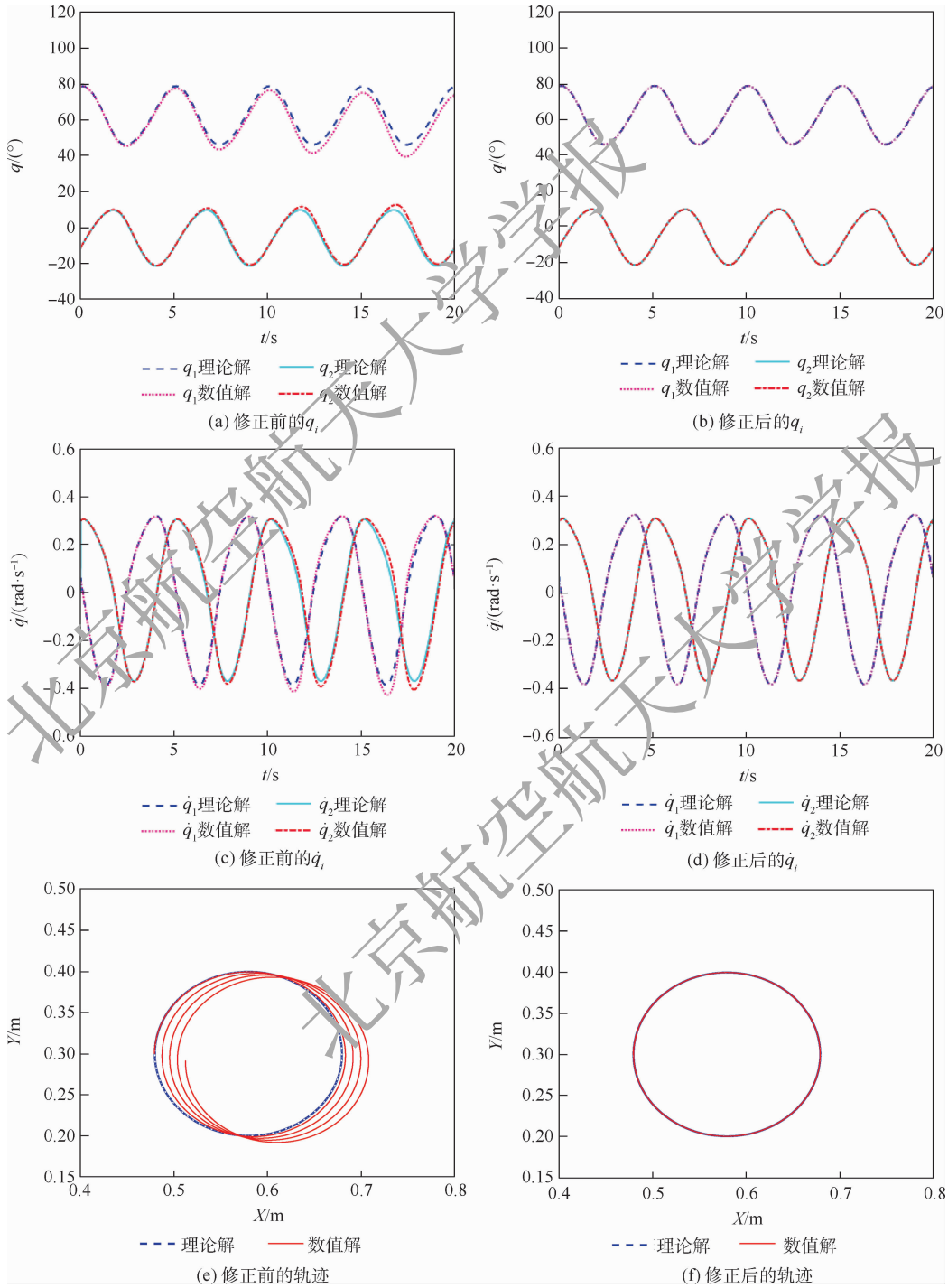
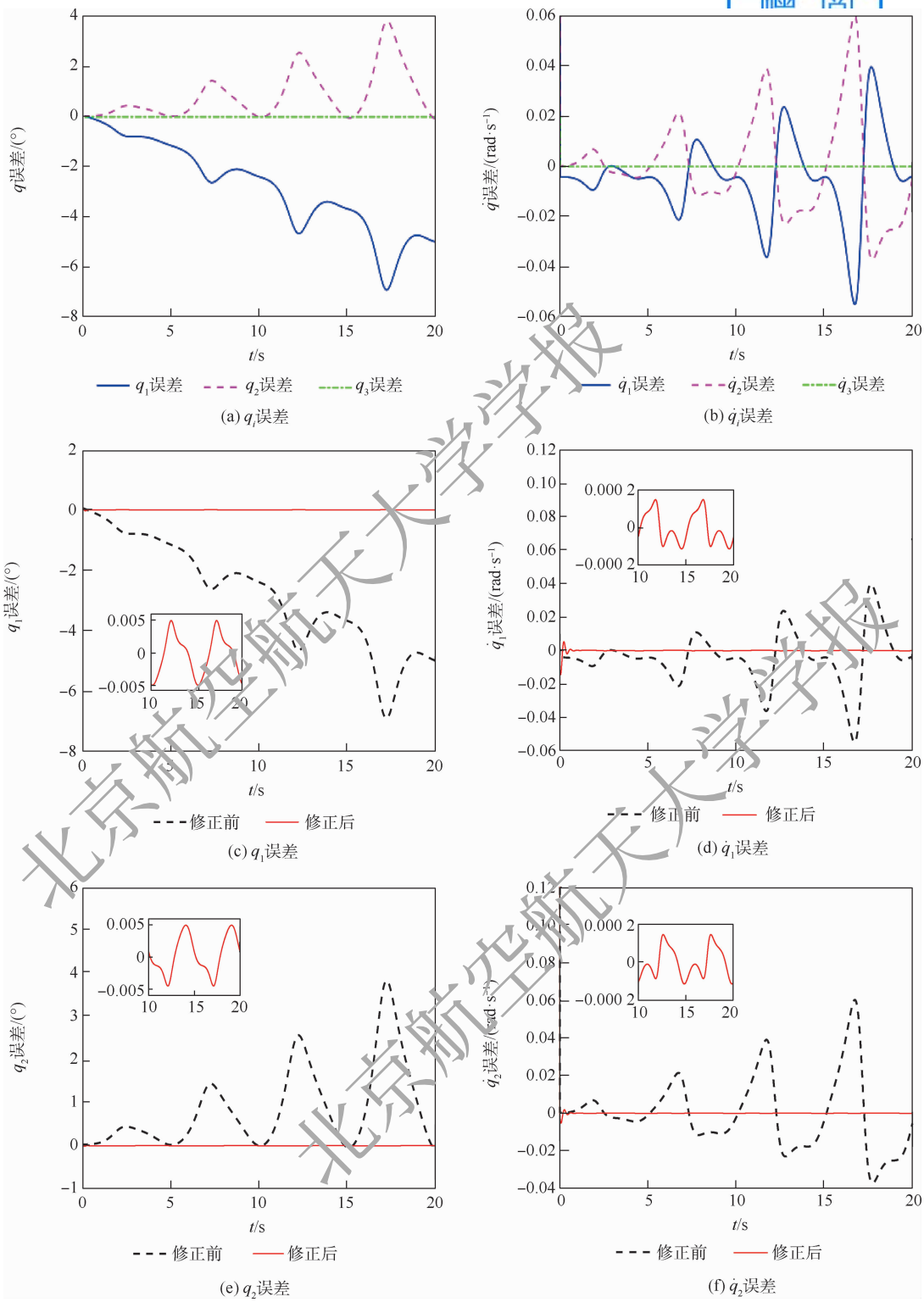


图 5 修正前后数值解与理论解的对比（$h=0.01$，$\alpha=10$，$\beta=300$）

Fig. 5 Comparison of numerical solutions and theoretical solutions before and after correction ($h=0.01$, $\alpha=10$, $\beta=300$)



北京航空航天大学学报

图 6　修正前后误差的对比（$h=0.01$，$\alpha=10$，$\beta=300$）

Fig. 6　Comparison of errors before and after correction ($h=0.01$, $\alpha=10$, $\beta=300$)

4　结　论

1）初始条件与约束方程相容或不相容均会导致约束违约，通常会使得仿真结果与实际情况相差较大，因此需要修正动力学方程。基于 U-K 理论通过 BSM 得到了稳定化的机器人系统动力学方程。

2）BSM 通过引入包含约束方程的修正项以达到抑制约束违约的目的，但是没有可靠的方法来选择 Baumgarte 参数，不恰当的选择可能导致错误的结果。基于经典的四阶 Runge-Kutta 法讨论了如何选取 Baumgarte 参数。

北航学报 醒阅

3）以下肢康复机器人为研究对象进行仿真分析，在不同步长和 Baumgarte 参数取值下所得的关节角度、关节角速度的修正数值解与理论值的曲线贴合良好；关节角度误差、关节角速度误差范围大幅减小。采用该方法可以效抑制机器人系统动力学方程在位置和速度层面上的约束违约，使得机器人末端执行器运行轨迹很好地贴近系统预定目标，具有较高精度。

但是，利用约束违约稳定法并不能通过选取稳定参数使违约误差的控制范围不加限制的减小。因此，下一步工作将分析探讨更精确的约束违约抑制方法。

参考文献（References）

[1] UDWADIA F E, KALABA R E. Analytical dynamics: A new approach[M]. Cambridge: Cambridge University Press, 2007.

[2] LIU J, LIU R. Dynamic modeling of dual-arm cooperating manipulators based on Udwadia-Kalaba equation[J]. Advances in Mechanical Engineering, 2016, 8(7): 1-10.

[3] HUANG K, SHAO K, ZHEN S C, et al. A novel approach for modeling and tracking control of a passive-wheel snake robot [J]. Advances in Mechanical Engineering, 2017, 9(3): 1-15.

[4] LYU G Z, LIU R. Determination of stability correction parameters for dynamic equations of constrained multibody systems [J]. Mathematical Problems in Engineering, 2018, 2018: 1-10.

[5] SUN H, ZHAO H, HUANG K, et al. A new approach for vehicle lateral velocity and yaw rate control with uncertainty[J]. Asian Journal of Control, 2018, 20(1): 216-227.

[6] ZHAO X M, CHEN Y H, ZHAO H, et al. Udwadia-Kalaba equation for constrained mechanical systems: Formulation and applications[J]. Chinese Journal of Mechanical Engineering, 2018, 31(6): 1-14.

[7] PAPPALARDO C M, GUIDA D. On the dynamics and control of underactuated nonholonomic mechanical systems and applications to mobile robots[J]. Archive of Applied Mechanics, 2019, 89(4): 669-698.

[8] BLAJER W. Elimination of constraint violation and accuracy aspects in numerical simulation of multibody systems[J]. Multibody System Dynamics, 2002, 7(3): 265-284.

[9] BRAUN D J, GOLDFARB M. Eliminating constraint drift in the numerical simulation of constrained dynamical systems[J]. Computer Methods in Applied Mechanics and Engineering, 2009, 198(37-40): 3151-3160.

[10] BAUMGARTE J. Stabilization of constraints and integrals of motion in dynamical systems[J]. Computer Methods in Applied Mechanics and Engineering, 1972, 1(1): 1-16.

[11] LIN S T, CHEN M W. A PID type constraint stabilization method for numerical integration of multibody systems[J]. Journal of Computational and Nonlinear Dynamics, 2011, 6(4): 044501.

[12] ZHANG X C, ZHANG X M, CHEN Z. Dynamic analysis of a 3-RRR parallel mechanism with multiple clearance joints[J]. Mechanism and Machine Theory, 2014, 78(78): 105-115.

[13] MARQUES F, SOUTO A P, FLORES P. On the constraints violation in forward dynamics of multibody systems[J]. Multibody System Dynamics, 2016, 39: 385-419.

[14] LIANG D, SONG Y M, SUN T. Nonlinear dynamic modeling and performance analysis of a redundantly actuated parallel manipulator with multiple actuation modes based on FMD theory [J]. Nonlinear Dynamics, 2017, 89(1): 391-428.

[15] GUO X, ZHANG D G, LI L, et al. Application of the two-loop procedure in multibody dynamics with contact and constraint [J]. Journal of Sound and Vibration, 2018, 427: 15-27.

[16] FLORES P, MACHADO M, SEABRA E, et al. A parametric study on the Baumgarte stabilization method for forward dynamics of constrained multibody systems[J]. Journal of Computational and Nonlinear Dynamics, 2011, 6(1): 1-9.

[17] 李健玲. 数值分析与实验[M]. 北京：科学出版社，2012：194-196.
LI J L. Numerical analysis and experiments[M]. Beijing: Science Press, 2012: 194-196(in Chinese).

[18] LIN S T, HUANG J N. Stabilization of Baumgarte's method using the Runge-Kutta approach[J]. Journal of Mechanical Design, 2002, 124(4): 633-641.

[19] OGATA K. Discrete-time control systems[M]. Englewood Cliffs: Prentice-Hall, 1994.

[20] SHENOY P, MILLER K J, CRAWFORD B, et al. Online electromyographic control of a robotic prosthesis[J]. IEEE Transactions on Biomedical Engineering, 2008, 55(3): 1128-1135.

北京航空航天大学学报

北航学报 赠阅

Constraint violation suppression for dynamics modeling of lower limb rehabilitation robot

XU Yaru[1], LI Kehong[1], LIU Jia[2], LIU Rong[3], ZHANG Jiancheng[1,*]

(1. College of Robotics, Beijing Union University, Beijing 100020, China;
2. College of Mechanical and Electrical Engineering, Shijiazhuang University, Shijiazhuang 050035, China;
3. Robotics Institute, Beihang University, Beijing 100083, China)

Abstract: The U-K theory provides a new concept for obtaining the explicit dynamic equation of constraint multibody system. However, one consequence of the numerical approximation and truncation error is the constraint violation of the dynamic equation at the position and velocity level. Baumgarte's constraint violation stability methods (BSM) provide a stable dynamic equation by constraint modification. Nevertheless, the selection of Baumgarte parameters usually involve a trial-and-error process, which may result in the failure of simulation results. Consequently, the Baumgarte parameters selection problem is studied by using the classical fourth-order Runge-Kutta method, and the explicit dynamic equation of robot system based on the modified U-K theory by BSM is established. Furthermore, the lower limb rehabilitation robot is taken as the research object for simulation analysis. The results show that the constraint violation can be effectively suppressed. The joint angle errors are controlled within the range of $-5\times10^{-3}(°)$ $-5\times10^{-3}(°)$, the joint angular velocity errors are controlled within the range of -2×10^{-4} rad/s -2×10^{-4} rad/s, and the operation trajectory of the robot end-effector can be well close to the predetermined target of the system.

Keywords: U-K theory; constraint violation; Baumgarte's constraint violation stability methods (BSM); Baumgarte parameters; lower limb rehabilitation robot

北京航空航天大学学报

Received: 2020-11-11; **Accepted**: 2021-02-05; **Published online**: 2021-02-24 12:06

URL: kns.cnki.net/kcms/detail/11.2625.V.20210224.1111.001.html

Foundation items: National Key R & D Program of China (2018YFB1307001); Beijing Municipal Science and Technology Project (KM202111417006); Science and Technology Project of Hebei Education Department (QN2020510); Academic Research Projects of Beijing Union University (ZK80202004)

* **Corresponding author**. E-mail: robotbuu@126.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0644

纤维素类生物质水热法制备航油前驱物能耗分析

全下文载

董升飞，黄星华，杨晓奕*
（北京航空航天大学 能源与动力工程学院 能源与环境国际中心，北京 100083）

摘 要：生物质制备航空替代燃料对于全球碳减排和控制温室气体排放发挥十分重要的影响。纤维素类生物质来源广、年产量大成为其作为生物质原料的显著优势。结合目前纤维素类生物质研究的最新成果，深入研究了纤维素类生物质制备航油前驱物（糠醛（FF）、5-羟甲基糠醛（5-HMF）、乙酰丙酸（LA））关键工艺单元的工艺参数和产率，通过 Aspen Plus 工艺模拟，研究比较了糠醛与乙酰丙酸制备工艺、糠醛与 5-羟甲基糠醛制备工艺过程的物质流和能流，得出了不同工艺参数对产率的影响，并进行了能耗分析，为提高平台化合物的产率和降低能耗提供了理论基础。

关 键 词：纤维素类生物质；航空替代燃料；水解；能耗分析；平台化合物
中图分类号：V312.3
文献标志码：A **文章编号**：1001-5965(2022)04-0620-12

世界上每年产生大量的木质纤维资源，如木材加工废弃物、玉米芯、玉米秸秆、小麦秸秆、棉籽壳、稻壳和甘蔗渣等，它们含有丰富的木质纤维，这些宝贵的资源一直没有得到高效利用及合适处理，造成了严重的环境污染及资源浪费。纤维素类生物质作为重要的生物质资源，年产量约 $1\ 700 \times 10^8$ t，主要由纤维素、半纤维素和木质素组成[1]。因此，纤维素类生物质资源化利用一直是生物质能源的研究热点之一。2006 年，Chheda 和 Dumesic[2] 提出了生物质衍生的碳水化合物生产液体烷烃的完整工艺流程。纤维素类生物质制备高附加值的平台化合物更是得到了快速发展。

不同生物质的成分含量有差异[3]。纤维素作为生物质中的主要成分，同时也是植物细胞壁的主要组成，性质稳定，通常与半纤维素、木质素结合在一起。纤维素常温下不溶于水及一般有机溶剂，在一定的酸催化作用下可水解成单糖，进一步脱水形成 5-羟甲基糠醛（5-HMF）和乙酰丙酸（LA）[4]。

半纤维素结构既有六碳糖基，也有五碳糖基，不如纤维素性质稳定，具有亲水性，容易在酸催化和碱催化条件下发生水解，水解产物主要为戊糖（阿拉伯糖、木糖）、己糖（半乳糖、葡萄糖、甘露糖）及乙酰基等成分[5]，其中木糖可转化成重要平台化合物糠醛（FF）。由于半纤维素较纤维素更容易水解，故一般半纤维素先发生水解生成单糖，之后脱水形成糠醛。

木质素为填充在植物细胞壁网状骨架之间的硬面物质，结构中含有氧代苯丙醇或其衍生物结构单元的芳香性高聚物，其对于液相分解的讨论较少。木质素分子中致密的网状芳环结构及复杂化学键合方式，使得其较碳水化合物更难转化，高碳含量使得其易在酸性催化剂表面积碳，导致催化剂失活，且当前多数解聚产物易重聚[6]。木质素由不同苯丙结构单元随机键组合而成，分子具有甲氧基、羟基和羰基等多种功能基团，以及不饱

收稿日期：2020-11-18；**录用日期**：2021-04-18；**网络出版时间**：2021-05-11 14:48
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210511.1123.002.html
基金项目：国家“863”计划（2018YFB1501505）
* **通信作者**. E-mail：yangxiaoyi@buaa.edu.cn

引用格式：董升飞，黄星华，杨晓奕. 纤维素类生物质水热法制备航油前驱物能耗分析[J]. 北京航空航天大学学报，2022，48(4)：620-631. DONG S F，HUANG X H，YANG X Y. Energy consumption for production of jet fuel precursors from cellulosic biomass by hydrothermal method[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：620-631 (in Chinese).

北京航空航天大学学报

北航学报 赠阅

和双键等活性位点[7],主要用于制备酚类单体等平台化合物。

目前,对纤维素类生物质制备平台化合物研究主要集中在定向调控和机理研究,缺乏对整体工艺的优化及能耗分析。本文系统分析了纤维素类生物质制备平台化合物的工艺过程和原理,通过 Aspen Plus 对系统进行能耗分析,确定了主要能耗单元,为实际工艺优化提供理论基础。

1　研究方法

收集国内外关于纤维素类生物质制备平台化合物的研究成果,根据产率高、能耗低、污染低指标参数选择糠醛、5-羟甲基糠醛(5-HMF)、乙酰丙酸(LA)制备方法,并组合设计可行的生产平台化合物的工艺,通过 Aspen Plus 模拟进行工艺流程的能耗分析。

纤维素类生物质制备航油前驱物的化工过程包含大量的设备和操作单元,通过化工过程模拟量化过程中设备的操作特性、设备参数及能量需求,可以判断工艺的可行性。Aspen Plus V11 提供了大量的组分物性参数,可以有效模拟出各工艺模型的物流结果和各单元的热量消耗结果。

1.1　物性方法选择

根据纤维素类生物质制备中间体各过程的反应条件,通常在水相或水相-有机相极性体系下进行,各组分存在交互作用参数。Wilson 模型属于活度系数模型的一种,适用于许多类型的非理想溶液,可以模拟极性和非极性化合物的混合物,应用广泛。因此,本文模拟计算的物性方法选用 Wilson。对于有聚合物参与反应的单元,模块物性方法选择 Polynrtl。

1.2　操作单元选择和参数设定

各个单元需要不同的温度条件,除了需要将体系加热到所需要的温度,还应考虑到反应热的影响,因此对有些单元需要使用 Aspen Plus 内置反应单元模块 RStoic,可规定此模块的反应条件、反应方程式和转化率。对于无化学反应的单元,可选用分离器 Sep 或换热器 HeatX 进行模拟。在模拟过程中,结合现有技术的成熟程度设计热回收系统,使工艺的总能耗分析更接近实际。各个单元的参数严格按照目前已有文献的条件进行设定。

为了在同一平台上进行对比,工艺流程中平台化合物出口均设置了纯度要求。

2　纤维素类生物质制备平台化合物工艺原理

2.1　生物质原材料组分特点

水稻秸杆、玉来秸杆、小麦秸杆、硬木、软木等纤维素类生物质原料均可以作为生物质液相分解的原料,其价格也相对较为便宜。

纤维素类生物质中具有复杂的纤维素-半纤维素-木质素结构,其通过热化学转化和生物化学转化技术可以转化为可燃气、生物油、乙醇等燃料和高附加值的化学品。典型的生物质中各种成分的含量如表 1 所示[8]。

表 1　生物质成分含量[8]

Table 1　Contents of structural components of biomass[8]

原料	成分含量/%		
	纤维素	半纤维素	木质素
大麦秸杆	48.6	29.7	21.7
亚麻杆	36.7	34.4	28.9
大豆秸杆	29.2	35.5	35.3
燕麦杆	44.8	33.4	21.8
油菜秸杆	54.8	23.2	22.0
油菜杆	59.4	19.2	21.4
油菜籽壳	55.9	20.0	24.1
水稻秸杆	52.3	32.8	14.9
黑麦秸杆	49.9	29.6	20.5
小黑麦秸杆	47.9	31.5	20.6
小麦秸杆	44.5	33.2	22.3
玉米秸杆	49.0	37.9	13.1
棉花秸杆	66.2	18.4	15.4
向日葵秸杆	71.6	12.3	16.1
烟草秸杆	44.6	30.2	25.2

注:干基、无灰,标准化为 100%。

2.2　工艺过程及原理

纤维素和半纤维素转化平台化合物(糠醛、5-HMF、LA)过程中可能发生的反应及副反应如图 1和图 2 所示[5,9]。

半纤维素又水解可以生成戊糖(主要为木糖)和少量己糖,半纤维素容易发生水解反应,且转化率很高,可以达到 90% 以上。戊糖可以脱水生成糠醛,糠醛自身容易降解,同时糠醛还会与木糖或自身发生聚合反应,产生胡敏素等聚合物,从而影响糠醛产率,为了减少这些聚合物的形成,一般快速提取出生成的糠醛。通常采用的工艺方法有水蒸气提取[10]和两相法的反应体系[11]。

纤维素可水解为葡萄糖(己糖),葡萄糖先异构化为果糖,果糖进一步脱水可转换为 5-HMF,5-HMF 与水可生成 LA 和甲酸。一部分葡萄糖会与

北京航空航天大学学报

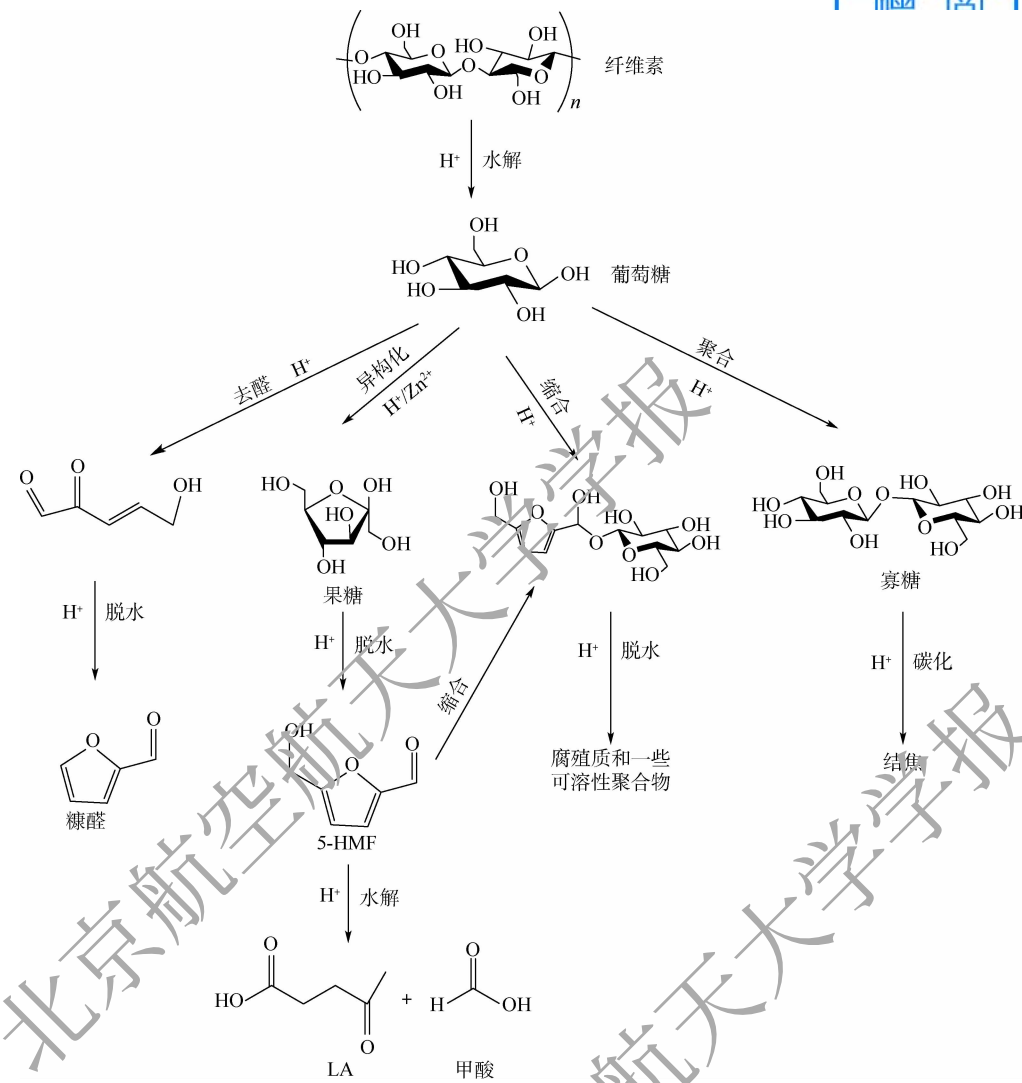


图 1 纤维素转化平台化合物可能发生的反应

Fig. 1 Possible reactions of cellulose converted to platform compounds

生成的 5-HMF 发生聚合生成聚合物,进一步转化为胡敏素及一些可溶性物质。还有一些葡萄糖也可以进行羟基间脱水再聚合生成具有高聚合度的多糖,聚合度较高时,会变成沉淀从溶液中分离出来,过程中会产生少量的糠醛。

2.3 产率汇总

针对半纤维素和纤维素的提取,目前不同方法的研究结果如表 2[12-18]所示。

纤维素、半纤维素及各种糖类转化为平台化合物的产率和转化率与反应条件和催化剂的选取息息相关,不同团队通过大量实验研究给出了不同的研究成果。

木聚糖水解为木糖:

$$(C_5H_8O_4)_n + nH_2O \rightarrow nC_5H_{10}O_5$$

半纤维素水解的产率如表 3[19-22]所示。

木糖制备糠醛:

$$C_5H_{10}O_5 \rightarrow C_5H_4O_2 + 3H_2O$$

木糖转化糠醛的产率如表 4[23-31]所示。

纤维素水解为葡萄糖:

$$(C_6H_{10}O_5)_n + nH_2O \rightarrow nC_6H_{12}O_6$$

纤维素水解的产率如表 5[19,32-33]所示。

葡萄糖转化 5-HMF 化学式:

$$C_6H_{12}O_6 \rightarrow C_6H_6O_3 + 3H_2O$$

葡萄糖转化 5-HMF 的产率如表 6[34-38]所示。

葡萄糖转化 LA 化学式:

$$C_6H_{12}O_6 \rightarrow C_5H_8O_3 + CH_2O_2 + H_2O$$

葡萄糖转化 LA 的产率如表 7[26,39-42]所示。

阿拉伯糖脱水生成糠醛的产率比木糖脱水生成糠醛产率低,产率如表 8[43-46]所示。半纤维素水解产物中仅产生很少(约 2%)的阿拉伯糖[19],故所选反应条件以木糖为主。

通过对流程中各部分进行分类归纳,可根据这些实验条件和数据设计出几种纤维素类生物质制备平台化合物产率较高、条件相对合理的工艺流程。根据所设计出的几种工艺进行物质流和能耗分析,通过对比得出不同工艺的优缺点。

北京航空航天大学学报

北航学报 赠阅

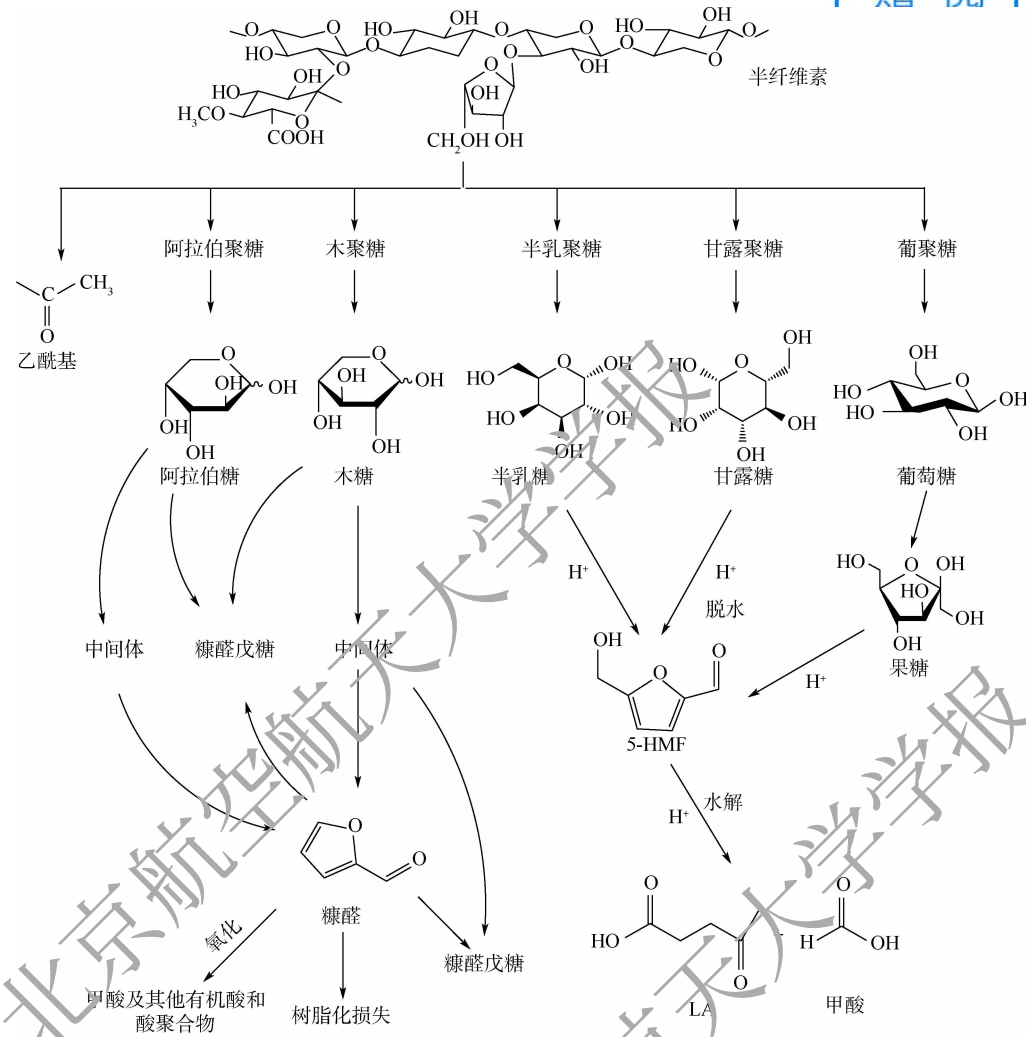


图 2　半纤维素转化平台化合物可能发生的反应

Fig. 2　Possible reactions of hemicellulose converted to platform compounds

表 2　纤维素、半纤维素的产率

Table 2　Yield of cellulose and hemicellulose

提取物	方法	产率/%	参考文献
半纤维素	60% 碱性乙醇,60℃,3 h	80.1	[12]
	甲酸/乙酸/水(3:6:1),0.1% HCl,85℃,4 h	76.5	[13]
	水,170℃,60 min	65	[14]
	水/GVL(1:4),120℃,75 mmol/L H_2SO_4	96	[15]
纤维素	甲酸/乙酸/水(3:6:1),90℃,4 h 5% H_2O_2,pH = 12,60℃,3 h	88.9	[16]
	漆酶系统,pH = 4.5,55℃,72 h	90.0	[17]
	醋酸-$NaClO_2$,9% NaOH,60℃,120 min	82.3	[18]

表 3　半纤维素水解为木糖的产率

Table 3　Yield of hemicellulose hydrolyzed to xylose

液态环境	催化剂	反应条件	质量产率/(g·g^{-1})	参考文献
3% H_2SO_4 固/液 = 1:10		120℃,1.5 h	77 ~ 94.2	[20]
0.1 mol/L $FeCl_3$	$FeCl_3$	140℃,20 min	87.7	[21]
HCl 质量分数 1%	20 CBU/g 纤维二糖酶	120℃,40 min	97	[19]
乳酸质量分数 1.2%		180℃,50 min	57.8	[22]

北京航空航天大学学报

北航学报 赠阅

表 4 木糖转化糠醛的产率

Table 4 Yield of xylose converted to furfural

液态环境	催化剂	反应条件	产率/%	质量产率/(g·g^{-1})	参考文献
水/GVL(1:9)	PSTA-POM	170℃,10 min	80.4	0.515	[23]
水/GVL(1:19)	Bronsted acid	170℃,30 min	78.5	0.502	[24]
[BMIM]Cl	$AlCl_3$	160℃,1.5 min	82.2	0.526	[25]
水/GVL(1:4)	SAPO-18 (B/L=0.11)	205℃,40 min	95.1	0.609	[26]
MIBK,pH=1.25	H^+	170℃,100 min	65.0	0.416	[27]
四氢呋喃(THF)	HCl	160℃,60 min	87.0	0.557	[28]
水/甲苯(1:2)	硅铝磷酸盐(SAPO-44)	170℃,6 h	93.0	0.595	[29]
水/THF=(1:2)	0.48 mol/L HCl	164℃	90.0	0.576	[30]
水/DCM(水/二氯甲烷)(1:2)	固体酸催化剂 SCC	170℃,60 min	81.4	0.521	[31]

表 5 纤维素水解为葡萄糖的质量产率

Table 5 Yield of cellulose hydrolyzed to glucose

液态环境	催化剂	反应条件	质量产率/(g·g^{-1})	参考文献
	CO:2 h;CU:2 h	120℃,4 h	0.898	[32]
0.1 mol/L $FeCl_3$(固/液=1:10)	纤维素酶,β-葡萄糖苷酶	170℃,30 min	0.819	[33]
pH=12.4,石灰:0.1 g/g	5FPU/g 纤维素酶	60℃,12 h	0.780	[19]

表 6 葡萄糖转化 5-HMF 的产率

Table 6 Yield of glucose converted to 5-HMF

液态环境	催化剂	反应条件	产率/%	质量产率/(g·g^{-1})	参考文献
水-diglyme(1:3)	ZrP	180℃,2 h	61.0	0.427	[34]
水-diglyme(1:3)	ZrP	180℃,3 h	63.0	0.441	[34]
Lewis + Bronsted	m-ZrP	155℃,6 h	46.1	0.323	[35]
[EMIM]Cl	$CrCl_3$	100℃,3 h	68.0	0.476	[36]
[EMIM]Br	Al_2O_3-b-0.05	140℃,3 h	49.7	0.348	[37]
DMSO	SnPCP@ MnO_2-PDA	150℃,5 h	55.8	0.391	[38]

表 7 葡萄糖转化 LA 的产率

Table 7 Yield of glucose converted to levulinic acid

液态环境	催化剂	反应条件	产率/%	质量产率/(g·g^{-1})	参考文献
水/GVL(1:4)	0.3 mol/L H_2SO_4	160℃,90 min	64.1	0.413	[39]
40 g/L 葡萄糖	磷钨酸银盐 0.7 g/25 mL	200℃,2 h	81.6	0.526	[40]
水/GVL(1:4)	SAPO-18 (B/L=0.17)	180℃,80 min	70.2	0.452	[26]
秸秆/水(1:15)	$S_2O_8^{2-}$/ZrO_2-SiO_2-Sm_2O_3:13.3%	200℃,10 min	70.0	0.451	[41]
玉米芯浓度 25 g/L	$SnCl_4$ 80 mmol/L	193℃,10 min	78.1	0.503	[42]

表 8 阿拉伯糖转化糠醛的产率

Table 8 Yield of arabinose converted to furfural

液态环境	催化剂	反应条件	产率/%	质量产率/(g·g^{-1})	参考文献
丁酮/水(4:1)	Lewis 酸性离子液	140℃,30 min	60.0	0.384	[43]
甲酸(质量分数 98%)		170℃,60 min	62.0	0.397	[44]
	草酸	200℃,50 min	53.0	0.339	[45]
微波辐射的水-环戊基甲基醚	Nafion NR50、NaCl	170℃,40 min	42.0	0.269	[46]

北京航空航天大学学报

北航学报 网阅

3　纤维素类生物质制备平台化合物工艺

根据 Zu 等[19]研究结果,玉米秸秆中各组分的质量含量为:葡聚糖 39.6%,木聚糖 19.5%,阿拉伯糖 2.0%,半乳糖体 1.2%,不溶酸木质素 19.1%,灰分 3.1%,其他 15.5%。由秸秆中的半纤维素和纤维素转化木糖和葡萄糖的产率分别为 97.0% 和 78.0%。本文模拟均以此组分含量数据为准。

针对纤维素类生物质转化成平台化合物(糠醛、LA、5-HMF)的方法,可根据不同的步骤及体系进行分类。半纤维素转化糠醛可以根据水解反应和脱水反应是否在同一个反应器中进行分为一步法和两步法。纤维素或半纤维素的催化转化可以根据反应体系分为单相体系和两相体系[11]。

目前,工业上使用较多的是一步法与单相体系,因为其结构简单,反应设备投资少,但缺点也十分明显,由于糠醛的不稳定性,容易在水相中与木糖或自身发生副反应,生成胡敏素等副产物,从而影响糠醛的产率。其中,中国科学院广州能源研究所[10,47]利用水蒸气及时提取出溶液中生成的糠醛,可以减小副反应的产生,得到选择制备航空燃料所需要的糠醛水溶液,是一个可行的解决方法。

目前,两步法和两相体系被广泛研究。两步法是指半纤维素低温水解为戊糖、戊糖高温脱水为糠醛在不同的反应器中进行。由于低温水解不破坏纤维素和木质素结构,可以在半纤维素低温预处理之后提取出固体残渣(纤维素和木质素),以产生其他的航油前驱物。两相体系则是通过加入不溶于水的有机溶剂,迅速将水相中生成的糠醛提取进入有机相中,以减少催化反应中的副反应,提高糠醛产率。

本文利用具有发展前景的两步法,并根据大量实验研究的结果,设计出不同的工艺流程制备航油前驱物糠醛、LA、5-HMF。使用 Aspen Plus 模拟软件对各流程进行模拟,以得到秸秆制备生物航油前驱物的几种工艺过程的能耗分析。

3.1　制备糠醛和 LA 工艺流程

3.1.1　蒸汽提取法

中国科学院广州能源研究所使用酸催化玉米秸秆,水蒸气提取糠醛,所使用的蒸汽为低压蒸汽(压力≤2.5 MPa,温度≤400 ℃)。进一步对固体残渣酸催化,得到 LA 溶液。蒸汽提取法流程如图 3 所示。

中国科学院广州能源研究所对该方法的能量消耗做了相应计算[10,47]。本文利用 Aspen Plus 进行模拟,并将糠醛和 LA 稀溶液进行浓缩,得到高纯度溶液的能耗值。蒸汽提取法的模拟流程如图 4 所示,各反应单元热量消耗结果如表 9 所示。

该方法的糠醛产率为 0.384 8 g/g 半纤维素,LA 产率为 0.259 0 g/g 纤维素,假设此系统 80% 的余热能够被利用,则热量消耗为 13 533.85 kJ/kg 玉米秸秆。

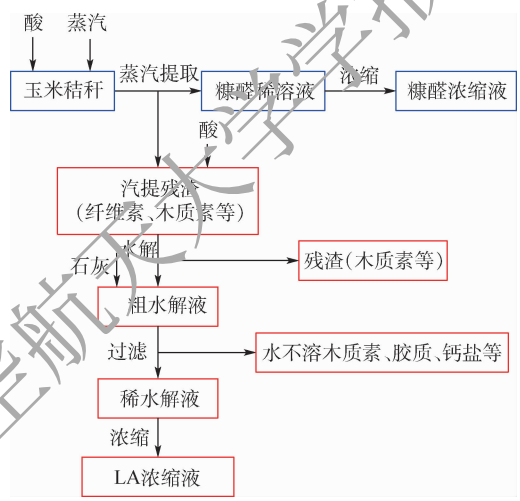


图 3　蒸汽提取法流程

Fig. 3　Flowchart of steam extraction method

3.1.2　预处理-催化转化法

先将玉米秸秆通过稀 HCl 预处理水解半纤维素得到木糖稀溶液,对糖溶液浓缩后经两相法(水/二氯甲烷体系)催化转化可得到糠醛溶液。通常糖类的浓缩有 2 种方法:①对葡萄糖水溶液进行加热,将水蒸馏出以提高糖溶液浓度;②进行

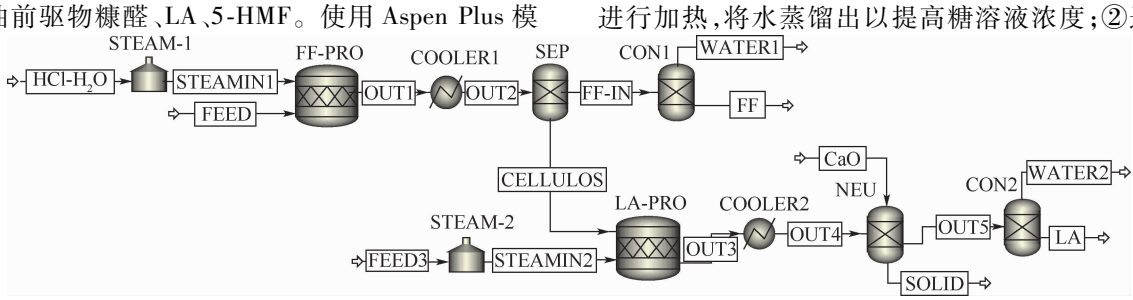


图 4　蒸汽提取法的模拟流程

Fig. 4　Simulation process of steam extraction method

北航学报 瞰阅

表 9 蒸汽提取法热量消耗

Table 9 Heat consumption by steam extraction method

单元名称	单元说明	使用模块	热量消耗/(kJ·kg^{-1})
STEAM-1	蒸汽生产单元 1	FURNACE	16 796.20
FF-PRO	糠醛生产单元	RStoic	−1 009.50
CON1	糠醛浓缩单元	Sep	15 490.80
STEAM-2	蒸汽生产单元 2	FURNACE	19 314.20
LA-PRO	LA 生产单元	RStoic	−1 237.87
NEU	CaO 中和单元	Sep	984.00
CON2	LA 浓缩单元	Sep	17 331.40

注:热量消耗为在此工艺条件下每千克玉米秸秆所消耗的热量。

膜浓缩,通过反渗透法可以将糖浓度浓缩 10 倍以上。膜分离技术操作条件温和,可以降低蒸发方法在过程中损失的热量。故本文工艺使用反渗透法对木糖溶液进行浓缩[48]。纤维素通过固体酸催化转化为 LA。预处理-催化转化法流程如图 5 所示。

使用 Aspen Plus 对此流程进行模拟,如图 6 所示。对各反应单元热流体的热量进行回热处理,充分利用热流中的余热,可以有效降低热量消耗。各反应单元热量消耗结果如表 10 所示。

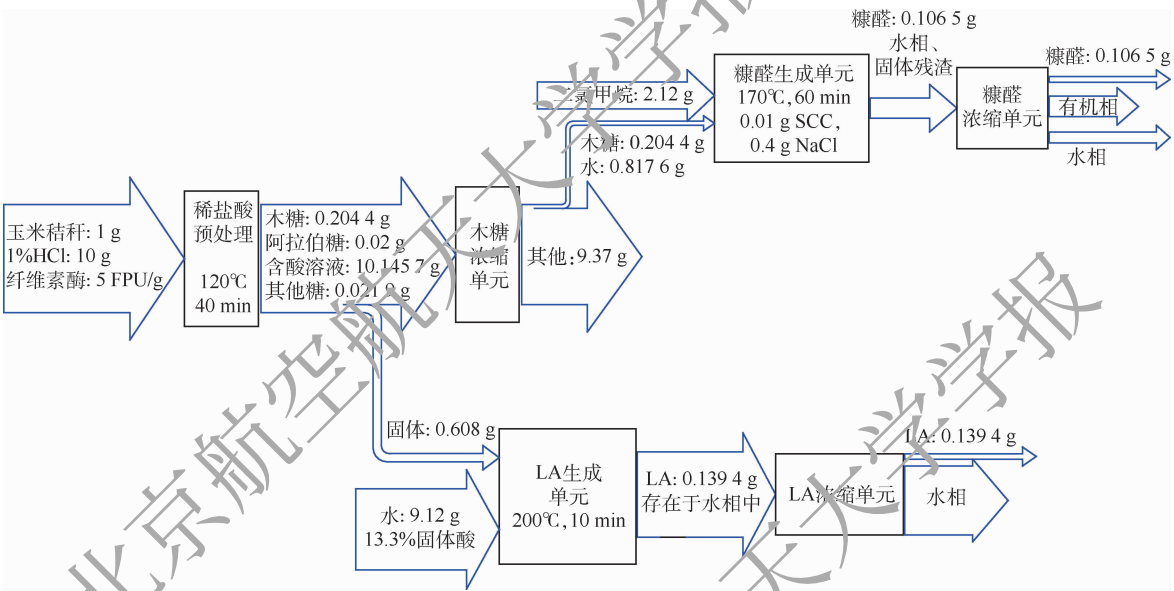


图 5 预处理-催化转化法流程

Fig. 5 Flowchart of pretreatment-catalytic conversion method

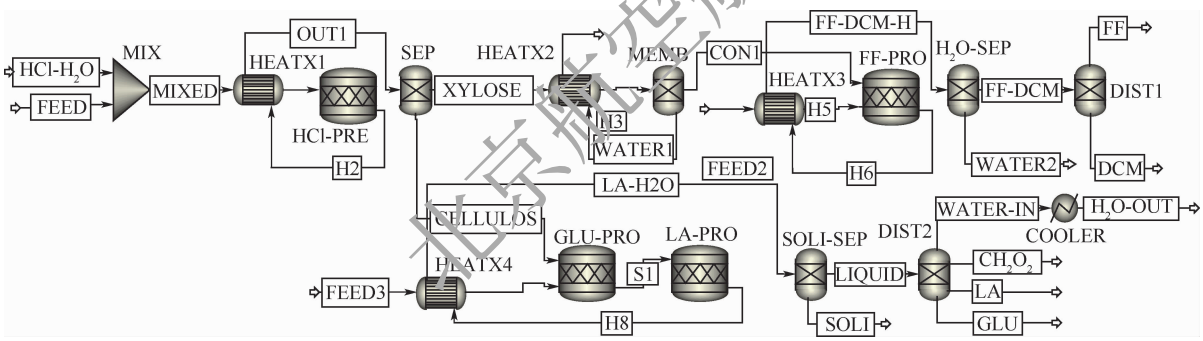


图 6 预处理-催化转化法的模拟流程

Fig. 6 Simulation process of pretreatment-catalytic conversion method

表 10 预处理-催化转化法热量消耗

Table 10 Heat consumption by pretreatment-catalytic conversion method

单元名称	单元说明	使用模块	热量消耗/(kJ·kg^{-1})
HCL-PRE	稀盐酸预处理	RStoic	1 473.57
MEMB	木糖浓缩单元	Sep	91.89
FF-PRO	糠醛生成单元	RStoic	695.69
DIST1	糠醛浓缩单元	Sep	701.72
GLU-PRO&LA-PRO	LA 生产单元	RStoic	1 463.75
DIST2	LA 浓缩单元	Sep	4 402.86

注:热量消耗为在此工艺条件下每千克玉米秸秆所消耗的热量。

该方法的糠醛产率为 0.546 2 g/g 半纤维素,LA 产率为 0.352 0 g/g 纤维素,热量消耗为 8 829.48 kJ/kg 玉米秸秆。

3.1.3 两级预处理转化方法

根据 Li 等[39]的研究结果,若对秸秆进行 2 次处理,再蒸馏提取,则分别经过高温条件(180℃,90 min)和低温条件(140℃,90 min),糠醛和 LA 的摩尔产率分别为 70.65% 和 57.7%。可以通过超临界 CO_2 将水-GVL 分成水相和有机相,糠醛

北京航空航天大学学报

和 LA 存在于有机相中，再通过减压蒸馏法将糠醛、GVL 和 LA 分离。两级预处理转化方法流程如图 7 所示。

该流程是一个循环过程，根据文献[39]的数据，在 Aspen Plus 中对其中一个循环过程进行模拟，对各反应单元热流体的热量进行回热处理，充分利用热流中的余热，模拟流程如图 8 所示，所得的结果如表 11 所示。

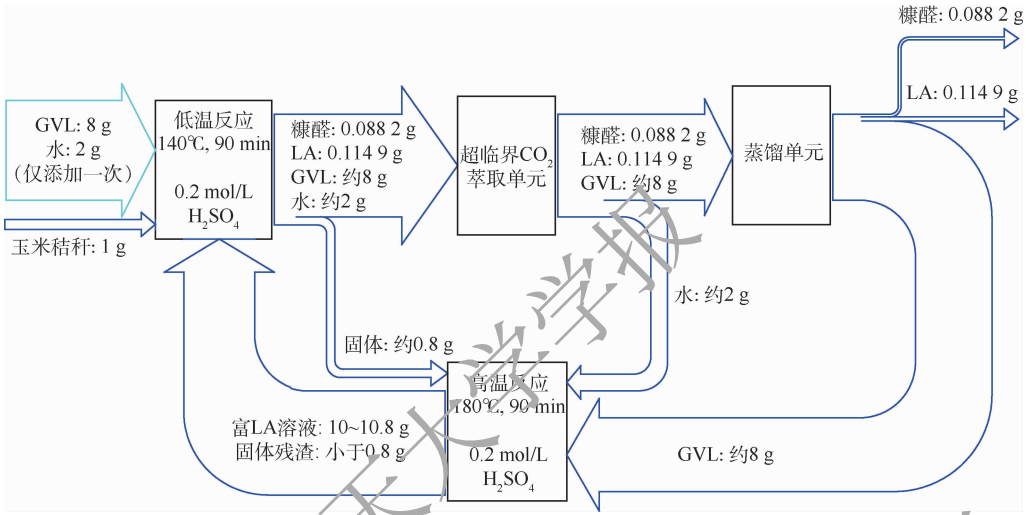


图 7　两级预处理转化方法流程

Fig. 7　Flowchart of two-stage pretreatment conversion method

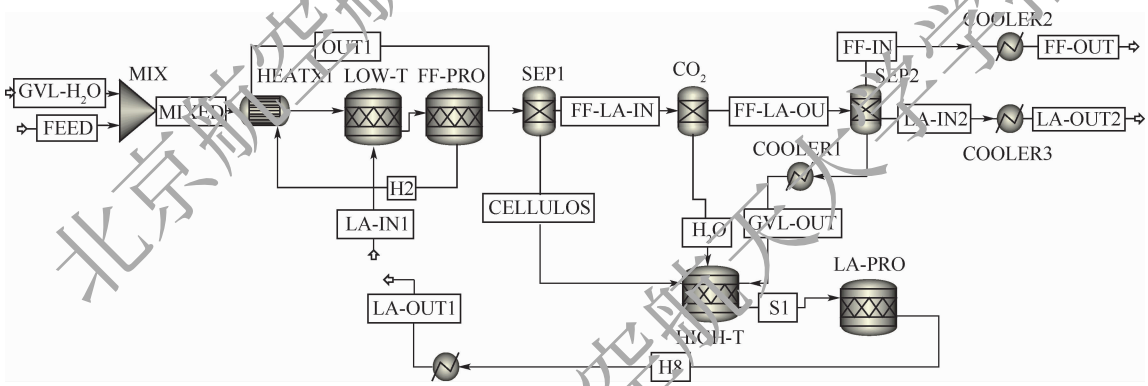


图 8　两级预处理转化方法的模拟流程

Fig. 8　Simulation process of two-stage pretreatment conversion method

表 11　两级预处理转化方法热量消耗

Table 11　Heat consumption by two-stage pretreatment conversion method

单元名称	单元说明	使用模块	热量消耗/(kJ·kg^{-1})
LOW-T&FF-PRO	低温反应单元	RStoic	639.13
HIGH-T&LA-PRO	高温反应单元	RStoic	688.32
CO_2	超临界 CO_2 萃取	Sep	0
SEP2	蒸馏提取单元	Sep	1 255.86

注：热量消耗为在此工艺条件下每千克玉米秸秆所消耗的热量。

该方法的糠醛产率为 0.438 6 g/g 半纤维素，LA 产率为 0.29 g/g 纤维素，热量消耗为 2 583.31 kJ/kg玉米秸秆。

3.2　制备糠醛和 5-HMF 工艺流程

由于 5-HMF 的产品价格更加昂贵，且 5-HMF 可以通过自缩合得到长碳链的含氧中间体，也可用于生产航空替代燃料，故本文设计一套生产糠醛和 5-HMF 的工艺路线。玉米秸秆先预处理后催化转化成为糠醛和 5-HMF，由于金属离子液体价格昂贵且有剧毒性，本文流程使用价格便宜且易于回收的有机溶剂与固体催化剂催化葡萄糖制备 5-HMF。纤维素类生物质催化转化糠醛和 5-HMF 如图 9 所示。

使用 Aspen Plus 对该流程进行模拟，模拟流程如图 10 所示。各反应单元热量消耗的结果如表 12 所示。

该方法的糠醛产率为 0.546 2 g/g 半纤维素，5-HMF 产率为 0.333 1 g/g 纤维素，热量消耗为 5 212.87 kJ/kg 玉米秸秆。

通过上述模拟，得到了 4 种不同工艺制备航油前驱物的产率、物质流及热量消耗。从模拟结果中可以得出各种工艺的目标产物得率及其主要能耗单元。

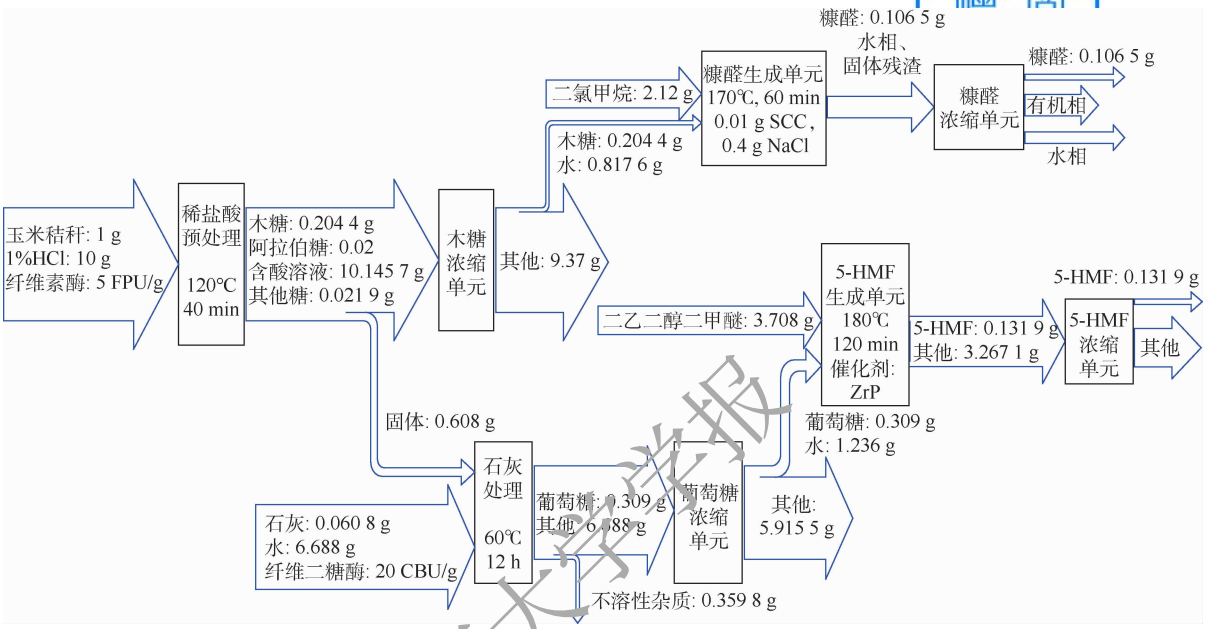


图 9 预处理后催化转化糠醛和 5-HMF

Fig. 9 Catalytic conversion of furfural and 5-HMF after pretreatment

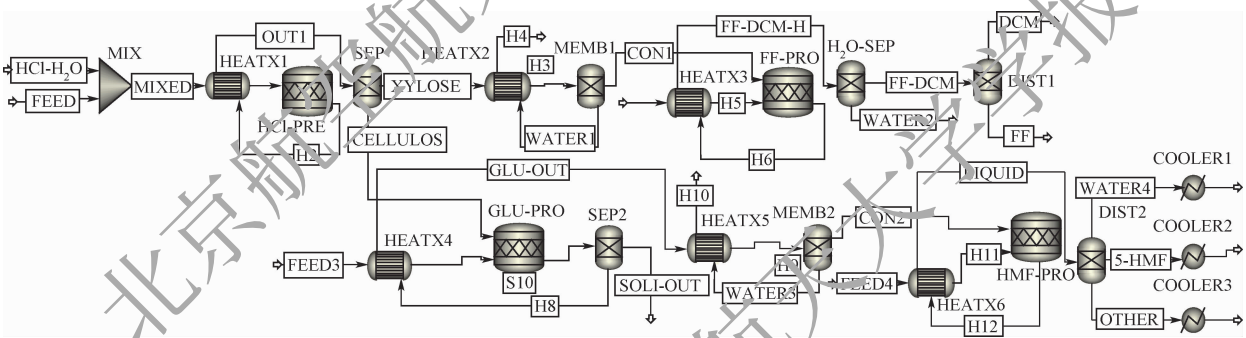


图 10 预处理后制备糠醛和 5-HMF 的模拟流程

Fig. 10 Simulation process for producing furfural and 5-HMF after pretreatment

表 12 预处理后制备糠醛和 5-HMF 的热消耗

Table 12 Heat consumption for production of furfural and 5-HMF after pretreatment

单元名称	单元说明	使用模块	热量消耗/(kJ·kg^{-1})
HCl-PRE	稀盐酸预处理	RStoic	1 463.56
MEMB1	木糖浓缩单元	Sep	91.88
FF-PRO	糠醛生成单元	RStoic	695.69
DIST1	糠醛浓缩单元	Sep	701.72
GLU-PRO	石灰处理单元	RStoic	178.74
MEMB2	葡萄糖浓缩单元	Sep	105.26
HMF-PRO	5-HMF 生成单元	RStoic	1 075.67
DIST2	5-HMF 浓缩单元	Sep	900.25

注:热量消耗为在此工艺条件下每千克玉米秸秆所消耗的热量。

4 结 论

通过分析与模拟秸秆制备生物航油前驱物的产率及物质流和能流,得到了纤维素类生物质制备平台化合物不同工艺的能耗和产率。

1)蒸汽提取法每生产 1 g 糠醛需要半纤维素含量为 2.598 7 g 的玉米秸秆,需要消耗热量 16 256.50 J;每生产 1 g 乙酰丙酸需要纤维素含量为 3.861 g 的玉米秸秆,需消耗热量 28 101.72 J。该工艺设备投资小,拥有低经济成本的优势。如果糠醛和 LA 是为了满足后续航空燃料的缩合与加氢工艺,则由于对溶液浓度无过高的要求,利用该方法可不考虑浓缩过程所产生的能量消耗。

2)运用预处理-催化转化法生产糠醛和 LA,每生产 1 g 糠醛需要半纤维素含量为 1.831 g 的玉米秸秆,需要消耗热量 5 424.51 J;每生产1 g乙酰丙酸需要纤维素含量为 2.841 g 的玉米秸秆,需消耗热量 16 666.51 J。产率高,能耗低。但对设备要求高,投资大,且需要使用大量有机溶剂,短期来看经济性没有优势,但具有应用前景。

3)两级预处理转化方法生产糠醛和 LA,每生产 1 g 糠醛需要半纤维素含量为 2.280 g 的玉米秸秆,需要消耗热量 4 320.54 J;每生产1 g乙酰

北航学报 赠阅

丙酸需要纤维素含量为 3.448 g 的玉米秸秆,需消耗热量 6 704.07 J。此方法具有低能耗和较高产率的优势,可以作为未来考虑的方法之一。

4)运用预处理-催化转化法来生产糠醛和 5-HMF,每生产 1 g 糠醛需要半纤维素含量为 1.831 g 的玉米秸秆,需要消耗热量 5 406.35 J;每生产1 g 5-HMF 需要纤维素含量为 3.002 g 的玉米秸秆,需消耗热量 6 784.51 J。5-HMF 的经济附加值很高,高纯度的 5-HMF 价格更加昂贵。因此运用该方法生产糠醛和 5-HMF 不仅为生产航油前驱物提供理论依据,同时也为生产高附加值的化工产品提供参考。

参考文献(References)

[1] LI C Z,ZHAO X C,WANG A Q,et al. Catalytic transformation of lignin for the production of chemicals and fuels[J]. Chemical Reviews,2015,115(21):11559-11624.

[2] CHHEDA J N,DUMESIC J A. An overview of dehydration,aldol-condensation and hydrogenation processes for the production of liquid alkanes from biomass-derived carbohydrates[J]. ChemInform,2007,38(39):59-70.

[3] 赵蒙蒙,姜曼,周祚万. 几种农作物秸秆的成分分析[J]. 材料导报,2011,25(16):122-125.
ZHAO M M,JIANG M,ZHOU Z W. The components analysis of several kinds of agricultural residues[J]. Materials Review,2011,25(16):122-125(in Chinese).

[4] SHI N,LIU Q Y,ZHANG Q,et al. High yield production of 5-hydroxymethylfurfural from cellulose by high concentration of sulfates in biphasic system[J]. Green Chemistry,2013,15(7):1967.

[5] 李丹,文飚,曹春昱. 植物原料中半纤维素预水解反应动力学及其机理研究进展[J]. 中国造纸,2019,38(2):61-66.
LI D,WEN B,CAO C Y. Research progress on hemicellulose pre-hydrolysis kinetics and mechanism of plant material[J]. China Pulp & Paper,2019,38(2):61-66(in Chinese).

[6] 龙金星,徐莹,王铁军,等. 木质素催化解聚与氢解[J]. 新能源进展,2014,2(2):83-88.
LONG J X,XU Y,WANG T J,et al. Catalytic depolymerization and hydrogenolysis of lignin[J]. Advances in New and Renewable Energy,2014,2(2):83-88(in Chinese).

[7] HEITNER C, DIMMEL D, SCHMIDT J. Lignin and lignans: Advances in chemistry[M]. Boca Raton:CRC Press,2010.

[8] VASSILEV S V,BAXTER D,ANDERSEN L K,et al. An overview of the organic and inorganic phase composition of biomass[J]. Fuel,2012,94:1-33.

[9] ENOMOTO K,HOSOYA T,MIYAFUJI H. High-yield production of 5-hydroxymethylfurfural from d-fructose,d-glucose,and cellulose by its in situ removal from the reaction system[J]. Cellulose,2018,25(4):2249-2257.

[10] 张琦,李宇萍,陈伦刚,等. 百吨/年规模生物质水相合成航油类烃过程的物质与能量转化[J]. 天津大学学报(自然科学与工程技术版),2017,50(1):13-18.
ZHANG Q,LI Y P,CHEN L G,et al. Material and energy conversion of integrated 100 t/a-scale bio-jet fuel-range hydrocarbon production system via aqueous conversion of biomass[J]. Journal of Tianjin University (Science and Technology),2017,50(1):13-18(in Chinese).

[11] 许文娟. 木质纤维素类生物质水解制取糠醛及乙酰丙酸的研究[D]. 上海:华东理工大学,2016.
XU W J. Study on the production of furfural and levulinic acid from lignocellulosic biomass[D]. Shanghai:East China University of Science and Technology,2016(in Chinese).

[12] PENG P,PENG F,BIAN J,et al. Isolation and structural characterization of hemicelluloses from the bamboo species Phyllostachys incarnata Wen[J]. Carbohydrate Polymers,2011,86(2):883-890.

[13] XU F,LIU C F,GENG Z C,et al. Characterisation of degraded organosolv hemicelluloses from wheat straw[J]. Polymer Degradation and Stability,2006,91(8):1880-1886.

[14] 袁梅婷,翟华敏,冯年捷,等. 麦草自水解过程中半纤维素和木质素的变化特性[J]. 纤维素科学与技术,2015,23(2):55-61.
YUAN M T,ZHAI H M,FENG N J,et al. Characterization of hemicelluloses and lignin change of wheat straw in autohydrolysis process[J]. Journal of Cellulose Science and Technology,2015,23(2):55-61(in Chinese).

[15] SHUAI L,QUESTELL-SANTIAGO Y M,LUTERBACHER J S. A mild biomass pretreatment using γ-valerolactone for concentrated sugar production[J]. Green Chemistry,2016,18(4):937-943.

[16] 赵绘婷,刘振,任秋鹤,等. 玉米秸秆纤维素高效分离工艺研究[J]. 河南科学,2019,37(8):1328-1333.
ZHAO H T,LIU Z,REN Q H,et al. Efficient separation technique of cellulose from corn straw[J]. Henan Science,2019,37(8):1328-1333(in Chinese).

[17] 张循海,宋贺明,贾宏葛,等. 漆酶系统提取玉米秸秆中的纤维素[J]. 齐齐哈尔大学学报(自然科学版),2020,36(5):43-44.
ZHANG X H,SONG H M,JIA H G,et al. The laccase extracting cellulose systems from corn straw[J]. Journal of Qiqihar University(Natural Science Edition),2020,36(5):43-44(in Chinese).

[18] 杨文玲,王妨茶. 玉米秸秆纤维素提取工艺优化[J]. 安徽农业科学,2019,47(1):198-201.
YANG W L,WANG F C. Optimization of cellulose extraction process of corn straw[J]. Journal of Anhui Agricultural Sciences,2019,47(1):198-201(in Chinese).

[19] ZU S,LI W Z,ZHANG M J,et al. Pretreatment of corn stover for sugar production using dilute hydrochloric acid followed by lime[J]. Bioresource Technology,2014,152:364-370.

[20] 张扬,王运红,邓立红,等. 稻秸半纤维素水解条件和水解液脱毒的研究[J]. 纤维素科学与技术,2005,13(2):38-44.
ZHANG Y,WANG Y H,DENG L H,et al. Study of hemicellulose hydrolysis of the rice straw and hydrolysate detoxification[J]. Journal of Cellulose Science and Technology,2005,13(2):38-44(in Chinese).

[21] LIU L,SUN J S,LI M,et al. Enhanced enzymatic hydrolysis and

北京航空航天大学学报

structural features of corn stover by $FeCl_3$ pretreatment[J]. Bioresource Technology,2009,100(23):5853-5858.

[22] 高文中,张佳,付存亭,等. 玉米秸秆的水热酸处理工艺技术研究[J]. 化工设计通讯,2019,45(8):80-81.
GAO W Z,ZHANG J,FU C T,et al. Study on hydrothermal acid treatment of corn stover[J]. Chemical Engineering Design Communications,2019,45(8):80-81(in Chinese).

[23] XU Z P,LI W Z,DU Z J,et al. Conversion of corn stalk into furfural using a novel heterogeneous strong acid catalyst in γ-valerolactone[J]. Bioresource Technology,2015,198:764-771.

[24] ZHANG T W,LI W Z,XU Z P,et al. Catalytic conversion of xylose and corn stalk into furfural over carbon solid acid catalyst in γ-valerolactone[J]. Bioresource Technology,2016,209:108-114.

[25] ZHANG L X,YU H B,WANG P,et al. Conversion of xylan,D-xylose and lignocellulosic biomass into furfural using $AlCl_3$ as catalyst in ionic liquid[J]. Bioresource Technology,2013,130:110-116.

[26] LI X Y,XU R,LIU Q L,et al. Valorization of corn stover into furfural and levulinic acid over SAPO-18 zeolites: Effect of Brønsted to Lewis acid sites ratios[J]. Industrial Crops and Products,2019,141:111759.

[27] KAUR I,NI Y H. A process to produce furfural and acetic acid from pre-hydrolysis liquor of kraft based dissolving pulp process[J]. Separation and Purification Technology,2015,146:121-126.

[28] XING R,SUBRAHMANYAM A V,OLCAY H,et al. Production of jet and diesel fuel range alkanes from waste hemicellulose-derived aqueous solutions[J]. Green Chemistry,2010,12(11):1933.

[29] BHAUMIK P,DHEPE P L. Exceptionally high yields of furfural from assorted raw biomass over solid acids[J]. RSC Advances,2014,4(50):26215.

[30] XING R,QI W,HUBER G W. Production of furfural and carboxylic acids from waste aqueous hemicellulose solutions from the pulp and paper and cellulosic ethanol industries[J]. Energy & Environmental Science,2011,4(6):2193.

[31] 邓奥杰. 玉米芯两步法高效制备糠醛的研究[D]. 广州:华南理工大学,2016.
DENG A J. Study on high-efficiency preparation of furfural from corncob by two-step method[D]. Guangzhou:South China University of Technology,2016(in Chinese).

[32] HOU X D,FENG G J,YE M,et al. Significantly enhanced enzymatic hydrolysis of rice straw via a high-performance two-stage deep eutectic solvents synergistic pretreatment[J]. Bioresource Technology,2017,238:139-146.

[33] CHEN L H,CHEN R,FU S Y. $FeCl_3$ pretreatment of three lignocellulosic biomass for ethanol production[J]. ACS Sustainable Chemistry & Engineering,2015,3(8):1794-1800.

[34] JAIN A,SHORE A M,JONNALAGADDA S C,et al. Conversion of fructose,glucose and sucrose to 5-hydroxymethyl-2-furfural over mesoporous zirconium phosphate catalyst[J]. Applied Catalysis A:General,2015,489:72-76.

[35] SARAVANAN K,PARK K S,JEON S,et al. Aqueous phase synthesis of 5-hydroxymethylfurfural from glucose over large pore mesoporous zirconium phosphates: Effect of calcination temperature[J]. ACS Omega,2018,3(1):808-820.

[36] ZHAO H B,HOLLADAY J E,BROWN H,et al. Metal chlorides in ionic liquid solvents convert sugars to 5-hydroxymethylfurfural[J]. Science,2007,316(5831):1597-1600.

[37] HOU Q D,ZHEN M N,LI W Z,et al. Efficient catalytic conversion of glucose into 5-hydroxymethylfurfural by aluminum oxide in ionic liquid[J]. Applied Catalysis B:Environmental,2019,253:1-10.

[38] LI K,DU M M,JI P J. Multifunctional tin-based heterogeneous catalyst for catalytic conversion of glucose to 5-hydroxymethylfurfural[J]. ACS Sustainable Chemistry & Engineering,2018,6(4):5636-5644.

[39] LI M H,LI W Z,LIU Q Y,et al. A two-step conversion of corn stover into furfural and levulinic acid in a water/gamma-valerolactone system[J]. BioResources,2016,11(4):8239-8256.

[40] 曾珊珊,林鹿,刘娣,等. 磷钨酸盐催化转化葡萄糖合成乙酰丙酸[J]. 化工学报,2012,63(12):3875-3881.
ZENG S S,LIN L,LIU D,et al. Catalytic conversion of glucose to levulinic acid by solid heteropolyacid salts[J]. CIESC Journal,2012,63(12):3875-3881(in Chinese).

[41] CHEN H,YU B,JIN S. Production of levulinic acid from steam exploded rice straw via solid superacid, $S_2O_8^{2-}/ZrO_2$-SiO_2-Sm_2O_3[J]. Bioresource Technology,2011,102(3):3568-3570.

[42] 卿青,郭琪,周琳琳,等. $SnCl_4$ 催化玉米芯高效制备乙酰丙酸的工艺研究[J]. 常州大学学报(自然科学版),2018,30(2):14-22.
QING Q,GUO Q,ZHOU L L,et al. Study on high efficiency catalytical preparation of levulinic acid from corncob by $SnCl_4$[J]. Journal of Changzhou University(Natural Science Edition),2018,30(2):14-22(in Chinese).

[43] ZHAO Y,XU H,LU K F,et al. Experimental and kinetic study of arabinose conversion to furfural in renewable butanone-water solvent mixture catalyzed by lewis acidic ionic liquid catalyst[J]. Industrial & Engineering Chemistry Research,2019,58(36):17088-17097.

[44] DUSSAN K,GIRISUTA B,LOPES M,et al. Conversion of hemicellulose sugars catalyzed by formic acid:Kinetics of the dehydration of D-xylose,l-arabinose,and D-glucose[J]. ChemSusChem,2015,8(8):1411-1428.

[45] HONGSIRI W,DANON B,DE JONG W. The effects of combined catalysis of oxalic acid and seawater on the kinetics of xylose and arabinose dehydration to furfural[J]. International Journal of Energy and Environmental Engineering,2015,6(1):21-30.

[46] LE GUENIC S,GERGELA D,CEBALLOS C,et al. Furfural production from d-xylose and xylan by using stable nafion NR50 and NaCl in a microwave-assisted biphasic reaction[J]. Molecules,2016,21(8):1102.

[47] LI Y P,CHEN L G,ZHANG X H,et al. Process and techno-economic analysis of bio-jet fuel-range hydrocarbon production from lignocellulosic biomass via aqueous phase deconstruction and catalytic conversion[J]. Energy Procedia,2017,105:675-680.

北航学报 赠阅

[48] 章茹,刘辉,冯斐. 低聚木糖的膜分离浓缩工艺研究[J]. 食品工业,2012,33(9):26-29.
ZHANG R,LIU H,FENG F. Study on the concentration of xylooligosaccharides by membrane separation technology[J]. The Food Industry,2012,33(9):26-29(in Chinese).

Energy consumption for production of jet fuel precursors from cellulosic biomass by hydrothermal method

DONG Shengfei, HUANG Xinghua, YANG Xiaoyi*

(Energy and Environment International Centre, School of Energy and Power Engineering, Beihang University, Beijing 100083, China)

Abstract: Biomass for production of alternative jet fuel has a very important impact on global carbon reduction and the control of greenhouse gas emissions. Wide sources and large annual output of cellulosic biomass have become its significant advantages as a biomass raw material to produce aviation alternative fuels. Based on the latest research results of cellulosic biomass, the process parameters and yield of the key process units for the production of jet fuel precursor (furfural (FF), 5-hydroxymethylfurfural (5-HMF) and levulinic acid (LA)) from cellulosic biomass were studied in depth in this paper. Through Aspen Plus process simulation, the material flow and energy flow of FF and LA, FF and 5-HMF were studied and compared. The influence of different process parameters on the yield was obtained, and energy consumption analysis of the advantageous process flow was carried out, which provides a theoretical basis for increasing the yield of platform compounds and reducing energy consumption of production of alternative jet fuel from biomass.

Keywords: cellulosic biomass; alternative jet fuel; hydrolysis; energy consumption analysis; platform compound

北京航空航天大学学报

Received: 2020-11-18; **Accepted**: 2021-04-18; **Published online**: 2021-05-11 14:48
URL: kns.cnki.net/kcms/detail/11.2625.V.20210511.1123.002.html
Foundation item: National High-tech Research and Development Program of China (2018YFB1501505)
* **Corresponding author**. E-mail: yangxiaoyi@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0647

基于 SINS/RFID 的隧道列车高精度定位方法

全下文载

张辰东[1,2], 王兆瑞[1,*], 金声震[1], 艾国祥[1]
(1. 中国科学院国家天文台, 北京 100101; 2. 中国科学院大学 天文与空间科学学院, 北京 100049)

摘 要: 为了获取高速铁路列车在隧道这种导航卫星不可见环境下的定位信息,提出一种基于捷联惯性导航系统(SINS)和射频识别技术(RFID)的组合定位方法。通过响应时间模型来计算标签的定位精度,依据实际轨道环境增加标签对列车姿态校准的能力,同时结合惯性导航系统解算得到连续的定位数据。仿真结果表明:在 30 km 长的隧道利用射频识别标签位置信息进行校准,可以很大程度地减小惯性导航系统的误差积累,提高定位精度。引入姿态信息后,可以在陀螺仪性能与标签间隔的多种组合中保持隧道全线定位精度在米级,最高能够达到 0.5 m。

关 键 词: 组合导航; 隧道定位; 射频识别技术(RFID); 捷联惯性导航系统(SINS); 动态姿态校准
中图分类号: TN967.2
文献标志码: A **文章编号**: 1001-5965(2022)04-0632-07

高速铁路(简称高铁)列车的定位技术是列车控制系统的基础,是保障铁路安全的核心。近年来铁路逐渐增大的运输压力,对列车在隧道这种事故高发区的控制管理提出了更高的要求,包括高精度的定位。实时高精度的定位结果不仅能提高铁路调度效率和运输能力,更能够保障列车人员和铁路工作人员的生命财产安全。

目前在卫星可视条件良好的室外情况下,利用全球导航卫星系统(global navigation satellite system, GNSS)可以得到米级甚至亚米级别精度的定位结果[1]。而单纯使用卫星定位技术来对高铁列车进行定位,由于定位结果较为依赖卫星的可见数目,在高铁站附近、城区、山区和隧道等卫星可见数目少甚至不可见的位置,定位结果往往与真实位置相差很大,而这些位置区域往往也是事故的高发地带,定位精度往往直接影响着列车控制、调度和相关人员的生命财产安全,因此在这些卫星信号较差甚至是卫星不可见的环境下,寻找一种适合于高铁列车的高动态、实时、高精度定位方法是非常必要的。

对于这种大范围卫星不可见环境内的高精度定位,许多国内外的学者们进行了相关研究[2-5]。文献[6]采用在隧道内布设漏泄同轴电缆的方式实现在大型环形隧道内 10 m 级精度的定位;文献[7]采用无线传感器网络,利用改进的接收信号强度加权定位算法在 2 000 m 隧道仿真环境内得到最高精度 3 m 的定位信息;文献[8]采用多种车内传感器进行误差补偿,结合混合定位策略和强跟踪滤波算法进行多点定位得到最高精度接近 2 m 的定位信息;文献[9]采用 GNSS 与惯性导航系统(inertial navigation system, INS)的组合系统,配合电子地图和射频识别技术(radio frequency identification, RFID)信息,通过数据后处理的方法在 GNSS 卫星不可见区域得到接近 1 m 精度的定位信息。而应用其他室内定位的技术手段,包括测量 Wi-Fi、Zigbee、RFID 和 Ultra 超宽带等信

收稿日期: 2020-11-20; **录用日期**: 2021-02-26; **网络出版时间**: 2021-03-25 18:41
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210324.1519.003.html
基金项目: 国家重点研发计划(2016YFB0501900); 国家自然科学基金(U1931125,11603041)
* **通信作者**. E-mail: zhaorui_w@nao.cas.cn

引用格式: 张辰东, 王兆瑞, 金声震, 等. 基于SINS/RFID的隧道列车高精度定位方法[J]. 北京航空航天大学学报, 2022, 48(4): 632-638. ZHANG C D, WANG Z R, JIN S Z, et al. High-precision positioning method based on SINS/RFID for trains in tunnel [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 632-638 (in Chinese).

北京航空航天大学学报

北航学报 网阅

号的到达时间、到达角度和信号强度值进行加权定位,以及近年来热门的信号指纹库匹配定位,在小范围内可以达到亚米级别甚至是厘米级别精度,但由于硬件成本问题、时钟同步问题及信号本身受环境影响等问题,目前还没有适用于诸如高铁隧道这种大面积环境定位的成熟方案。

针对以上问题,本文提出一种在隧道环境内对高铁列车进行高精度定位的方法,从响应时间和姿态角 2 个方面研究 RFID 标签的误差模型,应用捷联惯性导航系统(strapdown inertial navigation system, SINS)和 RFID 进行组合导航定位。

1　基于 RFID 系统的定位原理

1.1　RFID 系统的常见定位方法

射频识别系统一般包括 RFID 标签、RFID 阅读器和微计算机,标签按供电方式可分为有源、半有源和无源标签,无源标签因其成本低、被动式通信和使用寿命长等特点被广泛研究和应用。

利用 RFID 技术进行定位的方法,大致分为以下几种[10-11]:

1)接收信号强度指示(received signal strength indicator, RSSI)的 k 最邻近(k-nearest neighbor, KNN)算法,将接收信号强度代入信号传播模型从而计算收发端距离,再代入铺设的标签网络中进行加权定位。该方法受限于自由空间信号传播方程与实际环境的偏差,精度一般在米级。

2)基于信号到达角(angle of arrival, AOA),到达时间(time of arrival, TOA)或到达时间差(time difference of arrival, TDOA)的方法,需要使用天线阵得到精确的角度信息或高精度的时钟及同步设备得到精确的信号传播时间,才能够提高最终的定位精度,该方法所需成本过高。

3)空间指纹特征匹配法,通过在区域内设置若干个参考点形成定位区域的 RSSI 指纹库,继而寻找与待定位点特征相同或相近的参考点进行加权获得定位结果。该方法需要在前期获取大量参考位置的特征指纹信息,并且需要定期进行指纹库更新,准备工作量大且环境适应性较差。

以上 3 种较常见的 RFID 定位方法都只适用于小范围的室内定位环境。对于大面积的 GNSS 不可见环境,还需要寻找更合适的方法在提高定位精度的同时减小定位成本。

1.2　基于响应时间的 RFID 定位精度计算

目前,针对 RFID 的定位研究大多基于标签信号的传播特征,而很少针对 RFID 标签的响应过程。以无源 RFID 标签为例,将从阅读器发出激活信号到最终阅读器收到标签的反向散射信号这一过程的时间称为 RFID 标签的响应时间。对于某种特定环境下的特定标签、阅读器,响应时间都会处在一个特定范围内[12],即存在一个特定的上下界,文献[13]对几种标签做了大量的测试,其中响应时间最短的标签,对应的响应时间上下界分别为 17.4 ms 和 10.0 ms。

针对 RFID 标签响应时间的这种性质,结合具体环境可以对标签的定位精度做进一步的分析计算。以本文研究的隧道环境为例,如图 1 所示,RFID 标签水平放置于墙壁,对应的天线辐射角 $\angle AOC = 60°$。装配 RFID 阅读器的高铁列车通过 AC 段时,发出的激活信号可以被标签识别,而散射信号需要经过响应时间后才能返回并被仍旧处于这一范围的阅读器接收,即阅读器能够接收返回信号的区段为 BC 段。

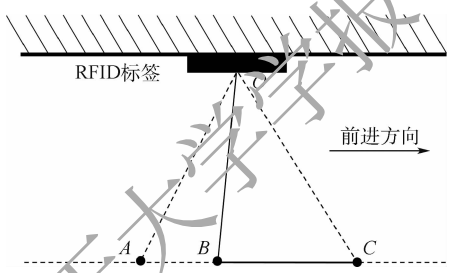


图 1　RFID 标签精度分析

Fig. 1　Precision analysis of RFID tag

对于高铁隧道,墙壁与车厢之间的距离 d 为 2.16 m,因此,这段信号可识别范围 AC 的距离为 $|\overrightarrow{AC}| = 2d\tan 30° \approx 2.494\ 1$ m。用 MATLAB 软件来仿真这一过程得到阅读器收到标签信号的位置,并进行蒙特卡罗仿真模拟得到这个位置相较于 A 点的距离分布。即以 A 为起点,$\overrightarrow{AC}$ 为正方向,携带阅读器的高铁列车速度为 100 m/s(略高于目前高铁最高时速 350 km/h),响应时间分别按正态分布和均匀分布进行 1 000 次蒙特卡罗仿真,最后得到位置的统计分布规律如表 1 所示。

从表 1 中可以看出,2 种分布的统计结果相近,距离起点 A 的均值位置对应的坐标位置可以设置为 RFID 标签内存储的具体定位信息,即后面组合导航中的位置校准点,而标准差代表这个

表 1　RFID 阅读器读取位置分布统计

Table 1　Distribution statistics of received signal position of RFID reader

响应时间分布	均值/m	标准差/m
正态分布	1.921 9	0.337 3
均匀分布	1.908 7	0.358 9

北京航空航天大学学报

位置点的定位误差。以响应时间为参考的定位方法在水平面位置误差不超过 36 cm，相较其他 RFID 定位方法，该方法定位精度较高、成本低且布设简单、灵活性高。

1.3 基于隧道轨道的 RFID 标签姿态角的精度计算

为了得到连续的定位信息，在得到 RFID 定位点位置信息后需要使用惯性导航系统进行推航，对于低精度或者长时间的惯性导航系统解算，由于失准角和惯性器件误差的不稳定，仅用位置信息进行校准往往不能很好地抑制误差[14]。利用 RFID 标签与惯性导航系统进行组合导航定位，校准信息不但可以包含位置信息，更可以根据实际环境下轨道特性，对姿态信息加以补充和校准，更好地减小定位误差。

在高铁隧道环境下，对于时速 350 km/h 的高铁列车，要求轨道曲率半径 ρ 不小于 7 km，因此在 RFID 标签的校准范围内，对应的航向角姿态变化最大值 $\delta\psi_{\max}$ 为

$$\delta\psi_{\max} = \frac{|\overrightarrow{AC}|}{\rho} = 0.000\,356\,3 \text{ rad} \tag{1}$$

即在对应区域内，航向角校准误差不超过最大变化值的一半，即 0.000 178 2 rad，俯仰角 θ 和横滚角 γ 的校准误差还要小于这个值。而在实际情况下，对应校准点的选择应该是一段区域内轨道变化较小的位置，因此姿态误差相对 $\delta\psi_{\max}$ 而言会更小一些。应用该方法，在一定程度上实现对高铁列车在动态环境下的姿态校准，减少惯性导航系统的姿态发散，进一步提高定位精度。

2 基于 SINS/RFID 定位系统的原理和方法

SINS 系统[15-16]利用自身三轴陀螺仪与三轴加速计测量数据进行推航，从而提供姿态、速度、位置等导航信息，短时间定位性能好，但其误差会随时间积累，利用 RFID 标签对惯性导航系统作校准，可以优势互补，得到全线路高精度定位信息。

2.1 隧道环境下列车 SINS/RFID 组合导航原理

针对隧道环境高铁列车的动态实时组合定位导航系统结构如图 2 所示。在隧道墙壁每隔一定距离布设一枚 RFID 标签，标签内部写入可以简单代表该处地点的 ID 编号，同时在微计算机中输入一张查找表，每个 ID 编号对应着该地点三维位置信息和列车通过该地点的姿态欧拉角信息（预

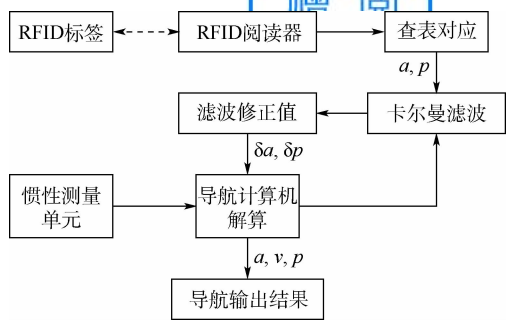


图 2 SINS/RFID 组合导航系统结构

Fig. 2 Structure of SINS/RFID integrated navigation system

先测量），列车运行时，惯性测量单元将陀螺仪与加速计的测量值输送给导航计算机解算并输出定位信息，当列车经过 RFID 标签时，RFID 阅读器接收到标签的散射信号并解析出对应 ID，通过查表找到 ID 所对应的地理信息，将这一信息与同一时间下的 SINS 解算信息做卡尔曼滤波，得到对应信息的改正值，修正并得到最终的导航结果。

2.2 卡尔曼滤波方程

组合导航系统需要对多个传感器信息作有效处理从而获得准确可靠的结果。卡尔曼滤波能够实现连续时间下多维状态的最优估计，因此被广泛应用于组合定位，尤其是动态环境下运载体的导航定位。

卡尔曼滤波针对线性系统的状态方程，通过观测状态输入输出建立量测方程，从而对系统状态进行最优估计[17]。式(2)～式(6)分别为状态一步预测、状态一步预测均方误差、滤波增益、状态估计和状态估计均方误差这 5 个卡尔曼滤波算法的基本公式。式中：$\hat{\boldsymbol{X}}_k$ 和 $\boldsymbol{P}_k$ 分别为 k 时刻的状态变量矩阵和误差协方差矩阵；$\hat{\boldsymbol{X}}_{k/k-1}$ 和 $\boldsymbol{P}_{k/k-1}$ 分别为状态预测矩阵和误差协方差预测矩阵；$\boldsymbol{\Phi}_{k/k-1}$ 为对应的状态转移矩阵；$\boldsymbol{Q}_{k-1}$ 为预测协方差；$\boldsymbol{K}_k$ 为对应卡尔曼滤波增益矩阵；$\boldsymbol{Z}_k$、$\boldsymbol{H}_k$ 和 $\boldsymbol{R}_k$ 分别为 k 时刻的量测向量矩阵、观测矩阵和量测噪声协方差矩阵；$\boldsymbol{I}$ 为单位矩阵。

$$\hat{\boldsymbol{X}}_{k/k-1} = \boldsymbol{\Phi}_{k/k-1}\hat{\boldsymbol{X}}_{k-1} \tag{2}$$

$$\boldsymbol{P}_{k/k-1} = \boldsymbol{\Phi}_{k/k-1}\boldsymbol{P}_{k-1}\boldsymbol{\Phi}_{k/k-1}^{\mathrm{T}} + \boldsymbol{Q}_{k-1} \tag{3}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k)^{-1} \tag{4}$$

$$\hat{\boldsymbol{X}}_k = \hat{\boldsymbol{X}}_{k/k-1} + \boldsymbol{K}_k(\boldsymbol{Z}_k - \boldsymbol{H}_k\hat{\boldsymbol{X}}_{k/k-1}) \tag{5}$$

$$\boldsymbol{P}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k/k-1} \tag{6}$$

卡尔曼滤波状态方程和量测方程如式(7)所示，其中 $\boldsymbol{X}$ 为 15 维状态变量，具体如式(8)所示。$\boldsymbol{\phi}$、$\delta\boldsymbol{v}^n$、$\delta\boldsymbol{p}$、$\boldsymbol{\varepsilon}$ 和 ∇ 分别表示三维失准角、三维速度误差、三维位置误差、陀螺测量零漂和加速度计测

北京航空航天大学学报

量零偏；$\boldsymbol{W}$ 包含陀螺角速度测量白噪声和加速度计比力测量白噪声；$\boldsymbol{V}$ 为量测信息中的白噪声；$\boldsymbol{Z}$ 为系统的实时更新量测方程。若进行位置组合，以 SINS 和 RFID 的位置差值作为量测值，量测矩阵 $\boldsymbol{H}$ 如式(9)所示。

$$\begin{cases}\dot{\boldsymbol{X}} = \boldsymbol{F}\boldsymbol{X} + \boldsymbol{G}\boldsymbol{W} \\ \boldsymbol{Z} = [\boldsymbol{p}_{\mathrm{SINS}} - \boldsymbol{p}_{\mathrm{RFID}}] = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{V}\end{cases} \tag{7}$$

$$\boldsymbol{X} = [\boldsymbol{\phi}^{\mathrm{T}} \quad (\delta \boldsymbol{v}^n)^{\mathrm{T}} \quad (\delta \boldsymbol{p})^{\mathrm{T}} \quad \boldsymbol{\varepsilon}^{\mathrm{T}} \quad \nabla^{\mathrm{T}}]^{\mathrm{T}} \tag{8}$$

$$\boldsymbol{H} = [\boldsymbol{0}_{3\times6} \quad \boldsymbol{I}_{3\times3} \quad \boldsymbol{0}_{3\times6}] \tag{9}$$

式中：$\boldsymbol{F}$ 和 $\boldsymbol{G}$ 分别为状态量和噪声。

若增加姿态校准信息，则量测值包含失准角和位置的差值信息，而正常测得的一般是姿态角信息，因此需要将包含校准误差的校准姿态角信息与 SINS 解算的姿态角信息转换至失准角，如式(10)所示，$\boldsymbol{C}_b^{n'}$ 即 SINS 的计算导航系下的方向余弦矩阵，$\boldsymbol{C}_b^n$ 即 RFID 中对应理想姿态角的方向余弦矩阵。增加了三轴失准角信息后，量测矩阵如式(11)所示。

$$\boldsymbol{\phi}\times = \boldsymbol{I} - \boldsymbol{C}_b^{n'}(\boldsymbol{C}_b^n)^{-1} \tag{10}$$

$$\boldsymbol{H} = \begin{bmatrix}\boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times6} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times6}\end{bmatrix} \tag{11}$$

3　仿真实验

3.1　隧道列车轨迹仿真

首先利用 MATLAB 软件仿真高铁列车的轨迹。由于实际隧道环境下，列车在高度方向变化很小，在轨迹仿真与后续的定位精度比较仅涉及二维水平方向。设定好起点与列车初始朝向，行进中保持 100 m/s (360 km/h)的速度，转弯速度为 0.8°/s(对应轨道曲率半径为 7 161.97 m)，隧道总长 30 km，具体行进方案按时间顺序如表 2 所示。

按上述轨迹方案进行仿真，记录理想轨迹下姿态、速度和位置信息。列车轨迹曲线如图 3 所示。

表 2　列车仿真轨迹的行进方案

Table 2　Scheme of train trajectory simulation

序号	路线方案	时间/s	重复次数
1	直行	50	1
2	左转，直行	5，5	7
3	左转，直行	45，5	1
4	右转，直行	5，5	1
5	左转，直行	5，5	2
6	右转，直行	5，5	5
7	左转，直行	5，5	1
8	右转，直行	5，5	4

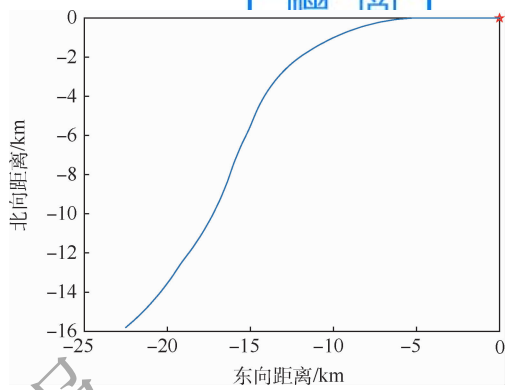


图 3　列车隧道轨迹曲线仿真

Fig. 3　Curve of train trajectory simulation in tunnel

3.2　SINS/RFID 组合导航精度分析

利用 MATLAB 软件，按第 2 节内容进行仿真，RFID 标签的定位误差设为 0.5 m，即水平面的总误差为 0.5 m，根据理想轨迹下的航向角在校准位置分配给北向和东向。标签的姿态角误差设为[0.000 3，0.000 3，0.000 3] rad。三向加速度计测量零偏为 $\nabla_E = \nabla_N = \nabla_U = 100\ \mu g$，速度随机游走系数为 $\omega_{dx} = \omega_{dy} = \omega_{dz} = 50\ \mu g/\sqrt{\mathrm{Hz}}$。

为了比较组合导航系统的定位精度，需要针对不同的情况做仿真并进行分析。关于 RFID 的校准点设置分为间隔 $D = 100$ m、$D = 200$ m、$D = 500$ m、$D = 1\ 000$ m 和不设置标签 5 种情况。而对三轴陀螺仪陀螺测量零漂 $\boldsymbol{\varepsilon}$ 和角度随机游走系数 $\boldsymbol{\omega}$ 的设置分以下 3 种情况，其中右下角标 E、N、U 分别表示东、北、天 3 个方向上的陀螺参数分量。

情况 1：

$$\begin{cases}\varepsilon_E = \varepsilon_N = \varepsilon_U = 0.1°/\mathrm{h} \\ \omega_E = \omega_N = \omega_U = 0.01°/\sqrt{\mathrm{h}}\end{cases}$$

情况 2：

$$\begin{cases}\varepsilon_E = \varepsilon_N = \varepsilon_U = 1°/\mathrm{h} \\ \omega_E = \omega_N = \omega_U = 0.1°/\sqrt{\mathrm{h}}\end{cases}$$

情况 3：

$$\begin{cases}\varepsilon_E = \varepsilon_N = \varepsilon_U = 10°/\mathrm{h} \\ \omega_E = \omega_N = \omega_U = 1°/\sqrt{\mathrm{h}}\end{cases}$$

3 种情况分别代表了目前市场高、中、低 3 种精度的惯性导航系统，也称惯性级、导航级和速率级惯性导航系统。按理想仿真轨迹下的惯性传感器数据，加上对应噪声误差，进行 SINS 更新算法和 SINS/RFID 的组合导航算法的仿真，得到的轨迹再与理想轨迹作对比，得到定位误差如图 4 和图 5 所示。

图 4 和图 5 分别为在 $\varepsilon = 1°/\mathrm{h}$，$\omega = 0.1°/\sqrt{\mathrm{h}}$，间隔 $D = 200$ m 的情况下的定位误差，图 5 相较于

北京航空航天大学学报

北航学报 赠阅

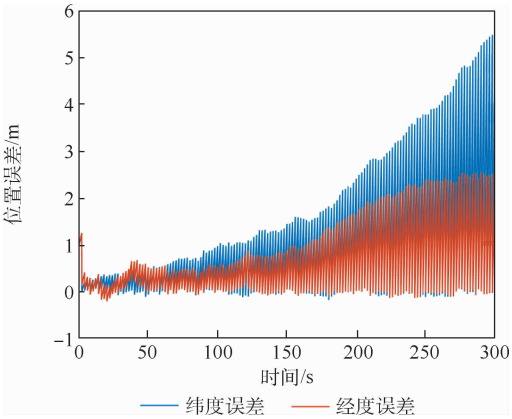


图 4 RFID 位置校准 SINS 的定位误差

Fig. 4 Positioning error of SINS calibrated by position of RFID

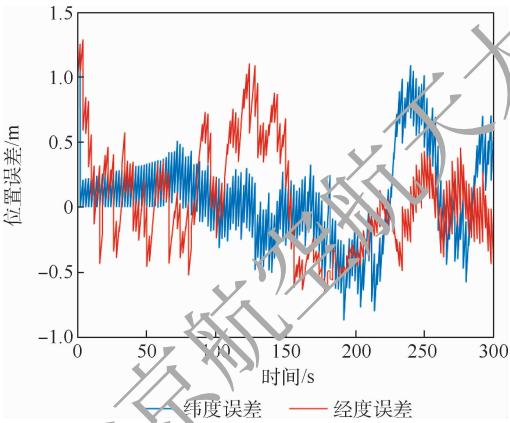


图 5 RFID 姿态位置信息校准 SINS 的定位误差

Fig. 5 Positioning error of SINS calibrated by position and attitude of RFID

图 4 在 RFID 标签中增加了姿态角信息。由图 4 和图 5 对比可以看出，增加了姿态角信息后，组合系统同时校准了位置和姿态，在校准间隔内抑制了定位误差的发散趋势，使定位误差明显减小。

对于动态的仿真轨迹，由于单次仿真的定位误差随机性较强，需要做多次仿真去得到定位误差的统计规律，针对上述不同精度陀螺仪（情况 1～情况 3）与 RFID 的组合，每种情况下的一次仿真只保留整个过程的最大误差值，即代表这一次通过隧道过程中能够达到的最大误差，然后每种情况做 1 000 次蒙特卡罗仿真，得到该情况下定位误差的统计规律（包括最大值和平均值）。SINS/RFID 位置组合误差统计数据如表 3 所示，SINS/RFID 位置与姿态组合的误差统计数据如表 4所示。

由表 3 可见，对于不用 RFID 只依靠 SINS 进行定位的场景，即使是高精度的陀螺仪也会产生百米级别的平均定位误差，而低精度的陀螺仪带来 3 000 m 以上的定位误差是无法接受的。引入 RFID 位置信息后，在标签间隔 100 m 的场景下，配合高精度 SINS 能够达到 0.75 m 的平均定位误差，随着陀螺仪精度的降低和 RFID 布设间隔的增加，定位误差也随之增加，对应于低精度 SINS 在 2 种场景下的平均定位误差高于 100 m。

由表 4 可见，在位置信息基础上增加姿态信息后，所有场景下的定位误差均减小，对应高精度 SINS 能够达到最高 0.5 m 的平均定位误差，而对应上文低精度 SINS 可能产生百米以上误差的情况，进行位置姿态校准后平均定位误差不超 30 m。

表 3 SINS/RFID 位置组合导航定位误差统计

Table 3 Navigation positioning error statistics of SINS/RFID integrated by positon

间隔/m	情况 1		情况 2		情况 3	
	平均值/m	最大值/m	平均值/m	最大值/m	平均值/m	最大值/m
100	0.752 8	1.153 3	3.104 4	5.682 5	27.824 4	51.267 9
200	1.374 4	1.993 8	6.150 7	11.315 7	56.235 7	107.712 2
500	3.187 6	4.505 5	15.423 3	27.781 1	141.115 7	254.971 2
1 000	6.202 6	9.020 6	30.243 2	52.392 2	276.417 9	538.789 2
不用 RFID	93.333 2	122.828 6	338.633 0	653.079 8	3 072.103 3	6 425.889 1

表 4 SINS/RFID 位置姿态角组合导航定位误差统计

Table 4 Navigation positioning error statistics of SINS/RFID integrated by positon and attitude

间隔/m	情况 1		情况 2		情况 3	
	平均值/m	最大值/m	平均值/m	最大值/m	平均值/m	最大值/m
100	0.510 4	0.743 4	0.573 8	0.851 9	0.625 2	1.002 6
200	0.845 7	1.413 4	0.968 0	1.427 1	1.183 1	1.987 6
500	1.812 3	2.936 7	2.158 1	3.580 0	3.789 5	6.411 3
1 000	3.497 4	6.041 7	4.507 6	9.521 8	11.982 4	24.171 1

北京航空航天大学学报

北航学报 赠阅

4　结　论

本文提出了一种针对高铁列车在隧道等卫星不可见环境下的 SINS/RFID 组合导航定位方法，结合隧道环境给出 RFID 标签的定位精度，对比多次蒙特卡罗仿真数据分析得出：

1）对于 30 km 长的隧道，列车单独依靠 SINS 定位，装备低精度陀螺仪将产生千米以上的平均定位误差，即使配备高精度陀螺仪也将产生百米级的定位误差。

2）引入 RFID 标签的位置信息进行组合导航，在最好的情况下能够达到 1 m 以内的定位误差，其他情况下相较于单独使用 SINS，在导航性能上也都有很大的提升，但定位误差会随时间而逐渐有发散趋势。

3）利用高铁轨道的特征在 RFID 标签信息中加入姿态信息，与 SINS 进行位置姿态的组合，在定位精度上相对于只用位置信息组合又有了很大的提升，在校准间隔为 100 m 的情况下，能够达到全程平均定位精度在 1 m 以内，最好精度能够达到全程最大误差的平均值 51 cm。

基于 SINS/RFID 的组合定位方法在大型的隧道环境内对于高铁列车位置信息的获取是行之有效的，引入姿态信息后，更能够在动态环境下对列车姿态进行校准，抑制位置误差的发散，进一步提高定位精度。

参考文献（References）

[1] 上官伟，王韦舒，张路，等. 北斗导航 RAIM 技术在列车定位的应用研究[J]. 铁道学报，2018，40(2)：73-81.
SHANGGUAN W，WANG W S，ZHANG L，et al. Application of BDS-based RAIM technology in train positioning[J]. Journal of the China Railway Society，2018，40(2)：73-81(in Chinese).

[2] WANG D S，LU Y J，ZHANG L，et al. Intelligent positioning for a commercial mobile platform in seamless indoor/outdoor scenes based on multi-sensor fusion[J]. Sensors，2019，19(7)：1696.

[3] 王晓薇. 基于 WIFI 的隧道人员定位算法的分析与研究[D]. 武汉：武汉邮电科学研究院，2017：18-24.
WANG X W. Analysis and research of tunnel personnel location algorithm based on WIFI[D]. Wuhan：Wuhan Research Institute of Posts and Telecommunications，2017：18-24(in Chinese).

[4] 黄文轩. 在 GPS 不完备条件下 RFID 列车定位研究[J]. 铁路通信信号工程技术，2018，15(1)：25-29.
HUANG W X. Research on RFID train location under incomplete GPS[J]. Railway Signalling & Communication Engineering，2018，15(1)：25-29(in Chinese).

[5] 陈伟. 卫星遮挡交通环境下车辆融合定位策略研究[D]. 南京：东南大学，2017：81-97.
CHEN W. Research on fusion positioning strategy for land vehicles in satellite signal-blocked traffic environments[D]. Nanjing：Southeast University，2017：81-97(in Chinese).

[6] PEREIRA F L. Positioning systems for underground tunnel environments[D]. Portugal：Faculdade de Engenharia da Universidade do Porto，2016：79-115.

[7] 方旺盛，高银. 狭长直隧道环境中 WSN 的 RSSI 加权质心定位算法[J]. 传感技术学报，2014，27(2)：247-251.
FANG W S，GAO Y. A weighted centroid localization algorithm based on RSSI for wireless sensor network in straight narrow tunnel[J]. Chinese Journal of Sensors and Actuators，2014，27(2)：247-251(in Chinese).

[8] SONG X，LI X，TANG W C，et al. A hybrid positioning strategy for vehicles in a tunnel based on RFID and in-vehicle sensors[J]. Sensors，2014，14(12)：23095-23118.

[9] KO H，SHIM Y，KONG S H. Realization and demonstration of enhanced Korean high-speed train navigation system with noise filtering schemes[J]. International Journal of Control Automation and Systems，2018，16(9)：769-781.

[10] CUI L，ZHANG Z H，GAO N，et al. Radio frequency identification and sensing techniques and their applications—A review of the state-of-the-art[J]. Sensors，2019，19(18)：4012.

[11] 彭国旗. 基于 RFID 的车辆定位技术研究[D]. 长春：长春理工大学，2018：7-18.
PENG G Q. Research on vehicle positioning technology based on RFID[D]. Changchun：Changchun University of Science and Technology，2018：7-18(in Chinese).

[12] ZHAO Y，LIU K H，MA Y T，et al. Similarity analysis based indoor localization algorithm with backscatter information of passive UHF RFID tags[J]. IEEE Sensors Journal，2017，17(1)：185-193.

[13] MATBOULY H，ZANNAS K，DUROC Y，et al. Analysis and assessments of time delay constrains for passive RFID tag-sensor communication link：Application for rotation speed sensing[J]. IEEE Sensors Journal，2017，17(7)：2174-2181.

[14] 严恭敏，邓瑀. 传统组合导航中的实用 Kalman 滤波技术评述[J]. 导航定位与授时，2020，7(2)：50-64.
YAN G M，DENG Y. Review on practical Kalman filtering techniques in traditional integrated navigation system[J]. Navigation Positioning & Timing，2020，7(2)：50-64(in Chinese).

[15] 严恭敏，翁浚. 捷联惯导算法与组合导航原理[M]. 西安：西北工业大学出版社，2019：78-105.
YAN G M，WENG J. Strapdown inertial navigation algorithm and principle of integrated navigation[M]. Xi'an：Northwestern Polytechnical University Press，2019：78-105(in Chinese).

[16] SAVAGE P G. Strapdown inertial navigation integration algorithm design Part 1：Attitude algorithms[J]. Journal of Guidance，Control，and Dynamics，1998，21(1)：19-28.

[17] 秦永元，张洪钺，王叔华. 卡尔曼滤波与组合导航原理[M]. 2 版. 西安：西北工业大学出版社，2012：5-67.
QIN Y Y，ZHANG H Y，WANG S H. Kalman filtering and principle of integrated navigation[M]. 2nd ed. Xi'an：Northwestern Polytechnical University Press，2012：5-67(in Chinese).

北京航空航天大学学报

北航学报 赠阅

High-precision positioning method based on SINS/RFID for trains in tunnel

ZHANG Chendong[1,2], WANG Zhaorui[1,*], JIN Shengzhen[1], AI Guoxiang[1]

(1. National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China;
2. School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: based on strapdown inertial navigation system (SINS) and radio frequency identification (RFID), an integrated positioning method is proposed for the position acquisition of high-speed trains in satellite-denied environment such as tunnels. The positioning accuracy of RFID tags is calculated by the response time model. The tags could also calibrate the attitude by adding actual railway information. Setting RFID tags on tunnel wall and at the same time combining with inertial navigation system provide continuous dynamic positioning data. Simulation results show that utilizing RFID tags in position calibration significantly decreases the error accumulation of inertial navigation system and increases positioning accuracy in 30 km tunnel. After the addition of attitude information, the positioning accuracy of the whole tunnel railway maintains at the level of meters in a variety of combinations of gyroscope performance and RFID tag interval, the best of which is 0.5 m.

Keywords: integrated navigation; tunnel positioning; radio frequency identification (RFID); strapdown inertial navigation system (SINS); dynamic attitude calibration

北京航空航天大学学报

Received: 2020-11-20; **Accepted**: 2021-02-26; **Published online**: 2021-03-25 18:41

URL: kns.cnki.net/kcms/detail/11.2625.V.20210324.1519.003.html

Foundation items: National Key R & D Program of China (2016YFB0501900); National Natural Science Foundation of China (U1931125, 11603041)

* **Corresponding author**. E-mail: zhaorui_w@nao.cas.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0625

一种基于地转风模型的平流层风场插值方法

全下文载

胡康[1,2], 周江华[1,2,*], 张晓军[1]
(1. 中国科学院空天信息创新研究院, 北京 100094; 2. 中国科学院大学 光电学院, 北京 100049)

摘 要: 平流层风场水平分量的数据精度和密度对高空气球实验的轨迹预测精度和飞行控制精度有着直接的影响。面向平流层风场,提出一种基于地转风模型的插值方法。该方法通过改进科里奥利频率公式,并用二维卷积计算地转风模型,提升了地转风模型的计算效率和低纬度地区风场精度,再用迭代二元线性回归和改进的自适应观测场偏差权重矩阵减小了观测误差对插值结果的影响。实验结果表明,所提插值方法能有效提升平流层风场插值计算速度,并对平流层风场水平分量的插值精度有显著提升。

关 键 词: 风场插值; 平流层风场; 地转风; 二元线性回归; 二维卷积
中图分类号: V321.2
文献标志码: A **文章编号**: 1001-5965(2022)04-0639-08

高空气球作为平流层科学实验的主要载体,具有有效载荷大、发放成本低等显著优点。但是作为一种无动力飞行器,高空气球在驻空阶段只能通过调节浮力和重力来控制垂直方向的运动,而无法主动调节飞行的水平方向,因此,平流层风场水平方向风速分量的数据精度和密度将直接影响高空气球的飞行轨迹[1]。目前,高空气球飞行仿真和实际实验使用的全球风场数据主要来自于欧洲中期天气预报中心(ECWMF)的 ERA 系列全球气象数据模型,其中最新的 ERA5 模型提供了全球气象再分析资料和未来 5 d 内的气象预报数据,每一时刻的数据水平分辨率为 0.25°×0.25°(经纬度),包涵从地面到距地面 80 km 处的37 个等压面风场模型[2]。然而在使用 ERA5 数据对高空气球进行飞行仿真和实际操控的过程中,通过控制气球高度处于合适的风速带来达成区域驻留目标,其有效驻留控制半径在 100 km 以上,控制精度差,对高空气球实验的轨迹预测、设备回收和空域申请等工作都带来很大程度的影响,因此,需要研究一种根据现有数据生成大范围、高精度、高密度风场数据的插值方法来提高高空气球的控制精度。

现有的风场插值方法可以分成 2 类:直接插值类方法和融合插值类方法。直接插值类方法是对风场的预报数据或者观测数据采用各种插值方法进行插值,从而获得最终结果。Jardin 和 Erbes-ser[3] 利用全局多项式插值方法和风场预报对风场格点数据进行了插值。Tao 等[4] 研究了 Hermite 插值在风场光谱表示中的误差传递和数值模型。汤新民等[5] 研究了 3 种不同变异模型下的克里格插值法在风场插值中的应用效果。董志南等[6] 对地面区域内的实测风场数据进行交叉验证,比较了反距离权重插值、全局多项式插值、局部多项式插值、径向基函数插值、最近邻插值及普通克里金插值 6 种方法。上述直接插值类方法[3-6] 是对风场的预报数据或者观测数据直接进行插值,具有计算速度快、数据质量要求低等优点。这些方法主要面对地面及近地风场,观测的时间和空间尺度范围小且采样难度低,通过密集的风场实测数据,补插出与实际风场高

收稿日期: 2020-11-09; **录用日期**: 2020-12-25; **网络出版时间**: 2021-01-22 11:19
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210122.1116.001.html
基金项目: 国家自然科学基金(61733017); 中国科学院战略性先导科技专项(XDA170400)
* **通信作者**. E-mail: zhoufma@aoe.ac.cn

引用格式: *胡康, 周江华, 张晓军. 一种基于地转风模型的平流层风场插值方法[J]. 北京航空航天大学学报, 2022, 48(4): 639-646. HU K, ZHOU J H, ZHANG X J. A stratospheric wind field interpolation method based on geostrophic wind model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 639-649 (in Chinese).*

北京航空航天大学学报

北航学报 赠阅

度拟合的插值函数。但是平流层风场时间和空间观测尺度大、采样成本高、数据稀疏且缺乏连续性特征[7],直接插值类方法难以给出有效的插值结果。

融合插值类方法则先对风场数据进行融合,在融合过程中分析风场特征并设计插值函数,从而对融合结果进行插值,获得最终结果。Gandin[8]采用最优插值法进行了风场融合插值,以最小方差作为约束条件,对观测风场和预报风场进行数据融合,从而获得对风场插值结果的最优估计。朱成阵等[9]通过风场的实测数据和气象预报数据,用内插外推法预测特定位置的风场信息。上述融合插值类方法[8-9]将小范围的实测数据作为观测场,将风场预报数据作为背景场,进行数据融合,分析风场特征,再根据特征进一步设计风场插值模型。但是统计数据表明,风场在时间和空间维度上有显著的差异性[10],只采用小范围的实测数据进行数据融合获得的插值方法面对更大的区域难以具有良好的泛化性能。同时,计算高精度的风场预报作为模型的背景场需要巨大的算力和时间开销。以国内全球/区域同化预测系统(global/regional assimilation prediction system, GRAPES)中的全球版本为例,作为序贯模型,对某一时刻 0.5°×0.5°(经纬度)网格 31 层模型的积分计算需要 200 个以上的 CPU 节点计算 10 d[11],面对需要更高数据密度的高空气球实验,传统的融合插值类方法很难符合其时效性要求。

为了解决平流层风场插值面临的插值精度低、计算时间长和算力开销大等问题,进行速度快、精度高的插值计算,本文提出一种基于地转风(based on geostrophic wind, BGW)模型的平流层水平方向风场插值方法,用 2019 年 7 月 ERA5 全球气象再分析数据对平流层的全球风场进行插值计算实验验证。相比于直接插值类方法,BGW 插值对平流层水平分量的插值效果有明显提升;相比于融合插值类方法,BGW 插值大幅度降低了计算复杂度,有效提高了平流层风场插值的计算速度。

1 基于地转风模型的风场插值

本文提出的 BGW 风场插值方法是一种改进的融合插值类方法,主要包含两部分:①为了提升计算效率,用重力位势数据计算地转风模型作为插值背景场;②为了减小背景场误差和观测误差对插值结果的影响,将已有的风速数据作为观测场,用迭代的二元线性回归对背景场数据进行修正,获得最终的插值结果。

1.1 地转风模型改进

一般的融合插值类方法需要使用气象预报模型计算的目标精度风场预报数据作为背景场。气象预报模型一般是基于大气运动方程组的数值模型,能够高精度地预测风速的东向、北向及垂直方向分量,对中短期风场都有着良好的预测效果。然而数值预报模型的实际计算过程需要大量的历史数据、算力和时间作为支撑。因此,当前的融合类插值方法需要大量的时间来进行计算,或依赖于外部输入的目标密度风场预报数据,难以在实验现场根据需求进行实时的离线计算。为了解决这一问题,本文提出一种用地转风模型替代风场预报作为背景场的方法。

地转风模型由白贝罗于 1857 年提出,是大气运动方程组在大尺度风场条件下推导而出,当大气运动满足科里奥利力和气压梯度力平衡,即地转平衡条件时的理论风[12],在局地直角坐标系下,其公式如下:

$$\begin{cases} u = -\dfrac{1}{f\rho}\cdot\dfrac{\partial p}{\partial y} \\ v = \dfrac{1}{f\rho}\cdot\dfrac{\partial p}{\partial x} \end{cases} \tag{1}$$

式中:u、v 分别为合风速在直角坐标系中的西风、南风分量;p 为气压;ρ 为大气密度;x、y 分别为坐标系的向东、向北坐标;f 为科里奥利频率,公式如下:

$$f = 2\Omega\sin\varphi \tag{2}$$

式中:Ω 为地球的旋转角速度,值为 7.29×10^{-5} rad/s;φ 为纬度。

地转风虽然是根据气压分布计算出的估计值,并非实际存在的风,但是根据实验验证,以地转平衡作为条件计算的中高纬度自由大气风场与真实风场十分接近[13],因此用地转风模型作为风场插值的背景场能够在提高计算速度的同时,准确地显示中高纬度地区的风速分布特征。

在纬度逼近赤道的过程中,由于科里奥利频率 f 的值逐渐趋近于 0,导致在低纬度地区地转风计算值趋于无穷,因此对 f 的计算公式进行改进。对于一般意义上的赤道附近区域,即南北纬 10°之间的区域,选取边界对科里奥利频率公式进行截断,截断边界的选择范围为(0°,10°],越接近 0 则赤道附近区域风场误差越大,越接近 10 则初值场和准地转运动偏离越多。综合考虑选取中值 5°为截断边界,将北纬 5°和南纬 5°之间的 f 值固定为 $2\Omega\sin5°$,改进后的公式如下:

北京航空航天大学学报

北航学报 瞰阅

$$f=\begin{cases}2\Omega\sin\varphi & |\varphi|\geqslant 5^\circ\\ 2\Omega\sin 5^\circ & |\varphi|<5^\circ\end{cases} \tag{3}$$

经过实验验证,相比于使用原始的地转风模型,导致最终插值精度较差,用式(3)的 BGW 插值在低纬度地区也能获得良好的插值效果。

1.2　基于二维卷积的地转风快速计算方法

式(1)的地转风模型需要大气密度 ρ 作为输入,而在实际实验中无法实时获得大气密度数据,因此对式(1)进行改进。另外,为了进一步提升背景场的计算速度,结合 GPU 的矩阵运算优势改进了背景场计算方法。

1.2.1　基于重力位势的地转风模型

由于气压 p 是局地直角坐标系 (x,y,z,t) 的单值函数,可以将坐标系中的高度 z 用气压 p 替代,从而对于任意坐标系 (x,y,z,t) 下的函数 $F(x,y,z,t)$ 有

$$F(x,y,z,t)=F[x,y,z(x,y,p,t),t] \tag{4}$$

对函数 $F(x,y,z,t)$ 进行不涉及时间 t 变化的计算时,时间可以作为常量,则在水平方向沿等压面对 x,y 求导则有

$$\begin{cases}\left(\dfrac{\partial F}{\partial x}\right)_p=\left(\dfrac{\partial F}{\partial x}\right)_z+\dfrac{\partial F}{\partial z}\left(\dfrac{\partial z}{\partial x}\right)_p\\ \left(\dfrac{\partial F}{\partial y}\right)_p=\left(\dfrac{\partial F}{\partial y}\right)_z+\dfrac{\partial F}{\partial z}\left(\dfrac{\partial z}{\partial y}\right)_p\end{cases} \tag{5}$$

如果函数 F 是气压函数,则在同一等压面上,即 $\left(\dfrac{\partial p}{\partial x}\right)_p=0$ 时,可得

$$\left(\frac{\partial p}{\partial x}\right)_z=\rho g\left(\frac{\partial z}{\partial x}\right)_p \tag{6}$$

同理可得

$$\left(\frac{\partial p}{\partial y}\right)_z=\rho g\left(\frac{\partial z}{\partial y}\right)_p \tag{7}$$

式中:g 为重力加速度。

将式(6)和式(7)代入式(1),可得

$$\begin{cases}u=-\dfrac{g}{f}\left(\dfrac{\partial z}{\partial y}\right)_p\\ v=\dfrac{g}{f}\left(\dfrac{\partial z}{\partial x}\right)_p\end{cases} \tag{8}$$

又有重力位势 $\phi=gz$,代入式(8)可得

$$\begin{cases}u=-\dfrac{1}{f}\left(\dfrac{\partial \phi}{\partial y}\right)_p\\ v=\dfrac{1}{f}\left(\dfrac{\partial \phi}{\partial x}\right)_p\end{cases} \tag{9}$$

式(9)把式(1)中的大气密度和气压这 2 个变量转换为只有重力位势这一个变量,在降低计算复杂度的同时,解决了地转风模型需要大气密度参与计算的问题。

1.2.2　基于二维卷积的优化计算

为了解决离散重力位势的微分无法计算这一问题,用中心差分近似计算微分,则对于 $n\times m$ 的网格中,任意格点 $n_{i,j}(x_i,y_j)$ 处有

$$\begin{cases}u(x_i,y_j)=-\left(\dfrac{\phi(x_i,y_{j+1})-\phi(x_i,y_{j-1})}{2f\Delta y}\right)_p\\ v(x_i,y_j)=\left(\dfrac{\phi(x_{i+1},y_j)-\phi(x_{i-1},y_j)}{2f\Delta x}\right)_p\end{cases} \tag{10}$$

式中:$i=1,2,\cdots,n$;$j=1,2,\cdots,m$。

对输入的二维格点重力位势数据进行预处理,如果有缺省则用三次样条插值进行补插。对处理后的重力位势数据进行如式(10)所示的计算,计算矩阵形式如下:

$$\begin{cases}[u(x,y)]=\dfrac{1}{f\Delta y}[\phi(x,y)]*\boldsymbol{A}\\ [v(x,y)]=\dfrac{1}{f\Delta x}[\phi(x,y)]*\boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A}=[-0.5\quad 0\quad 0.5]\end{cases} \tag{11}$$

式中:“$*$”表示矩阵的二维卷积运算;$\boldsymbol{A}$ 为提取的卷积核。

由于输入是全球重力位势数据的平面展开投影,而卷积计算对矩阵边缘有计算损失,考虑到投影过程是将地球沿 180°经线东西向分开,南北两极将纬度 90°处展开,在进行卷积计算时将重力位势矩阵左右进行循环扩展,上下进行复制扩展,这种改进有效抑制了卷积计算的边缘损失。

相比于传统模型沿等压线根据大气密度进行逐格计算来求解地转风[14],BGW 插值通过提取模型卷积核,用二维卷积进行快速的矩阵计算,有效提升了地转风模型的计算速度。

1.3　迭代二元线性回归确定插值函数

BGW 插值方法将地转风模型计算的目标密度风场格点数据作为背景场,用部分格点的实际风速观测数据作为观测场,用观测场对背景场进行修正以获得最终插值结果。每一个格点上的插值结果是由背景场值和修正值用迭代二元线性回归确定,修正值由一定范围内 N 个格点上的观测值与背景场值的偏差加权获得,最终获得如下线性回归方程式(12)和其约束条件式(13):

$$\boldsymbol{X}_k^a=m_0\boldsymbol{X}_k^b+m_1\sum_{i=1}^{N}w_i(\boldsymbol{X}_i^o-\boldsymbol{X}_i^b)+m_2 \tag{12}$$

$$\min(\boldsymbol{X}_k^a-\boldsymbol{X}_k^o)^2 \tag{13}$$

式中:$\boldsymbol{X}_k^a$ 为待插值点的插值结果矩阵;$\boldsymbol{X}_k^b$ 为待插值点的背景场矩阵;m_0、m_1 和 m_2 为线性回归方程的待定系数;$\boldsymbol{X}_i^o$ 和 $\boldsymbol{X}_i^b$ 分别为观测格点上的观测值矩阵和背景场值矩阵;N 为以待插值点为中

北京航空航天大学学报

心、半径为 r 的圆形搜索区域内的观测点个数，对不同的风场数据，参数 r 的值不同，为了保证在搜索区域内有一定数量的观测格点，r 的值要大于临近观测格点的间距 d，同时根据后续的计算验证，r 小于 $2d$ 时插值精度最高；w_i 为搜索范围内观测格点上背景场值与观测值的偏差权重。

在地理信息领域，传统的距离相关权重矩阵一般采用反距离权重矩阵。反距离权重受到观测点和待插值点的距离及幂次参数控制，需要根据数据密度、搜索范围和插值效果不断调整幂次参数[15]。在权重 w_i 中引入搜索半径参数 r 并进行重新设计，使得权重可以根据搜索范围进行自适应调整，解决了对不同搜索范围都要重新设计权重的问题，其公式如下：

$$w_i = \frac{\left(\frac{r-d}{rd}\right)^2}{\sum_{i=1}^{N}\left(\frac{r-d}{rd}\right)^2} \tag{14}$$

式中：d 为观测点到待插值点的距离。同时为了加快线性回归的收敛速度，对 w_i 进行归一化处理。

为了进一步提高插值精度，在完成一轮线性回归计算后，将插值结果矩阵作为背景场数据再一次输入回归模型进行迭代，重复迭代过程，直到插值精度收敛。经过验证，在两轮迭代时模型的插值精度已经收敛，取两轮迭代的模型作为最终的插值函数。

2 实验验证和对比分析

2.1 数据集

平流层风场插值方法的目的是为了获得尽可能贴近真实风场情况的插值结果，ECWMF 的 ERA 系列气象再分析模型使用了全球气象资料融合系统和完善的气象数据数据库，从 1970 年至今，对全球各地来自地面雷达测站、船载气象测站、海上浮标、探空气球、飞机和卫星的多源观测资料进行质量评价和数据融合，从而生成高度接近真实风场的完整数据集[2]。以往的高空气球飞行试验也以 ERA 系列模型作为参考，历史数据也验证了 ERA 系列模型的可靠性。因此，选择 ERA 系列最新的 ERA5 模型 2019 年 7 月的气象再分析数据进行插值计算和实验验证。由于高空气球的驻空阶段一般驻留在平流层的底部区域，距地面 20 km 左右，位于 50 hPa 大气压等压面附近，选用 50 hPa 等压面模型的 0.5°×0.5°（经纬度）网格的重力位势数据作为输入，用经纬度网格0.25°×0.25°的风速数据进行验证。

2.2 评价指标

通过实验对 BGW 插值方法中，背景场对实际风场特征的估计能力、插值模型的优劣程度和插值精度进行评价，分别使用相关系数矩阵、R^2 检验和平均绝对误差（mean absolute error，MAE）作为评价指标。

2.2.1 相关系数矩阵

背景场对实际风速分布特征的表达能力用背景场数据与再分析资料的相关系数矩阵表示，矩阵 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的相关系数矩阵 ρ_{XY} 的公式为

$$\rho_{XY} = \frac{\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})}{\sqrt{\mathrm{Var}(\boldsymbol{X})\cdot\mathrm{Var}(\boldsymbol{Y})}} \tag{15}$$

式中：Cov 为计算 2 个数据集的协方差矩阵；Var 为计算数据集的标准差。

相关系数矩阵常用来度量 2 个变量间的线性关系，值在 -1 ~ 1 之间，数值越大表明 2 个变量之间的线性相关性越强，一般 0 ~ 0.3 为弱相关，0.3 ~ 0.6 为中等相关，0.6 ~ 1 为强相关。

2.2.2 R^2 检验

插值模型的优劣程度用 R^2 检验进行评价。R^2 检验模型拟合优度，是最常用于评价回归模型优劣程度的指标，其结果表明了模型对目标真实值的决定系数。R^2 检验公式如下：

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}} = \frac{\sum_k (x_k^a - \overline{X})^2}{\sum_k (x_k^o - \overline{X})^2} \tag{16}$$

式中：ESS（explained sum of squares）为回归平方和；TSS（total sum of squares）为总离差平方和；x_k^a 为模型插值结果；x_k^o 为观测数据；$\overline{X}$ 为观测场平均值。

R^2 检验的结果在 0 ~ 1 之间，数值越大表明模型效果越好。

2.2.3 平均绝对误差

模型的插值精度用 MAE 作为评价指标。MAE 是插值结果与观测值的误差绝对值的平均值，常用来描述插值方法的误差情况，数值越小说明插值方法的误差越小，插值精度越高。对于 m 个数据，其 MAE 公式为

$$\mathrm{MAE} = \frac{1}{m}\sum_{k=1}^{m}\left|x_k^a - x_k^o\right| \tag{17}$$

2.3 BGW 插值背景场风场特征估计能力实验

为验证 BGW 插值的背景场模型对实际风场特征的估计能力，实验选择 ERA 模型 2019 年 7 月第一个星期的 50 hPa 等压面气象数据进行实验，对一周内每日 UTC 时间 8:00 的平流层数据，

计算其 BGW 插值的背景场和再分析风场的相关系数矩阵,计算时将风场数据沿纬度线剪开,每一纬度的数据作为一维特征。由于每一个完整的相关系数矩阵包涵 721 个相关系数,难以直接分析,取相关系数矩阵的中位数进行分析,结果保留 6 位有效数字,数据如表 1 所示。

表 1　相关系数矩阵中位数

Table 1　Median of correlation coefficient matrix

风速分量	7 月 1 日	7 月 2 日	7 月 3 日	7 月 4 日	7 月 5 日	7 月 6 日	7 月 7 日
西风	0.715 669	0.879 563	0.666 926	0.705 835	0.826 480	0.630 954	0.694 991
南风	0.627 580	0.708 596	0.517 900	0.453 280	0.649 182	0.630 954	0.478 832
合风	0.715 380	0.783 561	0.617 456	0.651 150	0.718 303	0.650 856	0.665 432

从表 1 中可以看出,西风、南风和合风的背景场与再分析风场的相关系数中位数都在 0.4 以上,属于中等相关和强相关范畴,而且合风的相关系数都在 0.6 以上,属于强相关,说明背景场数据能够一定程度上表达实际风场的风速分布特征。

由于表 1 只对风速的数值特征进行分析,为了更直观地分析风场的几何特征,以 ERA5 模型 2019 年 7 月 24 日 UTC 时间 8:00 的 50 hPa 等压面气象数据为例,用数据中经纬度网格 0.5°×0.5°的重力位势再分析数据作为输入计算经纬度网格 0.25°×0.25°的风场水平方向合风速的西风分量 u 和南风分量 v 背景场数据,风速数据用热图表示,西风分量和南风分量背景场风速热图分别如图 1(a)和图 1(b)所示。以 ERA5 气象再分析模型的 0.5°精度西风 u 和南风 v 数据作为实际风场进行对比,其风速热图分别如图 1(c)和图 1(d)所示。

图 1 中每幅子图的横坐标为经度,纵坐标为纬度。在同一幅西风分量图中,颜色越偏向黄色表示西风越大,颜色越偏向蓝色表示东风越大;在同一幅南风分量图中,颜色越偏向黄色表示南风越大,颜色越偏向蓝色表示北风越大。

从图 1(a)中可以发现,模型计算的背景场西风数据在南纬 60°附近有一个沿纬度圈环绕地球一周、风速 40 m/s 以上的大风速西风带,而其他区域的西风风速在沿纬度方向分布较为均匀。从图 1(b)可以发现,背景场在南纬 60°纬线圈上西经 60°和西经 140°附近有 2 处南风超过 20 m/s的集中分布区域。而从实际风场图 1(c)和图 1(d)中可以看出,背景场估计的西风分量和南风分量的大风速带和风速集中分布区域的主要特征与观测场基本一致,结合相关系数分析结果说明,用改进的地转风模型作为插值背景场能够替代风场预报数据,对风速分布特征进行估计。

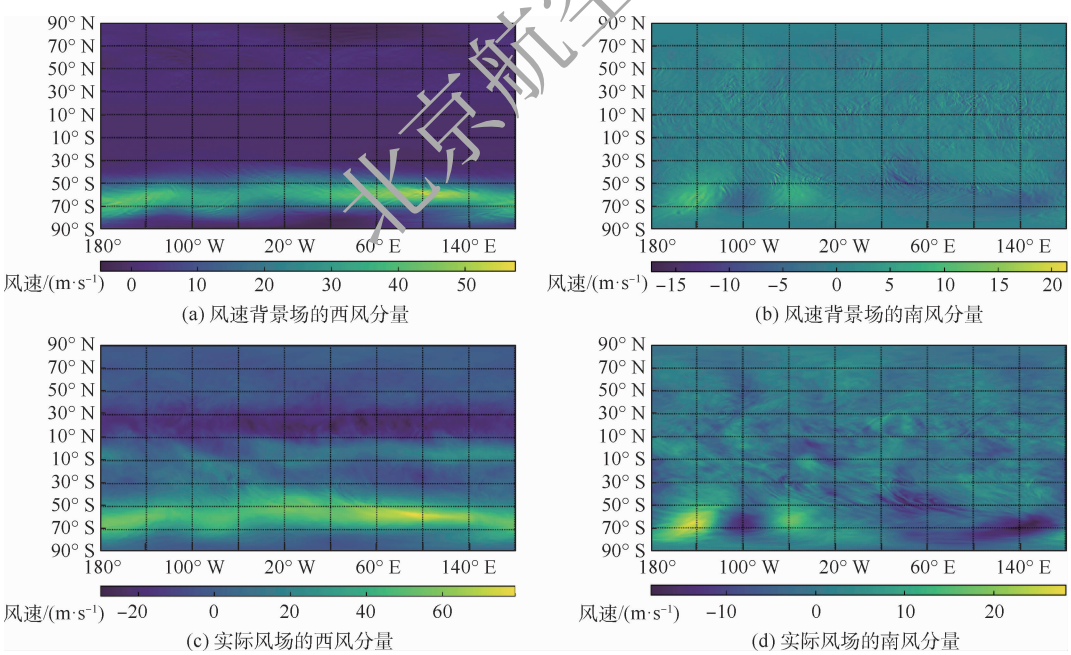


图 1　风速背景场和实际风场数据

Fig. 1　Background field and actual field of wind speed data

北航学报 赠阅

2.4 平流层风场插值实验

用搜索半径分别为 30 km、60 km 和 90 km 的 BGW 插值方法与直接插值类方法中使用最多的最近邻域插值法(nearest neighbor, NN)和反距离权重插值法(inverse distance weighted, IDW)进行对比,IDW 的幂次参数取 2,比较几种方法在 2019 年 7 月每天 24 个 50 hPa 全球风场数据上的模型插值性能,取 R^2 检验结果的均值作为评价指标,结果保留 6 位有效数字,数据如表 2 所示。

表 2 不同插值方法的 R^2 检验结果

Table 2 R^2 square test results of different interpolation methods

方法	R^2		
	西风	南风	合风
BGW(30)	0.999 937	0.997 788	0.999 770
BGW(60)	0.999 804	0.995 244	0.999 441
BGW(90)	0.999 726	0.993 489	0.999 138
NN	0.999 782	0.996 398	0.999 522
IDW	0.999 202	0.998 367	0.998 437

对于数据网格间距为 0.5°(经纬度)的输入数据,待插值点与观测点的最远平均距离约为 20 km,从表 2 的 R^2 检验结果可以发现,BGW 插值的搜索半径为 30 km 时模型性能最好,该结果验证了 1.3 节中搜索半径的最佳参数范围应该在待插值点与观测点最远距离的一倍到两倍之间的结论。之后的实验 BGW 插值默认搜索半径选择 30 km。

与 NN 和 IDW 插值方法相比,表 2 中的数据显示,BGW 插值对西风和合风的插值效果要明显优于这 2 种方法,对南风插值能力比较相近。用 MAE 数据进一步分析 3 种方法的西风、南风和合风插值效果,平均绝对误差如表 3 所示。

表 3 不同插值方法的平均绝对误差

Table 3 Mean absolute error of different interpolation methods

方法	MAE/(m·s^{-1})		
	西风	南风	合风
BGW	0.082 857	0.115 629	0.125 175
NN	0.195 772	0.238 833	0.211 241
IDW	0.398 948	0.134 958	0.429 27

从表 3 中可以发现,BGW 插值在西风、南风和合风上的 MAE 都要小于 NN 和 IDW 插值。在西风上,BGW 插值误差是 NN 插值误差的 42.32%,是 IDW 插值误差的 20.77%;在南风上,BGW 插值误差是 NN 插值误差的 48.41%,是 IDW 插值误差的 85.68%;在合风上,BGW 插值误差是 NN 插值误差的 59.26%,是 IDW 插值误差的 29.16%。

同时对 BGW 插值方法的计算速度进行实验,对于单等压面风场的平面插值问题,用 Windows 系统下的 Python 代码进行算法实现,用英特尔 7 代 i7CPU 进行计算的平均耗时为 1.37 s,用英伟达 GTX1050GPU 进行同样工作的平均耗时为 1.26 s。而且在实际的高空气球飞行仿真和外场实验中,需要对风场进行持续、实时的插值计算,GPU 在矩阵计算上的速度优势在面对大量矩阵计算时还会进一步扩大。另外,一般的融合类插值背景场采用气象预报模型,无法对每一层等压面风场数据解耦进行单独计算,而完整的预测模型需要大量的算力设备支持和天级的运算时间,BGW 插值有效解决了融合插值类方法背景场计算时间长、算力开销大的问题。

实验结果表明,相比于直接插值类方法,BGW 插值对一维的重力位势矢量进行插值,再用观测风速对背景场进行修正,误差绝对值要小于对三维的观测风速矢量进行直接插值,有效减小了插值方法的误差。而相比于融合插值类方法,BGW 插值是用当前时刻的重力位势数据结合地转风模型来计算风速的估计值作为背景场,与直接引入预报数据的方法相比,BGW 插值无需引入外部目标精度的气象预报数据,可以只通过已有的数据进行离线运算;与其他自主计算风场预报的方法相比,由于预报模型是一个高度耦合的数值系统,无法对风速的水平方向分量单独进行预测,导致其模型的运算时间和算力消耗极大,而 BGW 插值能只对单一等压面的背景场进行计算,且计算速度在秒级,有效提高了模型的计算效率。

3 结 论

面对稀疏的平流层风场数据,基于地转风模型和迭代二元线性回归提出一种针对全球平流层风场水平分量数据的 BGW 插值方法,有效提升了计算速度和精度。

1)提出了一种适用于全纬度地区地转风模型计算的科里奥利频率改进公式,同时推导出一种适用于高空气球实验且能有效表达真实风场特征的地转风快速计算方法。

2)通过迭代的二元线性回归和改进的自适应观测场偏差权重矩阵减小了观测场误差对插值结果的影响,并解决了面对不同密度的风场数据都要重新设计权重矩阵的问题。

3)实验结果表明,与 2 种直接插值类方法相比,NN 插值的平均绝对误差约为 0.21 m/s,IDW

北京航空航天大学学报

北航学报 赠阅

插值的平均绝对误差约为 0.43 m/s,BGW 插值将风速平均绝对误差减小到约 0.13 m/s,有效提升了插值精度;与融合插值类方法相比,一般使用的背景场模型需要 10 d 以上的计算时间,BGW 插值对单一等压面的风场模型的插值计算速度则提升到 1.26 s,有效提升了插值速度。

为使提出的平流层风场插值方法能对高空气球仿真和实验提供进一步支持,下一阶段工作应将目前的风场插值方法结合大气位温方程,研究时间维度上的风场插值方法。

参考文献(References)

[1] NOBUYUKI Y,TAKESHI I,NAOKI I,et al. Scientific ballooning technology and applications of exploration balloons floating in the stratosphere and the atmospheres of other planets[M]. Tokyo: Corona Publishing Co.,Ltd.,2004:21-22.

[2] HANS H,BILL B,PAUL B,et al. The ERA5 global reanalysis[J]. Quarterly Journal of the Royal Meteorological Society,2020,146(730):1999-2049.

[3] JARDIN M,ERZBERGER H. Atmospheric data acquisition and interpolation for enhanced trajectory prediction accuracy in the center:TRACON automation system[C]// Proceedings of the 34th Aerospace Sciences Meeting and Exhibit,1996:1-8.

[4] TAO T,WANG H,HU Y,et al. Error analysis of multivariate wind field simulated by interpolation-enhanced spectral representation method[J]. Journal of Engineering Mechanics,2020,146(6):04020049.

[5] 汤新民,陈平,韩云祥. 面向航空器战略航迹规划的风场信息插值方法[J]. 指挥信息系统与技术,2019,10(6):20-24.
TANG X M,CHEN P,HAN Y X. Wind field information interpolation method for aircraft strategic trajectory planning[J]. Command Information System and Technology,2019,10(6):20-24(in Chinese).

[6] 董志南,郑拴宁,赵会兵,等. 基于空间插值的风场模拟方法比较分析[J]. 地球信息科学学报,2015,17(1):37-44.
DONG Z N,ZHENG S N,ZHAO H B,et al. Comparative analysis of methods of mind field simulation based on spatial interpolation[J]. Journal of Geo-Information Science,2015,17(1):37-44(in Chinese).

[7] 李崇银,李琳,谭言科,等. 平流层气候[M]. 北京:气象出版社,2008:1-3.
LI C Y,LI L,TAN Y K,et al. Stratospheric climate[M]. Beijing:China Meteorological Press,2008:1-3(in Chinese).

[8] GANDIN L S. Objective analysis of meteorological field[M]. [S. l.]:Gidrometeorologicheskoe Izdatestvo,1963.

[9] 朱成阵,程农,李清,等. 终端区四维轨迹预测[J]. 系统仿真学报,2010,22(S1):163-165.
ZHU C Z,CHENG N,LI Q,et al. The four-dimensional trajectory prediction in terminal airspace[J]. Journal of System Simulation,2010,22(S1):163-165(in Chinese).

[10] 王文龙. 大气风场模型研究及应用[D]. 长沙:国防科学技术大学,2009:26-30.
WANG W L. Atmospheric wind field modeling and its application[D]. Changsha:National University of Defense Technology,2009:26-30(in Chinese).

[11] 伍湘君,金之雁,陈德辉,等. 新一代数值预报模式 GRAPES 的并行计算方案设计与实现[J]. 计算机研究与发展,2007,44(3):510-515.
WU X J,JIN Z Y,CHEN D H,et al. A parallel computing algorithm and its application in new generation of numerical weather prediction system (GRAPES)[J]. Journal of Computer Research and Development,2007,44(3):510-515(in Chinese).

[12] 丁一汇. 高等天气学[M]. 北京:气象出版社,2005:32-33.
DING Y H. Advanced climatology[M]. Beijing:China Meteorological Press,2005:32-33(in Chinese).

[13] 马瑞平,廖怀哲. 中国地区 20 – 80 km 高空风的一些特征[J]. 空间科学学报,1999,19(4):334-341.
MA R P,LIAO H Z. The characteristics of winds at height of 20 – 80 km in the Chinese area[J]. Chinese Journal of Space Science,1999,19(4):334-341(in Chinese).

[14] 肖琢静. 地转风和梯度风的实用计算[J]. 海洋预报,1995,12(3):78-86.
XIAO Z J. Practical calculation of geostrophic wind and gradient wind[J]. Marine Forecasts,1995,12(3):78-86(in Chinese).

[15] BABAK O,DEUTSCH C V. Statistical approach to inverse distance interpolation[J]. Stochastic Environmental Research & Risk Assessment,2009,23(5):543-553.

北京航空航天大学学报

北航学报 赠阅

A stratospheric wind field interpolation method based on geostrophic wind model

HU Kang[1,2], ZHOU Jianghua[1,2,*], ZHANG Xiaojun[1]

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
2. School of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The data accuracy and density of the stratospheric horizontal wind field have direct influence on the trajectory prediction accuracy and flight control accuracy of the high-altitude balloon experiment. Focusing on stratospheric wind field, this paper puts forward an interpolation method based on geostrophic wind model. The method improves the wind field accuracy in low-latitude area and calculation speed of the geostrophic wind model by improving the formula of Coriolis frequency and using two-dimensional convolution. Meanwhile, the method reduces the influence of observation error on interpolation result by using iterative binary linear regression and an improved adaptive bias weight matrix for the observed field. The experimental results show that the proposed interpolation method can effectively improve the calculation speed of the interpolation of stratospheric wind field and the interpolation accuracy of stratospheric horizontal wind field.

Keywords: wind field interpolation; stratospheric wind field; geostrophic wind; binary linear regression; two-dimensional convolution

北京航空航天大学学报

Received: 2020-11-09; **Accepted**: 2020-12-25; **Published online**: 2021-01-22 11:19

URL: kns.cnki.net/kcms/detail/11.2625.V.20210122.1116.001.html

Foundation items: National Natural Science Foundation of China (61733017); Strategic Priority Research Program of Chinese Academy of Sciences (XDA170400)

* **Corresponding author**. E-mail: zhoufma@aoe.ac.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0614

基于双重属性信息的跨模态行人重识别算法

全下文载

陈琳[1], 高赞[2], 宋雪萌[1], 王英龙[2], 聂礼强[1,*]

(1. 山东大学 计算机科学与技术学院, 青岛 266200;
2. 齐鲁工业大学(山东省科学院) 山东省人工智能研究院, 济南 250014)

摘 要: 通过对跨模态检索问题的研究,属性信息的使用可以增强所提取特征的语义表达性,但现有基于自然语言的跨模态行人重识别算法对行人图片和文本的属性信息利用不够充分。基于双重属性信息的跨模态行人重识别算法充分考虑了行人图片和文本描述的属性信息,构建了基于文本属性和图片属性的双重属性空间,并通过构建基于隐空间和属性空间的跨模态行人重识别端到端网络,提高了所提取图文特征的可区分性和语义表达性。跨模态行人重识别数据集 CUHK-PEDES 上的实验评估表明,所提算法的检索准确率 Top-1 达到了 56.42%,与 CMAAM 算法的 Top-1(56.68%)具有可比性,Top-5、Top-10 相比 CMAAM 算法分别提升了 0.45%、0.29%。针对待检索图片库中可能存在身份标签的应用场景,利用行人的类别信息提取属性特征,可以较大幅度提高跨模态行人图片的检索准确率,Top-1 达到 64.88%。消融实验证明了所提算法使用的文本属性和图片属性的重要性及基于双重属性空间的有效性。

关 键 词: 跨模态检索; 匹配算法; 行人属性信息; 特征表示; 特征融合
中图分类号: TP183
文献标志码: A **文章编号**: 1001-5965(2022)04-0647-10

在信息化时代,视频监控在维护公共安全方面发挥了不可取代的作用,行人重识别是视频监控场景下的一个重要子任务,其旨在不同的监控摄像头产生的图像数据中找到属于同一个行人的照片[1-3]。在一些特定的紧急场景下,如快速寻找走丢老人、孩童,或追捕逃犯时,人们无法及时提供与所寻找的行人相匹配的图片作为检索的依据,只能提供口头描述。因此,通过自然语言描述搜寻匹配的行人图片具有重要的现实意义[4]。

基于自然语言的跨模态行人重识别是指根据对行人的文本描述,在图片库中找到最符合文本描述信息的行人图片,解决这一问题的关键在于建立图片和文本之间的语义关联。研究者们已经提出了一些深度学习相关的跨模态行人重识别算法,大致可以分为计算语义亲密度值法[4-7]和子空间法[8-9]。然而,在研究中发现,该算法所提取特征的语义表达性还有待提高,其忽略了利用行人属性信息表现语义概念是否有效,或考虑的不够充分。目前,在解决这一问题中取得最好效果的跨模态属性辅助匹配(cross modal attribute aided matching, CMAAM)算法[10]利用文本描述中蕴含的语义信息作为图文属性标签辅助监督图文特征的提取过程,从而提高特征的语义表达性。然而,利用文本描述中提取的属性标签监督图片

收稿日期: 2020-11-04; **录用日期**: 2020-11-12; **网络出版时间**: 2021-03-02 17:01
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210302.1556.003.html
基金项目: 国家自然科学基金(61772310,61702300,U1936203,61872270); 国家重点研发计划(2018AAA0102502); 山东省自然科学基金(ZR2019JQ23); 山东省重大科技创新工程项目(2019JZZY010118); 山东省高等学校青年创新团队发展计划(2020KJN012); 济南市高校二十条创新团队(2018GXRC014)

* **通信作者**. E-mail: nieliqiang@gmail.com

引用格式: *陈琳, 高赞, 宋雪萌, 等. 基于双重属性信息的跨模态行人重识别算法[J]. 北京航空航天大学学报, 2022, 48(4): 647-656. CHEN L, GAO Z, SONG X M, et al. A cross-modal pedestrian Re-ID algorithm based on dual attribute information [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 647-656 (in Chinese).*

北京航空航天大学学报

特征的生成有失偏颇:①利用词频高的名词词组作为文本特征的辅助标签在一定程度上可以满足自然语言的自由度,但未经处理的文本属性标签可能存在冗余或重复的情况,会对图片特征的语义表达形成干扰,如“male”和“man”均属于词频较高的名词,但二者语义相同。②这一过程相当于行人图片属性识别,现有的行人属性重识别工作所构建的数据集的图片属性标签经过归纳整理,数目有限且满足唯一性,因此本文通过较成熟的行人属性重识别获取图片属性更为合理。本文将所采集的图文属性特征与图文特征直接融合,可以更加直接地增强特征的语义表达性。

基于自然语言描述的行人重识别的主要优势在于文本描述的“自由度”,但文本描述的过于“自由”为解决跨模态行人重识别带来了巨大挑战。Li 等[4]的实验结果表明,文本描述中的名词和形容词在图文匹配中具有更高的权重。因此,本文使用名词词组作为行人文本的属性信息控制文本描述的自由度。另外,行人图片属性通常由简单的名词词组构成,这与行人的文本属性之间可以形成语义关联。在现有的行人属性识别方法中[11-14],行人图片的属性数目有限,若仅用有限的行人图片属性作为辅助信息,会违背自然语言描述行人语义信息高自由度的特性。因此,本文考虑应用行人图片和行人文本描述两部分的属性信息,构建基于图片属性和文本属性的双重属性空间,提出了一种新的基于双重属性信息的跨模态行人重识别网络,实现了行人双重属性信息与图文特征的自适应融合,以提高图文特征的语义表达性。本文提出了一种新的基于文本属性和图片属性的双重属性空间构建和属性融合算法,对行人属性信息的应用更加充分。在公开的跨模态行人重识别数据集上的实验评估证明了该算法的有效性,同时证明了若将多视角的行人图片属性进行合并将进一步提升检索准确率。

1 相关工作

1.1 单一模态行人重识别

行人重识别自提出以来一直是计算机视觉领域的热点研究问题。传统的行人重识别仅考虑在图像数据库中检索从不同摄像头中采集到的相同身份信息(identification card,ID)的行人图片即可,使用表征学习[1,15]、度量学习[16-17]等策略就可取得较好的结果。然而,在依赖视觉特征完成行人重识别的方法中,对于穿着相似的行人很难找到其对应的图片特征之间的区分性,换句话说,单一模态的行人重识别更像是一个细粒度的图像检索问题。针对该问题,Yin 等[18]发布了细粒度的行人重识别数据集,将行人图片按照颜色先进行分类,在此基础上学习更加精细的图像特征。

行人重识别的实际应用场景复杂,有很多现实问题需要解决,如行人的姿态会随时发生变化,行人在行走的过程中可能会换衣服,可能会有遮挡情况出现等。这些应用上的困难催生了一系列行人重识别的研究子方向。为了解决遮挡问题,Gao 等[19]提出了 DCR 模型,将被遮挡后部分图片的分块由未被遮挡的整张图片的分块进行协同表示。为了解决换衣问题,Zheng 等[20]提出了生成行人换衣的图片来做辅助特征。Yang 等[21]考虑在短时间内行人穿着的厚度没有大的变化的情况下,通过处理行人图像的轮廓草图而不是通过衣服的颜色信息来提取视觉特征。

1.2 基于自然语言的跨模态行人重识别

现实生活中,在一些特定的紧急场景下,人们无法及时提供与所寻找的行人相匹配的图片作为检索的依据,只能提供口头描述,因此,常常需要实现基于自然语言描述的跨模态行人重识别。解决行人图片-文本跨模态检索问题的关键在于找到行人图文特征表达之间的语义关联。目前,研究者们已经提出了一些跨模态行人重识别算法,大致分为:①基于图片文本匹配亲密度值的跨模态行人重识别算法。例如,Li 等[4]率先提出了基于自然语言描述的行人检索问题,且提出了基于门控神经注意力机制的循环神经网络(recurrent neural network with gated neural attention,GNA-RNN)来获取行人图片和行人文本描述的亲密度,致力于提高同属于一个行人的图片和文本的亲密度值,然而该工作主要有 2 点不足:一是 GNA-RNN 考虑的是文本和整张图片之间的语义亲密度,对行人关键属性的具体空间分布不够敏感,二是在计算不同单词跟图片的匹配度时,没有设置阈值判断单词是否与图片存在关联。为了解决以上问题,Chen 等[6]提出了基于单词跟图片的局部区域之间的匹配模型,并设计了自适应阈值机制(adaptive threshold mechanism,ATH)确保匹配同一个单词的图片跟单词的亲密度值比较相近,在最终计算整个文本与图片的亲密度时,相关单词的贡献都比较公平。Chen 等[7]考虑到了文本与整张图片之间的全局联系,以及文本中所含的名词词组与图片的局部区域之间的联系,在局部联系的设计中,还使用名词词组加权重建了图

片的重点区域，且约束重建后的图片特征保持与输入词汇的一致性。②基于共同向量空间的跨模态行人重识别算法，构建共同子空间是跨模态检索中的常用方法，使用对抗学习机制是构建有效跨模态共享子空间的技巧[22]。特征映射器和模态分类器的相互对抗可以使得不同模态的数据分布差异变小，同时保持特征的可分性。Wang[8]和Zhang[9]等将特征映射到共同子空间后，使用相对熵（KLD）建立了图片特征和文本特征的联系。③基于辅助信息的跨模态行人重识别算法，为了获得更好的检索结果，也有一些算法利用了辅助信息（如行人骨架信息、属性信息等）增强特征的语义表达能力。例如，Jing 等[23]在跨模态行人重识别中使用了行人的姿态估计提取图片中与行人相关的区域，使得图片特征更加精确地描述行人；Aggarwal 等[10]使用从文本中提取出的词组作为文本的属性信息构建了属性空间增强特征的语义表达能力，且取得了非常优越的性能。然而，在以上算法中，对于属性信息的应用还不够充分，它们要么忽略属性信息的重要性，要么仅使用文本信息中提取的属性，缺乏对两者潜在关联的挖掘。本文提出的基于隐空间和属性空间的跨模态行人重识别端到端网络利用了行人文本描述和图片的双重属性，可以增强所提取特征的语义表达性。

2　模型设计

2.1　行人双重属性信息提取

2.1.1　行人文本属性提取

在文本描述中，对行人重识别起到决定性作用的通常是名词词组，每个行人都会由不同的名词词组描述，出现频率高的名词词组说明其可以同时描述多个行人，这些行人在语义上具有共性，因此，在本文中考虑将高频率的名词词组作为文本描述中蕴含的属性信息。如图 1 所示，为了获得名词词组，本文使用了自然语言处理工具包（NLTK）[24]，对数据集中的文本数据进行预处理，提取了形容词加名词、多个名词叠加这 2 种格式的名词词组，并将提取结果按照词频进行排序，低频词组被舍弃，仅保留排名前 400 的名词词组组成了属性表。每个行人的文本描述都可以与文本属性表进行对比，若含有属性表中的某一项，则该行人的属性表的该项就置为 1，否则置为 0，这样就获得了一个 400 维的文本属性特征向量 $\boldsymbol{c}^{T_k}$，形成了行人图片的文本描述与文本属性之间的映射。

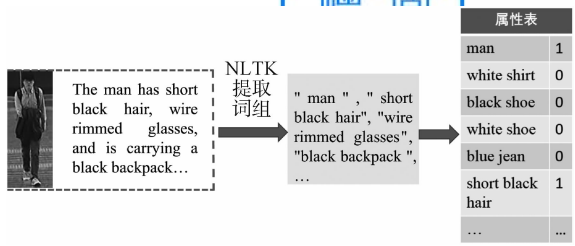


图 1　文本属性提取

Fig. 1　Text attribute extraction

2.1.2　行人图片属性提取

针对行人图片属性识别已有许多工作，且都取得了不错的效果，在已公开代码和模型的工作中，本文选择使用 Li 等[11]发布的行人属性识别模型，提取了数据集中行人图片所含的属性及可能性值。具体地，本文采用了在 PA-100K 数据集[25]上训练的属性识别模型，因此生成的属性标识一共有 26 种。根据属性预测可能性值的大小进行划分，可能性大于 0 的属性标为 1，表示该行人图片能反映出该属性所表示的语义，反之标为 0，表示该行人图片并不含有该属性所表示的语义。因此，行人图片属性特征向量 $\boldsymbol{c}^{I_k}$ 由 26 维特征向量表示。

将单独提取一张行人图片属性或文本描述属性的过程称为单元属性提取。此外，如图 2 所示，在某一时刻，同一行人在不同角度所呈现的图片属性可能是不同的，在一段固定时间内，关于行人的文本描述大多是固定且含多角度的。本文尝试将属于同一行人的图片属性或文本属性进行归并

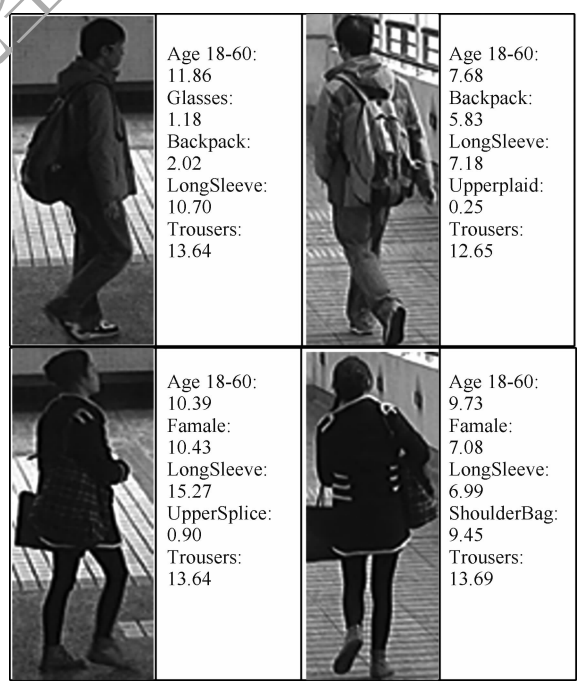


图 2　图片属性识别结果

Fig. 2　Result of image attribute recognition

北京航空航天大学学报

处理,从而使得属性特征具有多角度的丰富信息,并将这一过程称为多元属性提取。

2.2 双重属性信息模型的主要结构

本文所提出的模型如图 3 所示,主要由中间部分的隐空间网络和上下两部分组成的属性空间网络组成。具体地,在文本特征提取时,先对文本信息进行预处理,即将描述语句分词之后建立词频表,过滤掉低频词语,将单词使用独热编码方式编码,再采用可以充分考虑每个单词上下文的双向长短期记忆网络(bi-directional long-short term memory, Bi-LSTM)提取文本特征,使得学习到的文本特征更加丰富。为了降低文本属性与文本特征在融合过程中由于维度变化带来的信息损失,将 Bi-LSTM2 的维度设置为文本属性的维度,本文使用了 2 个 Bi-LSTM 来分别提取隐空间和属性空间中的文本特征。在图片特征提取时,使用了在 ImageNet 上做了预训练的 ResNet[26]。第 k 个行人图片经过 ResNet 后输出 2 048 维的图片特征 $\boldsymbol{e}^{I_k}$。为了更好地表达图片语义,还将行人图片作了语义分割[27],在属性空间用相同的 ResNet 模型进行特征提取,获得2 048维的图片特征 $\boldsymbol{e}^{I_k}$。

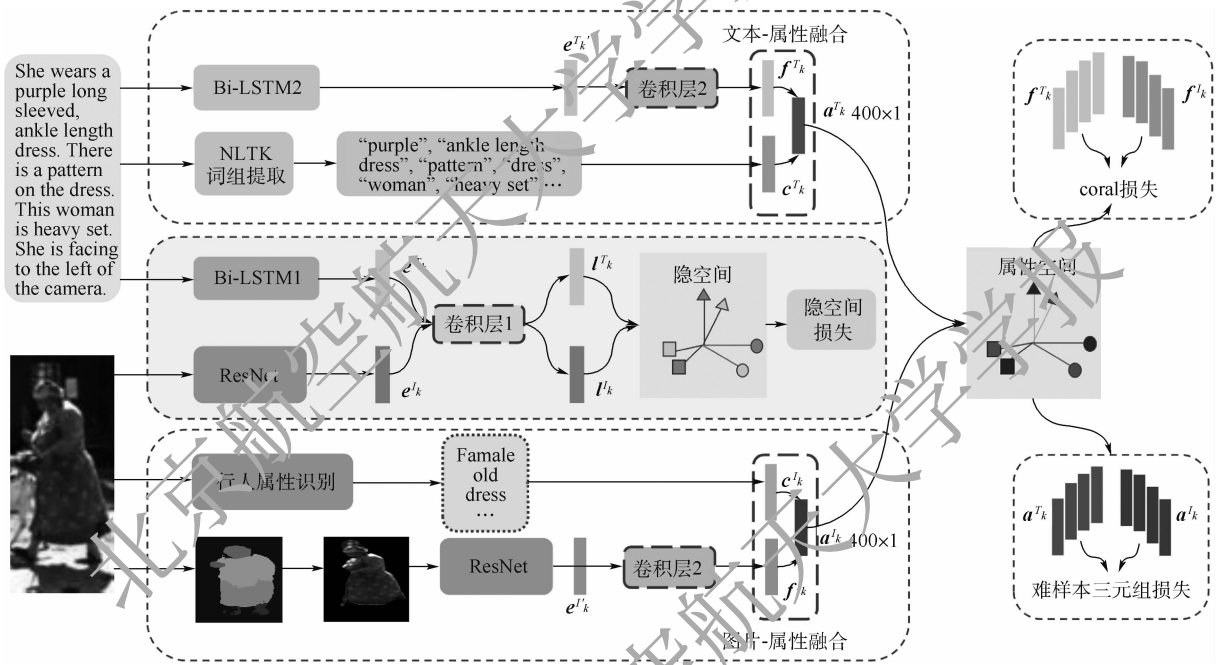


图 3 基于双重属性信息的跨模态行人重识别算法框架

Fig. 3 Framework of cross-modal matching algorithm based on dual attribute information

此外,图片特征 $\boldsymbol{e}^{I_k}$ 和文本特征 $\boldsymbol{e}^{T_k}$ 被送入隐空间前,先经过了卷积层 1,确保使得 2 种特征具有相同维度 dl,这也可以在一定程度上拉近 2 种模态数据的分布。经过卷积层后的图片特征用 $\boldsymbol{l}^{I_k}$ 表示,文本特征用 $\boldsymbol{l}^{T_k}$ 表示,$\boldsymbol{l}^{I_k}$ 和 $\boldsymbol{l}^{T_k}$ 被送入隐空间中。同理,图片特征 $\boldsymbol{e}^{I_k}$ 和文本特征 $\boldsymbol{e}^{T_k}$ 在与属性特征融合前,先经过了卷积层 2,使得 2 种特征具有相同维度 da,经过卷积层的 2 种特征分别用 $\boldsymbol{f}^{I_k}$ 和 $\boldsymbol{f}^{T_k}$ 表示,图片特征 $\boldsymbol{f}^{I_k}$ 与图片属性 $\boldsymbol{c}^{I_k}$ 融合得到的特征用 $\boldsymbol{a}^{I_k}$ 表示,文本特征 $\boldsymbol{f}^{T_k}$ 与文本属性 $\boldsymbol{c}^{T_k}$ 融合得到的特征用 $\boldsymbol{a}^{T_k}$ 表示,$\boldsymbol{a}^{I_k}$ 和 $\boldsymbol{a}^{T_k}$ 被送入属性空间中进行相关计算。

2.3 隐 空 间

隐空间的设置是为了使行人的图片特征和文本特征都具有行人 ID 的可分性及图文特征间具有基本的语义关联。在基于自然语言的跨模态行人重识别中,同一个行人对应着多张图片和多个相应的文本描述,损失函数的设计目标在于拉近同属于一个行人 ID 的图文特征之间的距离,拉远不属于同一个行人 ID 的图文特征的距离。本文使用了难样本三元组损失[17]函数,具体地,令其中一种模态中的数据作为锚,取另一种模态中与锚属于同一类的数据作为正样本,与锚属于不同类的数据作为负样本(难样本指的是与锚样本距离最大的正样本和与锚样本距离最小的负样本)。这样不仅起到了分类的作用,还在一定程度上使不同模态中具有相同语义的图片和文本描述之间形成对应,本文默认同一个行人图片和文本描述具有相同的语义信息。例如,给定一批数据,难样本三元组损失函数 L_{trip} 为

$$L_{\text{trip}}(\boldsymbol{I},\boldsymbol{T}) = \sum_{\boldsymbol{I}_k \in \boldsymbol{I}} [\Delta_1 + D(\boldsymbol{I}_k, \tilde{\boldsymbol{T}}_k^{p}) - D(\boldsymbol{I}_k, \tilde{\boldsymbol{T}}_k^{n})]_+ + \sum_{\boldsymbol{T}_k \in \boldsymbol{T}} [\Delta_1 + D(\boldsymbol{T}_k, \tilde{\boldsymbol{I}}_k^{p}) - D(\boldsymbol{T}_k, \tilde{\boldsymbol{I}}_k^{n})]_+ \quad (1)$$

$$D(\boldsymbol{I}_j,\boldsymbol{T}_k) = 1 - S_{\cos}(\boldsymbol{l}^{I_j},\boldsymbol{l}^{T_k}) \tag{2}$$

式中：$\boldsymbol{I}_k$ 为第 k 个图片的特征，且将其作为锚；$\tilde{\boldsymbol{T}}_k^n$ 为与其距离最近的异类文本特征；$\tilde{\boldsymbol{T}}_k^p$ 为与其距离最远的同类文本特征；使用第 k 个文本特征 $\boldsymbol{T}_k$ 做锚；用 $\tilde{\boldsymbol{I}}_k^n$ 表示与之相距最近的异类图片特征；$\tilde{\boldsymbol{I}}_k^p$ 表示相距最远的同类图片特征；$S_{\cos}$ 为计算图文特征之间的余弦相似度；D 为隐空间图文特征间的余弦距离；Δ_1 为三元组损失的边界。

为使隐空间中的特征更好地保持行人 ID 的可分性，本文定义了一个基于行人 ID 的分类器 $\boldsymbol{W}\in\boldsymbol{R}^{\mathrm{dl}\times C}$，$C$ 为行人 ID 的类别数目，dl 表示隐空间特征的维度，2 种模态的训练数据共用同一个分类器，使用 Softmax 损失函数得到的隐空间分类损失 L_{cls} 定义为

$$L_{\mathrm{cls}}(\boldsymbol{l}_i) = -\ln\left(\frac{\exp(\boldsymbol{l}_i^T\widehat{\boldsymbol{W}}_{y_i})}{\sum_{j=1}^{C}\exp(\boldsymbol{l}_i^T\widehat{\boldsymbol{W}}_j)}\right) \tag{3}$$

式中：$\boldsymbol{l}_i$ 为隐空间中的图文特征；y_i 为 $\boldsymbol{l}_i$ 的类别；$\widehat{\boldsymbol{W}}_j$ 为经过了正则化的 $\boldsymbol{W}$。

总之，隐空间的损失函数 $L_{\mathrm{latent}}(\boldsymbol{I},\boldsymbol{T})$ 表示为

$$L_{\mathrm{latent}}(\boldsymbol{I},\boldsymbol{T}) = \frac{1}{n}L_{\mathrm{trip}}(\boldsymbol{I},\boldsymbol{T}) + \frac{1}{n}\sum_i(\boldsymbol{L}_{\mathrm{cls}}(\boldsymbol{I}_i) + L_{\mathrm{cls}}(\boldsymbol{T}_i)) \tag{4}$$

式中：n 为一批数据中所含样本数；$L_{\mathrm{cls}}(\boldsymbol{I}_i)$ 为隐空间特征 $\boldsymbol{I}_i$ 产生的分类损失；$L_{\mathrm{cls}}(\boldsymbol{T}_i)$ 为 $\boldsymbol{T}_i$ 的分类损失。

2.4　属性空间

2.4.1　消除模态差异

设置属性空间的目的在于使得属性空间中学习到的特征具有更精准的语义概念。在文本特征和图片特征分别与属性特征融合之前，需要保证文本和图片 2 个模态的样本特征的空间分布没有较大的模态鸿沟，消除模态差异对于特征表达的影响，才能更好地学习融合属性特征的图片和文本特征之间的匹配。因此，采用迁移学习中的 coral 损失[28]消除图像特征和文本特征之间的模态差异。假定一批数据 $\beta=(\boldsymbol{I},\boldsymbol{T})$，$\boldsymbol{I}$ 由 $\boldsymbol{f}^{I_k}$ 组成，$\boldsymbol{T}$ 由 $\boldsymbol{f}^{T_k}$ 组成，其函数 L_{coral} 定义如下：

$$L_{\mathrm{coral}} = \frac{1}{4|\boldsymbol{v}|^2}\|\boldsymbol{C}_I - \boldsymbol{C}_T\|_{\mathrm{F}}^2 \tag{5}$$

式中：$\boldsymbol{C}_I$ 和 $\boldsymbol{C}_T$ 分别为图片特征 $\boldsymbol{I}$ 和文本特征 $\boldsymbol{T}$ 的协方差矩阵；$|\boldsymbol{v}|$ 为特征 $\boldsymbol{f}^{I_k}$ 和 $\boldsymbol{f}^{T_k}$ 的维度；$\|*\|_{\mathrm{F}}$ 表示 Frobenius 范数。

2.4.2　属性-特征融合模块

考虑到不同的属性在图文匹配时发挥的作用程度可能不同，因此，本文使用权重机制[29-30]促使特征融合时起关键作用的语义信息发挥更大的作用。具体地，假定一批数据 $\beta=(\boldsymbol{I},\boldsymbol{T})$，用 $\boldsymbol{C}$ 表示待融合的属性特征，由 $\boldsymbol{c}^{I_k}$ 或 $\boldsymbol{c}^{T_k}$ 组成，$\boldsymbol{F}$ 表示待融合的图片或文本特征，由 $\boldsymbol{f}^{I_k}$ 或 $\boldsymbol{f}^{T_k}$ 组成。其融合过程为

$$\tilde{\boldsymbol{C}} = \|\boldsymbol{C}\times\boldsymbol{W}_{\mathrm{g}}\|_2 \tag{6}$$

$$\tilde{\boldsymbol{F}} = \|\boldsymbol{F}\times\boldsymbol{W}_{\mathrm{f}}\|_2 \tag{7}$$

$$\boldsymbol{t} = \sigma(\boldsymbol{C}\times\boldsymbol{U}_{\mathrm{g}} + \boldsymbol{F}\times\boldsymbol{U}_{\mathrm{f}}) \tag{8}$$

$$\boldsymbol{A} = \boldsymbol{t}*\tilde{\boldsymbol{C}} + (1-\boldsymbol{t})*\tilde{\boldsymbol{F}} \tag{9}$$

式中：σ 表示 sigmoid 函数；$\boldsymbol{W}_{\mathrm{g}}\in\boldsymbol{R}^{s\times\mathrm{da}}$ 和 $\boldsymbol{W}_{\mathrm{f}}\in\boldsymbol{R}^{\mathrm{da}\times\mathrm{da}}$ 分别为属性特征和图文特征的投影矩阵，且对投影结果进行了 L_2 正则化；s 为图片属性或者文本属性的类别数，da 表示属性空间的特征维度；$\boldsymbol{U}_{\mathrm{g}}\in\boldsymbol{R}^{s\times\mathrm{da}}$ 和 $\boldsymbol{U}_{\mathrm{f}}\in\boldsymbol{R}^{\mathrm{da}\times\mathrm{da}}$ 分别为属性特征和图文特征的投影矩阵；投影结果相加后经过 sigmoid 函数得到特征融合时的权重 $\boldsymbol{t}$；$\boldsymbol{A}$ 为融合后的特征，由 $\boldsymbol{a}^{I_k}$ 或 $\boldsymbol{a}^{T_k}$ 组成。

2.4.3　属性空间损失函数

属性空间中，对于融合了语义信息的特征，同样使用难样本三元组损失使不同模态的特征之间建立语义联系。使用 $\boldsymbol{a}^{I_k}$、$\boldsymbol{a}^{T_k}$ 表示属性空间中融合属性信息后的具有语义信息的图文特征。属性空间的三元组损失 $\boldsymbol{L}_{\mathrm{a\text{-}trip}}$ 定义为

$$L_{\mathrm{a\text{-}trip}}(\boldsymbol{I},\boldsymbol{T}) = \sum_{I_k\in\boldsymbol{I}}[\Delta_2 + D_{\mathrm{a}}(\boldsymbol{I}_k^s,\tilde{\boldsymbol{T}}_k^{sp}) - D_{\mathrm{a}}(\boldsymbol{I}_k^s,\tilde{\boldsymbol{T}}_k^{sn})]_+ + \sum_{T_k\in\boldsymbol{T}}[\Delta_2 + D_{\mathrm{a}}(\boldsymbol{T}_k^s,\tilde{\boldsymbol{I}}_k^{sp}) - D_{\mathrm{a}}(\boldsymbol{T}_k^s,\tilde{\boldsymbol{I}}_k^{sn})]_+ \tag{10}$$

$$D_{\mathrm{a}}(\boldsymbol{I}_j,\boldsymbol{T}_k) = 1 - S_{\cos}(\boldsymbol{a}^{I_j},\boldsymbol{a}^{T_k}) \tag{11}$$

式中：Δ_2 为三元组损失边界；D_{a} 为图文特征之间的余弦距离；其他参数含义与式(1)和式(2)中的含义类似。

总之，属性空间中的损失函数 L_{attr} 定义为

$$L_{\mathrm{attr}}(\boldsymbol{I},\boldsymbol{T}) = \frac{1}{n}L_{\mathrm{a\text{-}trip}}(\boldsymbol{I},\boldsymbol{T}) + \boldsymbol{L}_{\mathrm{coral}} \tag{12}$$

式中：n 为一批数据所含样本数。

在模型学习的过程中，隐空间和属性空间的训练同时进行，双重属性空间的损失函数最终为

$$L(\boldsymbol{I},\boldsymbol{T}) = L_{\mathrm{latent}}(\boldsymbol{I},\boldsymbol{T}) + L_{\mathrm{attr}}(\boldsymbol{I},\boldsymbol{T}) \tag{13}$$

在测试过程中，为使隐空间中学到的行人身份信息和属性空间中学到的行人语义信息起到互补作用，将测试阶段的相似度 $S(\boldsymbol{I}_j,\boldsymbol{T}_k)$ 定义为

$$S(\boldsymbol{I}_j,\boldsymbol{T}_k) = S_l(\boldsymbol{l}^{I_j},\boldsymbol{l}^{T_k}) + S_{\mathrm{a}}(\boldsymbol{a}^{I_j},\boldsymbol{a}^{T_k}) \tag{14}$$

式中：S_l 为由隐空间中学到的特征 $\boldsymbol{l}^{I_j}$、$\boldsymbol{l}^{T_k}$ 所计算出的相似度；S_{a} 为由属性空间中学到的特征 $\boldsymbol{a}^{I_j}$、$\boldsymbol{a}^{T_k}$

北京航空航天大学学报

所计算的相似度。

3 实验评估与讨论

3.1 实验设置

3.1.1 数据集

本文实验采用的数据集是 CUHK-PEDES[4]，其是目前基于自然语言的跨模态行人重识别任务仅有的一个数据集。从 5 个不同的行人重识别数据集中收集了 13 003 个行人的图片，每张行人图片由 2 句话描述，这 2 个句子由 2 个独立的标注者标注，每个句子最少 15 个单词。下文实验沿用了该数据集的原始划分，11 003 个行人数据用于训练，1 000个行人数据用于交叉验证，1 000 个行人数据用于测试。

3.1.2 参数设置

实验中设置隐空间的特征维度 d_l 为 512，属性空间的特征维度 d_a 为 400，训练的迭代次数为 50，一批数据的大小设置为 16。实验中使用了 Adam 优化器[31]，学习率设为 0.000 2。超参数 Δ_1、Δ_2 均设为 0.3。

3.1.3 评价指标

本文使用余弦值度量文本特征和图片特征之间的相似度，且使用 Top-k(k = 1,5,10)作为评价指标，Top-k 表示输入文本描述后返回的图片检索结果中前 k 张图片中出现与文本描述属于同一个行人 ID 的概率。

3.2 定量分析

3.2.1 与现有模型的准确率对比

在定量分析实验中，先在 CUHK-PEDES 数据集上对本文算法进行评估，将其与现有算法进行比较，结果如表 1 所示。

表 1 在 CUHK-PEDES 数据集上的性能评估及与现有算法的性能比较

Table 1 Performance evaluation on CUHK-PEDES dataset and comparison of performance with existing algorithms %

算法		Top-1	Top-5	Top-10
GNA-RNN[4]		19.05		53.64
IATV[32]		25.94		60.48
PWM + ATH[6]		27.14	49.45	61.02
DPCE[33]		44.40	66.26	75.07
GLA[7]		43.58	66.93	76.26
CMPC + CMPM[9]		49.37		79.27
MCCL[8]		50.58		79.06
GALM[23]		54.12	75.45	82.97
CMAAM[10]		56.68	77.18	84.86
本文	无预处理	56.42	77.63	85.15
	预处理	64.88	85.02	90.55

从表 1 中可以看出，本文提出的使用双重属性信息的跨模态行人检索模型在未对数据进行预处理的情况下与取得最好结果的 CMAAM 算法相比，在检索准确率上依然具有竞争力，Top-5、Top-10 分别提升了 0.45%、0.29%，这也可以证明本文所提出模型的合理性。其他算法大多采用计算图文匹配亲密度值法或共享子空间法建立图片和文本之间的语义关联，忽略了属性信息在加强语义概念方面起到的辅助作用。取得最好匹配效果的 CMAAM 算法也只考虑到了文本信息中所蕴含的属性信息。本文算法充分考虑了图片和文本中所蕴含的属性信息，通过构建属性空间使得融合了语义限制的图片和文本特征之间形成对应，语义信息的补充使得本文提出的模型在检索准确率 Top-5、Top-10 上取得了更好的结果。

此外，考虑到待检索的图片库可能存在提供人员身份信息的应用场景，将训练集和测试集中的图片和文本进行了预处理，将同属于一个行人的图片属性和文本属性分别进行归并处理，同一个行人的图片和文本的属性信息分别保持一致，实验结果表示，利用归并的属性信息可以进一步提升检索准确率，其中 Top-1 可以达到 64.88%。

3.2.2 双重属性空间的重要性分析

本文提出的模型在属性空间和隐空间中进行了联合训练，为了分析双重属性信息的重要性，在计算文本和图像的相似度时，仅使用隐空间特征或双重属性空间特征得到的检索结果分别用 R_{latent} 和 R_{attr} 表示。通过式(14)将 2 个空间得到的特征相似度矩阵对应相加，最终得出的检索结果用 R_{total} 表示。

从表 2 中可以看出，若不利用属性空间中的特征提供语义信息，仅用隐空间中学习到的特征来作行人图文检索，准确率 R_{latent} 与以往算法相比没有得到明显提升。由于设置属性空间的目的在于使得图文特征具备更精准的语义概念，语义相同但不属于同一个行人的情况比较普遍，因此，仅仅利用属性空间中的特征完成行人图文检索得到的准确率结果 R_{attr} 也不理想。将隐空间和属性空间训练得出的隐空间特征和属性空间特征共同参与检索过程，分析表 2 中的结果可知，具备行人 ID 可分性的隐空间特征和具备语义信息限制的属性空间特征形成了互补，使得准确率 R_{total} 有了较大幅度的提升。此外，隐空间与属性空间联合训练过程中损失函数的收敛过程如图 4 所示。结果表明，联合训练下，各个空间中损失函数的收敛趋势大致相同，且都能够快速的收敛。

北航学报 网阅

表 2　双重属性空间和隐空间的重要性分析

Table 2　Analysis of importance of dual attribute space and latent space　%

对比策略	准确率类型	Top-1	Top-5	Top-10
单元属性提取	R_{latent}	54.34	75.42	82.94
	R_{attr}	39.26	65.50	75.29
	R_{total}	56.42	77.63	85.15
多元属性提取	R_{latent}	54.35	75.88	83.14
	R_{attr}	52.65	77.86	85.98
	R_{total}	64.88	85.02	90.55

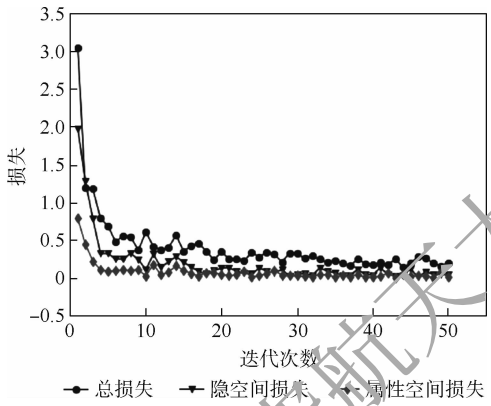


图 4　训练过程中损失函数变化情况

Fig. 4　Change of loss function during training

实验过程中，若待检索图片库中可以提供图片的行人类别信息，根据行人的类别信息提取并整合行人图片属性和文本属性，即经过多元属性提取使得属于同一个行人的图片具有相同的属性，属于同一个行人的文本具有相同的属性，检索准确率会得到大幅度提升。其中，R_{attr} 的提升尤为明显，R_{total} 的 Top-1 较仅在单一图片和文本中提取属性分别提升了 8.46%，Top-5、Top-10 分别提升了 7.39%、5.40%。

3.2.3　不同属性信息的重要性分析

为对不同属性信息的重要性进行进一步的分析，在图 2 的网络结构中，在属性特征与图文特征融合时，分别将图像属性和文本属性删除，未融合属性信息的图片特征或文本特征被送入属性空间中，并与隐空间进行联合训练，其结果如表 3 所示。其中，“无图片属性”表示图 2 中最下方的图像属性未嵌入到网络，而“无文本属性”表示图 2 中最上方的文本属性未嵌入到网络中。

从表 3 可以看出，仅使用 1 种属性信息对于最终的检索结果 R_{total} 都会有负面影响，检索结果相较同时使用 2 种属性信息都有不同程度的下降，其中不使用文本属性对于结果的影响更大。无文本属性相比无图片属性模型的检索准确率 Top-1 下降了约 3.36%，Top-5、Top-10 分别下降了 4.34%、3.11%。因为文本属性中所包含的属性信息更加丰富，缺乏文本属性中丰富的语义信息会使属性空间中特征的语义表达性降低。仅使用其中一种属性对最终的检索结果 R_{total} 影响较大，如属性空间中的文本特征若没有融合文本属性特征，相比使用双重属性构造的属性空间，R_{total} 的 Top-1 下降了 4.97%，Top-5、Top-10 分别下降了 4.48%、3.59%。对这一结果给出的解释是：本文所提出的属性空间中使用的损失函数是难样本三元组损失，跨模态三元组损失根据行人 ID 及样本模态选定正负样本，仅使用一种模态数据的属性做辅助信息增强特征的语义表达，如果没有另一模态的属性与之相对应，某一模态的特征融合属性信息相当于在原有特征基础上添加了冗余信息，因此，不能很好地发挥作用。

表 3　文本属性和图像属性的重要性分析

Table 3　Analysis of importance of text attribute and image attribute　%

对比策略	准确率类型	Top-1	Top-5	Top-10
无图片属性	R_{latent}	53.20	74.77	82.47
	R_{attr}	40.07	66.81	76.69
	R_{total}	54.81	77.49	84.67
无文本属性	R_{latent}	53.41	75.00	82.88
	R_{attr}	37.35	62.13	72.38
	R_{total}	51.45	73.15	81.56

3.2.4　模态差异消除的作用评估

输入到属性空间中的图文特征在融合属性信息之前，会先通过 coral 损失拉近 2 种模态的数据之间的特征分布。为了衡量模态差异消除的作用，实验中，在图 2 的网络结构中，分别删除或嵌入该模块，从而评估其重要性，实验结果如表 4 所示。可以看到，相比未使用 coral 损失，使用 coral 损失后模型的检索准确率在 Top-1、Top-5、Top-10 上分别提升了约 12.17%、17.01%、17.72%。因此，coral 损失的应用对准确率的提升起到了正向作用。

表 4　模态差异消除作用

Table 4　Role of modal difference elimination　%

对比策略	准确率类型	Top-1	Top-5	Top-10
无 coral 损失	R_{total}	44.25	60.62	67.43
有 coral 损失	R_{total}	56.42	77.63	85.15

3.3　可视化分析

为了进一步评估本文算法，在实验中对检索结果进行了可视化分析，结果如图 5 所示，其中左侧为文本信息，右侧为对应的图像检索结果，叉号标注的图片表示错误检索结果，无叉号标注的为正确检索。

可以看出，第 1 行和第 2 行结果表现比较好，

北京航空航天大学学报

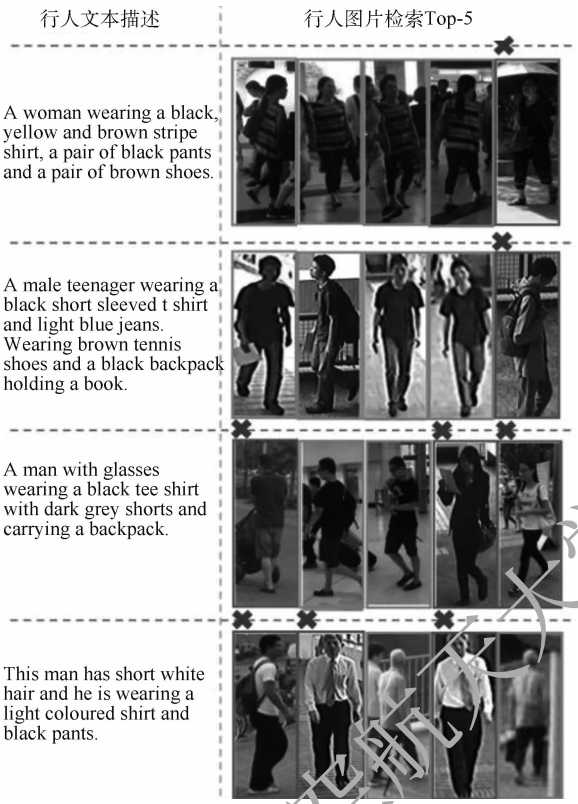


图 5 本文算法解决跨模态行人重识别的 Top-5 结果示例

Fig. 5 Examples of Top-5 results for cross-modal pedestrian re-identification by proposed algorithm

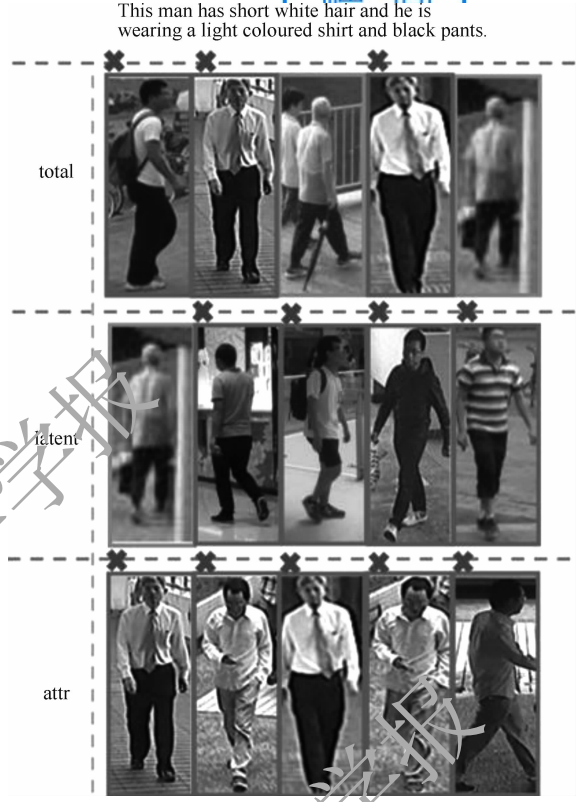


图 6 属性空间、隐空间及其融合后的检索结果

Fig. 6 Retrieval results using attribute space and latent space and after fusion

分析内容可知，第 1 行图文数据都体现出了“yellow and brown stripe”这种在数据集中少见的语义概念，第 2 行的图文数据中出现了“book”、“backpack”等语义概念，可见一些出现频率较低且具有标识性的语义对于提升检索结果有帮助。而第 3 行和第 4 行中的图文数据所体现出的语义概念比较常见，因此在检索时不免会发生混淆，但是第 3 行和第 4 行的检索结果也反映了本文所使用的属性辅助信息的作用，如第 3 行中前 3 张图片均含有语义“man”、“black tee shirt”，第 5 张图片也含有语义“glasses”、“backpack”，第 4 行中 5 张图片均含有语义“light coloured shirt”、“black pants”、“man”，而后 4 张图片也均体现了“short white hair”的语义。

图 6 进一步展示了属性空间和隐空间共同作用的效果，其中，attr 表示仅利用属性空间的结果，latent 表示仅利用隐空间的结果，total 表示两者的融合结果。可以看出，隐空间中的特征主要以行人 ID 为导向，检索到的图片统一的语义概念比较杂乱，但属性空间中的检索结果则都含有语义“man”、“light coloured shirt”，排名靠前的图片还能体现出语义“short white hair”。隐空间和属性空间同时作用时，由于属性空间可以提供更精准的语义信息，在 Top-5 中能够找到 2 张所期待的图像。

总之，经过可视化分析可以看出，本文提出的模型体现了属性作为语义概念在图文检索时能够起到辅助作用。

4 结 论

1）针对现有算法忽略了行人图片属性及其与文本属性的潜在关联的问题，本文提出了一种新的基于双重属性信息的跨模态行人重识别端到端网络。

2）在所提网络中，充分考虑了行人图片和行人文本描述的属性信息，构建了基于文本属性和图像属性的双重属性空间，将属性信息直接作用于图文特征以增强其语义表达性。

3）通过在公开的跨模态行人重识别数据集 CUHK-PEDES 上的实验评估表明，属性空间对跨模态行人检索起到了积极作用，跨模态行人重识别的准确率 Top-1 达到了 56.42%，Top-5、Top-10 相比 CMAAM 算法分别提升了 0.45%、0.29%。同时，针对可能存在身份标签的应用场景，属性的归并处理会对结果的提升起到较大的积极作用，检索准确率 Top-1 可以达到 64.88%。

在未来的工作中，将研究如何降低文本属性的冗余，在保证自然语言描述自由度的基础上使

得提取到的文本属性更加精炼。此外，在模型定性分析中所提出的，如何进一步利用低频却有效的属性信息也需要进一步的研究。

参考文献（References）

[1] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and future[EB/OL]. (2016-10-11)[2020-10-30]. http://arxiv.org/abs/1610.02984.

[2] 罗浩，姜伟，范星，等. 基于深度学习的行人重识别研究进展[J]. 自动化学报，2019，45(11)：2032-2049.
LUO H, JIANG W, FAN X, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049(in Chinese).

[3] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook. (2020-01-13)[2020-10-30]. https://arxiv.org/abs/2001.04193v1.

[4] LI S, XIAO T, LI H, et al. Person search with natural language description[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1970-1979.

[5] JI G, LI S J, PANG Y. Fusion-attention network for person search with free-form natural language[J]. Pattern Recognition Letters, 2018, 116: 205-211.

[6] CHEN T, XU C, LUO J. Improving text-based person search by spatial matching and adaptive threshold[C]// Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2018: 1879-1887.

[7] CHEN D, LI H, LIU X, et al. Improving deep visual representation for person re-identification by global and local image-language association[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 56-73.

[8] WANG Y, BO C, WANG D, et al. Language person search with mutually connected classification loss[C]// Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2019: 2057-2061.

[9] ZHANG Y, LU H. Deep cross-modal projection learning for image-text matching[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 707-723.

[10] AGGARWAL S, BABU R V, CHAKRABORTY A. Text-based person search via attribute-aided matching[C]// Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2020: 2617-2625.

[11] LI D, CHEN X, HUANG K Q. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios[C]// Proceedings of the Asian Conference on Pattern Recognition. Piscataway: IEEE Press, 2015: 111-115.

[12] DENG Y, LUO P, LOY C C, et al. Pedestrian attribute recognition at far distance[C]// Proceedings of the ACM International Conference on Multimedia. New York: ACM, 2014: 789-792.

[13] WANG J, ZHU X, GONG S, et al. Attribute recognition by joint recurrent learning of context and correlation[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 17467750.

[14] LIN Y, ZHENG L, ZHENG Z, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95: 151-161.

[15] MATSUKAWA T, SUZUKI E. Person re-identification using CNN features learned from combination of attributes[C]// Proceedings of the IEEE International Conference on Pattern Recognition. Piscataway: IEEE Press, 2016: 2428-2433.

[16] CHEN W, CHEN X, ZHANG J, et al. Beyond triplet loss: A deep quadruplet network for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 17355323.

[17] ALEXANDER H, LUCAS B, BASTIAN L. In defense of the triplet loss for person reidentification[EB/OL]. (2017-03-22)[2020-10-30]. https://arxiv.org/abs/1703.07737.

[18] YIN J, WU A, ZHENG W. Fine-grained person re-identification[J]. International Journal of Computer Vision, 2020, 128: 1654-1672.

[19] GAO Z, GAO L S, ZHANG H, et al. Deep spatial pyramid feature collaborative reconstruction for partial person reid[C]// Proceedings of the ACM International Conference on Multimedia. New York: ACM, 2019: 1879-1887.

[20] ZHENG Z, YANG X, YU Z, et al. Joint discriminative and generative learning for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 13.5-1344.

[21] YANG Q, WU A, ZHENG W, et al. Person re-identification by contour sketch under moderate clothing change[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(6): 2029-2046.

[22] WANG B, YANG Y, XU X, et al. Adversarial cross-modal retrieval[C]// Proceedings of the ACM International Conference on Multimedia. New York: ACM, 2017: 154-162.

[23] JING Y, SI C, WANG J, et al. Pose-guided joint global and attentive local matching network for text-based person search[EB/OL]. (2018-09-22)[2020-10-30]. https://arxiv.org/abs/1809.08440v2.

[24] LOPER E, KLEIN E, BIRD S. Natural language processing with python-natural language toolkit[CP/OL]. (2019-09-04)[2020-10-30]. http://www.nltk.org/book/.

[25] LIU X, ZHAO H, TIAN M, et al. HydraPlus-Net: Attentive deep features for pedestrian analysis[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 350-359.

[26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[27] LUO Y, ZHENG Z, ZHENG L, et al. Macro-micro adversarial network for human parsing[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 424-440.

[28] SUN B, SAENKO K. Deep coral: Correlation alignment for deep domain adaptation[C]// Proceedings of the European Conference on Computer Vision. Berlin, German: Springer, 2016: 443-450.

[29] SHI B, JI L, LU P, et al. Knowledge aware semantic concept ex-

北京航空航天大学学报

北航学报 赠阅

pansion for image-text matching[C]// Proceedings of the International Joint Conference on Artificial Intelligence. San Francisco: Margan Kaufmann, 2019: 5182-5189.

[30] HUANG Y, WU Q, SONG C, et al. Learning semantic concepts and order for image and sentence matching[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6163-6171.

[31] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2017-01-30)[2020-10-30]. https://arxiv.org/abs/1412.6980.

[32] LI S, XIAO T, LI H, et al. Identity-aware textual-visual matching with latent co-attention [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1908-1917.

[33] ZHENG Z, ZHENG L, GARRETT M, et al. Dual-path convolutional image-text embeddings with instance loss [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2020, 16(2): 1-23.

A cross-modal pedestrian Re-ID algorithm based on dual attribute information

CHEN Lin[1], GAO Zan[2], SONG Xuemeng[1], WANG Yinglong[2], NIE Liqiang[1,*]

(1. School of Computer Science and Technology, Shandong University, Qingdao 266200, China;
2. Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China)

Abstract: Through the investigation of cross-modal retrieval, the use of attribute information can enhance the semantic representation of extracted features. The attributes of the pedestrian image and text are not used adequately in the existing cross-modal pedestrian Re-ID algorithms based on natural language. To tackle the above issues, a novel cross-modal pedestrian Re-ID algorithm based on dual attribute information is proposed. Specifically, the attribute information of the pedestrian image and the attribute information of pedestrian text descriptions are fully and simultaneously explored, and the dual attribute space is also built to improve the distinguishability and semantic expression of extracted image and text features. Extensive experimental results on a public cross-modal pedestrian Re-ID dataset CUHK-PEDES demonstrate that the proposed algorithm is comparable with state-of-the-art algorithm CMAAM (Top-1 56.68%), the retrieval accuracy Top-1 of the proposed algorithm reaches 56.42%, and Top-5 and Top-10 are improved by 0.45% and 0.29% respectively. Besides, the retrieval accuracy of cross-modal pedestrian images can be significantly improved if the class information is provided in the gallery image pool and is used to extract attribute features, and Top-1 can reach 64.88%. The ablation study also proves the importance of the text attribute and image attribute used by the proposed algorithm and the effectiveness of the dual attribute space.

Keywords: cross-modal retrieval; matching algorithm; pedestrian attribute information; feature representation; feature fusion

北京航空航天大学学报 北京航空航天大学学报

Received: 2020-11-04; **Accepted**: 2020-11-12; **Published online**: 2021-03-02 17:01

URL: kns.cnki.net/kcms/detail/11.2625.V.20210302.1556.003.html

Foundation items: National Natural Science Foundation of China (61772310, 61702300, U1936203, 61872270); National Key R & D Program of China (2018AAA0102502); Shandong Provincial Natural Science Foundation (ZR2019JQ23); Shandong Provincial Key Research and Development Program (2019JZZY010118); Young Creative Team in Universities of Shandong Province (2020KJN012); Innovation Teams in Colleges and Universities in Jinan (2018GXRC014)

* **Corresponding author**. E-mail: nieliqiang@gmail.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0616

单元构架式抛物面天线张紧绳索多层设计方法

全下文载

王辉[1], 何天宇[1], 都显琛[1], 王春洁[1,2,*]

(1. 北京航空航天大学 机械工程及自动化学院, 北京 100083;
2. 北京航空航天大学 虚拟现实技术与系统国家重点实验室, 北京 100083)

摘 要: 星载抛物面天线在轨运行时由于姿态调整或环境因素变化,可能产生振动,从而降低其工作性能。为了减弱振动的影响,需采取加装张紧绳索的方法来提升天线刚度。提出了一种新型张紧绳索多层设计方法,通过将抛物面天线划分为多层,调整各层绳索的张紧力,从而提高天线结构刚度,同时尽量降低张紧力引起的形面误差。为了验证所提方法的有效性,建立了天线的有限元模型。在此基础上进行了有限元分析,研究了不同张紧力参数配置下的结构刚度和形面误差,并以提升结构刚度的同时降低张紧力引起的形面误差为目标,对张紧力参数进行了优化设计。为了提高计算效率,采用响应面法建立代理模型参与优化迭代计算。采用非劣排序遗传算法(NSGA-Ⅱ)完成优化迭代计算,优化后的张紧力参数使天线的性能得到了进一步的提升,为构架式可展开抛物面天线的设计提供理论指导。

关 键 词: 抛物面天线; 张紧绳索; 形面误差; 有限元分析; 优化设计
中图分类号: V443+.4
文献标志码: A **文章编号**: 1001-5965(2022)04-0657-08

随着通信、遥感测量等领域的发展,星载天线成为各类卫星上的重要功能组件。星载天线中,可展开抛物面天线因收纳比大、刚度高、相对于平板天线工作精度更高等特点,受到了广泛的研究和应用。

为了在理论层面上研究并提升抛物面天线的性能,各国学者开展了大量有意义的研究。文献[1]阐述了构建模块化抛物面天线的必要性,并分析论证了模块个数对抛物面精度的影响。文献[2]利用 ANSYS 软件,建立了桁架式抛物面天线的等效有限元模型,研究了均布的绳索张紧力对天线前 10 阶固有频率的影响。文献[3]针对大型可展开抛物面天线,利用 ANSYS 软件将绳索以交叉布置与平行布置 2 种形式对天线的模态振型影响进行了深入研究,得出绳索交叉布置形式对天线刚度提升作用优于平行布置的结论。文献[4-5]基于星载索杆式抛物面天线的特性,分析了其索杆结构,并提出了一种全新的索杆张拉构型。文献[6]对抛物面天线张紧绳索的热变形特性的影响因素进行了敏感度分析,探究了减小结构热变形的方法,并指出减小热变形应该注意的问题。文献[7]分析了 3 种抛物面天线的形面误差,并对 3 种误差的优缺点进行了对比和分析。文献[8]针对一种索网式抛物面天线,利用有限元法分析了绳索长度公差和绳索预张力对工作网面精度的影响。文献[9]采用有限体积法对一种索网式抛物面天线结构进行了静力分析和挠度分析。文献[10]在含绳索抛物面天线结构的有限元模型基础上,对结构的固有频率及模态振型进行了研究。文献[11]针对索网式抛物面天线

收稿日期: 2020-11-04; **录用日期**: 2020-12-21; **网络出版时间**: 2020-12-30 10:00
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20201229.1617.006.html
基金项目: 国家自然科学基金(51635002)
* **通信作者**. E-mail: wangcj@buaa.edu.cn

引用格式: 王辉, 何天宇, 都显琛, 等. 单元构架式抛物面天线张紧绳索多层设计方法[J]. 北京航空航天大学学报, 2022, 48(4): 657-664. WANG H, HE T Y, DU X C, et al. Multi-layer design method of tension rope for paraboloid antenna[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 657-664 (in Chinese).

北京航空航天大学学报

北航学报 赠阅

结构,基于力密度法,通过优化设计,使天线获得了较高的网面精度,并求解了最佳的索网预拉力。文献[12]通过实验测试,分析了控制天线展开绳索的寿命,并对影响展开绳索寿命的因素进行了分析。文献[13]对模块化抛物面天线的构型及外层模块存在的角度误差进行了详细推理分析。

上述文献对抛物面天线的索网构型设计、形面精度及振动特性进行了大量的研究,但是针对构架式抛物面天线张紧绳索的非均布布置方案的研究并不充分。对此,本文提出了一种新型的绳索分层布置方案,并利用有限元模型对方案的合理性进行了验证。基于响应面模型和多目标优化方法对绳索张紧力参数进行了优化。

1 模块化抛物面天线

1.1 天线构型

本文以一种构架式可展开抛物面天线为研究对象[1,13-14],其工作口径为 4 m。该天线的构架用于支撑金属反射网,提高天线整体刚度。金属反射网是天线的工作部分,负责反射传输信号。构架是金属反射网的支撑体系,其展开状态会影响金属反射网的精度。因此,保证构架展开后的形面精度也是非常重要的。鉴于此,本文重点以抛物面天线构架为研究对象,且下文所述的抛物面天线代指构架。构架主要由 7 个边长相等的基本模块构成,中心模块由 6 个角度间隔相等且完全相同的基本肋单元组成,周围 6 个模块由角度间隔近似相等的 6 个基本肋单元组成,整体结构如图 1 所示。

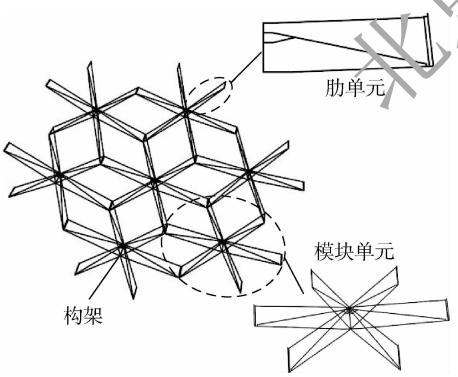


图 1 抛物面天线构架结构

Fig. 1 Structure diagram of paraboloid antenna frame

1.2 工作性能评估

为了减弱振动对工作性能的影响,有必要提升天线的结构刚度。提高天线的结构刚度,除了通过改善杆件材料性能和杆件截面尺寸以外,还可以在肋条间加装张紧绳索。但加装绳索的同时,会使背架发生形变,从而产生形面误差,其变形状况如图 2 和图 3 所示。为了更明显地显示出形面误差,图 3 将变形尺寸扩大了 500 倍。



图 2 构架变形前

Fig. 2 Frame before deformation

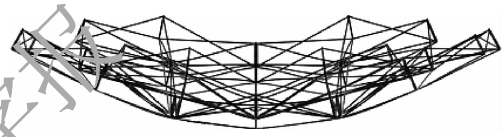


图 3 构架变形后

Fig. 3 Frame after deformation

为了定量分析形面误差,需要先给出其评价参数。在抛物面构架上选取若干个取样点,在绳索张紧力的作用下,每个取样点会偏离原有位置,使天线产生误差,单个取样点变形前后的距离及全部取样点偏移距离的均值表示为

$$\Delta r_i = \sqrt{(D_{\hat{x}_i^*} - D_{x_i})^2 + (D_{\hat{y}_i^*} - D_{y_i})^2 + (D_{\hat{z}_i^*} - D_{z_i})^2} \tag{1}$$

$$\overline{\Delta r} = \frac{\sum_{i=1}^{N} \Delta r_i}{N} \tag{2}$$

从而,形面误差的指标参数定义为

$$T = \sqrt{\frac{\sum_{i=1}^{N} (\Delta r_i - \overline{\Delta r})^2}{N}} \tag{3}$$

式中:D_{x_i}、D_{y_i}、D_{z_i}为样本点变形前坐标;$D_{\hat{x}_i^*}$、$D_{\hat{y}_i^*}$、$D_{\hat{z}_i^*}$为样本点变形后的新坐标;N 为样本点个数。

本文以天线的固有频率作为结构刚度的评价标准,并将天线的若干阶固有频率整合为一个性能评价参数,其表达式为

$$C = \left[\frac{1}{m}\sum_{i=1}^{m} (\lambda_i - \lambda_0)^{-1}\right]^{-1} \tag{4}$$

式中:C 为天线前 m 阶的平均固有频率;λ_i 为天线的第 i 阶固有频率;λ_0 为指定的基准频率。结合文献[15],1 阶固有频率对天线的性能影响权重较大,其他低阶固有频率也对天线的性能有一定的影响,综合考虑本文研究对象后,取 λ_0 为 0.2 Hz。上述方法解决了在以结构固有频率为研究目标的优化设计问题中,仍然存在目标函数振荡、收敛过程缓慢的问题。

加装绳索可以提升天线的平均固有频率,同

北航学报 赠阅

时会使抛物面产生形面误差,导致其形面精度降低,天线的 2 个性能之间存在矛盾,故需要细化绳索的加装方法。

2　张紧绳索加装方法

在绳索张紧力大小相同的情况下,相比于平行布置方案,交叉布置方案对提高天线高阶固有频率的效果更加明显[3],因此本文选择交叉布置方案。

以往的模块化抛物面天线绳索布置方案中,通常将所有绳索的张紧力大小设置为相同值[2],并没有考虑不同位置绳索加载不同大小张紧力的情况。传统布置方案虽然提高了天线刚度,但会使天线产生较大的形面误差。为了解决传统方案存在的问题,本文在已有研究的基础上提出了一种新的模块化抛物面天线绳索张紧力分层加载方案。将张紧绳索由内向外按层分成 4 个部分进行张紧力的添加,通过合理配置各层张紧力的大小,达到既保证天线的刚度,也保证天线的形面误差最小的目的。天线未展开时,绳索处于松弛状态;展开过程中,中心支柱根部弹簧驱动天线展开;即将到达指定展开状态时,各层绳索依次张紧,同时对绳索张紧力进行测量,保证同层绳索中每段受力大小相同,可以实现将中心模块连接的 6 个模块均匀展开,完成展开过程[13]。布置方案如图 4 所示。图中,A、B、C、D、E、F、G 分别为组成天线构架的 7 个模块单元。考虑到该方案各层绳索的张紧力会存在差异,因此在实际应用中需要将张紧绳索分段安装。

天线在展开过程中需保证中心模块的 6 个绳索张紧力大小一致,从而使每个肋单元之间的角度相等[13],新方案中心模块上的第一层绳索张紧力形成闭合力环,第二层张紧力对第一层模块的拉力矢量和为 0,满足要求。外层 6 个模块肋单元间隔不完全相同,张紧力存在差别,其角度可依靠第二层接头处自由度释放实现调整[13]。中心模块受力如图 5 所示。图中各力满足:

$$\sum_{i=1}^{6} \boldsymbol{F}_{1i} + \sum_{i=1}^{6} \boldsymbol{F}_{2i} + \sum_{i=1}^{6} \boldsymbol{F}_{3i} = 0 \tag{5}$$

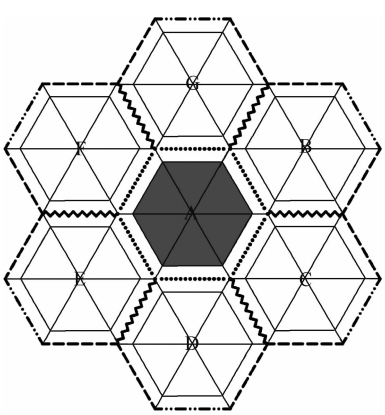


图 4　张紧力绳索布置方案

Fig. 4　Arrangement scheme of tension rope

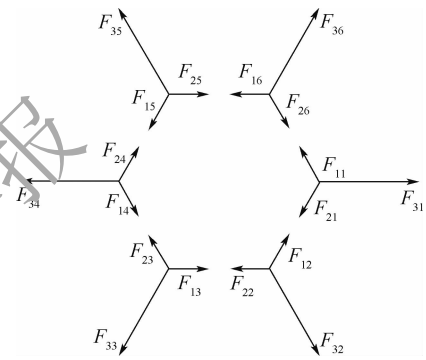


图 5　中心模块力矢量图

Fig. 5　Force vector diagram of central module

3　模块化抛物面天线结构性能分析

3.1　有限元模型

为了分析天线的结构刚度和形面误差,本文采用 ABAQUS 软件建立有限元模型,如图 6 所示。

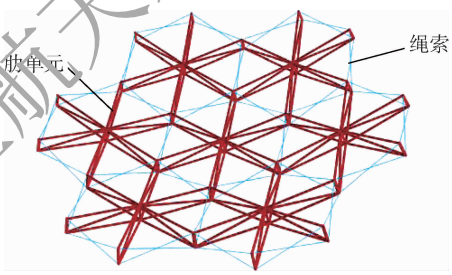


图 6　抛物面天线有限元模型

Fig. 6　Finite element model of paraboloid antenna

抛物面天线主要由多个杆件、铰链及绳索组成。有限元模型中采用 Bushing 元来代替铰链,可以很好地控制铰链的自由度,更接近铰链的真实状态。刚度设置如表 1 和表 2 所示,表中 D_{11}、D_{22}、D_{33}、D_{44}、D_{55}、D_{66}分别表示 Bushing 元控制沿 X 轴、Y 轴、Z 轴和绕 X 轴、Y 轴、Z 轴的刚度。铰链在天线展开过程中是绕其局部坐标系的 Z 轴转动[2],因此将其 Z 轴的刚度设置为较小值。

此外,结构杆件和绳索分别采用梁单元和桁架等效建模,建模方法的合理性在文献[2]中已进行了充分的验证。等效模型的关键参数值如表 3所示。

北京航空航天大学学报

北航学报 醒阅

表 1 Bushing 元位移刚度参数设置

Table 1 Parameter setting of displacement stiffness for Bushing coupling

位移刚度	数值/(N·mm^{-1})
D_{11}	10^8
D_{22}	10^8
D_{33}	10^8

表 2 Bushing 元转动刚度参数设置

Table 2 Parameter setting of rotational stiffness for Bushing coupling

转动刚度	数值/(N·mm·rad^{-1})
D_{44}	10^8
D_{55}	10^8
D_{66}	10

表 3 模型属性设置

Table 3 Model property setting

参数	杆	绳索
材料	铝合金	钢丝
弹性模量/MPa	7×10^4	1.5×10^5
密度/(t·mm^{-3})	2.84×10^{-9}	7.9×10^{-9}
泊松比	0.31	0.31
单元类型	Beam	Truss
网格类型	Beam	T3D2
截面尺寸/mm	直径:15,厚度:1	直径:1

考虑到铰链的质量对结构的性能也有一定的影响,在各个杆件的连接点处添加质量点,以模拟铰链质量[2]。

3.2 结构刚度分析

基于 3.1 节所述的有限元模型,本节对比了传统方案与本文提出的新方案二者对天线前 6 阶固有频率的提升能力,配置参数及分析结果如表 4 ~ 表 6 所示。表中,“方案 0”为不加张紧绳索的方案,“方案 1”为传统布置方案,“方案 2 ~ 方案 4”为新型分层布置方案的 3 种布置方式。

将表 6 中 4 种方案对应的各阶固有频率分别与表 4 进行对比,传统绳索张紧力布置方案 1 和新型绳索分层布置方案 2 ~ 方案 4 对抛物面天线的前 6 阶频率均有提升。说明相对于传统布置方案,新型布置方案同样可以起到提升天线刚度的作用,且提升作用不低于传统方案。

表 4 无绳索天线固有频率及平均固有频率

Table 4 Natural frequency and average natural frequency of cordless antenna

阶层	固有频率(方案 0)/Hz
1 阶	0.490
2 阶	0.800
3 阶	1.841
4 阶	4.350
5 阶	6.732
6 阶	8.203
C	0.962

表 5 张紧力布置方案

Table 5 Tension arrangement scheme

阶层	张紧力/N			
	方案 1	方案 2	方案 3	方案 4
1 层	200	140	220	290
2 层	200	300	270	250
3 层	200	170	190	110
4 层	200	190	120	150

表 6 不同方案各阶固有频率及平均固有频率

Table 6 Natural frequency and average natural frequency of each order for different schemes

阶层	固有频率/Hz			
	方案 1	方案 2	方案 3	方案 4
1 阶	0.503	0.504	0.504	0.505
2 阶	0.808	0.808	0.808	0.808
3 阶	1.830	1.829	1.829	1.830
4 阶	6.938	6.920	6.920	6.910
5 阶	10.078	10.046	10.046	10.029
6 阶	16.291	16.285	16.285	16.283
C	1.022	1.023	1.023	1.025

3.3 形面误差分析

本文提出的新方案可以在保证天线刚度的同时减小形面误差。首先,在抛物面天线上选取 30 个和金属反射网连接的样本点,作为天线形面误差的研究对象,如图 7 所示。

对于新型绳索预紧力布置方案,分别求得其对应的形面误差值,将结果与传统布置方案的误差值进行对比,如表 7 所示。

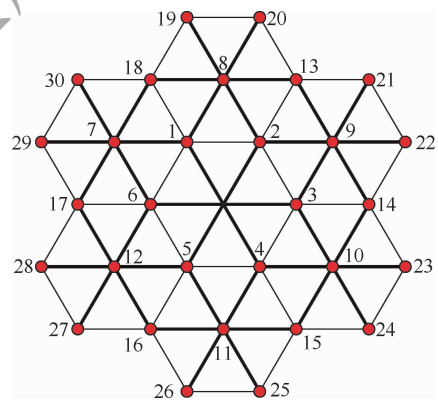


图 7 取样点分布

Fig. 7 Distribution of sampling points

表 7 不同方案形面误差

Table 7 Shape error of different schemes

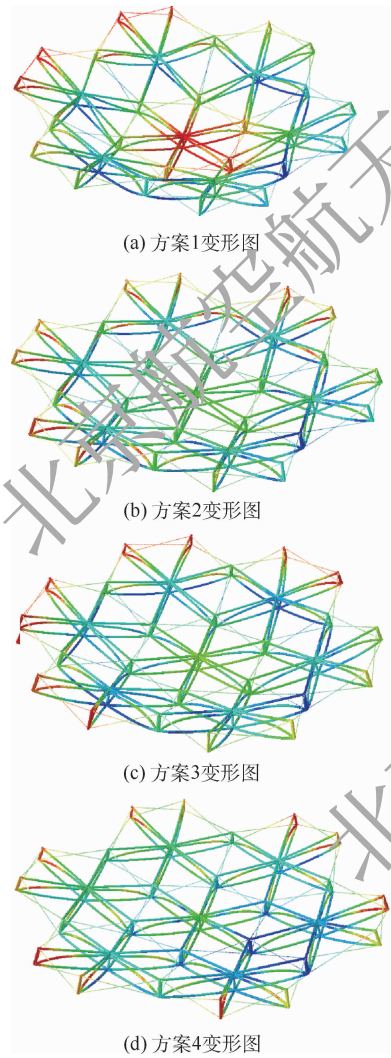
方案	张紧力/N				形面误差/mm
	1 层	2 层	3 层	4 层	
方案 1	200	200	200	200	0.177
方案 2	140	300	170	190	0.066
方案 3	220	270	190	120	0.109
方案 4	290	250	110	150	0.056

北京航空航天大学学报

北航学报 瞰阅

通过将传统均匀绳索张紧力布置方案 1 与新型绳索张紧力分层布置方案 2 ~ 方案 4 进行对比,方案 2 ~ 方案 4 条件下的形面误差均低于方案 1。在方案 2 的绳索张紧力布置条件下,形面误差降低了62.71%;在方案 3 的条件下,形面误差降低了 38.42%;在方案 4 的条件下,形面误差降低了 68.36%;上述 3 种方案对降低抛物面天线的形面误差有明显作用。

表 7 中各方案的天线变形云图如图 8 所示。可以看出,传统布置方案 1 和分层布置方案 2 ~ 方案 4 在相同变形放大系数下,后者的变形趋势明显比前者的变形趋势更加平缓。



(a) 方案1变形图

(b) 方案2变形图

(c) 方案3变形图

(d) 方案4变形图

图 8　抛物面天线变形图

Fig. 8　Deformation diagram of paraboloid antenna

4　张紧力参数优化

在新的绳索布置方案下,为使抛物面天线有较高结构刚度的同时保证形面误差最小,需要对每层绳索的张紧力进行优化。考虑有限元模型计算效率较低,如果在优化迭代计算中,仍用有限元模型计算抛物面的性能表征参数将花费大量时间。为此,本文采用完全四阶多项式建立以绳索力为输入、形面误差或平均频率为输出的响应面模型,利用响应面模型参与优化迭代计算将大幅度提高计算效率[16-18]。

4.1　建立绳索力优化模型

本文选取各层张紧力作为设计变量,选取平均固有频率和形面误差为目标函数。优化数学模型如下:

$$\begin{cases} \max\ C = f_1(F_1,F_2,F_3,F_4) \\ \min\ T = f_2(F_1,F_2,F_3,F_4) \\ \text{s.t.}\begin{cases} 100\ \text{N} < F_1 < 300\ \text{N} \\ 100\ \text{N} < F_2 < 300\ \text{N} \\ 100\ \text{N} < F_3 < 300\ \text{N} \\ 100\ \text{N} < F_4 < 300\ \text{N} \end{cases} \end{cases} \tag{6}$$

式中:F_1、F_2、F_3、F_4 为各层绳索张紧力。

4.2　代理模型

为了提高计算效率,本文采用完全四阶多项式建立 C、T 与 F_1、F_2、F_3、F_4 的映射关系,其一般表达式为

$$\begin{aligned} F(\boldsymbol{X}) = {} & \alpha_0 + \sum_{i=1}^{4}\alpha_i x_i + \sum_{i=1}^{4}\alpha_{ii}x_i^2 + \sum_{i=1}^{4}\alpha_{iii}x_i^3 + \\ & \sum_{i=1}^{4}\alpha_{iiii}x_i^4 + \sum_{1=i<j}^{4}\alpha_{ij}x_i x_j + \sum_{1=i<j}^{4}\alpha_{iij}x_i^2 x_j + \\ & \sum_{1=i<j}^{4}\alpha_{ijj}x_i x_j^2 + \sum_{1=i<j<k}^{4}\alpha_{ijk}x_i x_j x_k + \sum_{1=i<j}^{4}\alpha_{iiij}x_i^3 x_j + \\ & \sum_{1=i<j}^{4}\alpha_{iijj}x_i^2 x_j^2 + \sum_{1=i<j}^{4}\alpha_{ijjj}x_i x_j^3 + \sum_{1=i<j<k}^{4}\alpha_{iijk}x_i^2 x_j x_k + \\ & \sum_{1=i<j<k}^{4}\alpha_{ijjk}x_i x_j^2 x_k + \sum_{1=i<j<k}^{4}\alpha_{ijkk}x_i x_j x_k^2 + \\ & \sum_{1=i<j<k<l}^{4}\alpha_{ijkl}x_i x_j x_k x_l \end{aligned} \tag{7}$$

式中:n 为输入变量的个数;x_i 为不同输入变量 α_0、α_i、α_{ii}、α_{iii}、α_{iiii}、α_{ij}、α_{iij}、α_{ijj}、α_{ijk}、α_{iiij}、α_{iijj}、α_{ijjj}、α_{iijk}、α_{ijjk}、α_{ijkk}及 α_{ijkl}多项式的待定系数。利用实验设计方法生成 n 个随机输入变量样本点,并将所有样本点数据输入到仿真模型中,求得对应的输出值。将所有样本点代入式(7),可以得到矢量表达式为

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \hat{\boldsymbol{Y}} + \boldsymbol{\varepsilon} \tag{8}$$

式中:$\boldsymbol{Y}$ 为每组样本点对应的仿真模型计算的真实值组成的 n 维列向量;$\boldsymbol{X}$ 为样本点输入量组成的矩阵;$\boldsymbol{\alpha}$ 为由多项式待定系数组成的列向量;$\hat{\boldsymbol{Y}}$ 为响应面近似函数值向量;$\boldsymbol{\varepsilon}$ 为响应面近似函数误差向量。

根据最小二乘法,可得响应面的近似函数为

北京航空航天大学学报

$$\boldsymbol{\alpha} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \tag{9}$$

本文利用最优拉丁超立方设计方法，在 4 层张紧力的取值范围内随机均匀抽取 100 个样本点，代入仿真模型，计算出对应的 C、T，基于仿真结果，分别拟合出 C、T 关于 F_1、F_2、F_3、F_4 的响应面关系式。

本文利用均方根相对误差值 R_{MSE} 及决定系数 R^2 对 2 个优化目标对应的响应面表达式的拟合精度进行检测判断，有

$$R_{\mathrm{MSE}} = \frac{1}{n\bar{y}}\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{11}$$

式中：$0 < R_{\mathrm{MSE}}, R^2 < 1$；$n$ 为样本点的个数；y_i 为样本点代入到仿真模型中计算出的结果；$\hat{y}_i$ 为响应面对应的估计值；$\bar{y}$ 为 y_i 的平均值。若 R_{MSE} 的值越趋近于 0，表示拟合精度越高，反之越差；若 R^2 的值越趋近于 1，表示拟合精度越高，反之越差。表 8 为本文拟合出的 C、T 对应响应面模型的 R_{MSE} 及 R^2。

表 8 响应面精度系数

Table 8 Response surface accuracy coefficient

指标	R_{MSE}	R^2
C	0.000 076	0.992 176
T	0.003 468	0.999 587

结果显示，响应面模型精度较高，可参与后续的优化计算。

4.3 优化计算及结果分析

在解决多目标优化问题时，可以通过加权求和的方式将多个目标函数转化成单个目标函数，进而采用单目标优化算法完成优化计算，但该方法得到的优化结果比较依赖多个目标函数的权重系数。在本文的优化问题中，2 个目标函数的取值差异较明显，不易给出适当的权重系数。因此，直接采用非劣排序遗传算法（NSGA-Ⅱ）完成迭代计算。NSGA-Ⅱ基于帕累托最优理论，可获得多目标优化问题的最优解集，而非单个最优解。优化算法的参数设置如表 9 所示。

优化过程中，利用响应面模型计算出的 C、T 近似值来评价天线的性能。经过迭代计算，得到该模型的帕累托最优解集，绘制该模型对应的帕累托前沿曲线，如图 9 所示。

通过分析最优解集，选取形面误差最小的配置作为最优解，即（F_1，F_2，F_3，F_4）=（271.95，249.21，103.38，101.13）N，将最优解代入仿真模型进行验证，验证结果如表 10 所示。

从表 10 可知，经过优化后，新分层方案相较于传统方案使平均固有频率提升 0.1%，形面误差降低了 88.7%。

表 9 优化算法参数配置

Table 9 Parameter configuration of optimization algorithm

参数	数值
种群数	12
迭代代数	20
交叉指数	10
变异指数	20
交叉概率	0.9

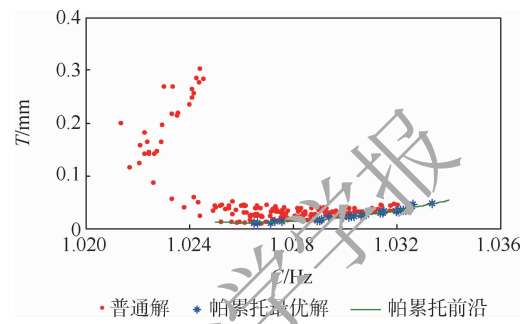


图 9 帕累托前沿曲线

Fig. 9 Curve of Pareto front

表 10 最优解验证

Table 10 Verification of optimal solution

名称	最优解	对照组
F_1/N	271.95	200
F_2/N	249.21	200
F_3/N	103.38	200
F_4/N	101.13	200
C/Hz	1.024	1.023
T/mm	0.020	0.177

5 结 论

1）本文以模块化抛物面天线为研究对象，提出一种分层布置张紧绳索的方法，并采用有限元模型对该方法的合理性进行了验证。仿真结果表明，该方法在保证结构刚度的同时，可以有效减小张紧力所引起的天线形面误差。

2）结合响应面模型和 NSGA-Ⅱ对本文方法的张紧力参数进行了优化设计，结果表明，优化后的张紧力参数可以进一步提升抛物面天线的性能。

3）本文方法可为构架式可展开抛物面天线的设计提供理论指导，且可以引入到其他星载天线的设计过程中。

北京航空航天大学学报

北航学报 聽阅

参考文献(References)

[1] ANDO K, MITSUGI J, SENBOKUYA Y. Analyses of cable-membrane structure combined with deployable truss[J]. Computers & Structures, 2000, 74(1): 21-39.

[2] 刘荣强,田大可,邓宗全,等. 多模块构架式空间可展开天线背架的模态分析[J]. 北京理工大学学报, 2011, 31(6): 685-690.
LIU R Q, TIAN D K, DENG Z Q, et al. Modal analysis of truss structure for deployable truss antenna with multi-module[J]. Journal of Beijing University of Technology, 2011, 31(6): 685-690(in Chinese).

[3] 崔吉. 大尺度可展开式抛物面天线机构的设计与展开性能分析[D]. 哈尔滨:哈尔滨工业大学, 2012.
CUI J. Design and analysis of large volume deployable parabolic antenna mechanism[D]. Harbin: Harbin Institute of Technology, 2012(in Chinese).

[4] 史创. 基于四棱柱可展单元的索杆张拉式天线机构研究[D]. 哈尔滨:哈尔滨工业大学, 2019.
SHI C. Research on cable-strut tension antenna mechanism based on quadriprism deployable unit[D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

[5] 刘望. 星载大型索网天线结构非线性动力学建模及形面误差特性研究[D]. 长沙:国防科学技术大学, 2013.
LIU W. Research on nonlinear dynamic modeling and surface error characteristics of spaceborne large cable-network antenna structures[D]. Changsha: University of National Defense Science and Technology, 2013(in Chinese).

[6] 杨东武,保宏. 非对称索网抛物面天线力平衡特性及预拉力设计[J]. 机械工程学报, 2009, 45(8): 314-318.
YANG D W, BAO H. Force balance characteristics and pretension design of asymmetric cable net paraboloid antenna[J]. Journal of Mechanical Engineering, 2009, 45(8): 314-318(in Chinese).

[7] 王朋朋,薛永刚,高博. 伞状天线张力绳索热变形敏感性分析[J]. 空间电子技术, 2014, 11(3): 20-24.
WANG P P, XUE Y G, GAO B. Sensitivity analysis of tensile cables thermal distortion in umbrella antenna[J]. Space Electronics Technology, 2014, 11(3): 20-24(in Chinese).

[8] 侯国勇,关富玲,赵孟良. 桁架式可展开天线精度计算方法[J]. 浙江大学学报(工学版), 2008, 42(9): 1469-1473.
HOU G Y, GUAN F L, ZHAO M L. Error computation method of deployable truss antenna[J]. Journal of Zhejiang University (Engineering Science), 2008, 42(9): 1469-1473(in Chinese).

[9] 范叶森,马小飞,李正军,等. 绳索长度误差对索网天线网面精度影响的分析方法[J]. 华南理工大学学报(自然科学版), 2014, 42(8): 64-69.
FAN Y S, MA X F, LI Z J, et al. Analysis method of influence of cable length errors on surface accuracy of cable-net antennas[J]. Journal of South China University of Technology (Natural Science), 2014, 42(8): 64-69(in Chinese).

[10] 高海燕,袁茹,王三民. 径射状可展天线反射器结构固有模态分析[J]. 机械设计与制造, 2007(5): 174-176.
GAO H Y, YUAN R, WANG S M. The natural modal analysis for reflectible device of space deployable antenna[J]. Mechanical Design and Manufacturing, 2007(5): 174-176(in Chinese).

[11] 李彬,胡国伟,黄芬. 周边桁架可展开天线索网结构力密度优化找形方法[J]. 桂林电子科技大学学报, 2009, 29(5): 373-376.
LI B, HU G W, HUANG F. Form-finding optimization of flexible cable net of AstroMesh deployable reflector by force-density method[J]. Journal of Guilin University of Electronic Technology, 2009, 29(5): 373-376(in Chinese).

[12] 方永刚,薛永刚,王勇. 卫星天线展开用 kevlar 绳索使用寿命研究[C]// 第三届中国国际复合材料科技大会, 2017: 1.
FANG Y G, XUE Y G, WANG Y. Research on the service life of kevlar ropes for satellite antenna deployment[C]// The 3rd China International Composite Science and Technology Conference, 2017: 1(in Chinese).

[13] 刘兆晶. 模块化可展开抛物面天线支撑机构设计与研制[D]. 哈尔滨:哈尔滨工业大学, 2011.
LIU Z J. Design and manufacture of supporting structure for modular deployable parabolic antenna[D]. Harbin: Harbin Institute of Technology, 2011(in Chinese).

[14] MEGURO A, ISHIKAWA H, TSUJIHATA A. Study on ground verification for large deployable modular structures[J]. Journal of Spacecraft & Rockets, 2006, 43(4): 780-787.

[15] MA Z D, KIKUCHI N, CHENG H C. Topological design for vibrating structures[J]. Computer Method in Applied Mechanics and Engineering, 1995, 121(1-4): 259-280.

[16] 吴宏宇,王春洁,丁宗茂,等. 两种着陆模式下的着陆器缓冲机构构型优化[J]. 宇航学报, 2017, 38(10): 1032-1040.
WU H Y, WANG C J, DING Z M, et al. Configuration optimization of landing gear under two kinds of landing modes[J]. Acta Astronautics, 2017, 38(10): 1032-1040(in Chinese).

[17] WU H Y, NIU W D, WANG S X, et al. Sensitivity analysis of input errors to motion deviations of underwater glider based on optimized response surface methodology[J]. Ocean Engineering, 2020, 209: 107400.

[18] WU H Y, NIU W D, WANG S X, et al. Prediction method of permissible error ranges of control parameters for underwater gliders under given operation accuracy[J]. Applied Ocean Research, 2020, 103: 102153.

北京航空航天大学学报

北航学报 赠阅

Multi-layer design method of tension rope for paraboloid antenna

WANG Hui[1], HE Tianyu[1], DU Xianchen[1], WANG Chunjie[1,2,*]

(1. School of Mechanical Engineering and Automation, Beihang University, Beijing 100083, China;
2. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100083, China)

Abstract: Satellite-borne paraboloid antenna may vibrate during orbit operation due to attitude adjustment or environmental factors, thus reducing its performance. In order to reduce the influence of vibration, it is necessary to add tension rope to enhance the antenna stiffness. Therefore, a new multi-layer design method of tension rope is proposed. In this method, the paraboloid antenna is divided into multiple layers and the tension of each layer is adjusted to improve the structural stiffness of the antenna and reduce the shape error caused by the tension force as far as possible. In order to verify the effectiveness of the above method, the finite element model of the antenna is established. On this basis, the finite element analysis is carried out, and the structural stiffness and shape error under different tension parameters are studied. In order to improve the stiffness of the structure and reduce the shape error caused by the tension force, the tension parameters are optimized. In order to improve the efficiency of optimization calculation, the response surface method is used to establish a surrogate model to participate in the optimization iterative calculation. Non-inferior sorting genetic algorithm (NSGA-Ⅱ) is used to complete the iterative calculation. The optimized tension parameters further improve the performance of the antenna. This research can provide theoretical guidance for the design of truss deployable paraboloid antenna.

Keywords: paraboloid antenna; tension rope; shape error; finite element analysis; optimization design

北京航空航天大学学报

Received: 2020-11-04; **Accepted**: 2020-12-21; **Published online**: 2020-12-30 10:00
URL: kns.cnki.net/kcms/detail/11.2625.V.20201229.1617.006.html
Foundation item: National Natural Science Foundation of China (51635002)
* **Corresponding author**. E-mail: wangcj@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0617

柔性翼气动力和变形特性的实验研究

全下文载

王涵斌，贺曦，王晋军*
（北京航空航天大学 航空科学与工程学院，北京 100083）

摘 要：微型飞行器在军、民用领域具有广阔的应用前景，柔性翼是提升微型飞行器的气动性能的有效方法。为了更好地对柔性翼进行控制，对柔性翼变形和振动特性及其对气动力的影响进行了同步测量。研究结果表明，相比于刚性翼，柔性翼使失速迎角推迟了6°，最大升力系数提升了47.4%，升阻比提高了17.8%。柔性翼的周期性振动除了迎角0°~2°呈现大振幅、小静变形特征外，振动的振幅随着迎角增加经历无明显波峰、三波峰到单波峰的转换。升力系数最大时对应的薄膜变形、振动振幅均达到最大。此外，变形最大的弦向位置随迎角的变化决定了俯仰力矩的特性。据此提出了施加弯度和特定频率的振动激励来提升气动性能的主动控制策略。

关 键 词：柔性翼；流固耦合；气动力测量；变形测量；振动分析
中图分类号：V211.41
文献标志码：A **文章编号**：1001-5965(2022)04-0665-09

微型飞行器(MAV)具有结构简单、尺寸小、成本低、隐蔽性好等诸多优点，在军事和民用领域都有较大的应用价值。在军事领域，可以用于通信、侦察、战场评估等方面；在民用领域，可以用于应急救灾、生态环境监测、地理信息测绘等方面。但是MAV飞行速度低，飞行雷诺数 Re 约为10^4~10^5，较低的雷诺数会造成边界层内的低能量流动，导致MAV气动性能差、飞行不稳定及对飞行环境敏感[1]，此外，低雷诺数还会造成飞行器升阻比下降[2]。因此，开展提高和改善MAV气动特性的研究具有重要的学术意义和应用价值。

自然界的蝙蝠与MAV的尺寸和工作环境相似，而蝙蝠具有更强的机动性和环境适应性，蝙蝠的可变形翅膀是造成优异气动特性的重要原因[3]。Swartz等[4-5]详细研究了蝙蝠翅膀的构造，指出蝙蝠翅膀中可以独立控制的结点多达12个，并且蝙蝠的骨头可以高度变形。蝙蝠可以通过动态改变翅膀的弯度来适应不同的迎角和来流工况，具有优异的飞行性能，进而能完成更加复杂的飞行动作[6]。He等[7]通过数值模拟发现蝙蝠可以通过动态改变翅膀的展长来实现高效飞行。可见，采用仿生、可变形翼型及机翼在提高MAV的气动效率设计中能起到关键作用。Shyy等[8]进行的柔性翼气动特性相关实验研究表明，低雷诺数下柔性薄膜翼的升阻比相较于刚性翼有所提升。Shyy和Levin[9]在改良的CLARK-Y翼型上进一步开展柔性翼实验，结果表明，柔性翼可以提高翼型的气动性能，延缓失速，并且对非定常飞行环境适应性较强。早期的柔性翼研究工作主要侧重对柔性翼的增升特性进行分析，很少有学者开展气动力与变形的综合实验来研究两者的内在联系。

Béguin等[10]采取可变后掠角的半翼展柔性翼模型开展实验，模型的支撑结构采用刚性前缘和弹性后缘，支撑结构表面覆盖柔性蒙皮，研究表

收稿日期：2020-11-04；**录用日期**：2020-12-25；**网络出版时间**：2021-01-12 13:47
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210111.1813.002.html
基金项目：国家自然科学基金国际合作项目（11761131009）；创新研究群体项目（11721202）
* **通信作者**. E-mail：jjwang@buaa.edu.cn

引用格式：王涵斌，贺曦，王晋军. 柔性翼气动力和变形特性的实验研究[J]. 北京航空航天大学学报，2022，48(4)：665-673.
WANG H B，HE X，WANG J J. Experimental study on aerodynamic and deformation characteristics of flexible membrane wing[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：665-673 (in Chinese).

北京航空航天大学学报

北航学报 赠阅

明,柔性翼的失速特性相较于刚性翼得到改善,且翼面的变形程度会随实验工况的改变而改变,如来流动压、迎角等,机翼气动特性也随之改变,同时还提出了通过主动控制柔性翼面的变形,使机翼在不同来流条件下达到更高气动效率的思路。Béguin 等[11]提出并制造了一种类似于翼龙的弹性柔性变形机翼结构,柔性膜被用于各种平面几何和机翼形状的空气动力表面,结果表明,可主动变形的翼面形状与可被动变形柔性翼面相结合,可在大范围的飞行条件下有效保持较高的升阻比。

He 等[12]研究了刚性机翼加柔性后缘平板模型变形与流场结构的耦合,发现柔性后缘平板相比于刚性平板可以减小尾流区的速度损失并抑制过失速分离。Guo 等[13]进行了柔性翼简化飞机模型气动力、流场及变形测量实验,发现柔性翼的振动使边界层掺混作用增强,分离被抑制,是升力系数增加的主要原因。

在这些研究工作的基础上,本文对柔性翼气动力及变形进行精细测量,并开展刚性翼对照实验,分析柔性翼改善气动特性的效果。将气动力与变形的瞬时和时均特性进行对比,揭示两者的内在联系,为主动控制变形提高飞行器气动特性提供控制策略。

1 实验设备、模型及方法

1.1 实验设备与模型

实验在北京航空航天大学的低速风洞进行。实验段横截面积为 430 mm × 500 mm,湍流度在实验速度范围内约为 0.3%。柔性薄膜翼(后文统称为柔性翼)通过刚性前后缘安装在实验段,前后缘固定在圆形端板上。圆形端板具有 3 个主要的作用:①固定柔性翼模型,起到迎角调节的作用;②利用端板厚度使柔性翼远离风洞壁面,减少风洞壁面处的边界层对实验测量造成的影响;③利用端板对翼尖涡的限制,提高柔性翼和周围流场的二维特性。柔性翼在实验段的安装如图 1 所示。

柔性翼展长 $b=330$ mm,弦长 $c=150$ mm,迎角范围 $\alpha=0°\sim28°$,自由来流速度 $U_\infty=12.5$ m/s,

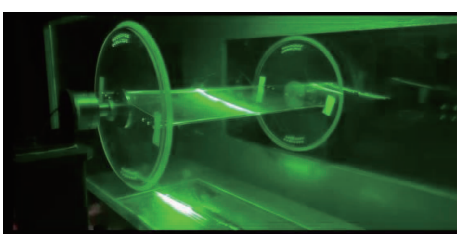


图 1 柔性翼示意图

Fig. 1 Schematic diagram of flexible membrane wing

基于弦长的雷诺数 $Re=12.5\times10^4$。柔性翼的截面图和几何尺寸如图 2 所示。柔性翼绕在直径为 3 mm 的前后缘钢柱上,为了降低前缘流动分离,薄膜可以在指定工况下绕前后缘自由旋转。柔性翼安装后保持薄膜平整,没有明显拉伸,几乎没有预应力。薄膜由透明度很高的热塑性聚氨酯弹性体橡胶(TPU)制成,厚度 $t=0.2$ mm,密度 $\rho=1.1$ g/cm^3,弹性模量 $E=31.2$ MPa。Rojratsirikul 等[14]的研究表明,当薄膜的弹性模量大于 2.2 MPa 时,柔性翼在重力作用下很难变形,因此,实验可以忽略重力对薄膜变形的影响。在实验来流条件下,柔性翼气动弹性参数 $\Pi=(Et/qc)^{1/3}=7.6$,q 为来流动压。

刚性翼对照实验中采用的机翼模型为 330 mm × 150 mm 的聚甲基丙烯酸甲酯聚合物(PMMA)平板,来流实验工况与柔性翼实验一致,刚性翼弹性模量 $E=3$ GPa,比柔性翼弹性模量大 2 个数量级,故可忽略刚性翼在实验工况下由气动力作用引起的变形。

流场坐标系如图 2 所示,x-z 平面与机翼平面垂直,并与自由来流平行,x-z 平面的原点在后缘中心,另一个用于气动力分析的坐标系是 x_{tran}-z_{tran}),其是由 x-z 平面旋转变换得到的,旋转角度为对应工况的迎角,新的坐标原点在前缘中心,如图 2 所示。

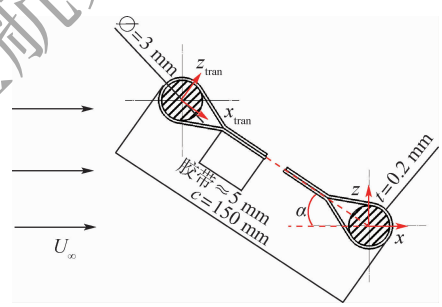


图 2 柔性翼细节与坐标系示意图

Fig. 2 Details of flexible membrane wing and schematic diagram of coordinate system

1.2 测量仪器

升力与阻力由 2 个量程为 40 N 的六轴测力传感器(ATI-Mini 40)测量,测力采样频率为 10 kHz,时长为 25 ~ 30 s,测力传感器的不确定度为 ±1.25%。

实验利用高速 PCO CS4 相机拍摄柔性翼的瞬时变形,相机的拍摄频率为 1 kHz,采样时长为 4.2 s,相机的放大率为 0.119 mm/pixel,图像的采样频率 1 kHz 比柔性翼振动的最大频率 107 Hz 高一个数量级,因此,薄膜变形测量的时间及空间

北航学报 赠阅

分辨率均较高。气动力与变形利用同步器控制,保证两者同步实时测量。柔性翼由 8 W 连续激光进行照明,由于薄膜透明度较高,激光能够穿透薄膜,照亮机翼的上下两侧,激光能量可以满足变形的识别需求,照明位置为展向 1/2 处,激光片光的厚度约为 1.5 mm。实验装置布置如图 3 所示。

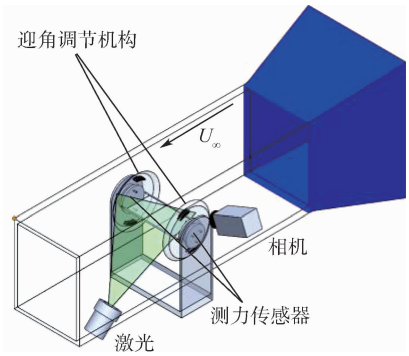


图 3　实验装置布置

Fig. 3　Experimental device arrangement

1.3　变形识别技术

利用 PCO CS4 相机对柔性翼变形位置进行拍摄,采用的边界识别算法如下:首先,选择柔性板的区域,该区域在不损失柔性翼变形信息的基础上,应尽可能小以减少计算成本。此外,前后缘光强的反射会导致靠近前后缘的部分光强不连续,因此选择区域应尽量避开前后缘。其次,利用最大类间方差方法(变形图像与黑色背景之间的灰度值方差最大)获取合适的灰度值阈值[15],使用二进制方法,将灰度值大于设定阈值的区域标记为值“1”,反之,标记该值为“0”。然后,保留柔性板的位置矩阵。最后,通过获取该位置矩阵的下边界,得到 x 方向的单值函数,通过插值和多项式拟合方法获得柔性板的曲线。该过程如图 4 所示,其中步骤 1 为区域选择与二值化过程,步骤 2 为柔性翼的轮廓识别曲线。该方法可以识别柔性翼某一确定截面的弦向变形,适用于二维变形测量。精度可达到单像素级别,分辨率为 0.119 mm/pixel,不确定度约为 0.08% c(0.119/150≈0.08%)。

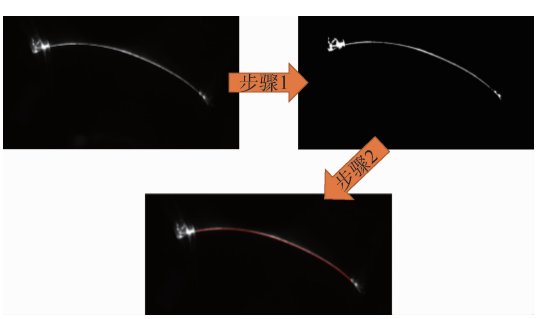


图 4　柔性翼识别

Fig. 4　Recognition of flexible membrane wing

2　柔性翼气动力特性

柔性翼与刚性翼的升力特性如图 5(a)所示,刚性翼的失速迎角 $\alpha = 14°$,对应的升力系数 C_L 为 1.35,柔性翼的失速迎角 $\alpha = 20°$,对应的最大升力系数为 1.99,柔性翼失速迎角相较于刚性翼延后了 6°,最大升力系数提升了 47.4%。当 $\alpha = 20°$时增升效果最好,升力系数提升 0.87(提升 76.9%)。

升力的增大伴随着升致阻力的增大,柔性翼的阻力在大多数迎角工况下是增大的,如图 5(b)

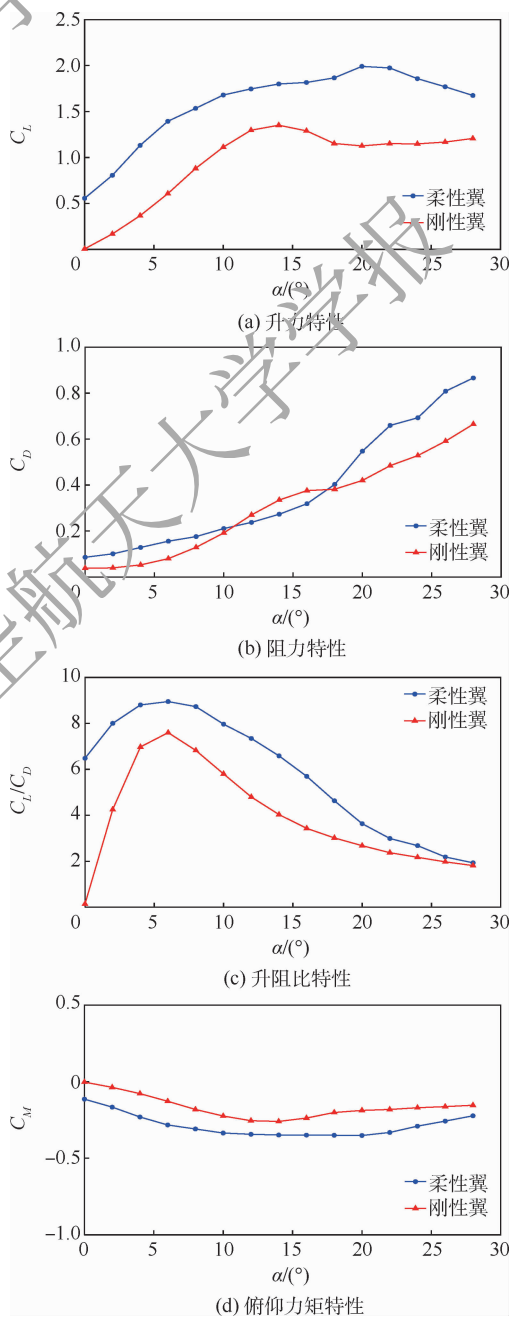


图 5　柔性翼气动力特性

Fig. 5　Aerodynamic characteristics of flexible membrane wing

北京航空航天大学学报

北航学报 赠阅

所示。全迎角工况下,阻力系数 C_D 增量普遍在 0.2 以下,增阻代价不大。柔性翼只有在 $\alpha=12°\sim16°$ 工况下是减阻的,$\alpha=14°$ 时减阻达到最大值,阻力系数下降 0.062(减阻 18.5%)。

升阻比 $K=C_L/C_D$ 表示机翼的气动效率,图 5(c)为柔性翼与刚性翼升阻比特性曲线,$\alpha=6°$ 时柔性翼与刚性翼的升阻比均达到最大值,分别为 8.95 和 7.60,柔性翼相较于刚性翼最大升阻比提升了 17.8%。除了迎角 $\alpha=0°$ 外,升阻比提升量最大的迎角是 $\alpha=14°$,提升了 2.55,升阻比提升比例最大的迎角是 $\alpha=16°$,提升了 66.0%。

柔性翼与刚性翼的俯仰力矩特性如图 5(d)所示,取矩点为距前缘 $c/4$ 的位置。刚性翼首次出现俯仰力矩系数 $C_{M\alpha}>0$ 的工况为 $\alpha=14°$,柔性翼首次出现 $C_{M\alpha}>0$ 的工况为 $\alpha=20°$ 迎角,延后了 6°。出现静不稳定的迎角工况与失速迎角相对应。另外 $C_{M\alpha}<0$ 时柔性翼的斜率有微弱的增幅,柔性翼的平均斜率为 $-1.394\ \mathrm{rad}^{-1}$,刚性翼的平均斜率为 $-1.279\ \mathrm{rad}^{-1}$。柔性翼在 $10°\sim20°$ 呈现近似的中立稳定性,可见柔性翼在一定程度上起到了增强静稳定性的效果。

对每个迎角下气动力的瞬时值进行快速傅里叶变换(fast Fourier transform, FFT),可得到气动力的频域特性,图 6 给出了 4 个典型工况的结果。对气动力的功率谱密度(power spectral density, PSD)值最大时对应的频率值进行无量纲化处理,得到主频的 St 数:

$$St=\frac{fc}{U_\infty} \tag{1}$$

式中:f 为气动力的主频;c 为柔性翼弦长;U_∞ 为自由来流速度。

图 7(a)为升力主频 St 数随迎角的变化,St 数的范围在 0.4~1.2 之间,与 Gordnier[16] 的研究很接近。在 $Re=2\ 500$、$\alpha=4°\sim20°$ 工况下得到柔性翼振动与流场结构的 St 数分布在 0.6~1.43 之间。刚性翼升力系数的主频在迎角 $\alpha\geqslant10°$ 后近似不变,稳定在 25 Hz,St 约为 0.3。但是柔性翼出现了主频的大幅度变化,在 $\alpha=12°\sim16°$ 呈现 100 Hz 的主频,St 达到 1.2,在 $\alpha=16°$ 以上呈现 40~50 Hz 的主频,St 数范围为 0.48~0.6。图 8(a)为各迎角工况下升力功率谱密度最大值 $\mathrm{PSD}_{\max}$,可以利用 $\mathrm{PSD}_{\max}$ 大小反映主频相较于其他频率的突出程度[17]。刚性翼的主频功率谱密度始终不突出,但是柔性翼的功率谱密度在失速后 $\alpha=26°$ 下达到最大。

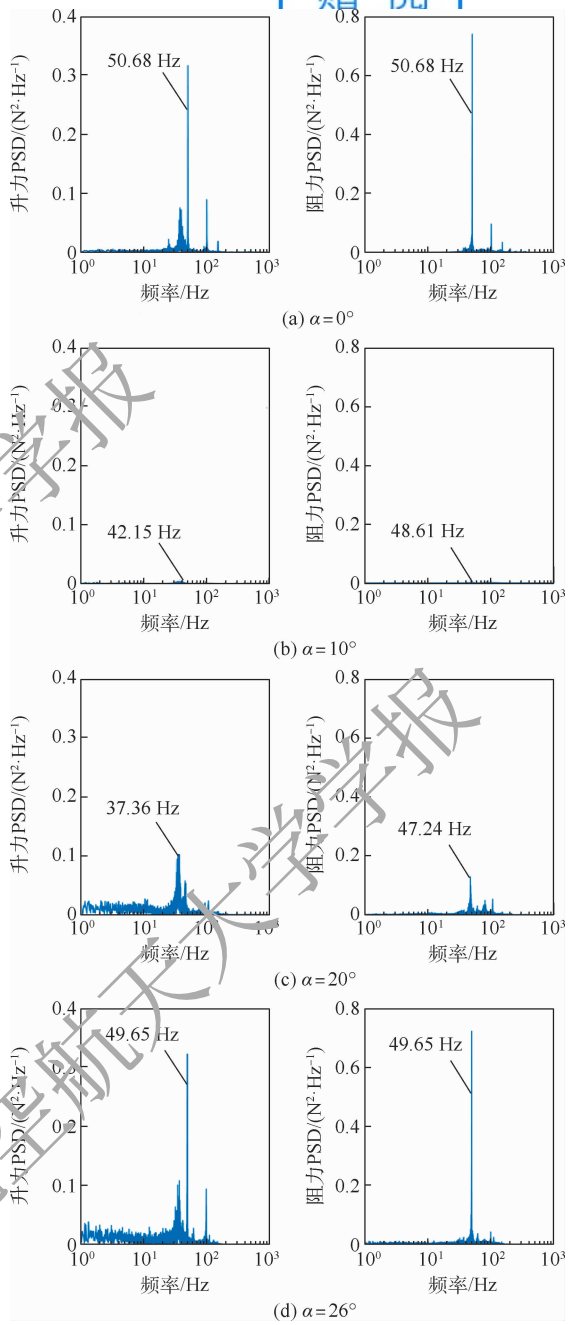


图 6 气动力频域特性

Fig. 6 Frequency-domain characteristics of aerodynamic forces

图 7(b)显示出柔性翼与刚性翼阻力系数主频的对比,与升力主频特性相近,但有所区别。刚性翼阻力主频在迎角 $\alpha\geqslant8°$ 后近似不变,同样稳定在 25 Hz,St 数约在 0.3。柔性翼阻力仅在 $\alpha=18°$ 迎角呈现 100 Hz 的主频,St 约为 1.2,在其他迎角(小迎角工况除外)呈现约 50 Hz 的主频,St 数约为 0.6。从图 8(b)也可以看出,阻力与升力的功率谱密度最大值趋势基本一致,柔性翼振动和变形可以改变气动力的主频,同时可以增加气动力的功率谱密度最大值。

北京航空航天大学学报

北航学报 网阅

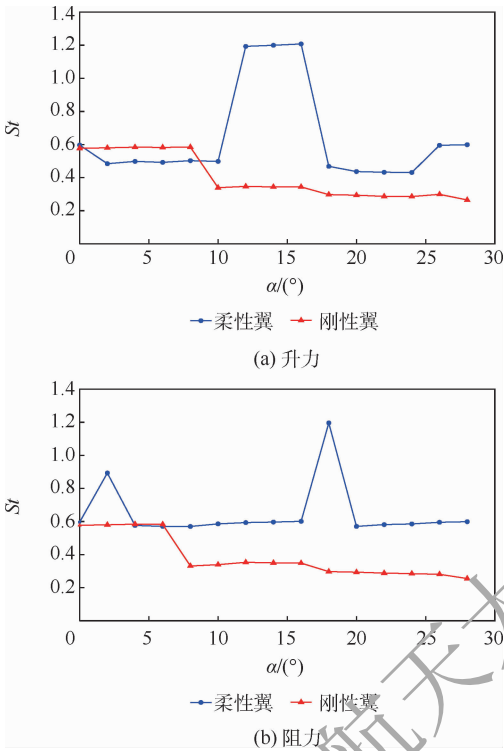


图 7　*St* 数随迎角变化关系

Fig. 7　Variation of *St* with angle of attack

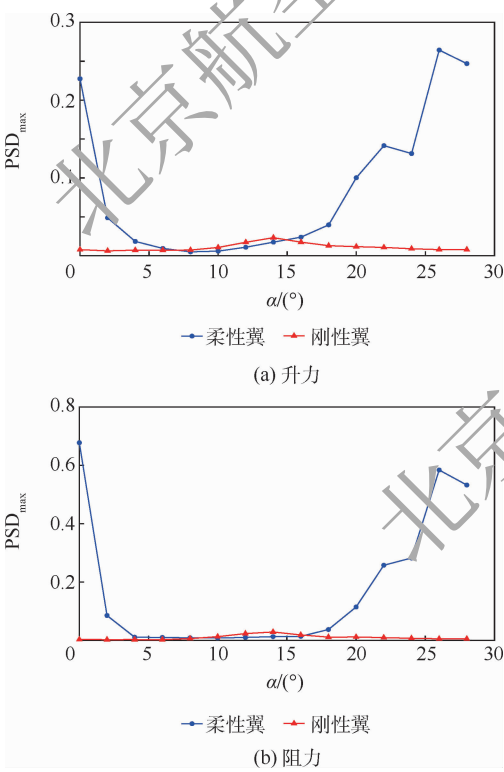


图 8　气动力 PSD_{max} 随迎角变化关系

Fig. 8　Variation of PSD_{max} of aerodynamic forces with angle of attack

3　柔性翼变形和振动特性

3.1　最大变形

对于柔性翼的最大变形计算,将在图 2 中介绍的 x_{tran}-z_{tran} 坐标系中进行统计计算,求出机翼弦向各点距水平位置 x_{tran} 的最大值。图 9 为旋转变形后获得各迎角工况最大变形结果,其中,变形最大对应的升力系数也最大。这一结果对主动变形改善飞行器气动特性提供了控制策略,可以通过增加柔性膜变形程度来提供较大的升力,其中,翼弯度约为 10% c 时具有最大升力。

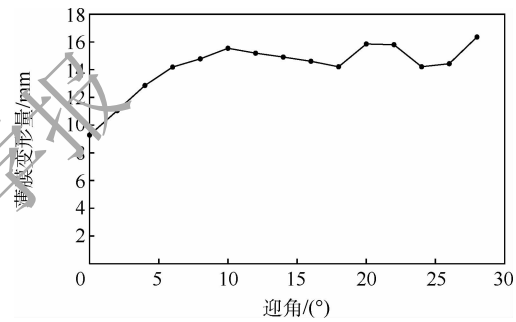


图 9　柔性翼最大变形

Fig. 9　Maximum deformation of flexible membrane wing

3.2　最大变形弦向位置

最大变形的弦向位置在一定程度上反映了俯仰力矩特性和压强的弦向分布。图 10 为最大变形的弦向位置随迎角的变化趋势。可见,随迎角的增大,最大变形的弦向位置从弦向的前半部分逐渐向后缘移动,然后再缓慢向前缘移动。这与俯仰力矩在中等迎角斜率平缓及大迎角后面出现不稳定工况的现象相吻合。在 10° ~ 20°范围内,最大变形弦向位置随迎角增加变化很微弱。俯仰力矩随迎角变化的导数为

$$\frac{\partial C_M}{\partial \alpha} = \frac{\partial C_M}{\partial C_L} \cdot \frac{\partial C_L}{\partial \alpha} = C_{L\alpha}(\bar{x}_{c\cdot g} - \bar{x}_{ac}) \tag{2}$$

式中:$C_{L\alpha}$为升力系数随迎角的变化率;$\bar{x}_{c\cdot g}$为机翼的质心位置;$\bar{x}_{ac}$为机翼的焦点位置。柔性翼在该范围出现明显中立稳定性的特征,是因为该部分气动力焦点在该迎角范围内近似不变,且接近于机翼的质心位置。这一发现给激励位置的选取提供了依据,变形的激励位置维持在弦向的中间位

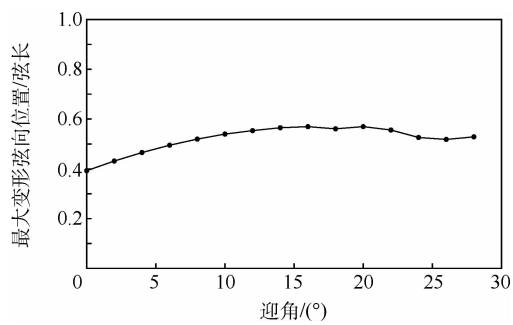


图 10　柔性翼最大变形的弦向位置

Fig. 10　Chordwise position of maximum deformation of flexible membrane wing

北京航空航天大学学报

北航学报 网阅

置附近，其中，升力最大的情况对应距前缘57.0% c的位置。

3.3 振动特性

柔性翼振动的幅值大小以柔性翼垂直位移的标准差来表示，定义式为

$$z_{\mathrm{std}}(x)=\sqrt{\frac{1}{T}\sum_{t=0}^{T}(z(x,t)-\bar{z}(x,t))^2} \tag{3}$$

式中：T为采样总时间；t为每次采样的瞬时时刻；$\bar{z}(x,t)$为柔性翼垂直位移的均值。

柔性翼垂直位移的标准差沿弦向的分布规律如图11所示。在振动幅值较大的迎角工况范围内，柔性翼振动的峰值呈现三波峰到单波峰的变换，这与Rojratsirikul等[18]的研究结果相一致。

振动随迎角的变化可以分成3个不同的区域：①$\alpha=0°\sim2°$为振动幅值较大但静变形较小的区域；②$\alpha=4°\sim16°$对应静变形明显但振动幅值较小的区域；③$\alpha\geqslant18°$为振动幅值与静变形都突出的区域，在该区域中振动幅值沿向分布规律呈现了三波峰到单波峰之间的转换，如图11所示。各区域对应的典型变形脉动图和频谱图如图12所示。

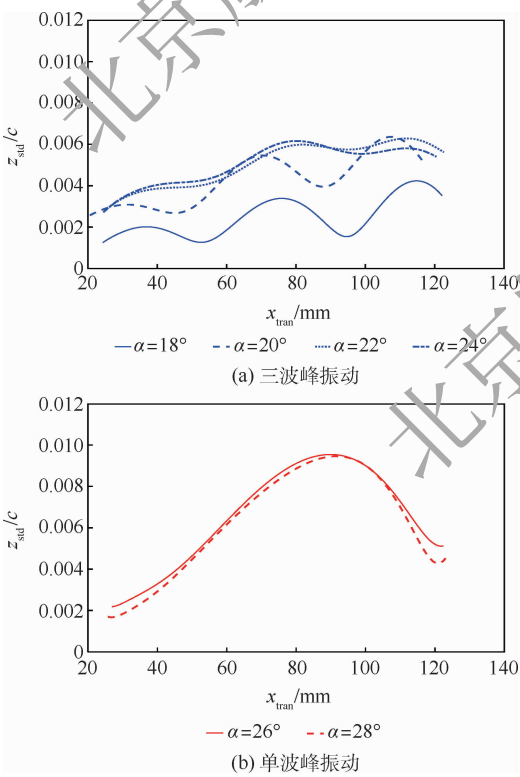


图11 柔性翼振动幅值沿弦向分布规律

Fig. 11 Vibration amplitude chordwise distribution law of flexible membrane wing

$\alpha=0°$附近呈现较大的振幅，可能是由于空气动力载荷低（缺乏膜张力）和柔性膜的双稳态特性造成的[19]。$\alpha=4°\sim16°$时，相位平均后的柔性翼变形脉动值比其他工况小一个数量级，导致在频谱图上高亮区域不明显，此时存在明显变形但是振动振幅并不突出，呈现只存在静变形的状态。$\alpha=18°\sim24°$时，因为气动力增加导致薄膜的张力增加，既具有较大变形，也具有较大的振动振幅。振动的外包络线具有三波峰时，$St=1.20$，PSD突出的区域集中在弦向距前缘20%、50%、80%的3处。然而，当迎角继续增加，柔性翼变形的脉动逐渐增强，当迎角达到$\alpha=26°$时，会出现波峰数下降为单波峰的情况，此时$St=0.58$，PSD突出的区域主要集中在弦向距前缘20%～80%的位置。

选取三波峰振动工况迎角$\alpha=20°$的工况与单波峰振动工况迎角$\alpha=26°$的工况，将振动与升力的频谱图进行对比，分别如图13和图14所示。

$\alpha=20°$时，升力与振动的PSD_{ma}所对应的主频存在一定差异性，但振动频谱上的每一个峰值在升力频谱图上都有对应的峰值。He和Wang[20]的研究也发现了类似的现象，流场频谱的峰值在振动频谱上都有对应的峰值，而流场与气动力密切相关。此时，柔性翼的振动受到气动力的影响，差异性源于柔性翼的刚度，在忽略预张力的情况下，刚度一部分依赖气动刚度，一部分依赖变形后产生的面内张力[21]。$\alpha=26°$时，升力与振动的PSD峰值所对应的主频高度一致，此时振动情况为气动力导致的受迫振动。

对各迎角工况下振动幅值的最大值大小进行统计。$\alpha=20°$时，振幅极大值对应失速迎角，$\alpha=26°$则对应振动波峰数呈现三波峰到单波峰的突变，即由高频小振幅转变为低频大振幅。柔性翼振幅、升力功率谱密度最大值对照如图15所示。可见，柔性翼振动与气动力直接耦合，柔性翼振幅的大小决定气动力的主频，进而达到调节气动力的效果。

这一结果同样对主动变形改善飞行器气动特性提供了控制策略。除了可以主动变形增加机翼弯度外，也可以提供振动激励来增加升力。在增升效果最好的$\alpha=20°$迎角工况下为具有三波峰的振动工况，振幅约为0.67% c。可以在机翼弦向20%、50%、80%处施加主动振动激励来实现更好的升力特性，振动的$St=1.20$，实验弦长对应的振动频率为100 Hz。

北京航空航天大学学报

北航学报 赠阅

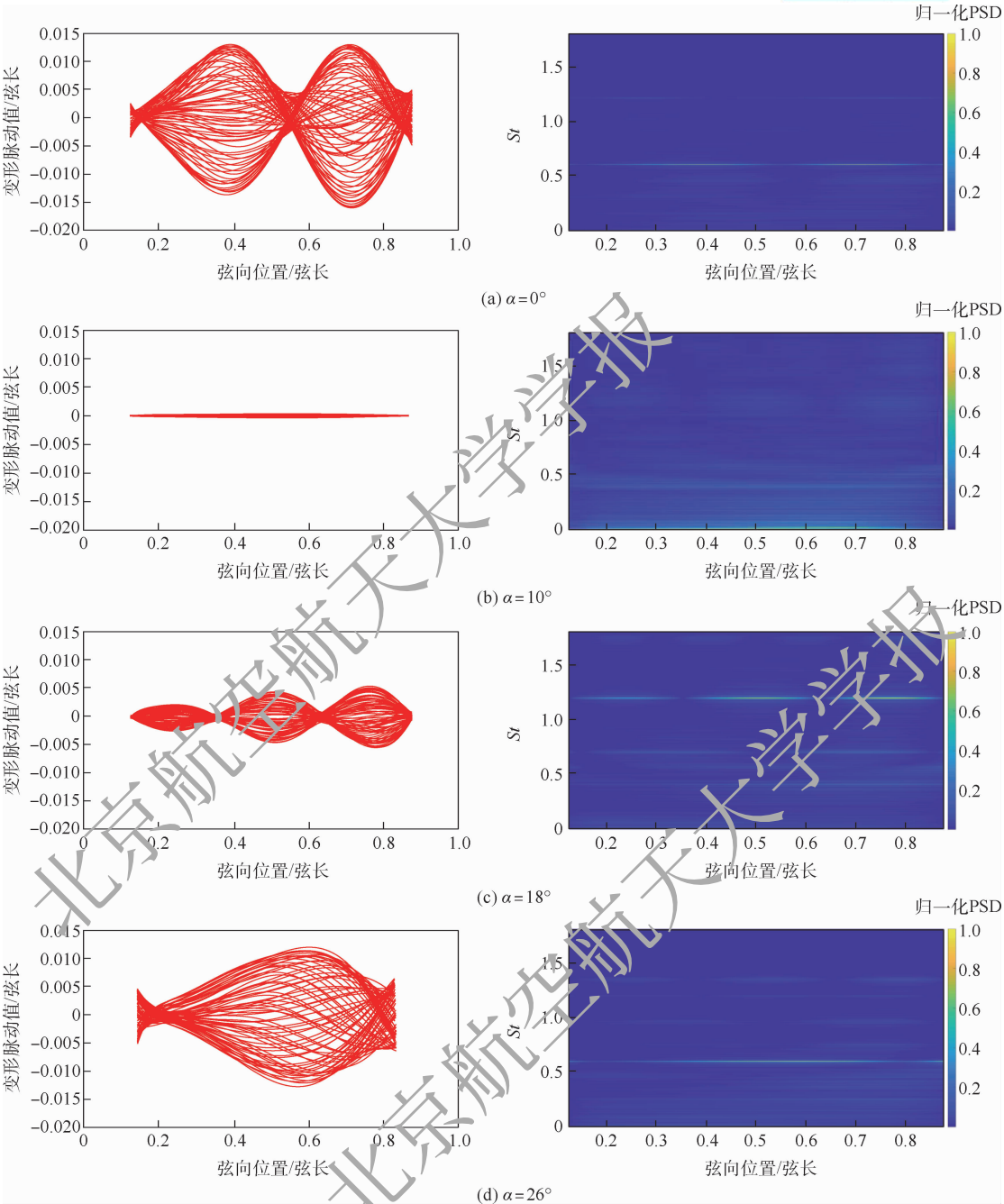


图 12　柔性翼相位平均后变形脉动值和频率特性

Fig. 12　Phase-averaged deformation fluctuations and frequency characteristics of flexible membrane wing

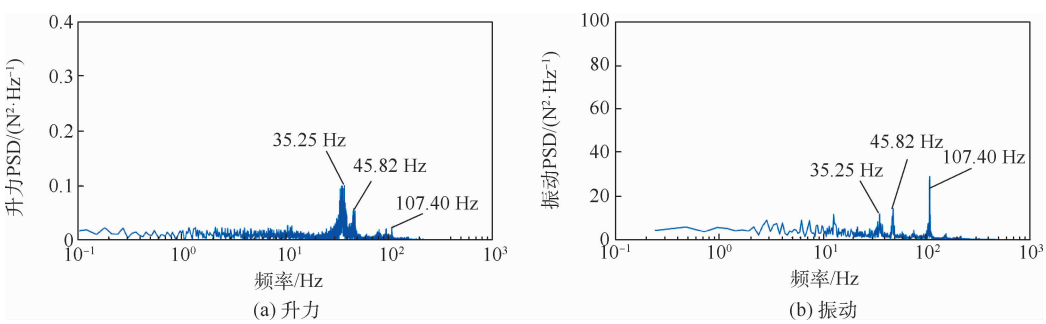


图 13　α = 20°时升力与振动频率特性

Fig. 13　Frequency characteristics of lift and vibration at α = 20°

北航学报 醒阅

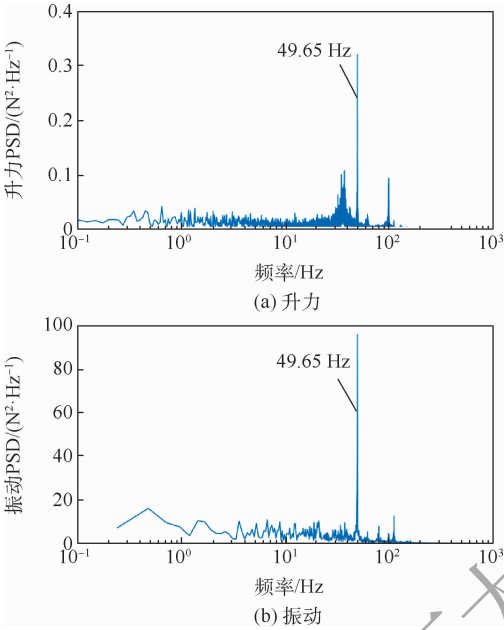


图 14 $\alpha=26°$时升力与振动频率特性

Fig. 14 Frequency characteristics of lift and vibration at $\alpha=26°$

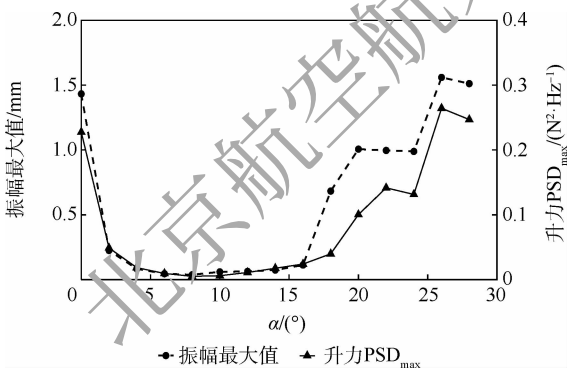


图 15 振幅与升力 PSD_{max} 比较

Fig. 15 Comparison between vibration amplitude and PSD_{max} of flexible wembrane wing

4 结 论

本文对柔性翼气动、变形等特性进行了同步测量,对气动力变化规律进行了分析,为柔性翼主动变形控制提供了依据。主要的结论如下:

1)气动力的测量结果表明,相比于刚性翼,实验条件下柔性翼显著提高了模型的气动特性,失速迎角推迟了 6°,最大升力系数和升阻比分别提高了 47.4% 和 17.8%。

2)柔性翼的弯度及最大变形的弦向位置随迎角变化,柔性翼的周期性振动除了 $\alpha=0°\sim2°$ 为大振幅、小静变形外,振动的振幅随迎角增加经历无明显波峰、具有三波峰到具有单波峰的转换,此外,柔性翼变形和振动使气动力的功率谱密度最大值增加。

3)当柔性翼的弯度增加或发生较明显的振动时,增升效果明显。主动控制策略可以参考增升效果最佳的工况,此时柔性翼在距前缘 57.0% c 位置处弯度最大,约为 10.0% c,相应的柔性翼振动的振幅约为 0.67% c。

参考文献(References)

[1] SHYY W, IFJU P, VIIERU D. Membrane wing-based micro air vehicles[J]. Applied Mechanics Reviews, 2005, 58(4): 283-301.

[2] LISSAMAN P B S. Low-Reynolds-number airfoils[J]. Annual Review of Fluid Mechanics, 2003, 15(1): 223-239.

[3] SWARTZ S M, IRIARTE-DIAZ J, RISKIN D K, et al. A bird? A plane? No, it's a bat: An introduction to the biomechanics of bat flight[M]. Cambridge: Cambridge University Press, 2012: 317-351.

[4] SWARTZ S M. Allometric patterning in the limb skeleton of bats: Implications for the mechanics and energies of powered flight[J]. Journal of Morphology, 1997, 234(3): 277-294.

[5] SWARTZ S M, BENNETT M B, CARRIER D R. Wing bone stresses in free flying bats and the evolution of skeletal design for flight[J]. Nature, 1992, 359(6397): 726-729.

[6] TIAN X D, IRIARTE J, MIDDLETON K, et al. Direct measurements of the kinematics and dynamics of bat flight[C]// 36th AIAA Fluid Dynamics Conference and Exhibit. Reston: AIAA, 2006, 1: 10-18.

[7] WANG S Z, ZHANG X, HE G W, et al. Lift enhancement by bats' dynamically changing wingspan[J]. Journal of the Royal Society Interface, 2015, 12(113): 2-11.

[8] SHYY W, JENKINS D, SMITH R. Study of adaptive shape airfoils at low reynolds number in oscillatory flows[J]. AIAA Journal, 1997, 35(9): 1545-1548.

[9] LEVIN O, SHYY W. Optimization of a low reynolds number airfoil with flexible membrane[J]. Computer Modeling in Engineering and Sciences, 2001, 2(4): 523-536.

[10] BÉGUIN B, BREITSAMTER C, ADAMS N. Experimental investigations of an elasto-flexible morphing wing concept[C]// 27th Congress of the International Council of the Aeronautical Sciences, 2010, 9: 1-11.

[11] BÉGUIN B, BREITSAMTER C, ADAMS N. Aerodynamic investigations of a morphing membrane wing[J]. AIAA Journal, 2012, 50(11): 2588-2599.

[12] HE X, GUO Q F, WANG J J. Extended flexible trailing-edge on the flow structures of an airfoil at high angle of attack[J]. Experiments in Fluids, 2019, 60(8): 122-143.

[13] GUO Q F, HE X, WANG Z, et al. Effects of flexible wing on the aerodynamic performance of an aircraft model[J]. Chinese Journal of Aeronautics, 2021, 34(9): 133-142.

[14] ROJRATSIRIKUL P, WANG Z, GURSUL I. Effect of pre-strain and excess length on unsteady fluid-structure interactions of membrane airfoils[J]. Journal of Fluids and Structures, 2010, 26(3): 359-376.

[15] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernet-

北京航空航天大学学报

ics,2007,9(1):62-66.

[16] GORDNIER R. High fidelity computational simulation of a membrane wing airfoil[C]// Proceedings of the 46th AIAA Aerospace Sciences Meeting And Exhibit. Reston:AIAA,2008,1:7-10.

[17] MATLIS E H,HE C,CORKE T C,et al. Sensing and control of flow separation using plasma actuators[J]. Philosophical Transactions of the Royal Society:Mathematical,Physical,and Engineering Sciences,2011,369(1940):1459-1475.

[18] ROJRATSIRIKUL P,WANG Z,GURSUL I. Unsteady fluid-structure interactions of membrane airfoils at low reynolds numbers[J]. Experiments in Fluids,2009,46(5):859.

[19] NEWMAN B G,LOW H T. Two-dimensional impervious sails:Experimental results compared with theory[J]. Journl of Fluid Mechanics,1984,144:445-462.

[20] HE X,WANG J J. Fluid-structure interaction of a flexible membrane wing at a fixed angle of attack[J]. Physics of Fluids,2020,32(12):127102.

[21] 叶正寅,张伟伟,史爱明. 流固耦合力学基础及其应用[M]. 哈尔滨:哈尔滨工业大学出版社,2010:86-109.
YE Z Y,ZHANG W W,SHI A M. Fundamentals of fluid-solid coupling mechanics and its application[M]. Harbin:Harbin Institute of Technology Press,2010:86-109(in Chinese).

Experimental study on aerodynamic and deformation characteristics of flexible membrane wing

WANG Hanbin, HE Xi, WANG Jinjun*

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: Micro air vehicles (MAV) have wide appications in both miltary and civilian fields, and flexible membrane wings are an effective method to improve the aerodynamic of MAV. In order to better control the flexible wing, the vibration and deformation characteristics of the flexible membrane wing and their impact on aerodynamic force are measured synchronously. Compared with the rigid wing, for the flexible membrane wing, the stall angle of attack is delayed by 6°, the maximum lift coefficient is increased by 47.4%, and the lift-drag ratio is increased by 17.8%. In addition to the characteristics of large amplitude and small static deformation at $\alpha = 0° \sim 2°$, the amplitude of the periodic vibration of the flexible membrane wing undergoes a transition from no obvious crest, three crests to one crest as the angle of attack increases. When the lift coefficient is maximum, the corresponding membrane deformation and vibration amplitude both reach their maxima. Besides, the chordwise position of the maximum deformation changes with the angle of attack, which determines the pitching moment characteristics. Based on these results, an active control method of improving aerodynamic performance by applying deformation and vibration excitation with specific frequency is proposed.

Keywords: flexible membrane wing; fluid-structure coupling; aerodynamic force measurement; deformation measurement; vibration analysis

Received: 2020-11-04; **Accepted**: 2020-12-25; **Published online**: 2021-01-12 13:47

URL: kns.cnki.net/kcms/detail/11.2625.V.20210111.1813.002.html

Foundation items: NSFC Projects of International Cooperation and Exchanges (11761131009); Science Fund for Creative Research Groups (11721202)

* **Corresponding author**. E-mail: jjwang@buaa.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0621

基于联合分布适配的智能电表运行状态评价

全下文载

许丹[1,*]，李游[1]，李贺龙[2]
北京航空航天大学 可靠性与系统工程学院，北京 100083； 2. 中国电力科学研究院有限公司，北京 100085）

摘 要：针对智能电表运行状态评价中经常存在带标签的数据采集困难且不同地区的数据分布不一致的问题，在智能电表的运行状态评价领域引入迁移学习中的联合分布适配（JDA）算法。该算法寻找一个最优化的变换矩阵，使得在变换后的空间中不同地区数据的边缘分布和条件分布距离最小化。针对条件分布适配中目标域没有数据标签的问题，采用伪标签迭代的方法使得目标域伪标签不断接近真实标签。变换后空间数据中训练得到的分类模型可以运用于新的地区，实现迁移。实验结果证明了 JDA 算法在智能电表运行状态评价中的有效性。

关 键 词：联合分布适配（JDA）；智能电表；运行状态评价；迁移学习；智能电网
中图分类号：TM744；TP312
文献标志码：A **文章编号**：1001-5965(2022)04-0674-08

随着电力在现代社会中越来越举足轻重，实现电网智能化对支撑国家和社会发展的意义越来越重大。智能电表是实现电网智能化的基础设备之一，据统计，截至 2017 年年底，国家电网公司安装智能电表超过 4 亿只[1]，电力公司通常定期用新智能电表替换已达到一定运行时间的旧智能电表，该方法将会替换一些仍可以正常工作的智能电表[2]，往往会造成维修资源的大量浪费。因此，国家电网公司迫切地需要一种能够快速准确评价智能电表运行状态的方法来对电表的检修和轮换计划提供指导，从而实现管理成本上的节约及更低的故障率。评价智能电表健康状态的主流方法是数据驱动方法，该方法利用多源信息融合技术及人工智能等相关算法来整合智能电表运行过程中产生的数据及与智能电表运行状态相关的各类数据，从而实现对智能电表运行状态评价的精细化和智能化。Cai 等[3]采用信息融合方法建立了智能电表运行状态模型，对电表的评价向量进行量化，结合熵权法得到电表的运行状态得分，将电表根据得分划分为状态等级。Li 等[4]在分析计量装置电压异常主要原因的基础上，采用决策树算法建立了计量装置电压判断模型，结果显示该模型能够较好地筛查出存在计量异常的电表。叶剑斌等[5]利用智能电表全寿命周期各业务管理系统的数据，确定评价指标，用熵权法确定权重，得出了评价分数并使用专家经验法将其划分为不同状态。祝宇楠等[6]采用数据挖掘算法建立了现场智能电表故障数据与检定数据之间的联系，验证了采用数据挖掘技术分析检定数据以提前获得智能电表故障信息的可能性及有效性。韩笑[7]针对国家电网的状态数据，选用遗传算法做特征选择，使用最大间隔贝叶斯网（maximum margin Bayesian network，MMBN）分类器对智能电表状态进行分类，得到了比支持向量机（support vector machine，SVM）更好的分类效果。Cheng 和 Shen 等[8-9]建立了能够检测智能电表运行状态的远程监测系统，通过数据交互和在线监测实时了解电表的运行状况，从而实现对电表的实时在线

收稿日期：2020-11-09；**录用日期**：2021-06-11；**网络出版时间**：2021-10-26 15:15
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20211026.1434.001.html
基金项目：国家自然科学基金（51875017）；中央高校基本科研业务费专项资金（YWF-20-BJ-J-726）
* **通信作者**. E-mail：xudan@buaa.edu.cn

引用格式：*许丹，李游，李贺龙. 基于联合分布适配的智能电表运行状态评价[J]. 北京航空航天大学学报，2022，48(4)：674-681.*
XU D，LI Y，LI H L. Operating status evaluation of smart electricity meters based on joint distribution adaption[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：674-681 (in Chinese).

北京航空航天大学学报

故障分析和远程维护。王新刚等[10]提出了一种基于关联规则挖掘的智能电表在线监测方法,使用传输层实时监测电表的异常运行情况,对数据进行分析后形成故障信息,应用层采用离线运行方式从历史数据库中抽取关联规则,并以在线方式将故障信息转换后进行匹配,实现故障诊断。

然而,使用数据驱动方法进行智能电表运行状态评价面临 2 个问题:①电表数据,尤其是带运行状态标签的电表运行数据往往难以获得,因此,如何充分利用和挖掘现有数据的内在信息将能够节约数据采集的成本,为决策提供更多的依据,具有一定的现实价值;②传统的有监督机器学习方法是建立在大量数据符合同分布的假设之上,而这个假设使得在使用不同分布数据时会导致分类或预测结果较差,从而使得不同地区的电表运行状态分类模型不能直接应用,大大限制了机器学习方法的应用。将已有分类模型应用到其他地区,将会降低新建分类模型的成本及新建分类模型的难度。

迁移学习是机器学习方法的一个分支,其主要思想是将在某一个领域中学到的知识采用某种方法运用到其他领域。本文引入迁移学习中的联合分布适配(joint distribution adaptation, JDA)算法[11]进行智能电表运行状态评价模型的迁移,解决了某地区电表运行状态评价模型无法直接运用到新地区的问题,使得能够利用已有的带标签数据来帮助评价其他地方的智能电表运行状态,扩大模型的适用性,提升数据利用效率。

1　智能电表运行状态评价方法框架

为实现智能电表运行状态评价模型的迁移,运用了 JDA 算法。该算法需要使用目标域的数据标签,而在该迁移场景下,目标域的数据标签不存在,因此需要先对目标域数据进行预标注。所谓预标注,即将目标域数据输入到由源域数据及标签训练出的分类模型中,得到目标域数据的初始标签。初始标签虽然不准确,但可以在迭代过程中一步步接近真实的标签。在给目标域数据进行预标注初始标签的过程中,采用合成少数类过采样技术(synthetic minority over-sampling technique, SMOTE)解决了数据的不平衡问题,降低了源域数据的不平衡问题给分类模型带来的负面影响。研究路线如图 1 所示。

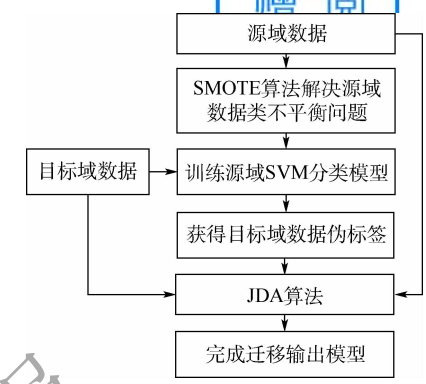


图 1　研究线路

Fig. 1　Procedure of study circuit

2　算法介绍

2.1　基于 SMOTE 算法的不平衡数据处理

数据的类不平衡问题是指在数据中多数类数据是少数类数据的几十倍甚至上百倍,往往会造成机器学习算法性能的下降。SMOTE[12]是一种解决数据类不平衡问题的有效算法,该算法采用已有少数类样本合成新的数据,使得已有少数类样本数量增加,从而解决数据不平衡问题。

若某个带标签的少数类有 m 个样本,每个样本有 n 个属性,其类别标签为 $l_{\min}$,则有少数类矩阵 $\boldsymbol{B}_{\min}=(b_{ij})_{m\times n}$:

$$\boldsymbol{B}_{\min}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n}\,|\,l_{\min} \\ b_{21} & b_{22} & \cdots & b_{2n}\,|\,l_{\min} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn}\,|\,l_{\min} \end{bmatrix}$$

SOMTE 算法基本步骤如下:

步骤 1　对于少数类中每一个样本 b_i,以欧氏距离为标准计算其到少数类样本集中其他所有样本的距离。

步骤 2　根据样本不平衡比例设置采样比例以确定采样倍率 N,对于每一个少数类样本 $b_i(i=1,2,\cdots,m)$,根据其与其他样本的距离,从中选择 k 个距离最近的样本,即 k 近邻,假设选择的近邻为第 j 个样本 $b_j(j=1,2,\cdots,k;i\neq j)$。

步骤 3　根据下式合成新的样本:

$$b_{sp}=b_{ip}+\mathrm{rand}(0,1)\cdot(b_{ip}-b_{jp})$$

$$i=1,2,\cdots,m;j=1,2,\cdots,k;p=1,2,\cdots,n$$

式中:b_p为新合成的样本在第 p 个属性下的值;b_{ip}为少数类样本 i 在第 p 个属性下的值;b_{jp}为样本 b_i 的某一个近邻 b_j 在属性 p 下的值;rand(0,1)为 0 ~ 1 之间的随机数。

2.2　基于 JDA 算法的智能电表状态评价

JDA 算法的核心思想在于通过寻找一个变换

北京航空航天大学学报

北航学报 网阅

矩阵,将源域和目标域的数据都变换到其他空间,在变换后的空间将源域与目标域之间的数据分布距离最小化。在寻找变化的过程中,需要给变换矩阵提供约束,以达到寻找最优矩阵的目的。因此,先用最大均值差异[13](maximum mean discrepancy,MMD)定义了源域和目标域之间的边缘分布和条件分布之间的差异,以此作为约束条件;然后,将其作为寻找变换矩阵中的优化目标;最后,将优化目标进行综合,求解得到变换矩阵的最优解。然而,在进行源域与目标域之间的条件分布适配时,需要用到目标域的数据标签。在当前的迁移环境下,目标域的数据不含运行状态标签,因此,采用数据伪标签进行代替。但是,由于源域与目标域之间数据分布的差异,伪标签是不准确的,采用迭代的方法使目标域的数据伪标签一步步接近真实标签,直到分类准确度收敛为止。研究的方法步骤如下:

1)特征变换方法。

在机器学习中,存在领域和任务 2 个概念。一个领域 D 由 m 维的特征空间 X 和边缘概率分布 $P(x)$ 组成,即 $D=\{X,P(x)\}$。给定领域 τ,任务由 C 类标签集 Y 和分类器 $f(x)$ 构成,即 $\tau=\{Y,f(x)\}$。在实际机器学习问题中,被迁移的领域成为源域,迁移的领域成为目标域,不同领域的数据分布往往不同。若存在带标签源域 $\mathcal{D}_s=\{(x_1,y_1),\cdots,(x_{n_s},y_{n_s})\}$ 和无标签目标域 $\mathcal{D}_t=\{x_{n_s+1},\cdots,x_{n_s+n_t}\}$,并设定源域和目标域的特征空间及类别空间相同,即 $X_s=X_t,Y_s=Y_t$,但两者的边缘分布和条件分布都不相同,即 $P_s(x_s)\neq P_t(x_t),Q_s(y_s|x_s)\neq Q_t(y_t|x_t)$。JDA 算法的核心思想是:将原始数据经过某个变换之后在其他空间能同时减少边缘分布 $P_s(x_s),P_t(x_t)$ 及条件分布 $Q_s(y_s|x_s),Q_t(y_t|x_t)$ 之间的差异,即实现源域和目标域联合分布的适配,在变换后的空间使用机器学习算法,输出训练好的模型,实现迁移。

设有待变换矩阵 $\boldsymbol{X}=[x_1,x_2,\cdots,x_n]\in\mathbf{R}^{m\times n}$,主成分分析(principal component analysis,PCA)方法是一种有效的特征变换方法[14],其能够通过优化式(1)学习到一种特征变换:

$$\max_{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}}\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}) \tag{1}$$

式中:$\boldsymbol{H}=\boldsymbol{I}-\frac{1}{n}\mathbf{1}$ 为协方差中心矩阵,$\boldsymbol{I}$ 表示元素全为 1 的 n 阶矩阵,$\mathbf{1}$ 代表单位矩阵。该优化目标是为了在进行变换时实施约束,最大化保留变换后数据的信息与结构。该优化问题已经可以通过特征分解的方式进行求解,其解为协方差矩阵 $\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}$ 前 k 个最大特征值对应的特征向量构成。因此,新的最优化 k 维特征可以表示为 $\boldsymbol{Z}=[z_1,\cdots,z_n]=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}$。

2)边缘分布适配。

要寻找出能够同时最小化目标域及源域的边缘分布、条件分布距离的变换矩阵,需要在进行变换的过程中给变换矩阵添加约束条件,从而求得变换的最优解。采用源域与目标域之间的分布差异作为优化目标。MMD 是一种能够有效度量概率分布之间矩的差异距离计算方法,由于其不需要对分布的概率密度进行估计,是一种非参数化的距离计算方式。源域的边缘分布 $P_s(x_s)$ 与目标域的边缘分布 $P_t(x_t)$ 的 MMD 距离计算公式为

$$\operatorname{MMD}_{\mathcal{H}}^2(P_s,P_t)=\left\|\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^{\mathrm{T}}x_i-\frac{1}{n_t}\sum_{j=n_s+1}^{n_s+n_t}\boldsymbol{A}^{\mathrm{T}}x_j\right\|^2 \tag{2}$$

式中:$\mathcal{H}$ 为无穷维希尔伯特空间。

根据文献[14],式(2)可以转化为

$$\left\|\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^{\mathrm{T}}x_i-\frac{1}{n_t}\sum_{j=n_s+1}^{n_s+n_t}\boldsymbol{A}^{\mathrm{T}}x_j\right\|^2=\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}_0\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}) \tag{3}$$

式中:

$$(\boldsymbol{M}_0)_{ij}=\begin{cases}\dfrac{1}{n_s n_s} & x_i,x_j\in D_s\\ \dfrac{1}{n_t n_t} & x_i,x_j\in D_t\\ \dfrac{-1}{n_s n_t} & \text{其他}\end{cases} \tag{4}$$

其中:n_s 和 n_t 分别为源域和目标域样本的数量。通过最小化式(3)即能够实现源域与目标域之间的边缘分布距离的最小化,从而实现两者边缘分布的适配[15]。在新的特征表达 $\boldsymbol{Z}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}$ 下,源域与目标域的边缘分布将被拉近。

3)条件分布适配。

条件分布适配即在变换之后最小化源域与目标域条件分布 $Q_s(y_s|x_s)Q_t(y_t|x_t)$ 的距离。然而,由于在本文迁移场景下,没有目标域的数据标签,无法直接获得目标域的条件分布。考虑贝叶斯公式:

$$Q(y|x)=\frac{Q(x|y)Q(y)}{Q(x)} \tag{5}$$

在式(5)中,一般可以认为源域和目标域的边缘分布 $Q(x)$ 和 $Q(y)$ 是相同的,因此在衡量源域和目标域的后验分布差异时,$Q(y)/Q(x)$ 可以忽略,从而可以用源域和目标域的类条件分布 $Q_s(x_s|y_s)$ 和 $Q_t(x_t|y_t)$ 代替条件分布。由于没有

北京航空航天大学学报

北航学报 赠阅

目标域的数据标签，$Q_t(x_t|y_t)$也无法直接获得，但是可以采用伪标签进行代替。所谓伪标签，即在源域中训练出一个基分类器，再将目标域的数据代入，从而得到目标域的数据标签，即目标域数据伪标签。在有了源域数据的真正标签及目标域数据的伪标签之后，便可以对类别空间 $c=\{1,2,\cdots,C\}$ 中的每个类别 c 分别进行类条件分布 $Q_s(x_s|y_s=c)$ 和 $Q_t(x_t|y_t=c)$ 的矩匹配。同样，采用 MMD 准则对源域和目标域类条件分布的距离进行度量[16]，可得到

$$\left\|\frac{1}{n_s^{(c)}}\sum_{x_i\in D_s^{(c)}}\boldsymbol{A}^{\mathrm{T}}x_i-\frac{1}{n_t^{(c)}}\sum_{x_j\in D_t^{(c)}}\boldsymbol{A}^{\mathrm{T}}x_j\right\|^2=\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}_c\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}) \tag{6}$$

式中：$\boldsymbol{M}_c$ 为类别 c 的 MMD 矩阵，其定义为

$$(M_c)_{ij}=\begin{cases}\dfrac{1}{n_s^{(c)}n_s^{(c)}} & x_i,x_j\in D_s^{(c)}\\ \dfrac{1}{n_t^{(c)}n_t^{(c)}} & x_i,x_j\in D_t^{(c)}\\ \dfrac{-1}{n_s^{(c)}n_t^{(c)}} & \begin{cases}x_i\in\mathcal{D}_s^{(c)},x_j\in\mathcal{D}_t^{(c)}\\ x_j\in\mathcal{D}_s^{(c)},x_i\in\mathcal{D}_t^{(c)}\end{cases}\\ 0 & \text{其他}\end{cases} \tag{7}$$

其中：$\mathcal{D}_s^{(c)}=\{x_i;x_i\in\mathcal{D}_s\wedge y(x_i)=c\}$ 为源域中属于类别 c 的样本集合；$y(x_i)$为样本 x_i的真实标签；$n_s^{(c)}=|D_s^{(c)}|$。相应地，$\mathcal{D}_t^{(c)}=\{x_j;x_j\in\mathcal{D}_t\wedge y(x_j)=c\}$ 为目标域中伪标签属于类别 c 的样本集合；$y(\boldsymbol{x}_j)$为样本 x_j的伪标签，$n_t^{(c)}=|D_t^{(c)}|$。通过最小化式(6)，得到新的特征表达 $\boldsymbol{Z}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}$，从而源域与目标域的边缘分布将被拉近。

综合考虑式(3)和式(6)，即可以得到边缘分布与条件分布的联合分布适配的正则项：

$$\mathcal{D}_{\mathcal{H}}(J_s,J_t)=\mathrm{MMD}_{\mathcal{H}}^2(P_s,P_t)+\sum_{c=1}^{C}\mathrm{MMD}_{\mathcal{H}}^2(Q_s^{(c)},Q_t^{(c)}) \tag{8}$$

在最小化式(8)后，边缘分布和条件分布的各阶矩统计量都会在无穷维希尔伯特空间 $\mathcal{H}$ 中得到适配。如果设源域与目标域的非线性特征映射为 T，则式(8)表示为

$$D_{\mathcal{H}}(J_s,J_t)=\mathrm{MMD}_{\mathcal{H}}^2(P_s,P_t)+\sum_{c=1}^{C}\mathrm{MMD}_{\mathcal{H}}^2(Q_s^{(c)},Q_t^{(c)})=\left\|\frac{1}{n}\sum_{i=1}^{n}T(x_i)-\frac{1}{m}\sum_{j=n+1}^{n+m}T(x_j)\right\|_{\mathcal{H}}^2+\sum_{c=1}^{C}\left\|\frac{1}{n^{(c)}}\sum_{x_i\in D_s^{(c)}}T(x_i)-\frac{1}{m^{(c)}}\sum_{x_j\in D_t^{(c)}}T(x_j)\right\|_{\mathcal{H}}^2 \tag{9}$$

4）模型问题综合与求解。

为了同时适配边缘分布和条件分布，需要同时最小化式(3)和式(6)，则 JDA 算法的优化目标可以总结为

$$\min\sum_{c=0}^{C}\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}_c\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})+\lambda\|\boldsymbol{A}\|_{\mathrm{F}}^2 \tag{10}$$

式中：$\lambda>0$ 为模型正则参数[17-18]，其作用是为了防止模型过于复杂。

为了保持数据结构及模型有解，需要式(1)也达到最大值，则综合优化式(1)和式(10)，得到统一的优化目标为

$$\min\frac{\sum_{c=0}^{C}\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}_c\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})+\|\lambda\boldsymbol{A}\|_{\mathrm{F}}^2}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}} \tag{11}$$

根据广义瑞利商[17]（generalized Rayleigh quotient），式(11)可以转化为式(10)的优化问题：

$$\min_{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}}\sum_{c=0}^{C}\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}_c\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})+\lambda\|\boldsymbol{A}\|_{\mathrm{F}}^2 \tag{12}$$

优化问题(12)是非线性问题，考虑使用核方法来进行求解。使用核函数映射：$\psi:X\to\psi(X)$ 及相应的核矩阵 $\boldsymbol{K}=\psi(X)^{\mathrm{T}}\psi(X)\in\mathbf{R}^{n\times n}$。将非线性映射 $T(\boldsymbol{x})=\boldsymbol{V}^{\mathrm{T}}f(\boldsymbol{x})=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}(:,\boldsymbol{x})$ 代入联合适配式(12)，可以得到核化后的联合适配正则化 PCA 的优化问题：

$$\min_{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}}\sum_{c=0}^{C}\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{M}_c\boldsymbol{K}^{\mathrm{T}}\boldsymbol{A})+\lambda\|\boldsymbol{A}\|_{\mathrm{F}}^2 \tag{13}$$

式(13)可以用拉格朗日乘子法进行求解，记 $\boldsymbol{\Phi}=\mathrm{diag}(\phi_1,\cdots,\phi_k)\in\mathbf{R}^{k\times k}$ 为拉格朗日乘子，则优化式(13)的拉格朗日形式为

$$L=\mathrm{tr}\left(\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{K}\sum_{c=0}^{C}\boldsymbol{M}_c\boldsymbol{K}^{\mathrm{T}}+\lambda\boldsymbol{I}\right)\boldsymbol{A}\right)+\mathrm{tr}((\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{\Phi}) \tag{14}$$

为求解上述最优化问题，令 $\partial L/\partial\boldsymbol{A}=0$，得到广义本征分解问题：

$$\left(\boldsymbol{K}\sum_{c=0}^{C}\boldsymbol{M}_c\boldsymbol{K}^{\mathrm{T}}+\lambda\boldsymbol{I}\right)\boldsymbol{A}=\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Phi} \tag{15}$$

通过求解式(15)，即可得到变换矩阵 $\boldsymbol{A}$ 的最优解，即经过矩阵 $\boldsymbol{A}$ 变换后同时适配源域和目标域的边缘分布和条件分布。

由于在进行源域与目标域的条件分布适配时使用的是目标域的伪标签，且数据分布不一致，其分类效果不好，需要采取不断迭代的方法使得伪标签不断接近真实标签，直到其收敛或达到迭代的次数。JDA 算法如下。

输入：数据样本矩阵 $\boldsymbol{X}$，标注矩阵 $\boldsymbol{y}$；子空间维度 k，模型正则化参数 λ。

北京航空航天大学学报

输出:变换矩阵 $\boldsymbol{A}$,特征表示 $\boldsymbol{Z}$,迁移分类模型 f。

开始:

由式(4)构造 MMD 矩阵 $\boldsymbol{M}_0$,令$\{\boldsymbol{M}_c:=0\}_{c=1}^{C}$

解广义本征分解,取 k 个最小本征向量构成矩阵 $\boldsymbol{A}$,令 $\boldsymbol{Z}:=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}$

在真实标注数据 $\{(\boldsymbol{A}^{\mathrm{T}}x_i,y_i)\}_{i=1}^{n_s}$ 上训练标准分类器 f,使用该分类器更新目标域的数据伪标签 $\{\hat{y}_j:=f(\boldsymbol{A}^{\mathrm{T}}x_i)\}_{j=n_s+1}^{n_s+n_t}$

由式(7)构造或更新 MMD 矩阵$\{\boldsymbol{M}_c\}_{c=1}^{C}$。

当算法收敛时停止

返回在标注数据 $\{(\boldsymbol{A}^{\mathrm{T}}x_i,y_i)\}_{i=1}^{n_s}$ 上训练的迁移分类模型 f

JDA 算法流程如图 2 所示。

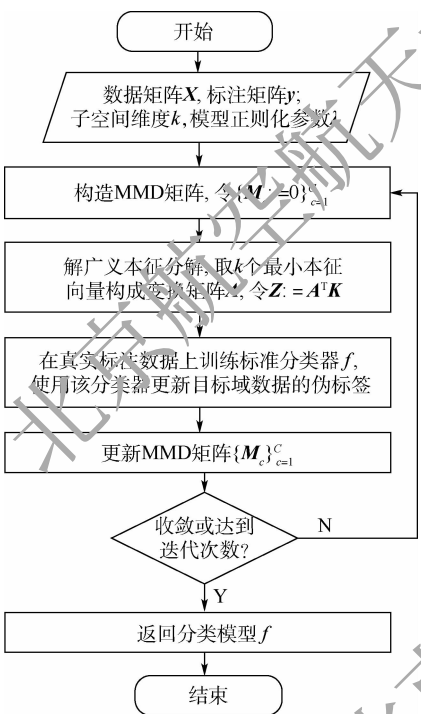


图 2 JDA 算法流程

Fig. 2 JDA algorithm flowchart

3 案例验证

本节采用数据集来自福建省和江苏省,其中福建省共有数据 507 条,江苏省共有数据 603 条,2 个地区具有同样的数据字段,包括有功准确度等级、额定负载点误差测量、最小负载点误差测量、最大负载点误差测量、第一次运行误差测量、第二次运行误差测量、第三次运行误差测量、运行时间、运行环境、家族缺陷、电量异常次数、时钟异常次数和运行等级等。其中,江苏省运行状态数据不含运行等级标签。福建省的数据的基本信息如表 1 所示。

福建省含电表运行等级标签分布如表 2 所示。

表 1 福建省电表数据集

Table 1 Dataset of electricity meters in Fujian province

属性	取值范围
有功准确度等级	0.5S,1.0,2.0
额定负载点误差测量	-0.054 ~ 0.25
最小负载点误差测量	-0.249 ~ 0.25
最大负载点误差测量	-0.265 ~ 0.25
第一次运行误差测量	0.05 ~ 0.4
第二次运行误差测量	0 ~ 0.45
第三次运行误差测量	0 ~ 0.5
运行时间	0 ~ 6.5
运行环境	1,1.2,2
家族缺陷	1
电量异常次数	0
时钟异常次数	0 ~ 7
运行等级	1,2,3

表 2 福建省电表等级分布

Table 2 Grade distribution of electricity meters in Fujian province

运行等级	数量	占比/%
1	13	2.56
2	418	82.44
3	76	15.00

从对数据的统计结果可以看出,福建省的运行等级标签中,多数类别比少数类别比例最高可达约 32∶1,存在类不平衡问题,因此,采用 SMOTE 算法进行不平衡数据处理。

由于 SMOTE 算法的原理,所合成的样本不一定符合原有的数据分布,所采用的手段是使用原始的数据样本训练出一个机器学习模型,使用该模型对合成的数据模型进行分类,从而使得新合成的数据样本符合原有的数据分布。采用该模型对新合成的样本进行分类,剔除掉不符合原始数据分布的合成样本,最终得到经过 SMOTE 算法扩充之后并符合原始数据分布的数据样本等级分布,如表 3 所示。

可以看出,经过 SMOTE 算法处理之后,数据量得到了扩充,数据的不平衡问题得到了较好的

表 3 基于 SMOTE 算法不平衡数据处理后的福建省数据分布

Table 3 Data distribution of Fujian province after imbalanced data processing based on SMOTE algorithm

运行等级	数量	占比/%
1	391	35.26
2	418	37.69
3	300	27.05
合计	1 109	100.00

北航学报 赠阅

解决。使用这些数据训练出 SVM 分类模型，并将目标域数据代入其中，得到目标域数据的初始预标注数据标签分布，如表 4 所示。

表 4　目标域数据初始伪标签分布

Table 4　Initial pseudo-label distribution of target domain data

运行等级	数量	占比/%
1	49	8.13
2	369	61.19
3	185	30.68
合计	603	100.00

由数据集包含的内容可以看到，江苏省和福建省的数据属性一致，但是福建省有数据标签而江苏省没有数据标签，根据对数据的分析显示两者数据的分布不一致。进行迁移过程中，数据迭代的初期，分类准确度的情况约为 0.89，这种情况表明 2 个地区的数据分布情况存在一定的偏差，印证了对 2 个地区数据分布情况的分析。普通的机器学习方法已经不再适用于这种情况，采用迁移学习方法来进行江苏省电表运行状态评价模型的构建。

对 JDA 算法采用 MATLAB 进行编程，根据初始伪标签对数据进行迭代。JDA 算法在使用正则化参数及不同的核函数时，得到的标签数据准确度会有不同。当正则化参数 $\lambda=0.1$，PCA 降维维度为 12，使用函数为径向基函数时，准确度达到最优，设置迭代次数为 15。分类准准确度随迭代次数的变化如图 3 所示。

从实验结果可以看出，在实验开始时，分类准确度约为 0.89，达不到进行智能电表运行状态评价的要求，反映了 2 个地区的数据分布不一致，由源域地区数据所训练的模型不能直接应用于目标域。随着迭代过程的进行，分类准确度逐步上升，最后稳定在一个较好的水平，证明了 JDA 算法适用于该类型数据的智能电表运行状态评价。迁移之后模型的混淆矩阵及主要评价指标如表 5 和表 6所示。

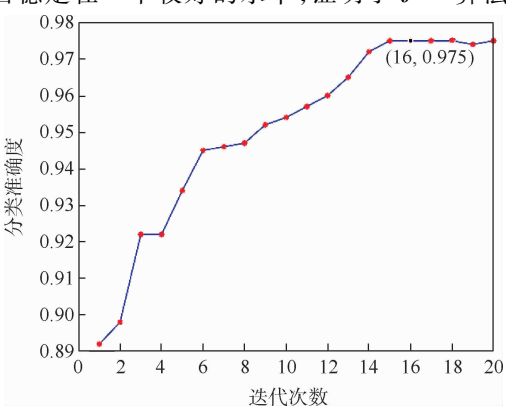


图 3　JDA 算法分类误差随迭代次数变化

Fig. 3　Classification error of JDA algorithm varies with number of iterations

从表 6 可以看出，主要评价指标结果均接近于 1，证实了采用 JDA 算法进行智能电表运行状态评价迁移的有效性，成功地使福建省的数据帮助江苏省的数据进行了运行状态评价模型分类。迁移之后的模型可以直接用于江苏省的智能电表运行状态评价，实现了从福建省到江苏省的评价模型迁移。

表 5　迁移后模型的混淆矩阵

Table 5　Confusion matrix of transferred model

混淆矩阵		预测类 1	预测类 2	预测类 3
真实类	1	39	3	0
真实类	2	10	364	0
真实类	3	0	2	185

表 6　迁移后模型的主要评价指标

Table 6　Main evaluation indicators of transferred model

评价指标	结果
准确度	0.975
F_1 分数	0.943

4　结　论

本文针对智能电表评价过程中不同地区数据分布不一致且某些地区没有数据标签的问题，将迁移学习中的 JDA 算法引入智能电表运行状态评价中，验证了能够在一定程度上有效利用有限的带标签数据来进行智能电表运行状态评价模型迁移。本文验证了 SMOTE 算法对于解决源域地区数据类不平衡问题的有效性。通过基于 JDA 算法的迁移学习，得到了迁移之后的智能电表运行状态分类模型，迁移后的模型分类准确率达到了 97.5%，验证了在智能电表运行状态评价领域应用迁移学习方法的可行性，提高了数据的利用率，扩大了已有模型的使用面，扩展了智能电表运行状态评价路径，具有一定的现实意义。

参考文献（References）

[1] 陈文娴. 基于深度学习的小样本异常用电数据检测技术研究[D]. 武汉：华中科技大学，2018.
CHEN W X. Abnormal electricity utilization detection based on deep learning and few-shot learning[D]. Wuhan: Huazhong University of Science and Technology, 2018(in Chinese).

[2] 商曦文，张颖，吉莹，等. 智能电表运行状态评估技术研究[J]. 自动化与仪器仪表，2018，11(3)：139-146.
SHANG X W, ZHANG Y, JI Y, et. al. Research on smart me-

北京航空航天大学学报

ters' running state assessment techniques[J]. Automation & Instrumentation,2018,11(3):139-146(in Chinese).

[3] CAI H,CHEN H Q,YE X Q,et al. An on-line state evaluation method of smart meters based on information fusion[J]. IEEE Access,2019,1(7):163665-163676.

[4] LI Y F,WU H,PANG S,et al. Application of supervised machine learning algorithms in diagnosis of abnormal voltage[J]. Electrical Measurement and Instrumentation,2016,53(1):58-62.

[5] 叶剑斌,朱东升,汪翊节,等.智能电能表状态检验技术研究[J].自动化仪表,2020,41(1):55-59.
YE J B,ZHU D S,WANG Y J,et al. Research on intelligent electric energy meter state inspection technology[J]. Process Aotumation Instrumentation,2020,41(1):55-59(in Chinese).

[6] 祝宇楠,徐晴,刘建,等.数据挖掘在智能电能表故障分析中的应用[J].电力工程技术,2016,35(5):19-23.
ZHU Y N,XU Q,LIU J,et al. Application of data mining in fault analysis of intelligent electricity meters[J]. Power Engineering Technology,2016,35(5):19-23(in Chinese).

[7] 韩笑.基于不确定性信息融合的智能电能表状态评价方法研究[D].长春:吉林大学,2018:22-58.
HAN X. Research on state evaluation method of smart electricity meter based on uncertainty information fusion[D]. Changchun: Jilin University,2018:22-58(in Chinese).

[8] CHENG Y Y,HOU X Z,XIAO J. Design and realization of condition management system of gateway electrical[J]. Electrical Measurement and Instrumentation,2013,50(8): 87-92.

[9] SHEN X,ZHAO D N,CAO M. Research on energy metering device remote online monitoring and analysis platform based on internet of thing[J]. Electrical Measurement and Instrumentation,2015,52(1):35-40.

[10] 王新刚,吴颖,张垠.基于数据挖掘的智能电能表在线监测方法[J].电测与仪表,2016,53(13):65-69.
WANG X G,WU Y,ZHANG Y. Online monitoring method of intelligent electricity meters based on data mining[J]. Electric Measurement and Instrumentation,2016,53(13):65-69(in Chinese).

[11] LONG M S,WANG J,DING G,et al. Transfer feature learning with joint distribution adaptation[C]//2013 IEEE International Conference on Computer Vision. Piscataway:IEEE Press,2013:1-8.

[12] CHAWLA N V,BOWYER K W,HALL L O,et al. SMOTE: Synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research,2011,16(1):321-357.

[13] BORGWARDT K M,GRETTON A,RASCH M J,et al. Integrating structured biological data by kernel maximum mean discrepancy[J]. Bioinformatics,2006,22(14):49-57.

[14] FILISBINO T A,GIRALDI G A,THOMAZ C E. Support vector machine ensembles for discriminant analysis for ranking principal components[J]. Multimedia Tools and Applications,2020,79(5):1-37.

[15] PAN S J,TSANG I W,KWOK J T,et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks,2011,22(2):199-210.

[16] GRETTON A,BORGWARDT K M,RASCH M J,et al. A kernel method for the two-sample problem[J]. Neural Information Processing Systems,2006,4(3):1-10.

[17] FAN J,KE Z T,LIU H,et al. QUADRO: A supervised dimension reduction method via Rayleigh quotient optimization[J]. Annals of Statistics,2015,4(43):1498-1534.

[18] 龙明盛.迁移学习问题与方法研究[D].北京:清华大学,2014.
LONG M S. Research on transfer learning problems and method[D]. Beijing:Tsinghua University,2014(in Chinese).

北京航空航天大学学报

Operating status evaluation of smart electricity meters based on joint distribution adaption

XU Dan[1,*], LI You[1], LI Helong[2]

(1. School of Reliability and Systems Engineering, Beihang University, Beijing 100083, China;
2. China Electric Power Research Institute Co., Ltd., Beijing 100085, China)

Abstract: In the operation status evaluation of smart electricity meters, it is often difficult to collect labeled data and the data distribution in different regions is inconsistent. To solve this problem, we introduce the joint distribution adaption (JDA) method in transfer learning in the field of operation status evaluation of smart electricity meters. This method tries to find an optimized transformation matrix to minimize the distance between edge distribution and conditional distribution in different regions in the transformed space. In order to solve the problem that there is no data label in the target domain when the conditional distribution adaptation is made, the JDA method uses pseudo-tag iteration method to make the target domain pseudo-label constantly approach the real tag. The classification model trained from the data in the post-transformation space can be applied to the new region to realize the transfer. The experimental results demonstrate the effectiveness of JDA method in the operation status evaluation of smart electricity meters.

Keywords: joint distribution adaption (JDA); smart electricity meters; operation status evaluation; transfer learning; smart grid

北京航空航天大学学报

Received: 2020-11-09; **Accepted**: 2021-06-11; **Published online**: 2021-10-26 15:15

URL: kns.cnki.net/kcms/detail/11.2625.V.20211026.1434.001.html

Foundation items: National Natural Science Foundation of China (51875017); the Fundamental Research Funds for the Central Universities (YWF-20-BJ-J-726)

* **Corresponding author**. E-mail: xudan@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0623

底凹结构减阻效应数值分析

全下文载

李彪，王良明*，杨志伟
（南京理工大学 能源与动力工程学院，南京 210016）

摘 要：为研究弹丸底凹结构的减阻机理，使用三维定常 CFD 方法对 M910 弹丸的流场特性进行了数值模拟。给出了零升阻力系数随马赫数的变化规律，所得结果与实验数据符合很好。在此基础上，为 M910 弹丸引入底凹结构并进行数值模拟。对比了不同弹底结构的底部流场特性，对底凹结构减阻效应的产生机理进行了分析。结果表明：亚声速下，底凹结构在底凹腔体内引入了高压“死水区”，并以“屈从”的流体边界代替了原固体底面，从而改变了尾部涡街的形成位置、形状和强度，最终增大底部压力，减小弹丸阻力；跨声速下，由于尾部涡街远离弹丸底面，固体底面与流体边界面的作用相同，使得底凹不再具有减阻效果；超声速下，底凹结构的减阻机理与底排弹丸减阻机理类似，即底凹结构中的流体为弹丸底部回流区添加质量从而达到减阻作用。

关 键 词：减阻效应；底部流场；数值模拟；底凹弹丸；零升阻力系数
中图分类号：TJ41；V211.3
文献标志码：A **文章编号**：1001-5965(2022)04-0682-09

弹箭阻力小不仅意味着弹箭的射程远、打击范围大，更能使弹箭在落点处保持较大的存速，增强毁伤能力，因此，减阻一直是弹箭设计者的首要任务。弹箭的零升阻力由摩阻、底阻和波阻 3 部分组成，且在中等超声速范围内，底阻占总阻的 40% ~50%[1]，因此减小底阻对减阻增程意义重大。常见的减小底阻方法有底部排气法和底凹减阻法。底凹是指在弹丸底部平面上向弹头方向开一个空腔，具有底凹结构的弹丸被称作底凹弹。虽然这种减阻方法已经在工程上实施并取得了成功，但是其减阻机理仍没有被完全理解，尤其是在亚跨声速范围内，数值计算和实验方法所得出的一些结论甚至相反。

Tanner[2]使用实验方法测得不同攻角下底凹弹和普通弹丸的压力分布，通过直接分析弹丸底部压力变化来研究攻角对底凹减阻效应的影响。Sahu 等[3-5]分别对超声速和亚跨声速范围内的底凹弹进行了数值仿真，讨论了底凹对弹丸飞行性能的影响。文献[2-5]都是通过比较阻力系数及弹底压力分布来分析底凹结构的减阻效应，并没有从弹丸底部流场结构方面来分析底凹减阻效应的产生机理。随后更高水平的技术，如纹影照相、镭射多普勒测速被引入到实验中用于测量超声速圆柱体[6]和跨声速锥形板底凹[7-8]的底部流场结构。对于超声速底部流场，文献[6]指出，在底部拐角处湍流边界层折转并膨胀，与圆柱体发生分离，并自此形成底部的自由剪切层。自由剪切层将外部无黏流和底部回流区隔离开。当自由剪切层到达对称轴时会再度发生压缩，之后附着在对称轴上。自由剪切层下边缘与对称轴的交点为驻点，最小断面处为喉部。对于亚跨声速底部流场，文献[7-8]指出，底凹减阻的主要原因是用屈从的流体边界替代了原有的固体壁面，使得尾涡形成的位置稍稍远离了底面，而且增大了涡脱离的频

收稿日期：2020-11-09；**录用日期**：2021-02-26；**网络出版时间**：2021-03-10 10:01
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210309.1034.005.html
*通信作者. E-mail：lmwang802@163.com

引用格式：李彪，王良明，杨志伟. 底凹结构减阻效应数值分析[J]. 北京航空航天大学学报，2022，48(4)：682-690.
LI B，WANG L M，YANG Z W. Numerical analysis on drag reduction effect of base cavity[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：682-690(in Chinese).

北京航空航天大学学报

北航学报 BUAA

率。这 2 点可能都是底凹结构减阻的原因,而这 2 点又与数值计算的结果相反。数值计算的结果指出,底凹的存在会增强尾部涡之间的相互作用,减小涡脱离的频率,另外尾涡还会部分进入到底凹中。这是因为计算模型是一个完美的二维流场,而实验永远不可能完全不受三维效应的影响。文献[7-8]中使用的风洞规模相对较小,一些三维效应是不可避免的。需要指出的是,文献[6-8]研究的对象均是锥形平板和圆柱体,而数值计算的对象是二维简化模型。文献[9]将粒子图像测速技术引入到实验中对亚声速锥形板底凹流场进行了实验测量,所得结论与文献[7-8]基本一致,且指出底凹结构不会对尾部涡街结构造成影响或改变。Fournier 等[10-11]对超高声速下底凹结构对 CAN-4 弹丸气动力的影响进行了数值研究,指出超高声速下底凹对气动力影响很小。Simon 等[12-13]研究了不同湍流模型对底部流场数值计算的影响和效果。

国内也已经对底凹弹丸的减阻效应进行了研究。谷嘉锦[14]通过风洞试验分析了不同底凹深度对减阻效应的影响,并指出在超声速条件下,底凹需从侧面开孔才能起到减阻的作用,否则减阻效果并不明显。文献[15-17]通过对超声速底凹弹侧壁开孔建立数学力学模型,获得了侧壁开孔底凹效应的工程计算方法。西北工业大学的 Pan 和 Cai[18]研究了底凹数目和形状对底凹减阻效应的影响。近年来,有关底凹效应的研究[19-21]逐渐增多,但对底凹效应的研究往往偏重于实验而忽视理论分析,并且国内只是对其减阻效果进行了数值模拟验证,关于底凹减阻的机理探讨尚不全面。另外,减小弹丸空气阻力对改善弹丸的弹道性能(如增大射程和存速、缩短飞行时间等)十分有利,而底阻在弹丸空气阻力中占比大,因此弹丸的底凹结构研究是十分有意义的。

本文采用标准 k-ε 湍流模型,对有无底凹结构的 M910 弹丸进行数值模拟,从底部流场特性,结合弹丸零升阻力系数对弹丸底凹结构在亚、跨、超声速范围下的减阻效应产生机理进行了分析。

1　数值计算方法

1.1　计算模型和网格

本文采用 M910 弹丸为对象进行数值计算,计算后的结果与实验数据[22]进行比较以验证计算的正确性。为 M910 弹丸引入圆柱底凹结构并将改造后的弹丸命名为 M910BC。计算模型的外形和底凹形状及位置尺寸如图 1 所示。

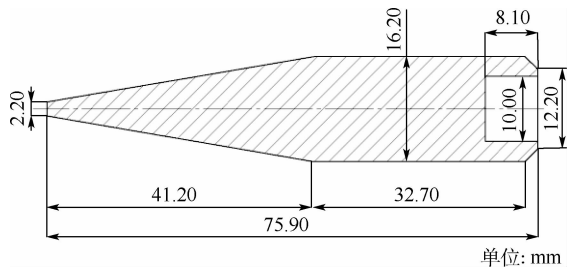


图 1　计算模型

Fig. 1　Computational model

实验和数值计算结果都表明,一旦底凹结构深度达到某一临界值后,继续增加底凹结构的深度将不会引起底凹结构减阻效应的任何改变,所有实验中测得的观测量也没有发生实质性的变化。文献[8]还指出,这一临界深度可能只有弹径的 1/4 或1/3,故本文取值为弹径的 1/2 以确保超过底凹结构的临界深度,从而不影响对计算结果的分析。

为了避免文献[7-8]中提到的计算模型采用二维结构网络出现的数值计算结果与实验结果存在差异,本文采用三维结构网格。为了对比分析弹丸有无底凹结构对底部流场的影响,应尽量保证不引入计算网格方面的差异,故本文均采用三维结构网格,并保证除在底凹内部,2 种弹丸的网格数目和形状都完全一致。图 2 为 M910BC 弹丸的底凹计算网格示意图。

为了验证网格无关性,选取来流速度为 1 190.7 m/s。对 M910 分别取网格数为 120 万、200 万和 310 万 3 组网格,对 M910BC 分别取网格数为 130 万、200 万和 320 万 3 组网格进行计算,比较阻力系数如表 1 所示。可以看出,2 组弹丸网格数选取中等粗细和细网格阻力系数差异很小,大大节省了计算时间。因此,通过对网格无关性的研究,获得了一套既不影响计算精度又能减少计算时间的最优网格。表 2 为 2 种弹丸的网格配置参数和最优网格数目,其中径向边界距离和

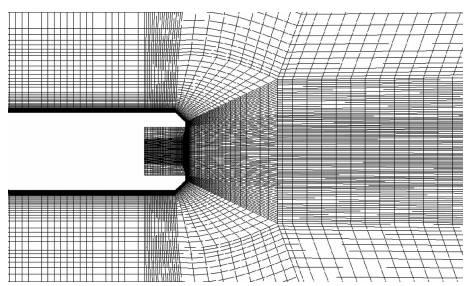


图 2　底凹计算网格

Fig. 2　Computational mesh of base cavity

北京航空航天大学学报

北航学报 网阅

表 1 网格无关性验证

Table 1 Grid independence verification

弹型	规格	网格数	阻力系数 C_D	计算时间/h
M910	粗	1 247 191	0.209 3	5
	中	1 974 720	0.227 4	7.9
	细	3 117 979	0.227 6	12.5
M910BC	粗	1 281 160	0.203 5	5
	中	2 028 504	0.218 1	8
	细	3 202 901	0.218 8	12.6

表 2 计算网格特性

Table 2 Computational mesh characteristics

弹型	计算值				最优网格数目
	径向边界距离	前向边界距离	后向边界距离	第一层网格厚度	
M910	45.8	24.7	39.8	4×10^{-5}	1 974 720
M910BC	45.8	24.7	39.8	4×10^{-5}	2 028 504

前向边界距离均指相应边界到弹丸顶点的距离，后向边界距离指后边界到弹丸底面的距离，表中距离的单位均为 1 倍弹径。

1.2 Navier-Stokes CFD

1.2.1 控制方程

完整的积分型三维可压缩 Navier-Stokes 方程可表示为

$$\frac{\partial}{\partial t}\int_{\Omega}\boldsymbol{W}\mathrm{d}\Omega+\oint_{\partial\Omega}(\boldsymbol{F}_{\mathrm{C}}-\boldsymbol{F}_{\mathrm{V}})\mathrm{d}S=\int_{\Omega}\boldsymbol{Q}\mathrm{d}\Omega \tag{1}$$

式中：$\boldsymbol{W}=[\rho\ \ \rho u\ \ \rho v\ \ \rho w\ \ \rho E]^{\mathrm{T}}$ 为守恒变量，ρ 为密度，u、v、w 分别为 3 个方向上的速度，E 为能量；$\boldsymbol{F}_{\mathrm{C}}$ 为对流通量；$\boldsymbol{F}_{V}$ 为黏性通量；$\boldsymbol{Q}$ 为源项；Ω 为控制体的体积；S 为控制体的各个表面。采用有限体积法解该方程。对流通量采用 Roe 格式离散，黏性通量采用中心差分格式离散。

1.2.2 湍流模型

标准 k-ε 湍流模型有较高的稳定性、经济性和计算精度，应用广泛[23]，故本文选取该湍流模型进行计算。该模型可表示为

$$\frac{\partial\rho K}{\partial t}+\frac{\partial}{\partial x_j}(\rho v_j K)=\frac{\partial}{\partial x_j}\left[\left(\mu_L+\frac{\mu_T}{\sigma_K}\right)\frac{\partial K}{\partial x_j}\right]+\tau_{ij}^{F}S_{ij}-\rho\varepsilon \tag{2}$$

$$\frac{\partial\rho\varepsilon^{*}}{\partial t}+\frac{\partial}{\partial x_j}(\rho v_j\varepsilon^{*})=\frac{\partial}{\partial x_j}\left[\left(\mu_L+\frac{\mu_T}{\sigma_\varepsilon}\right)\frac{\partial\varepsilon^{*}}{\partial x_j}\right]+C_{\varepsilon1}f_{\varepsilon1}\frac{\varepsilon^{*}}{K}\tau_{ij}^{F}S_{ij}-C_{\varepsilon2}f_{\varepsilon2}\rho\frac{(\varepsilon^{*})^{2}}{K}+\phi_{\varepsilon} \tag{3}$$

该模型的参数设置参照文献[24]：$C_\mu=0.09$，$C_{\varepsilon1}=1.44$，$C_{\varepsilon2}=1.92$，$\sigma_K=1.0$，$\sigma_\varepsilon=1.3$，$Pr_{\mathrm{T}}=0.9$。

因为壁面附近流场变量的梯度较大，所以壁面对湍流计算的影响很大。因此，在壁面附近要进行特殊处理，常用的方法有壁面函数法和近壁模型法。本文采用壁面函数法，即用半经验公式将自由流中的湍流与壁面附近的流动连接起来。文献[25]指出该方法的优点是可以保持雷诺应力与真实湍流一致，对旋流和分离流的模拟结果更符合真实情况。因此，该湍流模型适合于计算弹丸底部流场的分离和尾涡的形成。

近壁面阻尼函数为

$$\begin{cases}f_\mu=\exp\left[\dfrac{-3.4}{1+0.02Re_{\mathrm{T}})^{2}}\right]\\ f_{\varepsilon1}=1\\ f_{\varepsilon2}=1-0.3\exp(Re_{\mathrm{T}}^{2})\end{cases} \tag{4}$$

1.3 边界条件

流场的外边界设置为压力远场。来流的压力为 101 325 Pa，温度为 288.5 K。弹体表面为非滑移绝热壁面，非滑移即在固体边界上的流体速度等于固体表面速度（本文指固体表面静止），绝热即壁面处温度梯度为 0，故非滑移绝热的边界条件为

$$\begin{cases}u_{\mathrm{w}}=v_{\mathrm{w}}=w_{\mathrm{w}}=0\\ \left(\dfrac{\partial T}{\partial n}\right)_{\mathrm{w}}=0\end{cases} \tag{5}$$

式中：下标 w 代表壁面处；T 为温度；n 为垂直于壁面的法单位向量。

表 3 显示了所要计算的工况，还定义了马赫数所属的范围（参照文献[22]进行马赫数范围划分），为接下来的分析工作带来了便利。

表 3 来流条件与马赫数的关系

Table 3 Relationship between incoming flow conditions and Mach number

范围	马赫数	来流速度/(m·s^{-1})
亚声速	0.60	204.1
	0.70	238.1
跨声速	0.90	306.2
	0.98	333.4
	1.02	347.0
	1.20	408.3
超声速	2.00	680.4
	2.50	850.5
	3.00	1 021.1
	3.50	1 190.7

2 计算结果与分析

2.1 计算结果与实验数据对比

图 3 为计算所得的 M910 弹丸阻力系数 C_D 与 M910 弹丸实验数据[22]的对比结果。为了便于直观显示底凹结构的减阻效应，将 M910BC 弹

北京航空航天大学学报

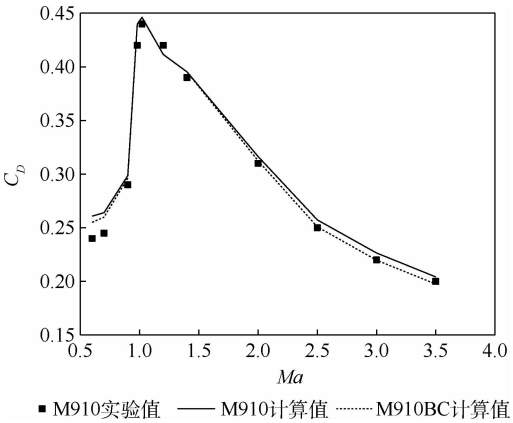


图 3　计算阻力系数与实验数据的对比

Fig. 3　Comparison between calculated drag coefficient and experimental data

丸的计算结果也放于图中。从图 3 可以看出,M910 弹丸的计算结果与实验值在跨声速段吻合很好,但在亚声速和超声速段,计算结果都略大于实验值。虽然标准 k-ε 湍流模型对旋流和分离流的模拟具有优势,但对强逆压梯度不敏感[24]。随着马赫数增大,这种强逆压梯度越明显,故计算误差越大。此外,标准 k-ε 湍流模型假设流动为完全湍流,忽略了分子黏性的影响。然而在亚声速段,弹体表面的流动是层流和湍流混合的,此时的计算误差也会相应较大。

对于 M910 和 M910BC 二种弹丸,强逆压梯度发生和存在于船尾开始处,有无底凹结构对其影响不大。另外,由于 2 种弹丸的外形完全一致,相同马赫数下弹体表面的流动特性也完全相同。综上所述,2 种弹丸在相同马赫数下由于湍流模型所带来的计算误差基本相等,图 3 中所显示的 M910BC 弹丸阻力系数减少的原因可基本确信为底凹结构的减阻效应,而并不是计算的误差。

此外,还可根据弹丸底凹结构的减阻效率随马赫数变化规律来判断计算的正确与否。底凹结构减阻效率为

$$E = \frac{C_D^{\mathrm{M910}} - C_D^{\mathrm{M910BC}}}{C_D^{\mathrm{M910}}} \tag{6}$$

图 4 给出了不同马赫数下底凹结构的减阻效率。可以看出,在亚声速段底凹结构减阻效率随马赫数增大而减小,这与文献[7-9]的实验结果规律一致。当马赫数增大到跨声速段,底凹结构减阻效应完全消失,这一现象在所有参考文献中都未曾给出。从图 3 可看出,在超声速段,底凹减阻的差值是先增大,后保持不变,这与文献[10]中实验结果一致。然而图 4 中减阻效率一直增大的原因是超声速下阻力系数随马赫数的增大而减小。结合底凹结构在亚声速和超声速段的减阻效

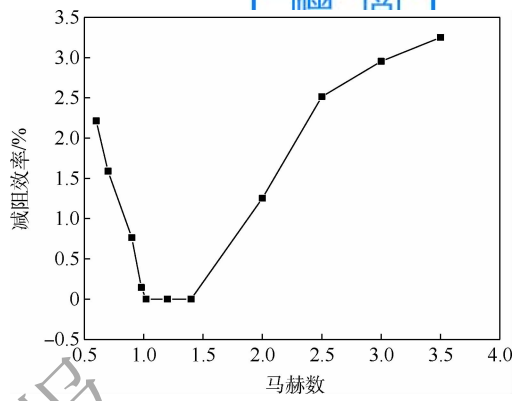


图 4　底凹结构的减阻效率

Fig. 4　Drag reduction efficiency of base cavity structure

率规律,可以确定在跨声速段底凹结构的减阻效率会逐渐到某一最小值然后缓慢增大,因此可以判断本文的计算结果是合理的。综上所述,本文所采用的数值计算方法基本捕捉到了各马赫数范围的流场物理规律,计算结果与实验结果吻合很好,故本文的计算结果可以用于底凹结构减阻效应的机理性分析。

2.2　亚声速下的底凹效应分析

亚声速下的弹丸底部流场结构复杂,不管是采用实验还是数值计算手段,获得真实的弹丸底部物理流场都是极具挑战的任务。大量研究指出,当亚声速气流从船尾流向弹底时,船尾处的横断面不断减小,在弹底面处突然消失为 0。由于物体横断面的减小,由一圈流线所围成的流管的横断面积 S 必然增大。根据连续方程,流体速度减小。再由伯努利方程可知,当流线上速度减小,压强 p 会增大。因此,从船尾起始点处起,后面的流体将被阻滞。当压强增大到一定程度,阻滞作用会使流体流动停止。随着压强继续增大,就形成了反压,在反压的作用下靠近弹体表面的部分流体会形成逆流。此时,附面层将与弹体表面分离,形成涡街。

由于本文采用的是定常 CFD 计算方法,计算所得的尾涡将是一个等效涡街。该等效涡街只能表征实际涡街的大小和形成位置,关于涡街脱落信息则由弹底横截面平均压力系数来推断。

图 5 和图 6 分别为弹丸底部中心线上的压力系数曲线及弹丸底部横截面上的平均压力系数曲线。压力系数 C_p 的定义如式(7)所示。弹丸底部横截面是指圆心在底部中心线上、以弹径为直径且平行于弹底面的一系列圆面。

$$C_p = \frac{p - p_\infty}{\frac{1}{2}\rho_\infty U_\infty^2} \tag{7}$$

式中:∞ 表示“远场”处。

北京航空航天大学学报

北航学报 赠阅

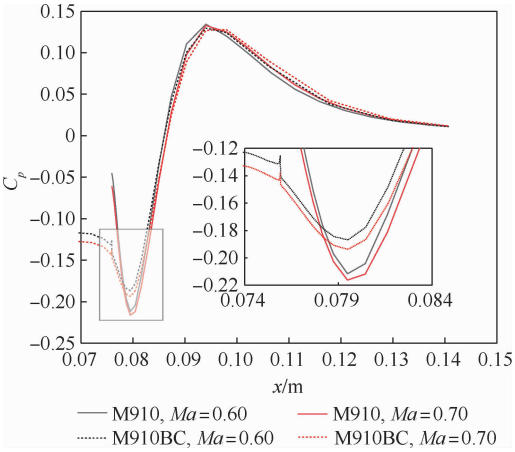


图 5 弹底中心线压力系数曲线

Fig. 5 Centerline pressure coefficient curves of projectile base

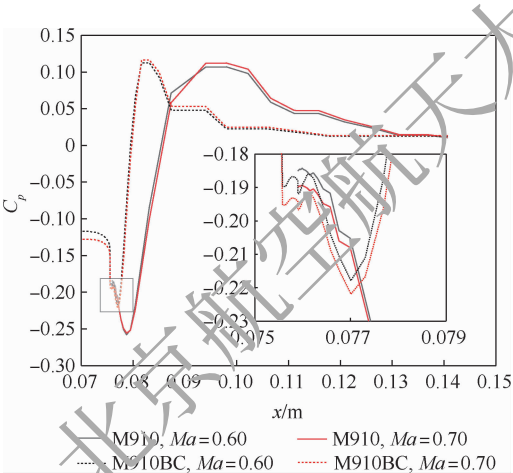


图 6 弹底横截面平均压力系数曲线

Fig. 6 Cross-sectional average pressure coefficient curves of projectile base

弹底中心线可从压力系数最大点处分为 2 段,如图 5 所示。压力系数最大点的左侧流场速度方向与来流相反,属于逆流,而右侧则属于顺流。对于 M910 弹丸,逆流是从压力高的地方向压力低的地方加速流动,在遇到固体底面后流体运动受到阻滞,流速降低,压力系数增大。这种流体运动遇到固体壁面受到阻滞的作用是在距壁面很小一段距离内发生的,从图 5 可以看出,在弹底尾部中心线上,压力系数在很短距离内快速减小,之后又迅速增大。而对于 M910BC 弹丸,由于底凹结构的存在,以流体边界取代原有的固体底面,逆流所受到的阻滞作用就没那么强,速度减小的少,在底面上,M910BC 的压力系数比 M910 弹丸要小。

需要指出的是,图 5 显示的只是弹底中心线上的压力变化规律,而图 6 则显示了整个弹底横截面上的平均压力变化规律。由图 6 可知,相同马赫数下,底凹底面上的平均压力系数明显大于固体底面上的平均压力系数,表明底凹结构的存在可以增大弹底压力,减小弹丸阻力。

图 7 为亚声速下弹底压力分布云图。由于计算结果在竖直平面内是轴对称的,为了便于比较分析,将相同马赫数下的 2 种弹丸计算结果各取一半放在图中。

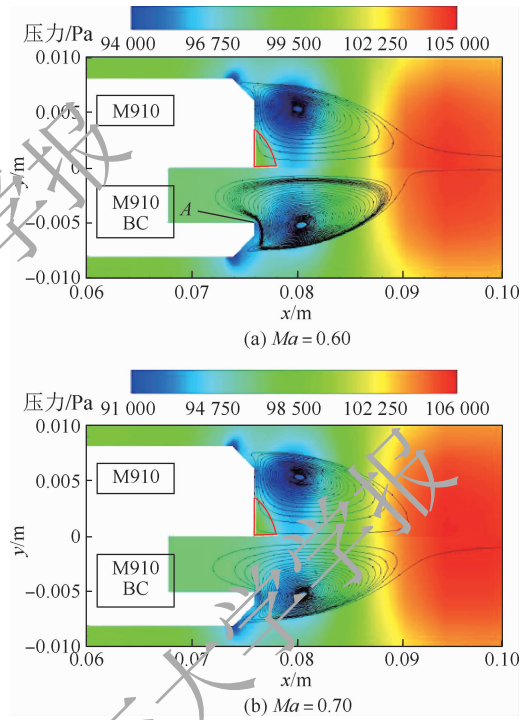


图 7 亚声速下弹底压力对比

Fig. 7 Comparison of projectile base pressure at subsonic speed

结合图 7 可知,对于 M910 弹丸,由于固体底面的阻滞作用,在底面中心处形成了一块面积不大的高压区。该高压区使得涡街内部的逆流方向发生改变,而并不是像固体底面那样产生阻滞作用,因此方向改变的逆流速度并没有减小。当逆流的方向改变 90°后,流体又加速从弹底中心流向弹底四周,这个过程中流体的压力在减小,因此在弹底四周形成了一片低压区,如图 7 所示。

底凹结构的引入给弹底流场带来了 2 个变化:①在底凹内部形成了高压“死水区”;②以“屈从”的流体边界代替了原来的固体底面边界。“屈从”的流体边界对逆流的阻滞作用不如固体底面那么强,因此在底面中心所形成的高压区不明显,面积也较小。与 M910 弹丸一样,涡街内部的逆流也会在底凹的流体边界上发生转向,流体由底面中心加速流向四周,从而导致沿弹径方向的压力减小。这样,底凹内外压差在拐点 A(见图 7(a))处达到最大。当内外压差达到一定值后,流体边界会被破坏,导致涡街部分进入底凹结构中,而边界破坏就发生在 A 点附近。由于在底

北京航空航天大学学报

面中心处流体边界内外的压差并不大，A 点处边界不会被破坏，涡街也不能由此进入到底凹结构内，如图 7 所示。因此，涡街只是部分进入到底凹结构中，这一结论与文献[7-9]中的实验结果一致，反驳了底部涡街会进入到底凹内部且涡街运动会受到底凹限制从而提高底部压力的假设。

图 5 和图 6 还显示出，M910BC 弹丸在底凹内部中心线上的压力系数变化不大，从约 0.076 m 处开始减小，这也证明了文献[8]中的结论，即底凹结构深度达到某一临界值后，底凹结构减阻效果不再随底凹结构深度的增加而变化。另外，“屈从”的流体边界在遇到涡街逆流时，在边界两侧会发生压力波动，如图 5 和图 6 所示，这种压力波动正是由流体边界“屈从”的特性所引起的。

随着马赫数的增大，弹底整体压力减小，形成的涡街会更强。涡街形成的位置也会更远离弹底面，如表 4 所示。无论是固体底面还是流体边界底面，涡街的逆流对其影响都会减弱。对于固体底面，涡街的远离就相当于在固体底面附近自然形成一个高压“死水区”，如图 7 中红色框图区域所示。这样因底凹结构所形成的高压“死水区”的作用就会减弱，在亚声速范围内，随着马赫数的增大，底凹结构减阻效率会降低，如图 4 所示。

表 4　亚声速下涡街中心信息

Table 4　Information of vortex street center at subsonic speed

弹型	马赫数	涡街中心坐标		涡街中心压力/Pa
		x/mm	y/mm	
M910	0.6	80.042 7	5.390 8	94 511
	0.7	80.181 0	5.344 7	92 193
M910BC	0.6	80.319 3	5.166 1	95 751
	0.7	80.457 6	5.212 2	93 620

2.3　跨声速下的底凹效应分析

跨声速是亚声速和超声速之间的过渡阶段，不仅同时具备了二者的特点，还具有自己的特点。从图 4 看，底凹结构的减阻效率在跨声速范围内先减弱至零，后慢慢增强。在该范围内低马赫数下，流场规律还是与亚声速流场规律一致，即随马赫数增大，底凹结构的减阻效果减弱。在该过程中，形成的尾部涡街越来越大，涡街形成位置也越来越远离弹底面，如表 3 和表 4 所示。

图 8 为跨声速下弹底压力对比。可以看出，当 $Ma=0.90$ 时船尾处就已经形成了膨胀波，此时的马赫角几乎接近 90°，附面层无法完成重新附着。随着马赫数的增大，马赫角逐渐减小，附面层向弹轴方向偏转的角度就越大。除此之外，经过膨胀波的流体会加速，附面层会快速流向弹底中心线完成重新附着，形成封闭的回流区。尾部涡街被“困”于回流区内，将不再脱离。此时定常计算的结果将与非定常计算的结果一样[22]。

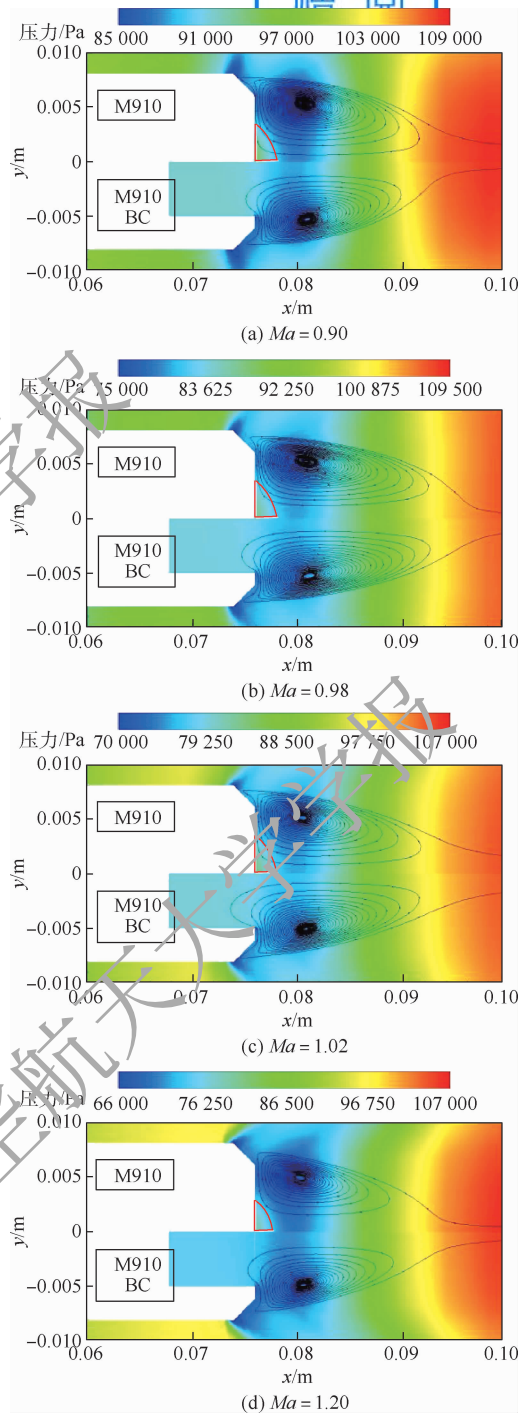


图 8　跨声速下弹底压力对比

Fig. 8　Comparison of projectile base pressure at transonic speed

由表 5 可知，对于 2 种弹型，当马赫数大于 0.98 后，涡街的中心位置不再向弹底后方移动，而是转而向弹头方向移动。这是由于回流区形成以后，涡街无法脱离便被其后更高的压力压向弹头方向，在涡街和固体壁面的相互作用下，最终涡

北京航空航天大学学报

表 5 跨声速下涡街中心信息

Table 5 Information of vortex street center at transonic speed

弹型	马赫数	涡街中心坐标		涡街中心压力/Pa
		x/mm	y/mm	
M910	0.90	80.642 0	5.483 0	86 368
	0.98	80.872 5	5.252 5	79 307
	1.02	80.411 5	5.160 3	74 054
	1.20	80.181 0	4.975 9	68 160
M910BC	0.90	80.872 5	5.304 3	87 852
	0.98	81.149 1	5.258 2	80 355
	1.02	80.780 3	5.073 8	75 926
	1.20	80.549 8	4.843 3	70 422

街中心会反向移动。在该过程中,因涡街位置越来越远而形成的弹底中心的高压“死水区”被回流区所破坏并成为回流区的一部分,最终形成新的稳定回流。

对于 M910 弹丸,弹底中心高压区与涡街中心的压差在亚声速段越来越大,而进入跨声速后压差又会减小,图 8 中的结果证明了该结论。对于 M910BC 弹丸,弹丸底部的高压“死水区”也随着涡街的位置变化而变化,然而这些变化并未影响底凹内的“死水区”,因此底凹结构并不能起到减阻作用,如图 4 所示。

2.4 超声速下的底凹效应分析

超声速下的弹丸固体底面底部流场结构已经清楚,下面讨论底凹结构如何影响超声速底部流场,从而实现减阻。

对于底凹弹丸,当回流区形成以后,回流区里的流体将与底凹结构中的流体混合,该混合随着马赫数的增大而加深。如图 9(a)、(b)所示,虽然自由剪切层和底凹壁面已经形成一个封闭的区域,但该区域内的流体并不全是回流。图 9(c)、(d)又显示,在高马赫数下流体充分混合。根据底排弹的工作原理可知,只要向低压回流区中“添质加能”,就能减小底阻。因此,回流区里的流体与底凹中的流体混合可以看成一种“添质”行为,底凹结构在超声速下也可以实现减阻。

一方面,底凹减阻与 2 种流体的混合程度有关,当完全混合后,减阻效果不会继续增大,如图 4所示;另一方面,在确定马赫数的情况下,2 种流体混合所造成的“添质”行为有一个上限,即无法将底凹结构中所有的流体都视作“添质”。这又证明了增加底凹结构深度到某一极限值后,底凹结构的减阻效果就不再改变。

图 9 为超声速下弹底压力对比云图。可以发现,底凹结构确实将驻点后移,自由剪切层的折转

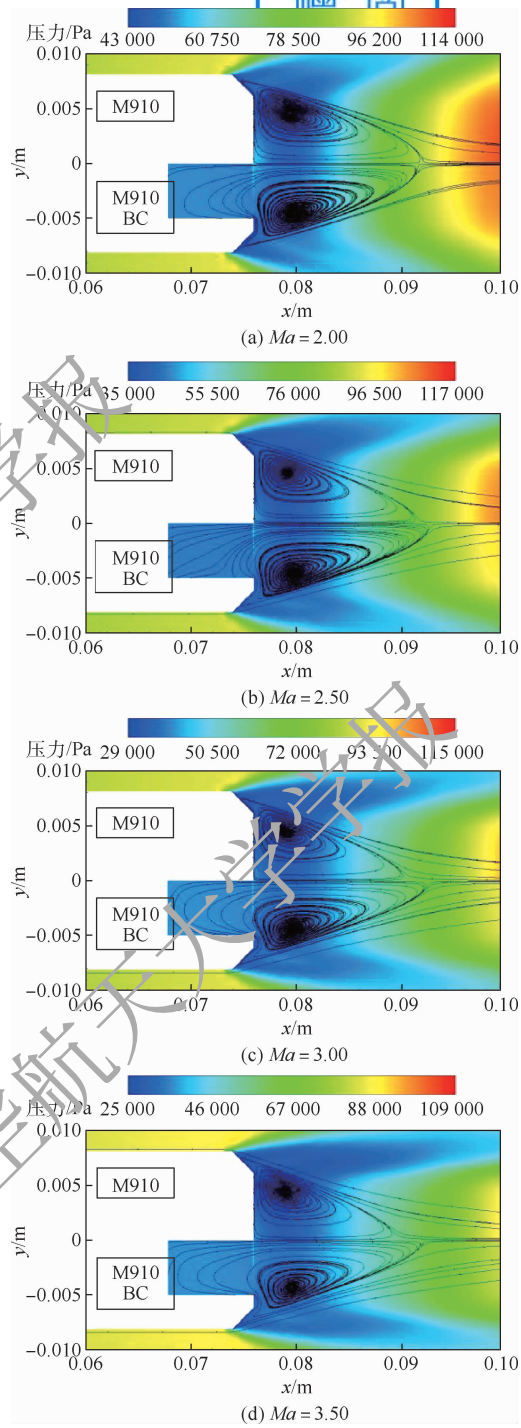


图 9 超声速下弹底压力对比

Fig. 9 Comparison of projectile base pressure at supersonic speed

角度减小,自由剪切层外缘抬高,喉部的高度增大,这些现象都说明了底凹结构起到了减小底阻的作用[26]。

3 结 论

通过对 M910 和 M910BC(M910 弹丸引入圆柱底凹结构)2 种弹丸全马赫数下的三维定常 CFD 仿真计算,比较 2 种弹丸的计算结果,分析了

北航学报 噌 阅

底凹结构减阻效应的产生机理。本文结论可为弹丸底凹结构的设计提供参考，还可为弹丸的弹道优化设计提供依据，具有一定的工程应用价值。本文得出以下结论：

1）马赫数一定时，底凹结构深度达到极限值后，底凹结构减阻效率将保持不变，且该极限深度随马赫数的增大而减小。

2）底凹结构在亚跨声速和超声速范围内的减阻机理不同。在亚跨声速下，底凹结构以“屈从”的流体边界代替了固体底面，且为弹丸引入了高压“死水区”从而实现减阻。

3）在跨声速下，随着马赫数的增大，尾部涡街会远离弹底面，而在弹底面与涡街之间形成高压区。该部分高压区与底凹结构中的“死水区”作用相似，底凹结构减阻作用逐渐减小，最后消失。随着马赫数进一步增大，弹底上下附面层会在弹底中心线上形成封闭区域，即底部回流区。底部回流区的形成又使得涡街中心前移从而影响之前生成的高压区，底凹减阻效果又将出现。

4）在超声速下，底凹结构通过底凹内部流体与回流区中的流体混合，以向低压回流区增加质量的方式来减阻，这种混合会随着马赫数的增大而加深，直到完全混合后底凹结构的减阻效果将不变。

参考文献（References）

[1] 臧国才，李树常. 弹箭空气动力学[M]. 北京：兵器工业出版社，1989：142-144.
ZANG G C, LI S C. Aerodynamics of projectiles and missiles[M]. Beijing: Publishing House of Ordnance Industry, 1989: 142-144(in Chinese).

[2] TANNER M. Base cavity at angles of incidence[J]. AIAA Journal, 1988, 26(3): 376-377.

[3] SAHU J. A computational study of the base region flow field for the M865 projectile: ARL-TR-109[R]. Reston: AIAA, 1993.

[4] SAHU J, NIRTUBICZ C J. Three dimensional flow calculations for a projectile with standard and dome bases: BRL-TR-3150[R]. Reston: AIAA, 1990.

[5] SAHU J, NIETUBICZ C J, HEAVEY K R. Computational study of the M825 projectile with standard and dome base: BRL-MR-3662[R]. Reston: AIAA, 1986.

[6] HERRIN J L, DUTTON J C. Supersonic base flow experiments in the near-wake of a cylindrical afterbody[J] AIAA Jornal, 1994, 32(1): 77-83.

[7] KRUISWYK R W, DUTTON J C. Effects of a base cavity on subsonic near-wake flow[J]. AIAA Journal, 1990, 28(11): 1885-1893.

[8] KRUISWYK R W, DUTTON J C. An experimental investigation of the effects of base cavity on the near-wake flowfield of a body at subsonic and transonic speeds[C] // Proceedings of 27th Aerospace Sciences Meeting. Reston: AIAA, 1989: 210.

[9] MOLEZZI M J, DUTTON J C. Study of subsonic base cavity flowfield structure using particle image velocimetry[J]. AIAA Journal, 1995, 33(2): 201-209.

[10] FOURNIER E Y, DUPUIS A D, EDWARDS J A. Base cavity effects on the aerodynamic characteristics of a hypersonic flared projectile[J]. Journal of Spacecraft and Rockets, 1997, 34(6): 737-743.

[11] FOURNIER E Y, DUPUIS A D. Computational fluid dynamic analysis of the base cavity interactions of the CAN-4 research projectile: DREV-TM-9620[R]. Quebec: Defense R&D, 1997.

[12] SIMON F, DECK S, GUILLEN P, et al. Zonal-detached-eddy simulation of projectiles in the subsonic and transonic regimes[J]. AIAA Journal, 2007, 45(7): 1606-1619.

[13] SIMON F, DECK S, GUILLEN P. Reynolds-averaged Navier-Stokes/large-eddy simulations of supersonic base flow[J]. AIAA Journal, 2006, 44(11): 2578-2590.

[14] 谷嘉锦. 底凹弹丸的风洞实验研究[J]. 南京航空航天大学学报，1979，12(1)：243-253.
GU J J. The study of wind tunnel experiments of a ball with hollow base[J]. Journal of Nanjing University of Aeronautics and Astronautics, 1979, 12(1): 243-253(in Chinese).

[15] 浦发，王中原. 超音速底凹弹减阻机理的研究[J]. 弹道学报，1989，9(1)：63-67.
PU F, WANG Z Y. The analysis of drag reduction in supersonic speed for projectiles with base cavity[J]. Journal of Ballisitics, 1989, 9(1): 63-67(in Chinese).

[16] 王中原. 超声速底凹弹侧壁开孔对飞行阻力的影响[J]. 空气动力学学报，1997，15(4)：502-506.
WANG Z Y. The effect about projectiles of base cavity with side holes to flight drag under supersonic flow[J]. Acta Aerodynamic Sinica, 1997, 15(4): 502-506(in Chinese).

[17] 王中原. 底凹弹侧壁斜孔减小底阻分析[J]. 南京理工大学学报，1997，21(1)：21-24.
WANG Z Y. Analysis of base drag reduced by side holes in the projectile with base cavity[J]. Journal of Nanjing University of Science and Technology, 1997, 21(1): 21-24(in Chinese).

[18] PAN S C, CAI J S. Investigation of vortical flow over bluff bodies with base cavities[J]. Acta Mechanica Sinica, 2012, 28(5): 1238-1247.

[19] 轩海彬，张文洁，于勇，等. 亚跨声速流动中底凹减阻的数值模拟[J]. 兵器装备工程学报，2017，38(11)：1-8.
XUAN H B, ZHANG W J, YU Y, et al. Numerical study on drag reduction of projectile in subsonic and transonic speed[J]. Journal of Ordnance Equipment Engineering, 2017, 38(11): 1-8(in Chinese).

[20] 陆海波，颛孙世周，孙凤文，等. 不同形状底凹结构对火炮弹丸飞行阻力影响研究[J]. 弹箭与制导学报，2016，36(5)：112-114.
LU H B, ZHUANSUN S Z, SUN F W, et al. Research on influence of base cavity with different shape on aerodynamic drag of projectile[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2016, 36(5): 112-114(in Chinese).

[21] 陆海波，高洁. 侧倾底凹结构对火炮弹丸阻力的影响[J]. 兵器装备工程学报，2019，40(6)：5-9.

北京航空航天大学学报

LU H B, GAO J. Research on shrinkage shape of base cavity upon aerodynamic drag of projectile[J]. Journal of Ordnance Equipment Engineering, 2019, 40(6): 5-9(in Chinese).

[22] DESPIRITO J, HEAVEY K. CFD computation of magnus moment and roll-damping moment of a spinning projectile[C]// AIAA Atmospheric Flight Mechanics Conference and Exhibit. Reston: AIAA, 2004: 1-24.

[23] 于勇. Fluent 入门与进阶教程[M]. 北京: 北京理工大学出版社, 2008: 23-25.
YU Y. Fluent introductory and advanced tutorials[M]. Beijing: Beijing Institute of Technology Press, 2008: 23-25 (in Chinese).

[24] LAUNDER B E, SHARMA B I. Application of the energy dissipation model of turbulence to the calculation of flow near a spinning disc[J]. Letter in Heat and Mass Transfer, 1974, 1(11): 131-138.

[25] 雷娟棉, 李田田, 黄灿. 高速旋转弹丸马格努斯效应数值研究[J]. 兵工学报, 2013, 34(6): 718-725.
LEI J M, LI T T, HUANG C. A numerical investigation of Magnus effect for high-speed spinning projectile[J]. Acta Armamentarii, 2013, 34(6): 718-725(in Chinese).

[26] 卓长飞, 武晓松, 封峰. 超声速流动中底部排气减阻的数值研究[J]. 兵工学报, 2014, 35(1): 18-26.
ZHUO C F, WU X S, FENG F. Numerical research on drag reduction of base bleed in supersonic flow[J]. Acta Armamentarii, 2014, 35(1): 18-26(in Chinese).

Numerical analysis on drag reduction effect of base cavity

LI Biao, WANG Liangming*, YANG Zhiwei

(School of Energy and Power Engineering, Nanjing University of Science and Technology, Nanjing 210016, China)

Abstract: In order to investigate the drag reduction mechanism of the base-cavity projectile, the flow field characteristics of M910 projectile are numerically simulated through the 3-dimensional steady CFD method. The zero-lift drag coefficient variation with Mach number are presented. The computational results have a good agreement with the experimental data. On this basis, a base cavity is introduced for M910 projectile that is named M910BC in this paper and then numerically simulated. The base flow field characteristics of the projectile with different base structures are compared and the drag reduction mechanism of the base cavity is analyzed. The results show that at subsonic speed, the drag reduction of the base-cavity projectile is found to be mainly due to the introduction of the high-pressure "dead zone" in base cavity and the displacement of the solid base with the compliant fluid boundary of the cavity base. Because of that, the forming location, shape and strength of the wake vortex are slightly changed. At transonic speed, the drag reduction effect of base cavity is vanished since the wake vortex is further from the base of projectile and the effects of the solid base and fluid boundary are the same. At supersonic speed, the drag reduction mechanism of the base cavity is that the mass of the recirculation region is increased by the flow of the base cavity, which is similar to the drag reduction mechanism of the base bleed projectile.

Keywords: drag reduction effect; base flow field; numerical simulation; base-cavity projectile; zero-lift drag coefficient

Received: 2020-11-09; **Accepted**: 2021-02-26; **Published online**: 2021-03-10 10:01

URL: kns.cnki.net/kcms/detail/11.2625.V.20210309.1034.005.html

* **Corresponding author**. E-mail: lmwang802@163.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0630

基于 QZSS L1 信号的岸基 GNSS-R 码延迟海面测高

全下文载

储俪[1,2], 贺凯飞[1,*], 高凡[2,3], 贺匀峤[3], 孟馨悦[3], 徐天河[2]

(1. 中国石油大学(华东) 海洋与空间信息学院, 青岛 266580; 2. 山东大学 空间科学研究院, 威海 264209;
3. 山东大学 空间科学与物理学院, 威海 264209)

摘 要: 为了探究 GNSS L1 波段中信号不同调制方式的 GNSS-R 码相位延迟海面的测高性能,开展了双天线岸基 GNSS-R 海面测高实验,收集了采样率为 40 MHz 的原始中频数据。利用自主开发的 GNSS-R 测高软件接收机和反演软件对实验数据进行了处理分析,同时获得了基于 QZSS L1C/A 码和 L1C 码的 GNSS-R 海面测高结果,并分别与岸基同步观测的雷达高度计的测高值进行对比,以对反演结果进行精度评定。实验结果表明:基于 QZSS L1C/A 码和 L1C 码的海面测高精度最优可达 0.63 m 和 0.4 m,L1C 码的延迟 GNSS-R 测高精度明显高于 L1C/A 码。此外,GNSS-R 测高精度会随着卫星高度角的增加而有所增加。

关 键 词: GNSS-R; QZSS; L1C; L1C/A; 测高
中图分类号: V221.3; TB553
文献标志码: A **文章编号**: 1001-5965(2022)04-0691-07

GNSS-R(global navigation satellite system-reflectometry)海面测高作为一项新兴的对地观测技术,主要优势在于能够对海面以较低成本进行高时空分辨率的观测,以期在海啸预警、海平面上升研究等领域发挥重要作用。双天线 GNSS-R 测高是通过测量直射信号与反射信号之间的路径差,从而获取反射天线到海面的垂直高度,并结合 GNSS 定位信息,反演得到海面测高结果。其路径差可以通过载波相位[1]和码相位延迟[2]2 种方式获取。文献[1,3-9]分析了基于 GPS L1/L2 频段的载波相位延迟测高结果,但在海面波动较大的情况下,接收机无法稳定跟踪反射载波相位信号。文献[2,10-14]分析了基于 GPS L1C/A 码和 BDS B1I 码的海面测高结果,但 GPS L1C/A 码和 BDS B1I 码调制方式简单且带宽较窄,制约了其 GNSS-R 测高精度。

近年来,GNSS 信号编码结构和调制方式不断升级,出现了多种抗多径性能更优、精度更高的民用测距码。文献[15]分析了更高码率的 GPS L5 码和 GALILEO E5 码的信号特性及可能获得的测高精度,而 L1C 码作为新一代 GPS 高精度民用码,采用了 BOC 调制方式,可使其有效提高测距精度及抗多路径能力[16]。目前,GPS 系统仅在 BLOCK III 卫星上播发 L1C 信号,且 BLOCK III 在轨卫星仅有 3 颗,其卫星信号难以覆盖中国东部沿海区域。QZSS 系统作为 GPS 在亚太地区的增强系统,在 L1 频段上调制有 L1C 码和传统的 L1C/A 码[17],而在中国东部沿海区域均可接收来自 QZSS 卫星的反射信号,为探究 L1C 码的 GNSS-R 海面测高精度提供了可能。

收稿日期: 2020-11-12; **录用日期**: 2021-01-17; **网络出版时间**: 2021-02-22 14:03
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210222.1326.001.html
基金项目: 国家自然科学基金(41604027,41604003,42074028);青岛海洋科学与技术国家实验室开放基金(QNLM2016ORP0401);中央高校基本科研业务费专项资金(18CX02054A)
* **通信作者**. E-mail: kfhe@upc.edu.cn

引用格式: *储俪, 贺凯飞, 高凡, 等. 基于 QZSS L1 信号的岸基 GNSS-R 码延迟海面测高[J]. 北京航空航天大学学报, 2022, 48(4): 691-697. CHU T, HE K F, GAO F, et al. Coastal GNSS-R ocean altimetry based on code delay of QZSS L1 signal[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 691-697 (in Chinese).*

北京航空航天大学学报

岸基 GNSS-R 测高技术可作为现有验潮站等近海海面高监测方式的一种有效补充。在传统验潮站的建设中,需要建造验潮井等施工过程复杂、耗费人力、物力较多的基础设施。若采用 GNSS-R 技术进行观测,仅需将设备固定在沿海岸边高处即可,具有成本低、架设方便等特点。通过在沿岸地区较为密集地布设 GNSS-R 监测设备,可为海洋科学和大地测量学研究提供重要的基础数据。因此,本文在山东省威海市开展了双天线岸基 GNSS-R 海面测高实验,同时,基于自主开发的 GNSS-R 测高软件接收机,采用了码相位延迟测高方法,系统分析了 QZSS L1C/A 码和 L1C 码的海面测高精度。

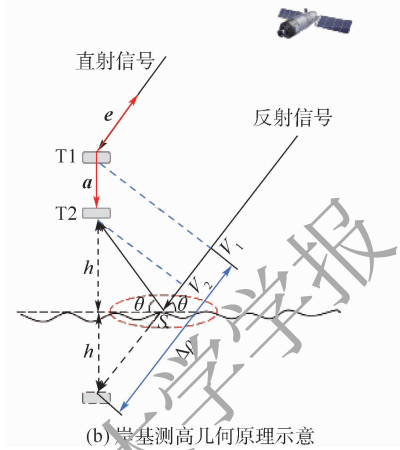
1 岸基 GNSS-R 测高理论

1.1 岸基测高几何模型

本文采用的岸基 GNSS-R 测高几何模型如图 1所示。图中:h 为反射天线相位中心到海面的垂直距离;θ 为卫星高度角;$\Delta\rho$ 为反射信号相对于直射信号的路径延迟;$\boldsymbol{a}$ 为直射和反射天线相位中心间基线矢量;$\boldsymbol{e}$ 为直射天线相位中心与卫星连线方向的矢量;T1、T2 分别为垂直向上、向下的极化天线;V_1、V_2 分别为天线 T1 和 T2 在卫星与直射天线相位中心连线方向的垂直投影。反射天线到海面的垂直距离小于 6 m,因此不考虑地球曲率影响。



(a) 卫星信号接收天线



(b) 岸基测高几何原理示意

图 1 岸基测高几何模型

Fig. 1 Coastal altimetry geometric model

$\Delta\rho$ 通常可表示为

$$\Delta\rho = \rho_r(t_r - t_e) - \rho_d(t_d - t_e) + \varepsilon_\rho \tag{1}$$

式中:ρ_r、ρ_d 分别为反射、直射信号传播路径长度;t_r、t_d、t_e 分别为反射天线信号接收时刻、直射天线信号接收时刻、卫星信号发送时刻;ε_ρ 为信号传播路径相关误差。

通过路径延迟 $\Delta\rho$ 与卫星高度角 θ 的几何关系,可求得反射天线到海面的垂直高度 h,h 可表示为[1-2]

$$h = \frac{\Delta\rho - V_1V_2}{2\sin\theta} + \varepsilon_h \tag{2}$$

$$V_1V_2 = |\boldsymbol{a}\cdot\boldsymbol{e}| \tag{3}$$

式中:V_1V_2 为矢量 $\boldsymbol{a}$ 在卫星与直射天线相位中心连线方向的投影;ε_h 为噪声。

1.2 QZSS L1 波段信号

QZSS 系统由 1 颗 GEO 和 3 颗 IGSO 卫星组成。目前,GPS 和 QZSS 在 L1 波段上同时调制 L1C/A 码和 L1C 码,以确保同现代化 GNSS 信号的兼容性及互操作性。L1 波段信号结构及其自相关函数如表 1 和图 2 所示[17],L1C/A 码由测距码和数据码进行异或相加后,并对载波进行 BPSK 调制而成。为了避免信号在实现测距及传输导航电文时可能出现的性能冲突,L1C 信号采用数据通道加导频通道的结构[16]。该信号由相互正交的数据分量和导频分量信号复合调制而成,数据分量由数据码和测距码异或相加后实现扩频,导频分量由测距码子码和测距码异或相加后实现扩频。扩频后的数据分量和导频分量分别对相互正交的载波进行 BOC 调制后复合成为复包络信号,其中,数据分量存在数据码周期跳变。

表 1 L1C/A 码和 L1C 码信号结构

Table 1 Signal structures of L1C/A and L1C code

信号类型	信号分量	载波频率/MHz	码率/MHz	调制方式	码长	功率比
L1C/A	数据	1 575.42	1.023	BPSK	1 023	
L1C	数据	1 575.42	1.023	BOC (1,1)	10 230	1/4
	导频			TMBOC (6,1,4/33)	10 230	3/4

北京航空航天大学学报

北航学报 赠阅

由表 1 和图 2 可知,L1C 码与 L1C/A 码的码率同为 1.023 MHz,并且归一化后的自相关函数皆存在主峰。与 L1C/A 码相比,新一代 L1C 信号在数据通道和导频通道中分别采用了 BOC(1,1)和 TMBOC(6,1,4/33)调制方式,其中,BOC(m,n)中的 m 和 n 分别是以 1.023 MHz 为基数归一化后的子载波速率和扩频码速率,TMBOC(6,1,4/33)为相互正交的 BOC(6,1)与 BOC(1,1)分量[16]。L1C 信号子载波的调制,使其自相关函数主峰较 L1C/A 码更尖锐,且主峰宽度约为 L1C/A 码的一半,因此,L1C 码的测距性能将优于传统的 L1C/A 码[18]。

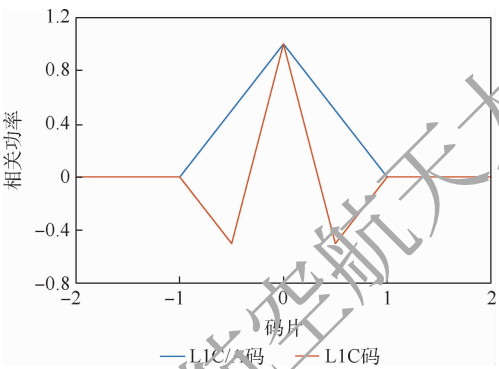


图 2　L1C/A 码和 L1C 码自相关函数

Fig. 2　Autocorrelation function of L1C/A and L1C code

1.3　码相位路径延迟获取

本文将路径延迟的获取分为 2 步:

步骤 1　通过射频前端采集数字中频信号。

步骤 2　利用软件接收机处理数字中频信号并获取码相位路径延迟。

1.3.1　中频信号获取

软件接收机射频前端信号处理流程如图 3 所示,流程如下:

1)右旋和左旋圆极化天线(PHCP 和 LHCP)分别接收到的直射信号和反射信号进入滤波器和放大器。信号经过滤波器后,除 L1 波段外的高频噪声被滤除,通过放大器对 L1 波段信号进行功率放大。

2)信号经过调整和放大后,进入混频器。信号经过混频后可获得 0.42 MHz 的低频信号,其他高频噪声将被滤除。

3)0.42 MHz 的低频信号经过 A/D 转换器后将从模拟信号采样为数字信号。其采样率为 40 MHz,符合香农采样定理[19]。数字中频信号进行 4 Bit 量化,且按高位到低位传输到存储设备中以便进一步处理。

1.3.2　中频信号处理

存储设备中的 4 Bit 数字中频信号,采用基于 C 语言平台开发的 GNSS-R 软件接收机进行处理。

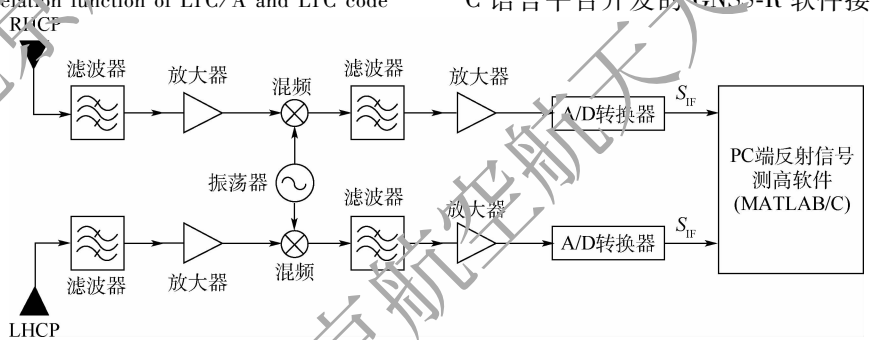


图 3　射频前端处理流程

Fig. 3　Processing flow of RF front-end

由于 L1C 信号中的数据分量和导频分量功率比为 1:3,导频分量的信号强度远高于数据分量,只处理 L1C 信号中的导频分量。中频信号经过软件处理后可分别获得 L1C/A 码和 L1C 码的海面测高结果。GNSS-R 软件接收机处理流程如图 4 所示,其流程如下:

1)粗捕获反射信号确定可见卫星。基于频域并行码相位捕获算法计算卫星相关功率是否超过捕获阈值,超过阈值则输出卫星 PRN、码相位及多普勒频移信息。

2)以当前卫星码相位及多普勒频移信息为基准,同时对直射信号和反射信号重新进行精捕获,获取直射和反射信号间码相位延迟。

3)基于图 1 中几何模型,通过码相位延迟及

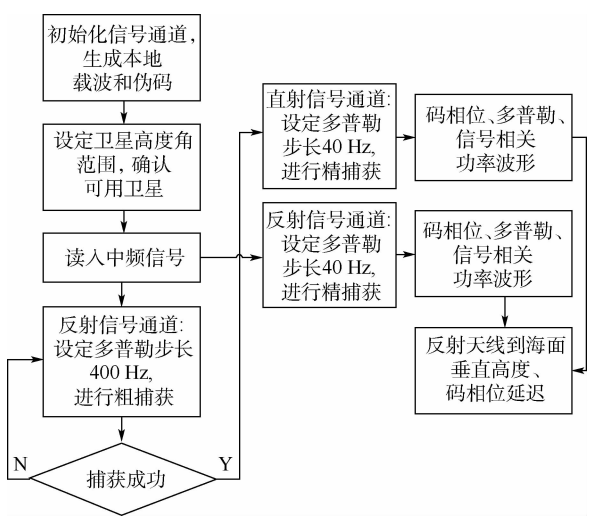


图 4　GNSS-R 软件接收机处理流程

Fig. 4　Processing flow of GNSS-R software receiver

北京航空航天大学学报

卫星高度角可计算出反射天线到海面的垂直高度。

2 算例分析

2.1 实验设计

本文实验地点位于山东省威海市的某一海边栈桥,实验时间为北京时间 2020 年 7 月 10 日 8:00—15:00。实验场地附近无高大建筑物遮挡,观测时段内风速约为 2 m/s,海面较为平静。图 5 为实验场地及设备位置,栈桥最前端距离海岸线约 100 m。图 5(a)中,竖直向上的直射天线和竖直向下的反射天线保持在同一铅垂线上,且反射和直射天线同时朝向南面海域,可保证更多可用卫星。为了进行外部验证,在距离反射天线水平距离约为 15 m 的固定位置处安置工作频率为 26 GHz的地基雷达测高仪,其到反射天线的垂直距离为 1.153 m,输出频率为 1 Hz,测距精度为 ±3 mm,可用于 GNSS-R 海面测高结果的精度检核。



(a) 地基雷达测高仪

(b) 设备相对位置

图 5 实验场地及设备位置

Fig. 5 Experimental site and equipment location

2.2 实验结果与分析

2.2.1 伪码相关功率波形

观测时段内捕获到 J1、J2、J3 共 3 颗卫星,通过可见卫星的直射信号和反射信号间码相位延迟来计算出反射天线到海面垂直高度。以可见卫星 J1 为例,L1C/A 码和 L1C 码的码相位延迟及相关功率波形变化如图 6 所示。可知,L1C 码的相关功率波形较 L1C/A 码更平滑,且 L1C 码的主峰更尖锐。卫星信号经海面反射后会出现功率损耗,因此,L1C/A 码和 L1C 码的反射信号相关功率值小于直射信号。图 6(b)中,由于 L1C 码中采用了 BOC 调制方式,使其相关功率波形的主峰附近存在 2 个副峰,且带宽大于 L1C/A 码。

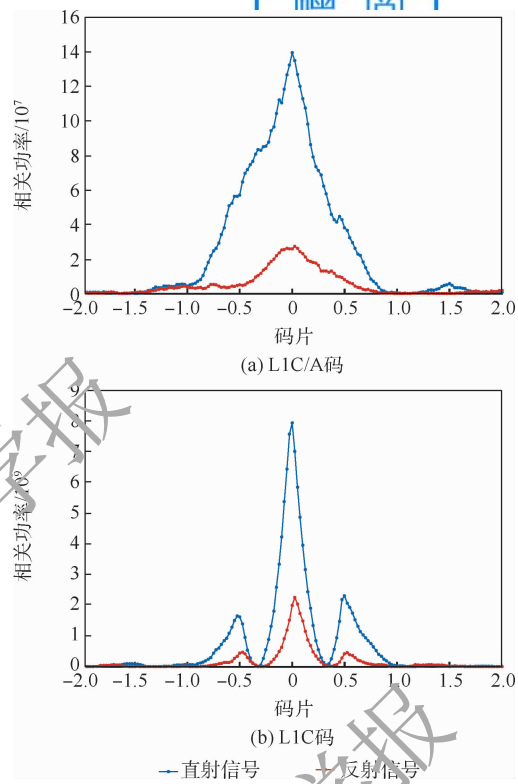


图 6 L1C/A 码和 L1C 码相关功率波形对比

Fig. 6 Correlation power waveform comparison between L1C/A and L1C code

2.2.2 测高结果分析

观测时段内,测站位置处不同卫星高度角的变化如图 7 所示。

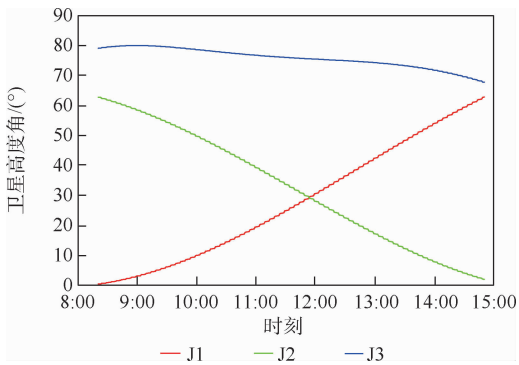


图 7 J1、J2、J3 卫星高度角变化(北京时间)

Fig. 7 J1, J2 and J3 satellite elevation changes (Beijing time)

由图 7 可知,J1 的可用卫星高度角变化范围为 50° ~ 65°,J2 的可用卫星高度角变化范围为 65° ~ 50°,而 J3 的可用卫星高度角变化范围为 80° ~ 68°,J3 卫星高度角要始终大于 J1、J2 卫星。进一步分析可知,在该观测时间段内,J3 的运行

北京航空航天大学学报

轨迹在远地点附近,而 J1、J2 的运行轨迹在近地点附近,因此,J3 的卫星高度角的变化速率小于 J1、J2。根据 GNSS-R 软件接收机输出的码相位路径延迟并结合卫星高度角,可分别计算出 L1C/A 码和 L1C 码的反射天线到海面的垂直高度。两者测高结果对比如图 8 所示,空白区域表示该观测时段内无数据。

由图 8 可知,海面的潮位最大变化量约为 2 m,L1C/A 码和 L1C 码测高结果均在雷达测高仪观测值上下波动,测高结果反映了潮位变化,验证了 GNSS-R 软件接收机的可用性。此外,J1、J2、J3 的 L1C/A 码的测高结果的离散程度均大于 L1C 码,且 L1C/A 码测高结果的系统偏差大于 L1C 码。进一步分析,相同时刻可见卫星的 L1C/A 码和 L1C 码延迟岸基测高残差序列对比结果如图 9、图 10 和表 2 所示。

由图 9、图 10 和表 2 可知,在 8:00—10:00 观测时段内,J2 的码相位观测值结果中,L1C/A 码、L1C 码的均方根误差分别为 0.917、0.695 m,标准差分别为 0.908、0.514 m;J3 的 L1C/A 码、L1C 码的均方根误差分别为 0.634、0.563 m,其标准差分别为 0.597、0.313 m。J2 码相位测高结果的误差范围约为 ±3 m,J3 码相位测高结果的误差小于 J2,但其测高结果存在部分系统偏差。结果表明,J2 和 J3 的 L1C 码相位测高结果相较于 L1C/A 码更稳定,且 L1C 码的测高精度相对 L1C/A 码精度分别增加了 24%、11%。

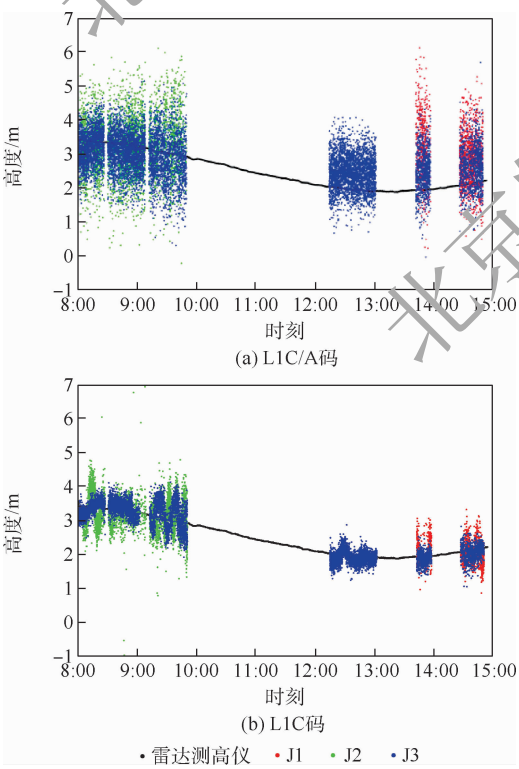


图 8　L1C/A 和 L1C 码 GNSS-R 测高结果对比

Fig. 8　Comparison of L1C/A and L1C code GNSS-R altimetry results

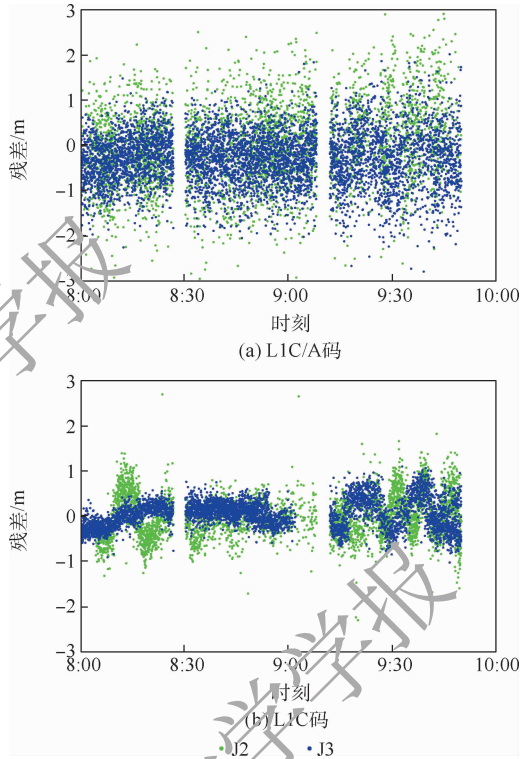


图 9　J2、J3 岸基测高结果残差序列对比

Fig. 9　Comparison of residual sequence of coastal altimetry results of J2 and J3 satellites

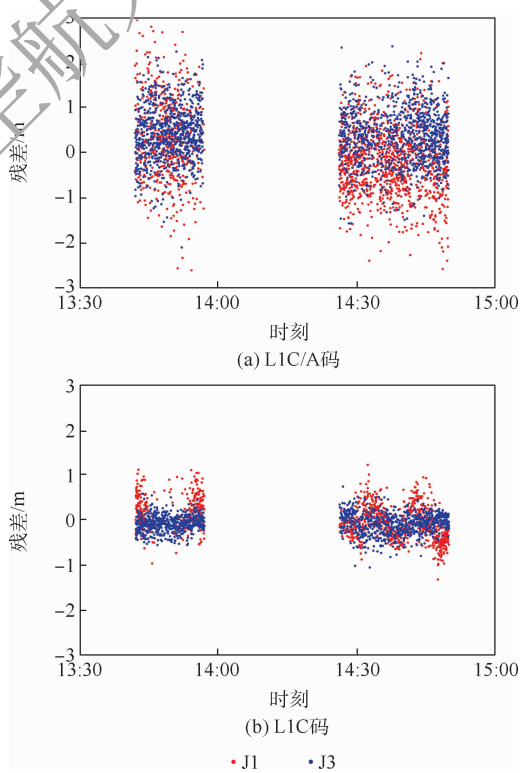


图 10　J1、J3 码延迟岸基测高结果残差序列对比

Fig. 10　Comparison of residual sequence of code delay coastal altimetry results of J1 and J3 satellites`

北京航空航天大学学报

北航学报 赠阅

表 2 J1、J2、J3 码延迟和岸基测高结果精度对比

Table 2 Comparison of the accuracy of code delay coastal altimetry results of J1, J2 and J3 satellites

观测类型	时刻	PRN	均方根误差/m	标准差/m	最大误差/m	最小误差/m
L1C/A	8:00—10:00	J2	0.917	0.908	2.977	-2.937
		J3	0.634	0.597	2.134	-2.673
L1C		J2	0.695	0.514	2.602	-2.341
		J3	0.563	0.313	1.527	-0.609
L1C/A	13:30—15:00	J1	1.226	0.855	2.985	-1.728
		J3	0.794	0.618	2.891	-1.977
L1C		J1	0.726	0.391	1.846	-0.701
		J3	0.395	0.226	1.164	-0.643

在 13:30—15:00 观测时段内，J1 的 L1C/A 码、L1C 码的均方根误差分别为 1.226、0.726 m，标准差分别为 0.855、0.391 m；J3 的 L1C/A 码、L1C 码的均方根误差分别为 0.794、0.395 m，标准差分别为 0.618、0.226 m。J1 码相位测高结果的误差范围约为 ±3 m，J3 码相位测高结果的误差范围约为 ±2 m，J1 和 J3 的码相位测高结果均存在部分系统偏差。结果表明，J1 和 J3 的 L1C 码测高相位结果较 L1C/A 码更稳定，L1C 码的测高精度相比 L1C/A 码精度可分别增加了 41%、50%。进一步分析，在相同时段内，J3 卫星高度角明显大于 J1、J2，J3 的 L1C/A 码、L1C 码测高精度比 J2 的分别增加了 31%、19%，比 J1 的分别增加了 35%、45%。文献[13]分析了码相位测高精度与卫星高度角之间联系，卫星高度角越大，码相位路径延迟中的噪声影响越小，卫星信号信噪比随之增大，海面测高精度相应提高。同时，图 9 和图 10 结果表明，卫星高度角较大时，残差序列比较稳定；卫星高度角较小时，残差序列在一定范围内上下波动。这是由于直射信号进入了反射天线，不同信号间相互叠加产生的多路径效应会对测高结果产生影响，进一步分析并削弱该误差因素可提高海面测高精度。

3 结 论

本文开展了双天线岸基海面测高实验，基于自主开发的 GNSS-R 测高软件接收机对 QZSS L1 波段信号的码相位延迟测高性能进行了分析，可得以下结论：

1）通过 GNSS-R 岸基实验验证了 QZSS L1C 码相位岸基测高精度，同时也验证了自主开发的 L1C/A 码和 L1C 码 GNSS-R 软件接收机的可用性。

2）海况良好时，L1C 码、L1C/A 码的测高精度分别为 0.60、0.94 m，L1C 码、L1C/A 码最优精度分别为 0.4、0.63 m。L1C 码的测高精度相对于 L1C/A 码精度增加了 32%。

3）在相同观测时段内，卫星高度角越大，路径延迟中的噪声越小，卫星信号信噪比随之增大，测高精度相应提高。

参考文献（References）

[1] MARTIN-NEIRA M. A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry[J]. ESA Journal, 1993, 17(4): 331-355.

[2] MARTIN-NEIRA M, CAPARRINI M, FONT-ROSSELLO J, et al. The PARIS concept: An experimental demonstration of sea surface altimetry using GPS reflected signals[J]. IEEE Transactions on Geoscience and Remote Sensing, 2001, 39(1): 142-150.

[3] HELM A. Ground-based GPS altimetry with the L1 OpenGPS receiver using carrier phase-delay observations of reflected GPS signals[R]. Potsdam: Deutsches GeoForschungsZentrum GFZ, 2008: 67-85.

[4] LOWE S T, ZUFFADA C, CHAO Y, et al. 5-cm-precision aircraft ocean altimetry using GPS reflections[J]. Geophysical Research Letters, 2002, 29(10): 1375-1378.

[5] 白伟华，孙越强，朱光武，等. 利用岸基 GNSS-R 信号反演湖面高度[C]// 第一届中国卫星导航学术年会，2010: 1646-1652.
BAI W H, SUN Y Q, ZHU G W, et al. Lake surface height retrieve using coastal GNSS-R signals[C]// Proceedings of the First China Satellite Navigation Conference, 2010: 1646-1652 (in Chinese).

[6] CARDELLACH E, RIUS A, MARTIN-NEIRA M, et al. Consolidating the precision of interferometric GNSS-R ocean altimetry using airborne experimental data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(8): 4992-5004.

[7] PASCUAL D, CAMPS A, MARTIN F, et al. Precision bounds in GNSS-R ocean altimetry[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(5): 1416-1423.

[8] SEMMLING A M, BECKHEINRICH J, WICKERT J, et al. Sea surface topography retrieved from GNSS reflectometry phase data of the GEOHALO flight mission[J]. Geophysical Research Letters, 2014, 41(3): 954-960.

北京航空航天大学学报

[9] 王娜子,鲍李峰,高凡. 逐历元 GNSS-R 测高单差和双差算法[J]. 测绘学报,2016,45(7):795-802.
WANG N Z,BAO L F,GAO F. Improved water level retrieval from epoch-by-epoch single and double difference GNSS-R algorithms[J]. Acta Geodaetica et Cartographica Sinica,2016,45(7):795-802(in Chinese).

[10] LOWE S T,ZUFFADA C,LABRECQUE J L,et al. An ocean-altimetry measurement using reflected GPS signals observed from a low-altitude aircraft[C]//2000 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press,2000:2185-2187.

[11] ZHANG Y,TIAN L,MENG W,et al. Feasibility of code level altimetry using coastal BeiDou-reflection (BeiDou-R) setups [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2015,8(8):4130-4140.

[12] ZHANG Y,LIU F,GU Q,et al. Study of accurate ocean-altimetry with GNSS-R[C]//2013 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2013:1410-1413.

[13] GAO F,XU T H,WANG N Z,et al. A shipborne experiment using a dual-antenna reflectometry system for GPS/BDS code delay measurements[J]. Journal of Geodesy,2020,94(9): 1-18.

[14] D'ADDIO S,BUCK C,MARTINNEIRA M. Paris altimetry precision prediction with Galileo signals-in-Space[C]// 2008 IEEE International Geoscience and Remote Sensing Symposium. Piscataway:IEEE Press,2008:1410-1414.

[15] 杨东凯,张其善. GNSS 反射信号处理基础与实践[M]. 北京:电子工业出版社,2012:127-142.
YANG D K,ZHANG Q S. GNSS reflected signal processing: Basics and practice [M]. Beijing: Publishing House of Electronics Industry,2012:127-142(in Chinese).

[16] 姚铮,陆明泉. 新一代卫星导航系统信号设计原理与实现技术[M]. 北京:电子工业出版社,2016:82-114.
YAO Z,LU M Q. The signal design principle and realization technology of the new generation satellite navigation system [M]. Beijing:Publishing House of Electronics Industry,2016: 82-114(in Chinese).

[17] Japan Cabinet Office. Performance standard and interface specification:IS-QZSS-PNT-003[S/OL]. (2018-11-05)[2020-11-01]. https://qzss.go.jp/en/technical/ps-is-qzss/ps-is-qzss.html.

[18] BETZ J W. Binary offset carrier modulations for radionavigation [J]. Navigation,2001,48(4):227-246.

[19] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal,1948,27(3):379-423.

Coastal GNSS-R ocean altimetry based on code delay of QZSS L1 signal

CHU Ti[1,2], HE Kaifei[1,*], GAO Fan[2,3], HE Yunqiao[3], MENG Xinyue[3], XU Tianhe[2]

(1. College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China;
2. Institute of Space Science, Shandong University, Weihai 264209, China;
3. School of Space Science and Physics, Shandong University, Weihai 264209, China)

Abstract: In order to explore the altimetry performance of GNSS-R code phase delay measurement with different modulation methods in GNSS LI band signals, a coastal dual-antenna GNSS-R ocean altimetry experiment was carried out, and the original intermediate frequency data with a sampling rate of 40 MHz was collected. The experimental data was processed using the self-developed GNSS-R altimetry software receiver and inversion software, and the GNSS-R ocean altimetry results of QZSS L1C/A code and L1C code were obtained simultaneously. Then we compare them with the measured values of the radar altimeter that were simultaneously observed coastally, and used to evaluate the accuracy of the GNSS-R inversion result. The experimental results show that the ocean altimetry accuracy based on L1C/A code and L1C code is about 0.63 m and 0.4 m, respectively. It can be seen that the GNSS-R measurement accuracy of the L1C code delay is significantly higher than that of the L1C/A code. In addition, the accuracy of GNSS-R will be improved as the satellite elevation increases gradually.

Keywords: GNSS-R; QZSS; L1C; L1C/A; altimetry

Received: 2020-11-12; **Accepted**: 2021-01-17; **Published online**: 2021-02-22 14:03

URL: kns.cnki.net/kcms/detail/11.2625.V.20210222.1326.001.html

Foundation items: National Natural Science Foundation of China (41604027, 41604003, 42074028); Open Fund of Qingdao National Laboratory of Marine Science and Technology (QNLM2016ORP0401); the Fundamental Research Funds for the Central Universities (18CX02054A)

* **Corresponding author**. E-mail: kfhe@upc.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0636

复合材料层板冰雹高速冲击损伤预测及失效分析

全下文载

张超*，方鑫，刘建春
（江苏大学 机械工程学院，镇江 212013）

摘 要：针对冰雹冲击对复合材料结构安全造成的潜在风险，提出了一种基于连续介质损伤力学的非线性有限元模型，研究了碳纤维复合材料层板冰雹高速冲击力学行为。综合采用拉格朗日法和光滑粒子流体动力学（SPH）法对冰雹进行建模，引入水的状态方程描述冰雹破碎后的流动特性；考虑应变率的单向复合材料本构模型，根据三维 Hashin 失效准则及材料刚度折减方案，进行复合材料层内损伤预测；引入界面单元结合双线性内聚力模型模拟层间分层现象；编写用户材料 VUMAT 子程序，实现基于 ABAQUS/Explicit 显式模块的数值求解。模拟了冰雹高速冲击复合材料层板的瞬态过程，分析了材料的损伤特性和失效机理。探讨了冰雹冲击速度、冲击角度对层板冲击损伤性能的影响，为复合材料结构冰雹冲击问题数值分析提供参考。

关 键 词：碳纤维复合材料；冰雹冲击；高速冲击；损伤；数值模拟
中图分类号：TB332
文献标志码：A **文章编号**：1001-5965(2022)04-0698-10

碳纤维增强复合材料层板具有比强度高、比模量大、可设计性强等性能优势，在现代飞机结构中应用非常广泛。飞机在飞行、停靠期间常会遇到冰雹等恶劣天气，冰雹冲击属于软体冲击事件，对复合材料层板往往会产生肉眼难以观察的隐形损伤，如层间分层、基体开裂等，这对飞机关键部位的结构承载非常危险。因此，需要深入开展碳纤维增强复合材料层板冰雹高速冲击力学响应及损伤机理研究。

复合材料层板冰雹冲击是一个高度非线性的瞬态过程，冲击过程中涉及接触分析、冰雹破裂、层板损伤等问题，研究方法主要是实验研究和数值模拟。实验研究成本高、周期长且受限于具体的实验条件；而有限元数值模拟可以较好地解决冰雹冲击这一瞬态动力学问题，详细获取冲击过程中冰雹破裂、流动特性和层板损伤演化过程，具有边界适应性强、求解精度及求解效率高等优点，已成为研究复合材料高速冲击问题的首选方法[1-2]。

在实验研究方面，学者们通过开展冰球高速冲击复合材料层板实验[3-6]，发现冰球冲击动能和冲击力峰值之间呈线性关系，层间损伤能量阈值和冰雹直径与层板厚度的比值呈线性关系，并利用微观扫描技术获得复合材料层板的损伤形貌及破坏阈值。数值模拟方面，冰雹数值模型主要采用传统的拉格朗日法和光滑粒子流体动力学（smoothed particle hydrodynamics，SPH）法。早期，学者 Kim 和 Kedward[3] 使用拉格朗日网格并

收稿日期：2020-11-15；**录用日期**：2021-03-12；**网络出版时间**：2021-04-07 14:13
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210407.1353.001.html
基金项目：江苏省普通高校研究生科研创新计划（KYCX20-3076）；江苏省自然科学基金（BK20180855）；机械结构力学及控制国家重点实验室开放课题（MCMS-E-0219Y01）
*通信作者. E-mail：zhangchao@ujs.edu.cn

引用格式：张超，方鑫，刘建春. 复合材料层板冰雹高速冲击损伤预测及失效分析[J]. 北京航空航天大学学报，2022，48(4)：698-707. ZHANG C，FANG X，LIU J C. Damage prediction and failure mechanism of composite laminates under high-velocity hailstone impact[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：698-707(in Chinese).

北京航空航天大学学报

采用弹塑性模型对冰雹进行建模,模型中未考虑冰雹的应变率效应及压力与体积的非线性关系,在冰雹破碎后的物理特征模拟方面不够理想。SPH 法可以有效解决高速冲击下网格大变形的问题,冰雹单元失效后转化为粒子,可以有效再现冰破碎、流动的整个力学行为。

Coles 等[7]对金属弹丸和冰球冲击复合材料层板进行了数值对比研究。Pernas-Sanchez 等[8]采用 SPH 法建立了冰雹高速冲击复合材料层板的有限元模型,采用界面接触(cohesive contact)来模拟层间分层行为,预测了不同冰雹直径和层板厚度下分层面积和冲击速度之间的定量关系。文献[5,9-10]在冰雹冲击复合材料层板有限元模拟中,采用内聚力界面单元(cohesive element)结合 Traction-Separation 本构关系模拟层间分层现象,发现层板的分层开裂损伤形貌与实验结果较为吻合,说明界面接触和界面单元均可以有效模拟冰雹冲击下复合材料层板的层间分层现象,但界面接触只是一种接触属性,不是材料属性,没有创建额外单元代表界面这种组分实体,难以直接表征材料的界面力学行为。

张晓晴等[11]建立了冰雹冲击复合材料加筋结构的有限元模型,预测了冲击位置、冲击能量及冲击入射角对复合材料加筋壁板冲击动力学响应的影响。Dolati 等[12]从能量吸收和破坏模式角度,对复合材料/波纹铝芯夹芯板开展了冰雹冲击数值模拟,讨论了冰雹弹形状、面板堆叠顺序和加固类型对夹芯板冲击响应的影响。

综上,国内外学者对冰雹冲击复合材料结构开展了一定的实验和模拟工作。但现有数值模型在冰雹破碎后的物理行为模拟方面还不理想,对冰雹高速冲击下纤维增强复合材料应变率效应的认识不足,相关的数值模拟和分析工作还需进一步深入。本文建立冰雹高速冲击复合材料层板有限元模型,模拟冰雹高速冲击复合材料层板的瞬态过程,分析材料的损伤失效机理并探讨冲击速度、角度对层板损伤性能的影响。

1　损伤本构模型

1.1　冰雹本构模型

在高速冲击条件下,冰的弹塑性模型与应变率具有明显的相关性,低应变率下冰材料表现为韧性,而高应变率下表现为脆性。其中,冰的拉伸性能受应变率影响不大,而压缩屈服强度受应变率影响较大[13]。基于 Tippmann[14]建立的考虑应变率效应的弹塑性模型,通过定义张力失效准则来描述冰雹破裂后的流动特性。达到拉伸失效应力时,偏应力分量设为零且材料只承受压缩应力,失效后单元不删除。冰雹本构的偏应力失效类型设置为脆性,压应力失效类型设置为韧性。

引入水的 Gruneisen 状态方程来控制冰破碎后压力与体积之间的非线性关系,该方程描述了压力、密度和能量之间的关系。

压缩状态时($\mu>0$),冰雹模型的状态方程可以表示为[12]

$$P=\frac{\rho_0 C_S^2\mu\left[1+(1-0.5\gamma_0)\mu-0.5\alpha\mu^2\right]}{\left[1-(S_1-1)\mu-S_2\frac{\mu^2}{\mu+1}-S_3\frac{\mu^3}{(\mu+1)^2}\right]}+(\gamma_0+\alpha\mu)E \tag{1}$$

冰雹处于膨胀状态时($\mu<0$),方程表示为[10]

$$P=\rho_0 C_S^2\mu+(\gamma_0+\alpha\mu)E \tag{2}$$

式中:C_S为 U_s-U_p 曲线(U_s和 U_p分别为冲击波速度和粒子速度)的截距即体积声速;S_1、S_2、S_3为 U_s-U_p曲线斜率的第一、二、三系数;E、α和 γ_0 分别为起始内能、体积修正系数和 Gruneisen 常数;$\mu=\rho/\rho_0-1$,ρ 为当前密度,ρ_0为初始密度。

1.2　单向复合材料损伤本构模型

碳纤维增强复合材料由纤维和基体 2 种组分材料构成,纤维通常被认为是线弹性、横观各向同性材料,可以用线弹性单元来描述;基体可以看作为黏弹性、均匀各向同性材料,可以用黏弹性单元来描述。它们的应力-应变关系分别为

$$\sigma_f=E_f\varepsilon \tag{3}$$

$$\sigma_m(t)=\int_0^t E_r(t-\tau)\dot{\varepsilon}\mathrm{d}\tau \tag{4}$$

式中:$\dot{\varepsilon}$ 为应变率;E_f为纤维弹性模量;E_r 为基体松弛模量。

Karim 和 Fatt[15]给出了 E_r 的表达式为

$$E_r(t)=E_m+E_1\mathrm{e}^{\frac{-tE_1}{\eta_1}}+E_2\mathrm{e}^{\frac{-tE_2}{\eta_2}} \tag{5}$$

式中:E_m为基体的弹性模量;E_1、E_2为 Maxwell 体中的弹性系数;η_1、η_2为 Maxwell 体中的阻尼系数。

单向纤维增强复合材料轴向力学行为,通过线弹性单元和黏弹性基体并联的模型来表征。结合等应变假设,有

$$\bar{\sigma}_{11}(t)=\int_0^t E_{11}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau=V_f E_{f1}\varepsilon_{11}+(1-V_f)\int_0^t E_r(t-\tau)\dot{\varepsilon}_{11}\mathrm{d}\tau \tag{6}$$

式中:E_{f1}为纤维轴向弹性模量;V_f为单向复合材料的纤维体积含量。

将式(5)代入式(6),并由 $t=\varepsilon_{11}/\dot{\varepsilon}_0$,积分得

$$\bar{\sigma}_{11}(\varepsilon_{11}) = [E_{\mathrm{f1}}V_{\mathrm{f}} + E_{\mathrm{m}}(1-V_{\mathrm{f}})]\varepsilon_{11} + E_1V_{\mathrm{m}}\theta_{e1}\dot{\varepsilon}_0\left(1-\mathrm{e}^{-\frac{\varepsilon_{11}}{\dot{\varepsilon}_0\theta_{e1}}}\right)+E_2V_{\mathrm{m}}\theta_{e2}\dot{\varepsilon}_0\left(1-\mathrm{e}^{-\frac{\varepsilon_{11}}{\dot{\varepsilon}_0\theta_{e2}}}\right) \tag{7}$$

式中:$\dot{\varepsilon}_0$ 为常应变率;$\theta_{e1}=\eta_1/E_1$ 和 $\theta_{e2}=\eta_2/E_2$ 为 Maxwell 体的松弛时间;$V_{\mathrm{m}}=1-V_{\mathrm{f}}$。

基于等应变或等应力假设,单向复合材料在其他材料主方向上的本构关系表达如下[15]:

$$\begin{cases}\bar{\sigma}_{22}(\varepsilon_{22}) = \dfrac{E_{\mathrm{f2}}E_{\mathrm{m}}}{E_{\mathrm{f2}}(1-V_{\mathrm{f}}')+E_{\mathrm{m}}V_{\mathrm{f}}'}\varepsilon_{22} + \dot{\varepsilon}_0\dfrac{Q}{M}\left(1-\mathrm{e}^{-M\frac{\varepsilon_{22}}{\dot{\varepsilon}_0}}\right)+\dot{\varepsilon}_0\dfrac{R}{N}\left(1-\mathrm{e}^{-N\frac{\varepsilon_{22}}{\dot{\varepsilon}_0}}\right)\\ \bar{\tau}_{12}(\gamma_{12}) = \dfrac{G_{\mathrm{f12}}G_{\mathrm{m}}}{G_{\mathrm{f12}}(1-V_{\mathrm{f}}')+G_{\mathrm{m}}V_{\mathrm{f}}'}\gamma_{12} + \dot{\gamma}_0\dfrac{Q_{12}}{M_{12}}\left(1-\mathrm{e}^{-M_{12}\frac{\gamma_{12}}{\dot{\gamma}_0}}\right)+\dot{\gamma}_0\dfrac{R_{12}}{N_{12}}\left(1-\mathrm{e}^{-N_{12}\frac{\gamma_{12}}{\dot{\gamma}_0}}\right)\\ \bar{\tau}_{23}(\gamma_{23}) = \dfrac{G_{\mathrm{f23}}G_{\mathrm{m}}}{G_{\mathrm{f23}}(1-V_{\mathrm{f}}')+G_{\mathrm{m}}V_{\mathrm{f}}'}\gamma_{23} + \dot{\gamma}_0\dfrac{Q_{23}}{M_{23}}\left(1-\mathrm{e}^{-M_{23}\frac{\gamma_{23}}{\dot{\gamma}_0}}\right)+\dot{\gamma}_0\dfrac{R_{23}}{N_{23}}\left(1-\mathrm{e}^{-N_{23}\frac{\gamma_{23}}{\dot{\gamma}_0}}\right)\end{cases} \tag{8}$$

式中:G_{m} 为基体剪切模量;E_{f2} 为纤维横向弹性模量;$V_{\mathrm{f}}'=\sqrt{V_{\mathrm{f}}}$ 为修正的纤维体积含量;G_{f12}、G_{f23} 为纤维面内剪切模量;M、N、Q、R、M_{12}、N_{12}、Q_{12}、R_{12}、M_{23}、N_{23}、Q_{23}、R_{23} 分别为基于不同假设的不同模型中 Maxwell 体的迟滞参数,其具体表达式可参考文献[15]。

正交各向异性材料的应力-应变本构关系可表示为

$$\boldsymbol{\sigma}=\boldsymbol{C}\boldsymbol{\varepsilon}=\boldsymbol{U}^{-1}\boldsymbol{E}\boldsymbol{\varepsilon} \tag{9}$$

式中:$\boldsymbol{C}$ 为刚度矩阵;$\boldsymbol{U}$ 和 $\boldsymbol{E}$ 定义分别为

$$\boldsymbol{U}=\begin{bmatrix}1 & -\mu_{21} & -\mu_{31} & 0 & 0 & 0\\ -\mu_{12} & 1 & -\mu_{32} & 0 & 0 & 0\\ -\mu_{13} & -\mu_{23} & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix} \tag{10}$$

$$\boldsymbol{E}=\begin{bmatrix}E_{11} & 0 & 0 & 0 & 0 & 0\\ 0 & E_{22} & 0 & 0 & 0 & 0\\ 0 & 0 & E_{33} & 0 & 0 & 0\\ 0 & 0 & 0 & G_{23} & 0 & 0\\ 0 & 0 & 0 & 0 & G_{31} & 0\\ 0 & 0 & 0 & 0 & 0 & G_{12}\end{bmatrix} \tag{11}$$

根据单向纤维增强复合材料横观各向同性特性,其应变率相关的三维本构关系可表示为

$$\begin{bmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\\ \sigma_{23}\\ \sigma_{31}\\ \sigma_{12}\end{bmatrix}=\boldsymbol{U}^{-1}\begin{bmatrix}\int_0^t E_{11}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau\\ \int_0^t E_{22}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau\\ \int_0^t E_{22}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau\\ \int_0^t G_{23}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau\\ \int_0^t G_{12}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau\\ \int_0^t G_{12}(t-\tau)\dot{\varepsilon}(\tau)\mathrm{d}\tau\end{bmatrix}=\boldsymbol{U}^{-1}\begin{bmatrix}\bar{\sigma}_{11}\\ \bar{\sigma}_{22}\\ \bar{\sigma}_{33}\\ \bar{\sigma}_{23}\\ \bar{\sigma}_{31}\\ \bar{\sigma}_{12}\end{bmatrix} \tag{12}$$

对于冰雹高速冲击下的复合材料层板,层内具有纤维拉伸、纤维压缩、基体拉伸、基体压缩 4 种典型失效模式。三维 Hashin 失效准则[16]能够很好地预测这 4 种失效模式,在此作为复合材料单层的失效判据,具体表达式如下:

纤维拉伸失效($\sigma_{11}\geqslant 0$):

$$F_{\mathrm{ft}}=\left(\frac{\sigma_1}{X_{\mathrm{T}}}\right)^2+\beta\left(\frac{\sigma_{12}}{S_{12}}\right)^2+\beta\left(\frac{\sigma_{13}}{S_{13}}\right)^2\geqslant 1 \tag{13}$$

纤维压缩失效($\sigma_{11}<0$):

$$F_{\mathrm{fc}}=\left(\frac{\sigma_1}{X_{\mathrm{C}}}\right)\geqslant 1 \tag{14}$$

基体拉伸失效($\sigma_{22}+\sigma_{33}\geqslant 0$):

$$F_{\mathrm{mt}}=\left(\frac{\sigma_2+\sigma_3}{Y_{\mathrm{T}}}\right)^2+\frac{1}{S_{23}^2}(\sigma_{23}^2-\sigma_{22}\sigma_{33})+\left(\frac{\sigma_{12}}{S_{12}}\right)^2+\left(\frac{\sigma_{13}}{S_{13}}\right)^2\geqslant 1 \tag{15}$$

基体压缩失效($\sigma_{22}+\sigma_{33}<0$):

$$F_{\mathrm{mc}}=\left(\frac{\sigma_2+\sigma_3}{2S_{12}}\right)^2+\frac{\sigma_2+\sigma_3}{Y_{\mathrm{C}}}\left[\left(\frac{Y_{\mathrm{C}}}{2S_{12}}\right)^2-1\right]+\frac{1}{S_{23}^2}(\sigma_{23}^2-\sigma_2\sigma_3)+\left(\frac{\sigma_{12}}{S_{12}}\right)^2+\left(\frac{\sigma_{13}}{S_{13}}\right)^2\geqslant 1 \tag{16}$$

式中:β 为剪切修正因子;X_{T}、X_{C} 分别为轴向拉伸、压缩强度;Y_{T}、Y_{C} 分别为横向拉伸、压缩强度;S_{12},S_{13} 和 S_{23} 为剪切强度。

为了考虑应变率对强度参数的影响,将复合材料各个方向上的强度值与应变率之间的关系统一表示为

$$S^0=S_{\mathrm{r}}^0\left(1+\xi\ln\frac{|\dot{\varepsilon}|}{\dot{\varepsilon}_0}\right) \tag{17}$$

式中:S^0 为当前应变率下的强度;S_{r}^0 为参考应变率 $\dot{\varepsilon}_0$ 下的强度;ξ 为应变率修正系数。

上述 4 种典型失效模式贯穿于冰雹高速冲击

北京航空航天大学学报

复合材料层板的整个过程。当复合材料层板出现局部破坏后,层板不会直接失去全部承载能力,而是材料局部性能出现下降。因此,当复合材料层板开始发生局部损伤后,根据损伤模式的不同,将纤维和基体的弹性常数分别进行相应的折减,折减方案如表 1 所示。

表 1　材料刚度折减方案

Table 1　Material stiffness reduction scheme

失效模式	刚度折减系数					
	E_{f1}	E_{f2}	G_{f12}	G_{f23}	E_m	G_m
纤维拉伸失效	0.01	0.2	0.01	1	1	1
纤维压缩失效	0.01	0.2	0.01	1	1	1
基体拉伸失效	1	0.2	0.2	0.2	0.01	0.01
基体压缩失效	1	0.2	0.2	0.2	0.01	0.01

1.3　层间损伤模型

为了模拟冰雹冲击下复合材料层板的层间分层现象,引入界面单元结合双线性内聚力模型来描述界面的力学行为。

界面初始为线弹性阶段,本构关系为

$$t_i = K\delta_i \qquad i = 1,2,3 \tag{18}$$

式中:t_i 和 δ_i 分别为法向、切向应力和法向、切向相对位移;K 为界面罚刚度。

定义混合模式下相对位移 δ_m 的表达式为

$$\delta_m = \sqrt{\langle\delta_1\rangle^2 + \delta_2^2 + \delta_3^2} = \sqrt{\langle\delta_1\rangle^2 + \delta_{shear}^2} \tag{19}$$

式中:δ_{shear} 为界面切向相对位移;〈 〉为 Macaulay 算子。

界面起始损伤位移 δ_m^0 为[17]

$$\left(\frac{\langle t_1\rangle}{N}\right)^2 + \left(\frac{t_2}{S}\right)^2 + \left(\frac{t_3}{T}\right)^2 = 1 \tag{20}$$

式中:N 为界面法向强度;S、T 为界面剪切强度。

界面完全失效位移 δ_m^f 为[17]

$$\left(\frac{G_I}{G_{IC}}\right)^2 + \left(\frac{G_{II}}{G_{IIC}}\right)^2 + \left(\frac{G_{III}}{G_{IIIC}}\right)^2 = 1 \tag{21}$$

式中:G_I、G_{II}、G_{III}分别为 I、II、III 型裂纹应变能释放率;G_{IC}、G_{IIC}、G_{IIIC}分别为 I、II、III 型裂纹临界应变能释放率。

界面损伤起始后,损伤演化为

$$d = \frac{\delta_m^f(\delta_m^{max} - \delta_m^0)}{\delta_m^{max}(\delta_m^f - \delta_m^0)} \qquad d \in [0,1] \tag{22}$$

式中:考虑界面损伤的不可逆性,混合模式下界面最大相对位移 δ_m^{max} 定义为

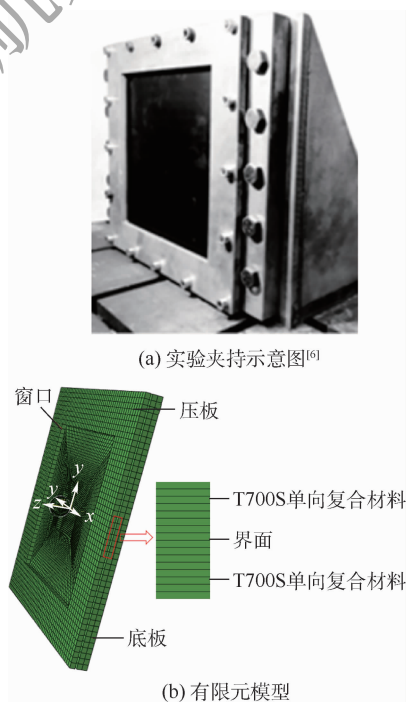
$$\delta_m^{max} = \max\{\delta_m^{max}, \delta_m\} \tag{23}$$

2　有限元模型及验证

2.1　冰雹冲击复合材料层板有限元模型

为了分析冰雹高速冲击下复合材料层板的损伤特性和失效机理,选取文献[6]的实验结果作为对比验证数据,复合材料层板夹持方式为四周固支,通过拧紧螺栓压紧底板和压板对实验件起到稳固作用(见图 1(a))。窗口尺寸为 200 mm × 200 mm,压板和底板尺寸为 300 mm × 300 mm,层板尺寸为 250 mm × 250 mm。复合材料层板的铺层方式为正交铺层$[0°/90°]_{4S}$和准各向同性铺层$[0°/45°/90°/-45°]_{2S}$,共 16 层,每层厚度为 0.125 mm。冰雹直径为 25.4 mm,采用具有沙漏控制的八节点 C3D8R 单元离散,单元尺寸为 1 mm。达到失效应力时自动转化为 SPH 粒子。复合材料层板每层均采用 C3D8R 单元离散,并且建立局部坐标系来定义每层纤维的局部方向。

为模拟冰雹冲击下层板分层现象,在复合材料各层之间插入零厚度内聚力界面单元 COH3D8。在冰雹高速冲击过程中,层板的损伤主要集中在冲击中心区域。在层板网格划分时,在中心冲击



(a) 实验夹持示意图[6]

(b) 有限元模型

图 1　冰雹高速冲击复合材料层板实验装置及有限元模型

Fig. 1　Experimental device and finite element model of composite laminates under high-velocity hailstone impact

区域细化网格,单元尺寸为 1.74 mm × 1.74 mm。从中心向外网格密度由密到疏,采用了 bias 功能从内到外布置种子,比例为 2,且该网格尺度下计算结果收敛,这样既保证了计算精度又提高了运算效率。

冰雹和层板之间采用通用接触,接触属性为无摩擦的硬接触。层板与压板、底板之间建立绑定约束,压板和底板均为离散刚体,并通过固定压板与底板的参考点来模拟实验中的夹持条件。所建立的冰雹冲击复合材料层板有限元模型如图 1(b)所示,该模型共由 165 331 个节点、6 176 个 R3D4 单元、87 200 个 C3D8R 单元和 67 500 个 COH3D8 单元组成。

水的 Gruneisen 状态方程参数、冰雹基本力学性能参数及基于应变率的冰雹屈服强度参数如表 2[12]、表 3[18]和表 4[14]所示,界面单元材料参数和单向复合材料的材料参数如表 5[19-20]和表 6[6,15,21]所示。

表 2 Gruneisen 状态方程参数[12]

Table 2 Parameters of Gruneisen state equation[12]

参数	数值
C/(cm · μs^{-1})	0.148
S_1	2.559
S_2	−1.98
Γ	0.228
γ	0.493
a	1.39
E	2.895 × 10^{-6}
V/(Pa · S)	10^{-3}

表 3 冰雹基本力学性能参数[18]

Table 3 Basic mechanical performance parameters of hailstone[18]

参数	数值
密度/(kg · m^{-3})	900
弹性模量/MPa	9 380
剪切模量/MPa	3 460
泊松比	0.33
压缩屈服强度/MPa	5.2
拉伸失效应力/MPa	0.517

表 4 基于应变率的冰雹屈服强度[14]

Table 4 Strain rate based yield strengths of hailstone[14]

应变率/s^{-1}	屈服因子	应变率/s^{-1}	屈服因子
0	1	500	3.62
0.1	1.01	10^3	3.84
0.5	1.5	5×10^3	4.33
1	1.71	10^4	4.55
5	2.2	5×10^4	5.04
10	2.42	10^5	5.25
50	2.91	5×10^5	5.75
100	3.13	10^6	5.96

表 5 界面单元材料参数[19-20]

Table 5 Material parameters of interface element[19-20]

参数	数值
密度/(kg · m^{-3})	1 440
K/(N · mm^{-3})	1 × 10^6
N/MPa	30
$S = T$/MPa	75
G_{IC}/(N · mm^{-1})	0.3
$G_{IIC} = G_{IIIC}$/(N · mm^{-1})	0.6

表 6 单向复合材料的材料参数[6,15,21]

Table 6 Material parameters of unidirectional composite[6,15,21]

参数	数值	参数	数值
E_{f1}/GPa	230	G_2/GPa	0.041
$E_{f2} = E_{f3}$/GPa	15	θ_{g2}/ms	12 000
$G_{f12} = G_{f13}$/GPa	2.35	ρ/(kg · m^{-3})	1 570
G_{f23}/GPa	24	V_f	0.6
E_m/GPa	2.31	$\mu_{12} = \mu_{13}$	0.25
E_1/GPa	0.971	μ_{23}	0.38
θ_{e1}/ms	0.041	s	0.1
E_2/GPa	0.104	X_T/MPa	2 100
θ_{e2}/ms	121 000	X_C/MPa	1 050
G_m/GPa	0.857	Y_T/MPa	71
G_1/GPa	0.401	Y_C/MPa	132
θ_{g1}/ms	0.077	S/MPa	75

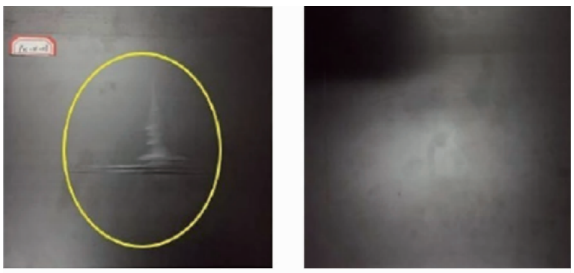
2.2 有限元模型验证

为了验证所建有限元模型的准确性和可靠性,将冰雹高速冲击下 2 种不同铺层方式层板的数值模拟结果与文献[6]中编号为 P010101 和 P020101 的实验结果进行对比。

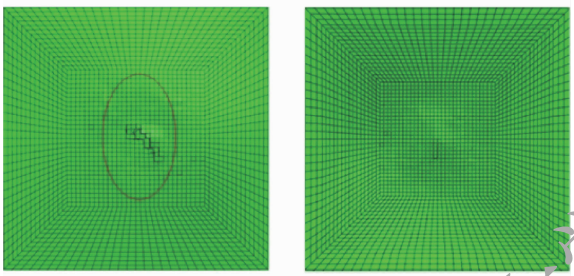
图 2 为铺层方式[0°/90°]$_{4S}$的层板在冰雹以 206 m/s 速度冲击下宏观损伤的实验和模拟对比情况。从图 2(a)、(c)可以发现,层板正面出现少量纤维断裂和基体开裂现象;从图 2(b)、(d)可以发现,层板背面出现鼓包现象。实验与模拟的宏观损伤形貌比较吻合,初步验证了所建有限元模型的有效性。

图 3 为铺层方式[0°/90°]$_{4S}$的层板在冰雹以 206 m/s 速度冲击后层间分层的 C 扫描和数值模拟结果对比。选中心 200 mm × 200 mm 区域进行查看,如图 3(a)所示,中间为发生分层区域;如图 3(b)所示,其为每一层的分层区域投影轮廓。可以发现,分层主要从撞击中心处沿 0°纤维方向展开,故层板在 206 m/s 冰雹冲击下损伤主要是层间分层这种肉眼难以观察到的内部损伤。

北京航空航天大学学报



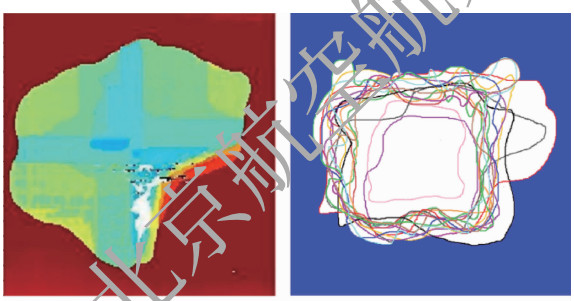
(a) 实验 (层板正面)　(b) 实验 (层板背面)



(c) 数值模拟 (层板正面)　(d) 数值模拟 (层板背面)

图 2　$[0°/90°]_{4S}$层板实验[6]与模拟宏观损伤对比

Fig. 2　Comparison of macroscopic damage of $[0°/90°]_{4S}$ laminates between experiment[6] and simulation

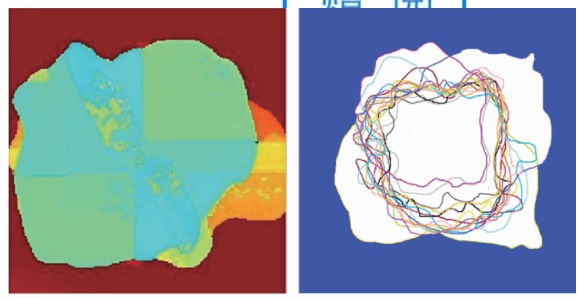


(a) 超声C扫描　(b) 数值模拟

图 3　$[0°/90°]_{4S}$层板分层情况实验[6]与模拟对比

Fig. 3　Comparison of delamination of $[0°/90°]_{4S}$ laminates between experiment[6] and simulation

图 4 为铺层方式$[0°/45°/90°/-45°]_{2S}$的准各向异性层板在冰雹 206 m/s 速度冲击后层间分层的 C 扫描和数值模拟结果对比。可以看出，其分层现象较正交铺层更为严重，主要从冲击中心



(a) 超声C扫描　(b) 数值模拟

图 4　$[0°/45°/90°/-45°]_{2S}$层板分层情况实验[6]与模拟对比

Fig. 4　Comparison of delamination of $[0°/45°/90°/-45°]_{2S}$ laminates between experiment[6] and simulation

沿着 45°、−45°方向扩展。整体的分层形状为梅花形，与实验较为吻合，进一步验证了该有限元模型的有效性。

3　数值结果分析与讨论

3.1　冰雹冲击过程分析

冰雹高速冲击复合材料层板是一个瞬时的非线性动态过程，图 5 为初速度 206 m/s 冰雹冲击$[0°/90°]_{4S}$复合材料层板的动态过程。结合层板中心处位移历程图可以看出，冰雹高速冲击层板过程中层板变形主要分为 3 个阶段：①冲击初期，层板的快速弯曲变形阶段(0 ~ 0.2 ms)；②达到最大位移后，层板的快速回弹阶段(0.2 ~ 0.4 ms)；③冰雹大面积沿层板扩散的回弹阶段(0.4 ~ 0.6 ms)。

由图 5 可以看出，该有限元模型很好地描述了冰雹的破碎流动特征，冲击过程中，应力波主要沿纤维铺层方向和层板厚度方向扩散而形成明显的圆形鼓包。冲击初期，层板中心处快速弯曲变形，t = 0.2 ms 时，层板背面中心处产生最大拉应

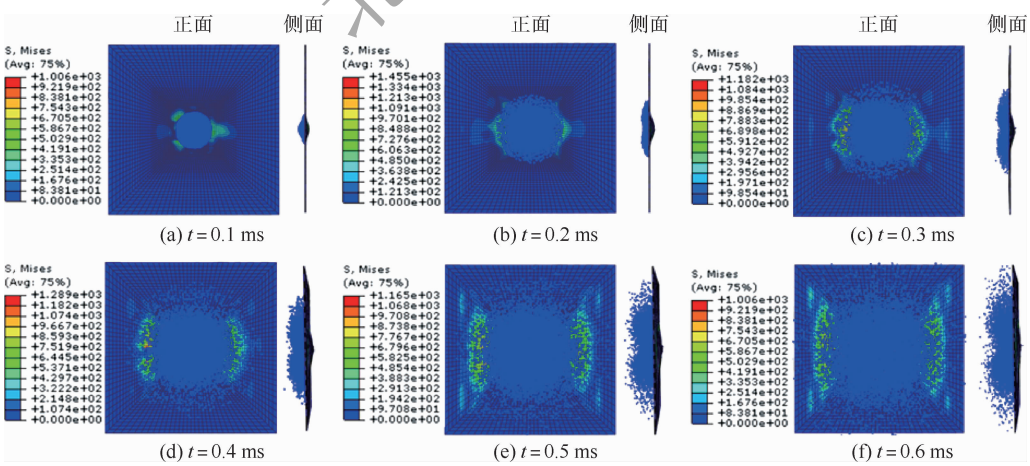


图 5　206 m/s 冰雹冲击速度下$[0°/90°]_{4S}$复合材料层板动态过程

Fig. 5　Dynamic process of $[0°/90°]_{4S}$ composite laminates under hailstone impact at velocity of 206 m/s

北航学报 赠阅

力而出现局部纤维拉伸断裂现象，并伴随中心区域基体开裂，层间分层面积快速增长；然后，冰雹完全失效破碎，以粒子形式沿层板四周继续冲击，冲击力急剧减小，层板快速回弹变形，层间分层面积逐渐增加到最大，层内出现了不连续的基体拉伸损伤；冲击后期，冰雹继续以粒子形式沿层板大范围扩散，该阶段层板缓慢回弹变形，层间分层面积基本不再变化，层内不连续的基体拉伸损伤继续小幅扩展。冲击结束后，复合材料层板呈轻微凸起状态，出现不可恢复的凹坑。

3.2 复合材料层板损伤分析

3.2.1 层间损伤

冰雹等类柔性体冲击复合材料层板时，与其他损伤模式相比，层间损伤最为严重，这是考虑冲击后层板剩余力学性能的一个重要因素。本节选取正交铺层和准各向同性层板为研究对象，在冰雹冲击速度为 206 m/s 时进行层间分层损伤分析。

图 6 为 2 种不同铺层方式层板分层面积随冲击时间的变化。数值模拟中，当界面的损伤变量达到 0.97 时，认为界面单元完全损坏即发生层间分层。可以发现，层间分层大面积扩散主要集中于冲击的起始阶段(0 ~ 0.27 ms)，在 0 ~ 0.2 ms 之间，分层扩展速率较快，且 2 种铺层方式层板分层总面积随时间变化趋势较为一致。冲击结束后，正交铺层层板的分层总面积约为 36 506 mm^2，准各向异性层板分层总面积约为 38 008 mm^2，准各向异性层板分层总面积稍大于正交铺层层板。

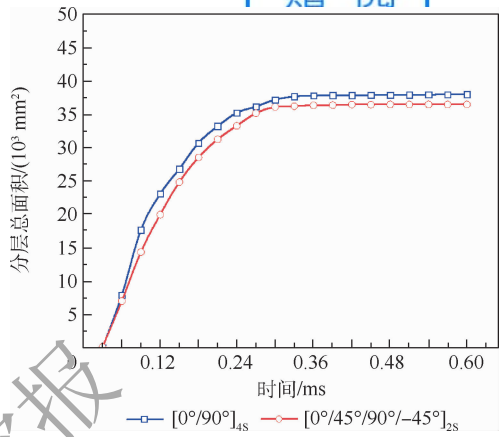


图 6 两种典型层板分层面积随时间变化历程

Fig. 6 Evolution of delamination area of two typical laminates with time

3.2.2 层内损伤

本节选取正交铺层层板为研究对象，在冰雹冲击速度为 206 m/s 时进行层内损伤分析。图 7 和图 8 为冲击后复合材料层板中 8 个单层基体拉伸开裂和纤维拉伸断裂的损伤分布情况，其中，白色部分为纤维和基体均完全失效，相关单元已从模型中删除。可以发现，在冰雹高速冲击过程中，基体开裂失效面积较大，基体开裂区域主要沿着层板中心以矩形向外扩展；而纤维拉伸断裂失效分布于层板中心处，面积相对较小。纤维主要沿 0°和 90°方向发生断裂，部分层纤维沿 45°和 −45°方向发生断裂。随着冲击的进行，纤维损伤面积不再扩展，但由于冰雹冲击过程中的流动特征，在冲击的后段每层的边缘处继续产生了一些基体拉伸开裂现象。

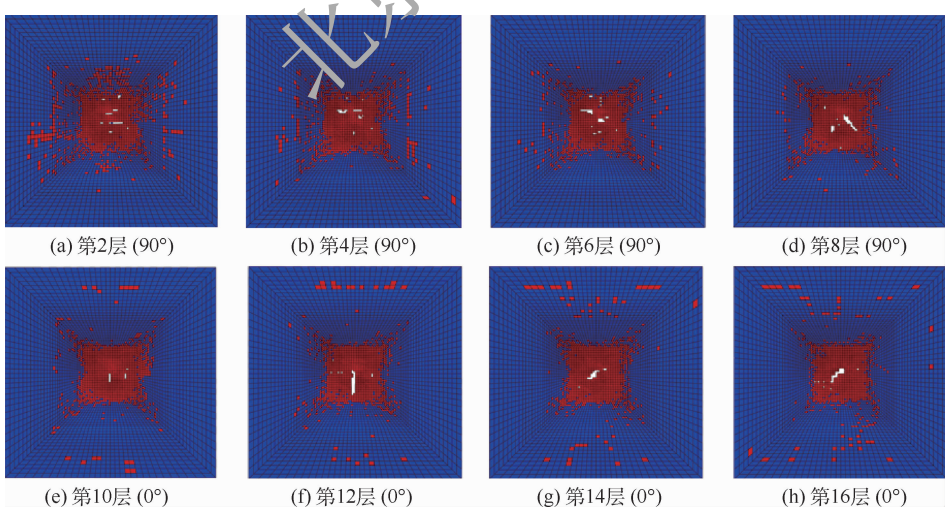


图 7 206 m/s 冰雹冲击速度下层板典型层基体拉伸损伤分布

Fig. 7 Distribution of matrix tensile damage in typical layers of laminates under hailstone impact at velocity of 206 m/s

北京航空航天大学学报

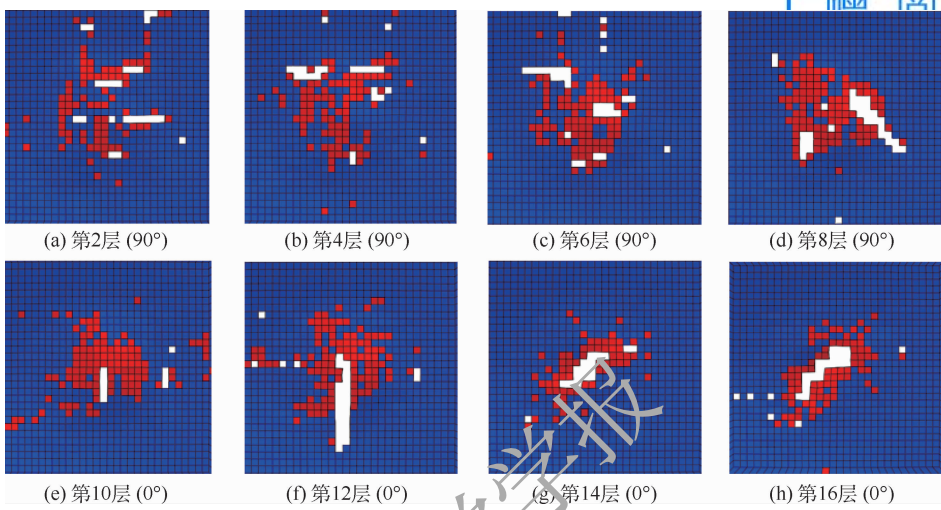


图 8　206 m/s 冰雹冲击速度下层板典型层纤维断裂损伤分布

Fig. 8　Distribution of fiber breaking damage in typical layers of laminates under hailstone impact at velocity of 206 m/s

3.3　冲击速度、冲击角度对复合材料损伤的影响

为了详细研究不同冰雹冲击载荷工况下碳纤维复合材料层板的力学响应和损伤特性,本节选取 25.4 mm 直径冰雹,以不同冲击速度、不同冲击角度冲击$[0°/90°]_{4S}$正交铺层层板,开展了多组数值模拟工作。

图 9 为不同冲击速度、冲击角度下层板所受到的最大冲击力变化曲线。可以发现,最大冲击力在冲击角度为 45°、60°时随着冲击速度的增大呈线性增大趋势。在同一冲击速度下,最大冲击力随冲击角度的增大而减小,很大程度上是由于在带有偏角的冲击条件下,层板所受的法向压应力小于垂直冲击条件下,同时受到一部分切应力作用。冲击角度为 0°(垂直冲击层板)时,在 160 m/s和 200 m/s 的冲击速度下层板所受到的最大冲击力比较接近,此时层板也承载着最大的法向压应力及弯曲变形,一般认为层板发生纤维断裂现象后力学性能大幅降低,故可以将200 m/s 冲击速度时对应的能量作为此层板可承受的最大冲击能量阈值(154.44 J)。同时,根据数值模拟得出的冲击力时间曲线可以发现,最大冲击力出现在冲击的初始阶段,在冲击角度为 45°时,最大冲击力出现的时间最早;在同一冲击速度下,0°冲击角度下最大冲击力出现的最早。

图 10 为不同冲击速度、冲击角度下层板分层总面积的变化曲线。可以发现,随着冲击角度的增加,分层面积逐渐减小,即冰雹垂直冲击层板时分层总面积最大,法向冲击力是造成层间分层现象的主要因素。垂直冲击下,分层面积随着冲击速度的增大而近似呈现出线性增加趋势。其中冲击角度为 0°和 30°时,分层面积的差值相对较小,冲击角度从 45°增加到 60°时,分层总面积减小幅度较大。此外,在较低的冲击速度下(低于 120 m/s),冲击角度对分层面积的影响较小,分层

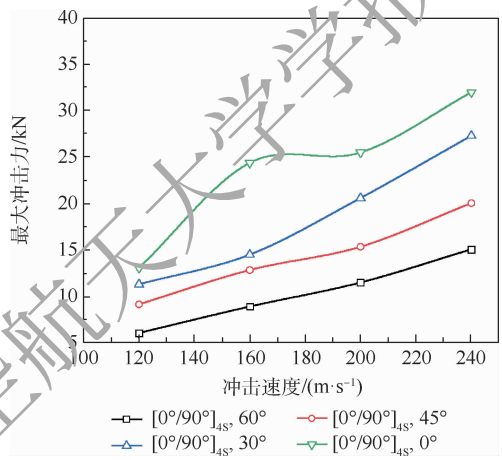


图 9　冲击速度和冲击角度对最大冲击力的影响

Fig. 9　Influence of impact velocity and impact angle on maximum impact force

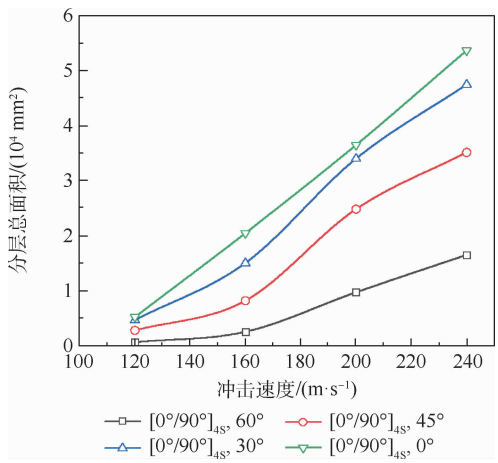


图 10　冲击速度和冲击角度对分层面积的影响

Fig. 10　Influence of impact velocity and impact angle on delamination area

北京航空航天大学学报

面积低于 5 000 mm^2。

为了有效分析冰雹冲击速度对层板分层面积的影响,本文进行了冰雹垂直冲击层板情况下,不同冲击速度下的多组数值模拟工作。结果发现,层板分层面积在较高范围内(大于 120 m/s)随冲击速度的增大而近似线性增大;当冲击速度为 80 m/s时,层板开始出现层间分层损伤。由此可以推测,此种复合材料层板在直径 25.4 mm 冰雹的冲击下,分层损伤能量阈值约为 24.71 J。

4 结 论

1)本文考虑了冰雹破裂后的流动特性和冰雹高速冲击下复合材料层板的应变率效应,建立了复合材料层板冰雹高速冲击有限元分析模型。模拟了冰雹高速冲击下复合材料层板的损伤过程,分析了材料的损伤特性和失效机理,通过与已有实验结果对比分析,验证了模型的有效性。

2)层间分层是冰雹高速冲击下层板的主要损伤形式。在 25.4 mm 冰雹 200 m/s 垂直冲击速度下,层板出现少量的纤维拉伸断裂,大面积基体裂纹及层间分层现象,分层总面积达到 54 457 mm^2。该冲击速度对应能量可以作为层板此种工况下可承受的最大冲击能量。冰雹 80 m/s 冲击速度时,层板开始出现分层损伤,故此复合材料层板的分层损伤能量阈值约为 24.71 J。

3)层板分层总面积随着冲击角度的增加而减少,在 200 m/s 速度垂直冲击下,分层面积随冲击速度增大而线性增大,而在一定冲击角度下(45°、60°),最大冲击力与冲击速度呈线性关系。冲击速度较低时,冲击角度对分层总面积的影响较小。相同冲击条件下,准各向异性铺层层板分层面积大于正交铺层层板,分层损伤主要发生在冲击的初始阶段,冰雹破碎后,层板卸载回弹阶段分层面积基本不变。

4)所建立的有限元模型能够有效获取高速冲击过程中冰雹的破裂、流动特性,模拟复合材料层板的应变率效应和层内、层间损伤过程,且所采用的损伤模型和损伤准则具有良好的通用性,研究结果可以为其他复合材料结构冰雹冲击问题数值分析提供参考。

参考文献(References)

[1] 张超.三维多向编织复合材料宏细观力学性能及高速冲击损伤研究[D].南京:南京航空航天大学,2013.
ZHANG C. Research on macro-meso-mechanical properties and high velocity impact damage of 3D multi-directional braided composites[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2013 (in Chinese).

[2] 朱倩.纤维金属层板抗高速冲击性能及损伤机理研究[D].镇江:江苏大学,2020.
ZHU Q. Study on impact resistance and damage mechanism of fiber metal laminates under high velocity impact[D]. Zhenjiang: Jiangsu University, 2020 (in Chinese).

[3] KIM H, KEDWARD K T. Modeling hail ice impacts and predicting impact damage initiation in composite structures[J]. AIAA Journal, 2000, 38(7): 1278-1288.

[4] RHYMER J D. Force criterion prediction of damage for carbon/epoxy composite panels impacted by high velocity ice[D]. San Diego: University of California, 2012.

[5] TANG E L, WANG J R, HAN Y F, et al. Microscopic damage modes and physical mechanisms of CFRP laminates impacted by ice projectile at high velocity[J]. Journal of Materials Research and Technology, 2019, 8(6): 5671-5686.

[6] 廖光兰.冰高速冲击作用下复合材料层合板的动态响应及损伤研究[D].南京:南京航空航天大学,2018.
LIAO G L. The dynamic response and damage research of laminates according to the high velocity ice impact[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2018 (in Chinese).

[7] COLES L A, ROY A, SILBERSCHMIDT V V. Ice vs. steel: Ballistic impact of woven carbon/epoxy composites. Part II: Numerical modelling[J]. Engineering Fracture Mechanics, 2020, 225: 106297.

[8] PERNAS-SÁNCHEZ J, ARTERO-GUERRERO J A, LÓPEZ-PUENTE J, et al. Numerical methodology to analyze the ice impact threat: Application to composite structures[J]. Materials & Design, 2018, 141: 350-360.

[9] 周述冰.层合复合材料冰雹和硬物冲击损伤研究[D].南京:南京航空航天大学,2019.
ZHOU T L. Study on damage of composite laminate experienced impactions of hail and rigid impactor[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019 (in Chinese).

[10] 王计真.复合材料层合板抗冰雹冲击性能研究[J].兵工学报,2017,38(S1):89-95.
WANG J Z. Research on anti-hailstone impact behavior of laminated composite panel[J]. Acta Armamentarii, 2017, 38(S1): 89-95 (in Chinese).

[11] 张晓晴,丁铁,龙舒畅,等.复合材料加筋壁板的抗冰雹冲击动力响应及损伤预测[J].华南理工大学学报(自然科学版),2017,45(5):120-128.
ZHANG X Q, DING T, LONG S C, et al. Dynamic response and damage prediction of composite stiffened panel under hail impact[J]. Journal of South China University of Technology (Natural Science Edition), 2017, 45(5): 120-128 (in Chinese).

[12] DOLATI S H, REZAEEPAZHAND J, SHARIATI M. Numerical simulation of hail impact response of hybrid corrugated core sandwich panels[J]. Journal of Reinforced Plastics and Composites, 2019, 38(14): 643-657.

[13] CARNEY K S, BENSON D J, DUBOIS P, et al. A phenomenological high strain rate model with failure for ice[J]. International Journal of Solids and Structures, 2006, 43(25-26): 7820-

北航学报 网阅

7839.

[14] TIPPMANN J D. Development of a strain rate sensitive ice material model for hail ice impact simulation[D]. San Diego: University of California,2011.

[15] KARIM M R,FATT M S H. Rate-dependent constitutive equations for carbon fiber-reinforced epoxy[J]. Polymer Composites,2006,27(5):513-528.

[16] HASHIN Z. Failure criteria for unidirectional fiber composites [J]. Journal of Applied Mechanics,1980,47(2):329-334.

[17] CAMANHO P P,DAVILA C G,DE MOURA M F. Numerical simulation of mixed-mode progressive delamination in composite materials[J]. Journal of Composite Materials,2003,37(16): 1415-1438.

[18] TIPPMANN J D,KIM H,RHYMER J D. Experimentally validated strain rate dependent material model for spherical ice impact simulation[J]. International Journal of Impact Engineering,2013,57:43-54.

[19] ZHANG C,CURIEL-SOSA J L,BUI T Q. A novel interface constitutive model for prediction of stiffness and strength in 3D braided composites[J]. Composite Structures,2017,163:32-43.

[20] 莫袁鸣,赵振华,罗刚,等.复合材料层合板冰雹冲击损伤研究[J].重庆理工大学学报(自然科学),2020,34(3):112-121.
MO Y M,ZHAO Z H,LUO G,et al. Investigation on damage of composite laminates subject to hail impact[J]. Journal of Chongqing University of Technology (Natural Science),2020,34(3):112-121(in Chinese).

[21] WANG S X,WU L Z,MA L. Low-velocity impact and residual tensile strength analysis to carbon fiber composite laminates [J]. Materials & Design,2010,31(1):118-125.

Damage prediction and failure mechanism of composite laminates under high-velocity hailstone impact

ZHANG Chao*, FANG Xin, LIU Jianchun

(School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: Aiming at the potential risk of hailstone impact on the safety of composite structures, a continuum damage mechanics based nonlinear finite element model was developed to study the mechanical behavior of carbon fiber composite laminates under high-velocity hailstone impact. The Lagrangian method and smoothed particle hydrodynamics (SPH) method were used together to model the impact of hailstone, and the equation of state of water was introduced to describe the flow characteristics of the hailstone after breaking. A rate-dependent constitutive model of unidirectional composite, as well as 3D Hashin failure criteria and material stiffness reduction rule, was applied to predict the in-plane damage in composite layers. Interface elements governed by bilinear cohesive model were employed to simulate the inter-laminar delamination phenomena induced by impact. A user material subroutine VUMAT was coded and implemented to obtain the numerical solution based on ABAQUS/Explicit solver. The transient process of composite laminates under hailstone impact was reproduced and the damage characteristics and failure mechanism were analyzed in detail. The effects of impact velocity and impact angle of hailstone on the impact properties of composite laminates are discussed, which provides proper reference for numerical investigation of hailstone impact problems in composite structures.

Keywords: carbon fiber composites; hailstone impact; high-velocity impact; damage; numerical simulation

Received: 2020-11-15; **Accepted**: 2021-03-12; **Published online**: 2021-04-07 14:13

URL: kns.cnki.net/kcms/detail/11.2625.V.20210407.1353.001.html

Foundation items: Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20-3076); Natural Science Foundation of Jiangsu Province (BK20180855); Research Fund of State Key Laboratory of Mechanics and Control of Mechanical Structures (MCMS-E-0219Y01)

* **Corresponding author**. E-mail: zhangchao@ujs.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0638

账户型区块链可扩展存储模型

全下文载

张栩豪，牛保宁*，弓婷
（太原理工大学 信息与计算机学院，太原 030600）

摘 要：区块链数据持续线性增长，最终导致单个节点无法提供足够的存储资源，产生存储可扩展性问题。轻节点模型虽然极大地降低了对存储资源的需求，但是也导致全节点减少、去中心化程度降低，威胁区块链网络的安全。账户型区块链尚未有成熟的可扩展存储方案提出。考虑到状态数据的使用频率远远高于区块数据，提出一种基于账户型区块链的可扩展存储模型(SSMAB)。SSMAB 以完全冗余的方式保存状态数据，保证其交易验证功能；以分片方式存储区块数据，降低冗余度；采用经济激励机制，在保证数据可用性的前提下，减少存储的消耗。实验结果表明：SSMAB 模型能够保证数据的可靠性与可用性，将存储数据量降低到全节点模型的 13%。

关 键 词：区块链；存储模型；以太坊；分片；经济激励
中图分类号：TP315
文献标志码：A **文章编号**：1001-5965(2022)04-0708-08

随着比特币、以太坊等加密货币的流行，区块链作为其底层技术受到了越来越多的关注[1-2]。区块链具有去中心化、防篡改和可追溯的特性，在版权、医疗、金融等多个领域[3]发挥重要作用。

区块链采用链表结构，区块线性增加，受限于区块链存储可扩展性，节点的存储资源消耗也在线性增长。以以太坊为例，根据 Etherscan 数据[4]，截至 2020 年 7 月 4 日，链上认证地址 99 914 268 个，存在将近 960 万个区块，包含 7.4 亿笔交易，全节点(保存区块链全部信息)的数据 280.65 GB，相比去年同期分别增加 53.88% 和 47.42%。

随着数据量的增加，部分全节点因无法存储区块链全部数据，选择成为轻节点(保存区块链中区块头数据)或退出网络。全节点是区块链去中心化特性和网络安全的保证，在全节点减少后，区块链网络中可对区块进行验证的节点数量也随之减少，导致网络安全性降低，容易遭到攻击。在 2017 年 2 月，以太坊全网拥有近 20 000 个全节点，但是随着以太坊数据量的增加，全节点的数量也逐步下降至 7 000 个左右[4]。此外，随着集中化挖矿机制的出现，一部分全节点因自身算力不足，选择加入矿池，为矿池贡献算力。几大矿池占据了以太坊网络中大部分算力，逐渐成为以太坊事实上的中心。以太坊中心化趋势日益显现。

区块链存在 2 种数据模型，分别为比特币为代表的交易型区块链和以太坊为代表的账户型区块链。交易型区块链中，每笔交易的输入为上一笔交易的输出，所有的交易通过输入和输出链接在一起，使所有交易可追溯；账户型区块链中，记录账户余额和状态。针对区块链中因存储问题导致可扩展性降低的问题，采取数据分片存储技术来缓解该问题。在交易型区块链中，通用的方法是节点缓存所有未花费交易输出(unspent transaction outputs，UTXO)来快速验证交易；采用区块

收稿日期：2020-11-16；**录用日期**：2021-01-17；**网络出版时间**：2021-03-16 13:56
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210316.1345.002.html
基金项目：国家自然科学基金（62072326）；山西省重点研发计划（201903D421007）
*通信作者. E-mail：niubaoning@tyut.edu.cn

引用格式：张栩豪，牛保宁，弓婷. 账户型区块链可扩展存储模型[J]. 北京航空航天大学学报，2022，48(4)：708-715.
ZHANG X H，NIU B N，GONG T. Account-based blockchain scalable storage model[J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：708-715(in Chinese).

北京航空航天大学学报

北航学报 网阅

分片存储技术,对区块数据进行分片存储设计,减少单节点所存储的数据量。账户型区块链与交易型区块链采用 UTXO 验证交易不同,可验证交易的状态数据需要通过执行以往所有交易来生成。因此,账户型区块链在分片存储下实现交易验证功能,不仅需要实现区块分片存储,也需要对全节点中存储的状态数据进行保存。

在账户型区块链当中,每个全节点都存储区块链的完整数据,如区块数据和状态数据等。状态数据包含所有账户的余额及交易次数等。因此,在全节点与轻节点之外,提出一种新节点模式,通过存储状态数据,使节点可以参与交易验证,成为区块链网络的维护者,维持区块链的去中心化特性。同时采取分片存储技术,对区块链中区块数据进行分片处理,每个节点存储一部分数据,降低存储资源消耗,提高区块链系统的可扩展性。此外,在现有的区块链中,全节点采取完全冗余存储数据,为轻节点义务提供服务,资源消耗无法得到补偿,不利于全节点长时间保留在网络中。因此,本文采取经济激励机制,节点可通过自己存储的区块数据来获得收益,激励节点存储区块链数据,保证网络的安全性。

本文提出 SSMAB。保存所有账户的状态数据,使 SSMAB 节点拥有验证交易的能力,不依靠全节点来保证区块链网络的安全性,维护区块链的去中心化水平,保证区块链的可扩展性及安全性。对区块数据进行分片处理。在保证数据安全性及可用性的情况下,减少存储空间的消耗。SSMAB 节点通过本地存储的区块头数据进行校验,保证获取数据的真实性。提出经济激励机制。SSMAB 节点通过交易市场获得存储区块的价格,并在其他节点请求区块时,通过区块数据获得收益,通过经济模型来调节区块的可用性。

1　相关工作

区块链存储模型有传统的全节点、轻节点模型及分片存储模型。全节点模型存储区块链网络自创始以来所有的区块及根据交易记录构建出的唯一的状态树。该状态树存储所有账户的状态数据,可快速验证每个账户的余额。轻节点模型通过分布式哈希表来追踪存储最新区块的节点,为了保持轻量性,只存储所有区块的区块头部分,无法独立对交易进行验证。当轻节点需要进行相关操作,如检查账户余额、验证转账确认及获取智能合约运行结果时,才会向全节点请求相应数据[2],通过默克尔树的哈希认证路径,确认相关消息。ZILLIQA[5]同传统模型相比,将节点分为普通节点和种子节点,实现了网络分片、交易分片及智能合约分片,普通节点存储用户状态数据,种子节点存储完整账本。由于未实现存储分片,该方案中节点面临较大的存储压力。本文模型利用 ZILLIQA 的状态存储思想,通过存储用户状态数据,实现账户验证功能。

分片存储对区块链数据进行划分,每个节点只需要存储一部分数据,不需要存储区块链完整数据。如果某分片遭到攻击后脱离区块链网络,依赖于该分片数据的交易将无法得到验证,从而导致整个区块链系统性能的下降甚至停滞。ElasticChain 模型[6]及区块链增强型轻量级节点模型 ESPV[7]对于交易型区块链的区块数据按照安全性及区块利用程度为依据对区块进行划分,降低节点及区块链网络的存储资源消耗。但 ElasticChain 模型未能对区块划分提出合适的理论,ESPV 模型无法对交易进行打包,对早期旧区块支持较差。CUB[8]及 RapidChain[9]将节点划分为不同分区,每个分区内部节点分片存储区块数据,同时,对于利用共识算法来决定每个分区生成区块的顺序,保证全网交易顺序的一致性,但是部署在公链上存在查询频繁、网络开销较大等缺点。Rchain[10]及 Dai 等[11]提出的隐私数据分片存储方案将一件事务/数据切分为多个子集进行处理或存储,当事务处理完毕,再将子集结合,形成完整的事务或子集,该方案虽然提高了事务的处理速度,保证了隐私数据的安全性,但是对于节点的协同性要求高,增加了事务的复杂度。以太坊 2.0 版本[12]中引入链上状态分区(on-chain state partition)概念,将以太坊网络分为主链(main chain)和分片链(shared chain),主链同现有链基本一致,分片链则在各自独立空间进行交易处理,通过内部验证器(shared validator)进行交易验证。主链只存储分片链提交的校验块头信息,减少节点存储压力。以上区块链存储方案的对比分析如表 1 所示。本文模型通过使用分片存储技术,通过理论推导得到区块链中最少分片的数量,减少节点存储资源的消耗,保证区块链系统的存储可扩展性。

经济激励常见于对等(peer to peer, P2P)网络,但在区块链中也存在经济激励制度。在 P2P 网络中,数据传输依赖于节点间的相互合作,而现实中存在大量“搭便车”的行为,即忽视了节点拥

北航学报 赠阅

表 1 区块链存储方案

Table 1 Blockchain storage scheme

分类	解决方案	适用区块链类型	方案不足
全节点 + 轻节点	比特币 SPV[1]、以太坊[2,12]	交易型/账户型区块链	轻节点不能独立对交易进行验证
全节点 + 增强型轻节点	ESPV[7]	交易型区块链	不能对区块打包
普通节点 + 种子节点	ZILLIQA[5]	账户型区块链	节点存储压力较大
弹性存储	ElasticChain[6]	交易型/账户型区块链	无法保证区块可用性
事务切分	Rchain[10]、隐私数据分片[11]	交易型/账户型区块链	对节点协同性要求较高,增加事务复杂度
分片	CUB[8]、RapidChain[9]	账户型区块链	交易验证查询频繁,网络开销较大

有者的意愿和需求等私利性因素,不能充分利用各类闲置的资源。P2P 网络主要采用基于信誉和基于奖励 2 种机制来激励节点提供更多更好的服务。主流的 P2P 软件如 Bitcomet[13] 等基于信誉机制,通过对节点进行评级,为评级高的节点提供更好的服务。在区块链中,为了鼓励节点参与交易验证,对于验证交易并通过工作量证明将已验证交易打包为区块的节点,给予虚拟代币进行奖励。IPFS[14] 系统基于奖励机制,为提供存储空间及带宽的节点提供 Filecoin 虚拟代币,节点提供的存储空间及带宽越大,得到的 Filecoin 就越多。本文模型采用经济激励思想,通过交易市场及经济模型来调节区块的手续费,激励节点来存储区块数据。

2 SSMAB 节点模型

2.1 账户型区块链可扩展存储模型构建思路

在现有账户型区块链中,全节点需要实现交易验证、数据存储及数据共享 3 方面功能。在存储可扩展模型 SSMAB 当中,以实现这 3 方面功能为目标。账户型区块链中,全节点在存储区块的基础上,存在由执行所有交易形成的唯一状态树,状态树中存储所有账户的余额、交易次数等信息,通过该状态树,即可验证由相关账户发起的交易。因此,本文提出 SSMAB 节点保存状态树,通过查询状态树中交易双方账户余额数据即可验证相应的交易,具有交易验证功能。在现有账户型区块链中,用户对于区块中交易数据,查询的次数相对较少,因此,不必采取完全冗余存储的方式来存储。SSMAB 节点通过采用分片存储技术,根据网络中节点数量、区块总数据量等参数对区块进行划分保存。为了激励区块分片后的区块链网络中节点存储区块数据,保证数据可用性, SSMAB 节点模型引入经济激励机制,通过分片路由表访问其他节点,加快区块数据检索,并获取本地未存储的数据。对于新区块产生后的状态更新,SSMAB 节点可通过其余已更新节点进行同步,或通过获取区块数据来进行更新。

2.2 状态树构建共享策略

状态树是一颗倒排树,具有唯一构建方式,树根的哈希值具有唯一性。在现有账户型区块链全节点中,状态树的构建有 2 种方式:①获取区块链上所有的区块,节点根据交易顺序,从创世区块开始对每一笔交易进行重放执行直至最新区块,以此生成状态树,在该方式下,每次交易都会得到重新验证;②节点直接通过区块链网络同步其他节点现有最新的状态树,在该方式下,节点不会对之前的数据进行验证,只会对之后产生的区块及交易进行验证。

在分片存储下,节点不再存储全部区块,无法依靠节点自身数据来构建状态树。状态树的构建依赖于其他节点存储的其他部分区块或最新状态树,因此本文提出 SSMAB 节点构建状态树的 2 种方法:

1)通过路由表顺序获得所有区块分片,在获取顺序区块后依照交易顺序重放执行区块,生成状态树。

2)根据路由表,获取其他节点信息,向存储最新状态树的 SSMAB 节点发送请求进行同步,获得区块链网络最新状态树。

方法 1 同全节点的账户状态树构建方法相同,由于区块数据量庞大,获取区块受限于网络带宽等原因需要消耗较长时间,更推荐使用方法 2 来进行状态树的同步及更新操作。

区块可分为区块头和区块体 2 部分,区块头存储区块相关校验信息,区块体存储交易数据。为了防止状态树被修改并保证区块链全网状态相一致,SSMAB 节点通过本地保存的最新区块头中存储的根哈希值来确定状态树是否经过修改、是否为最新状态树。如果状态树未经修改且不是最新状态树,SSMAB 节点将获取该区块高度至最新的区块来验证交易并更新状态树,或同拥有最新状态树的节点进行同步。

2.3 区块可扩展存储策略

在账户型区块链中,当新区块产生时,全节点会对该区块当中的交易进行逐一验证,并将区块

北航学报 赠阅

当中交易所涉及到账户信息的更改操作,在本地状态树当中执行。而在节点对区块进行验证之后,区块一般只需应对较小的查询请求,只有当其他节点向该节点请求同步区块数据时,节点才需要从底层数据库中调取区块进行传输。部分区块链如以太坊等还部署有智能合约模块。智能合约毫无规律分布在区块当中,执行智能合约要先在合约所在区块调取合约字节码,因此,在分片方案中需要保证区块的可靠性与可用性。

综合区块数据的访问特性,SSMAB 节点对区块数据采取分片存储的方式,对所有区块数据进行划分,单个 SSMAB 节点可提供的存储容量要大于单个区块分片,所有 SSMAB 节点的存储能力必须超过区块链数据大小。

为方便理解推导过程,符号含义如表 2 所示。

表 2　符号含义

Table 2　Description of symbols

符号	含义
P_r	单节点可靠性
T_w	单节点故障时间
t	请求响应固定时间
P_{se}	相同副本节点全部故障的概率
N_a	区块链中所有节点的数量
N_m	持有相同副本的节点数量
s	序号,表示 N_m 个以内的节点
T_w^s	N_m 个节点中第 s 个节点的平均故障时长
P_a	P2P 系统可用性
D_s	单分片数据量
D_a	区块数据总量
C_i	分片变动价格
C_b	分片基准价格

在区块链底层使用的 P2P 系统中,单节点的可靠性为

$$P_r = \frac{1}{T_w}\int_0^t (1 - P[T_w < u])\,du \tag{1}$$

式中:T_w 为单节点故障时间;t 为其他节点向该节点发送请求的固定时间。

与该节点存储相同副本节点同时全部故障的概率为

$$P_{se} = C_{N_a}^{N_m}\prod_{s=1}^{N_m}\left(1 - \frac{1}{T_w^s}\int_0^t (1 - P[T_w^s < u])\,du\right) \tag{2}$$

式中:N_a 为系统总节点数量;N_m 为系统中持有相同副本的节点数量;T_w^s 为 N_m 个节点中第 s 个节点的平均故障时长。

结合式(1)和式(2),得到 P2P 系统的可用性为

$$P_a = 1 - C_{N_a}^{N_m}\prod_{s=1}^{N_m}\left(1 - \frac{1}{T_w^s}\int_0^t (1 - P[T_w^s < u])\,du\right) \tag{3}$$

可解得 N_m 的值为

$$N_m = \frac{\log(1 - P_a)}{\log(1 - P_r)} \tag{4}$$

式中:根据 Borel 定律[15],将概率低于 $1/10^{50}$ 的事件定义为不可能事件,因此,$P_a = 1/10^{50}$ 为理想取值。

在区块链中,区块的大小并不均衡,早期受限于规模及功能,区块较小,之后随着规模扩大及功能增多,区块的大小也在逐渐增大。为了节点获取区块、同步数据更便利,采用连续区块、固定大小的方式进行存储划分。为了保证可用性,在极端情况下,假设每个节点只存储一个分片数据。为了保证每个分片都可得,节点总数与单分片备份节点数的比值应大于分片的个数,得到

$$N_m < \frac{D_s N_a}{D_a} \tag{5}$$

式中:D_a 为所有区块数据总量;D_s 为单分片数据量;分片大小可通过以上 3 个参数进行分析确定。

在固定分片大小下,从创世区块开始,达到分片大小即开始下一个分片打包直至最新的区块,同时,将每个分片的起始结束高度进行统计记录。这些值记录在节点的附加文件当中,一旦产生便不会发生变化。每个节点依据这些值来进行分片的同步存储或删除操作。在新节点加入网络中后,先从其他节点获取到的附加文件中获得分片信息,根据自身可提供存储容量来确定存储分片数量。

2.4　经济激励机制

使用定价模型来激励节点存储区块数据,提供更好的服务。在分片存储下,SSMAB 节点不会存储全部的区块数据。而对于节点而言,占用了存储空间无法从中获得收益,势必会影响到节点的积极性,造成节点在线时间越长所付出的电力及带宽成本呈线性增长。因此,SSMAB 节点在获取区块时需要向提供节点支付一定的以太币,价格由定价机制决定。

定价机制针对存储分片的 SSMAB 节点进行统计。通过式(4)计算得到 N_m,存储单一分片的节点数应大于该值,因此将 N_m 设为定价模型可接受的最小值,此时的手续费最高,而存储节点增多后又会带来高冗余度,随着节点增多手续费为

北京航空航天大学学报

北航学报 瞰阅

下降趋势。定价公式如下：

$$C_i = \frac{N_m^2 C_b}{x^2} \qquad x \geqslant N_m \tag{6}$$

式中：C_i 为本次分片定价；C_b 为分片基准价格，设定为最新 5 个分片价格的平均值；x 为统计所得存储某分片的节点数。

考虑到网络同步及需求波动性，将每次价格改变的间隔设置为 2 h。为了确保区块链网络中存储节点对所有节点可见，在定价机制下设计分片数据库，数据库中包含分片编号、本次价格、本次定价时间（采取世界标准时间）及存储节点地址，规定每个 SSMAB 节点存储该数据库。当节点存储或删除分片时，在分片数据库写入或删除对应分片下的该节点地址，保证分片数据库中数据真实性。

当节点需要某分片时，先访问分片数据库，得到分片当前价格及存储节点列表，随机访问存储节点列表中节点，建立连接并支付以太币后，即可得到此分片。

2.5 SSMAB 节点模型框架

SSMAB 节点模型框架如图 1 所示，分为底层及核心层。底层包含 P2P 网络、分片数据库、可扩展存储机制及数据库；核心层包含区块链、智能合约及共识算法 3 部分。

底层中 P2P 网络是区块链节点之间联通的方式；可扩展存储机制保证区块链中数据可用性，通过分片数据库获取存储相应区块的节点信息；数据库存储区块数据。核心层中，区块链模块产生和验证交易，打包区块，提供经济激励；根据交易构建状态树，为上层应用提供消息接口；智能合约使得区块链系统具有图灵完备的特性；共识算法模块提供分布式一致性协议算法，工作量证明（proof of work，PoW）及权益证明（proof of stake，PoS）是常用的共识算法。

相比于现有账户型区块链存储框架，SSMAB 模型在底层增加分片数据库及可扩展存储机制，在核心层增加经济激励机制。状态树及区块的同步需要通过分片数据库及路由表获取其他节点信息来完成相关操作。通过状态树，SSMAB 节点可以在不保存区块的情况下验证余额转移型交易，而 SSMAB 节点通过获取的区块，可以对智能合约进行执行，完成对智能合约交易的验证。在区块分片存储当中，SSMAB 节点减少了存储空间的消耗，但是默认存储的区块头数据依然可以保证获取区块的真实性及系统的安全。

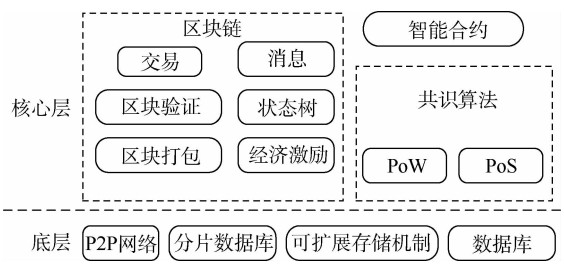


图 1 SSMAB 节点框架

Fig. 1 SSMAB node framework

3 以太坊下的 SSMAB 模型实现

采用以太坊来实现 SSMAB 节点模型，通过以太坊来验证 SSMAB 模型的各项性能指标。使用 2 台 Intel(R) Xeon(R) E3-1225@3.20 GHz CPU 和 64 GB 内存的服务器，虚拟出 60 台操作系统为 CentOS 的虚拟机，带宽随机虚拟为 4 ~ 100 Mbit/s 之间，利用以太坊现有的区块数据进行模拟实验。

3.1 以太坊状态树构建共享

在以太坊中，全节点存在一颗由全网以太坊账户通过递归长度前缀（recursive length prefix，RLP）编码构成的状态树（state tire），这棵树储存了外部账户及合约账户的余额（balance）及交易次数（nonce）等信息。截至 2020 年 5 月 26 日，以太坊状态树的占用空间大小为 2.8 GB。

对于 SSMAB 节点状态树构建，将分片数据随机分配在所有 SSMAB 节点，确保区块在网络中存在至少一个备份。选取 5 个节点执行 2.2 节中方法 1；选取 15 个节点在执行方法 1 的节点构建完状态树后执行方法 2；剩余节点则在以上 20 节点构建完后依次加入网络，并以随机方法获取状态树。

在状态树构建中，以太坊全节点（Ethereum full node，EFN）采用公共网络进行同步数据，耗时 6 d 15 h 完成状态树构建；使用方法 1 和方法 2 的 SSMAB 节点，分别平均耗时 19 h 43 min、1 h 56 min，完成了状态树构建。方法 2 不需要下载区块，也不需要计算并构建状态树，因此消耗时间少，构建时间呈线性。对于普通用户而言，方法 2 是更好的选择。

3.2 以太坊区块划分存储

使用开源项目 Ethstats[16] 对以太坊网络中全节点进行监测统计。截至 2020 年 8 月 14 日，以太坊中全节点最多同时在线数量为 10 295 个，最少为 4 466 个，平均同时在线数量为 7 418 个。同

时在线节点的数量不会骤然增多变少，具有一定的稳定性。随机选取 500 个全节点进行追踪统计，把每天在线 12 h 以上节点视为可靠节点。节点可靠性定义为可靠节点与选取节点数量之间的比值，结果如图 2 所示。在区块链网络中，节点可靠性大于 0.7，表示节点拥有存续性，即节点倾向于长期存在于网络当中。

根据以太坊中区块生成的速度及采取分片存储后节点获取区块的速度，设置超时时间为 5 s，在 5 s 内未收到应答可视为超时，同时，将以太坊网络节点总数设为 6 000 个，采用图 3 的数据，设置节点的可靠性 P_r 为 0.5，通过式(4)等解得副本节点 N_m 为 47。

使用 PeerSim[17] 软件，编写实现以太坊使用的 P2P 网络 Kademlia 协议，按照以上参数，虚拟 6 000个节点，每个节点可靠性设为 0.5。随机标记 0 ~ 80 个节点，这部分节点拥有特殊格式数据包，为了模拟区块分片获取情况，每个节点以随机概率需要该数据包。在每轮运行中，根据节点需求及得到的概率，分析数据包的可用性。实验结果如图 3 所示。

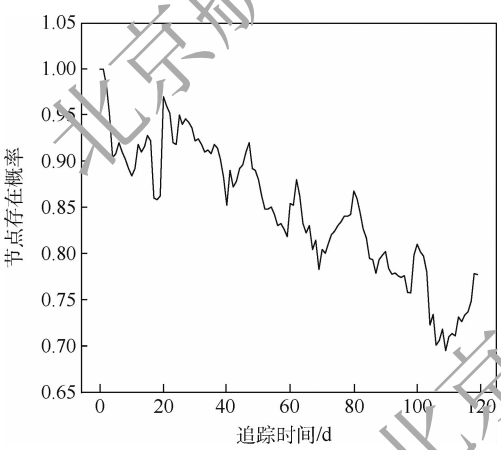


图 2　EFN 节点可靠性

Fig. 2　EFN node reliability

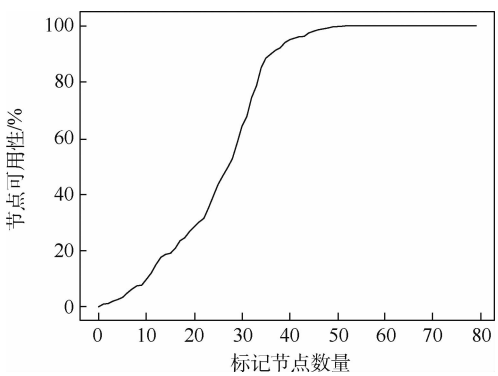


图 3　分片可用性

Fig. 3　Shard availability

实验结果显示在 6 000 个节点的模拟网络中，52 个副本节点下分片可用性可达到 100%。由于网络堵塞等因素，副本节点在模拟网络中高于理想值，在实际网络中受限于网络堵塞、路由转发及带宽等多种因素影响，副本节点数应更多。

对于以太坊区块数据，根据式(5)对区块数据进行划分，生成 73 个分片，每个分片大小为 4 GB。对区块获取进行测试，同时 SSMAB 模型针对交易验证及确认、区块平均生成时间及交易吞吐量同 EFN 进行对比测试，测试结果如图 4 和图 5所示。

本文对随机产生的 10 000 笔交易进行验证，由图 4 可以看出，SSMAB 节点验证交易的时间相比于 EFN 没有太大差异。对打包的 500 个区块进行统计，由图 5 看出，生成区块平均时间也相差不大。

本文统计了自 2020 年 5 月 26 日至 2020 年 7 月16 日 2 个模型的每秒交易次数。由于 SSBAB 节点在执行智能合约类型交易时需要获取相关区块，由图 6 可得 SSMAB 模型，相比于 EFN，每秒交

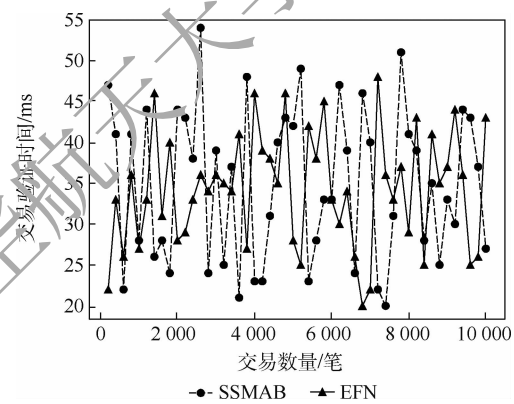


图 4　交易验证时间

Fig. 4　Transaction verification time

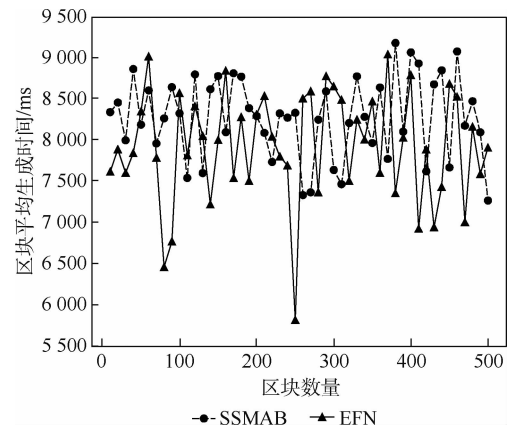


图 5　区块平均生成时间

Fig. 5　Average block generation time

北京航空航天大学学报

易次数相差 1 ~2 笔,性能有所衰退。

而在存储资源消耗上,通过图 7 可以得到 SSMAB 节点的存储容量,相比于 EFN 有明显降低,总数据量及年数据增长量为 EFN 的 13% 。

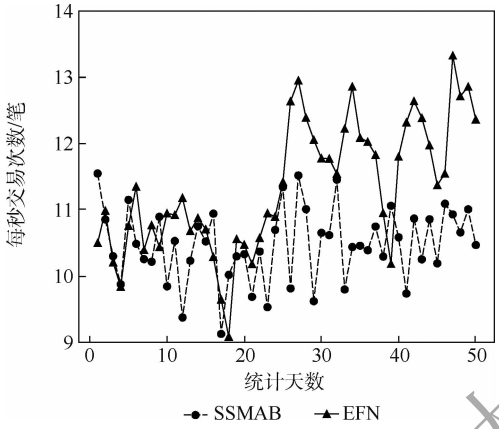


图 6 每秒交易次数

Fig. 6 Transactions per second

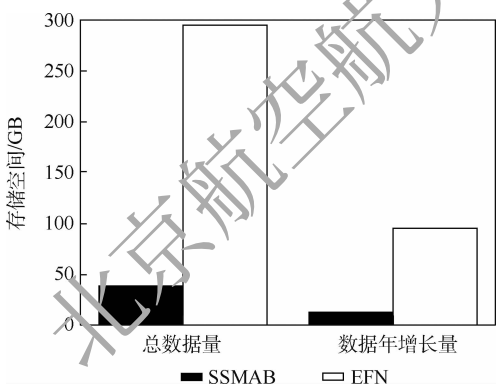


图 7 存储量对比

Fig. 7 Storage comparison

3.3 经济激励机制实现

在以太坊下实现经济激励制度。分片数据库使用 LevelDB 数据库进行实现,同时实现定价模型。在以太坊网络中,当某分片的存储节点减少,定价模型提高区块手续费。而获取到该分片的节点增多之后,该分片的手续费会保持下降,节点会综合考虑自己的存储成本来决定是否继续存储该分片,最终达到分片备份数同获取价格的动态平衡。保证节点能够通过本身存储的数据来获得收益。

在网络规模为 6 000 节点的模拟网络中,将初始分片备份数设为 83,价格设为 10,区块分片的价格同备份节点数动态平衡如图 8 所示。

当存储分片的节点数量上升后,区块价格随之下降,在图 8 中呈现出动态平衡关系,实现预定经济激励机制。

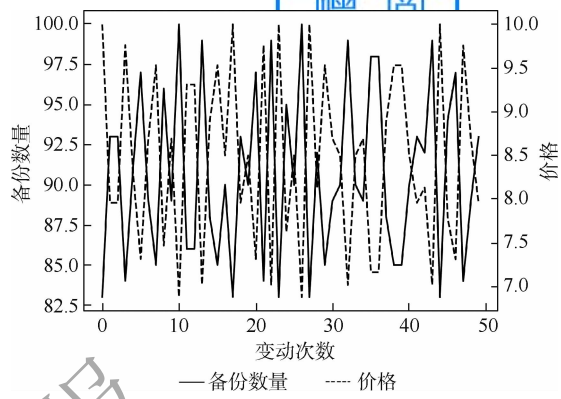


图 8 价格与备份数量关系

Fig. 8 Relationship between price and number of backups

4 结 论

1)本文针对区块链数据量线性增长,区块链系统存储可扩展性下降进而导致全节点减少、算力下降及区块链系统安全性降低的问题,基于账户型区块链提出一种可扩展存储节点模型。通过分离状态数据及区块数据,使节点在拥有状态数据的情况下可以对交易进行验证。

2)通过对区块数据进行分析,在确保区块可靠性及可用性的情况下,对区块进行分片存储,减少节点存储资源的消耗,增强系统存储可扩展性及安全性。

3)为了激励节点在网络中存在并维护系统安全性,设计并使用定价机制,通过经济的方式调节区块在网络中的备份数,调节区块的可靠性与可用性。

4)可扩展存储节点模型不能很好地支持智能合约的执行,由于智能合约无序分布在区块之中,运行智能合约需要消耗较长时间来获取智能合约相关数据,此外,经济激励机制较为简单,无法防范恶意攻击,需要在未来进行进一步优化。

参考文献(References)

[1] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system [EB/OL]. [2020-04-31]. https://bitc oin. org/ bitcoin. pdf.

[2] BUTERIN V. A next-generation smart contract and decentralized application platform [EB/OL]. [2020-04-31]. https://github. com/ethereum/wiki/wiki/White-Paper.

[3] CASINO F, DASAKLIS T K, PATSAKIS C. A systematic literature review of blockchain-based applications: Current status, classification and open issues [J]. Telematics and Informatics, 2019, 36: 55-81.

[4] Etherscan [EB/OL]. [2020-07-04]. https://eth erscan. Io.

[5] ZILLIQA. The ZILLIQA technical whitepaper [EB/OL]. [2020-05-15]. https://github. com/Zilliqa /docs/blob/master/whitepaper. pdf.

北京航空航天大学学报

北航学报 赠阅

[6] JIA D,XIN J,WANG Z,et al. ElasticChain:Support very large blockchain by reducing data redundancy[C]// Proceedings of Asia-Pacific Web. Berlin:Springer,2018:440-454.

[7] 赵羽龙,牛保宁,李鹏,等. 区块链增强型轻量级节点模型[J]. 计算机应用,2020,40(4):942-946.
ZHAO Y L,NIU B N,LI P,et al. Blockchain enhanced light-weight node model[J]. Journal of Computer Applications,2020,40(4):942-946(in Chinese).

[8] XU Z,HAN S,CHEN L. CUB,a consensus unit-based storage scheme for blockchain system[C]// 2018 IEEE 34th International Conference on Data Engineering (ICDE). Piscataway:IEEE Press,2018:173-184.

[9] ZAMANI M,MOVAHEDI M,RAYKOVA M. RapidChain:Scaling blockchain via full sharding[C]// Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York:ACM,2018:931-948.

[10] Rchain. The Rchain whitepaper[EB/OL]. [2020-05-12]. https:// github. com/rchain/reference/blob/master/docs/ RChain Whitepaper. pdf.

[11] DAI Q,XV K,GUO S,et al. A private data protection scheme based on blockchain under pipeline model[C]// 1st IEEE International Conference on Hot Information-Centric Networking (HotICN). Piscataway:IEEE Press,2018:37-45.

[12] Github[EB/OL]. [2020-06-11]. https:// git hub. com/ethereum/eth2.0-specs.

[13] Bitcomet[EB/OL]. [2020-06-11]. http:// www. bit comet. com/en.

[14] BENET J. IPFS-Content addressed,versioned,P2P file system[EB/OL]. (2014-07-14)[2020-06-11]. arxiv. org/abs/1407.3561v1.

[15] POLEL J. Probabilities and life[M]. New York:Dover Publications Inc,1962:23-87.

[16] Github[EB/OL]. [2020-06-12]. https:// git hub. com/swarmpit/ethstats/.

[17] PeerSim[EB/OL]. [2020-07-24]. http:// peersim. sourceforge. net/.

Account-based blockchain scalable storage model

ZHANG Xuhao, NIU Baoning*, GONG Ting

(College of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China)

Abstract: Blockchain data is growing linearly, eventually reaching a point where single node cannot provide sufficient storage capacity, and resulting in storage scalability problems. Although the light node model greatly reduces the demand for storage capacity, it also leads to a reduction in full nodes and decentralization, which threaten the security of blockchain networks. At present, there is no mature and scalable storage solution proposed for account-based blockchains. Considering that the access frequency of state data is much higher than that of block data, this paper proposes the scalable storage model for account-based blockchain (SSMAB). SSMAB saves state data in a completely redundant manner to guarantee its transaction verification function, stores block data via sharding storage to reduce redundancy, and adopts an economic incentive mechanism to ensure data availability while reducing storage consumption. Experiments show that SSMAB can ensure data reliability and availability, while reducing storage data volume to 13% of the full node model.

Keywords: blockchain; storage model; Ethereum; sharding; economic incentive

Received: 2020-11-16; **Accepted**: 2021-01-17; **Published online**: 2021-03-16 13:56

URL: kns. cnki. net/kcms/detail/11.2625. V.20210316.1345.002. html

Foundation items: National Natural Science Foundation of China (62072326); Key R & D Plan of Shanxi Province (201903D421007)

* **Corresponding author**. E-mail: niubaoning@ tyut. edu. cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2020.0642

低复杂度自适应容积卡尔曼滤波算法

全下文载

李春辉，马健*，杨永建，甘轶
（空军工程大学 航空工程学院，西安 710038）

摘 要：确定采样型滤波算法中的容积卡尔曼滤波(CKF)算法滤波性能优良，但是却难以克服目标模型不确定性或者目标状态突变带来的影响。构造强跟踪 CKF 能有效改善算法的自适应性，但是在求解渐消因子时大大增加了计算量。为此，提出一种低复杂度自适应 CKF 算法，通过设立基于新息的自适应修正判决准则和修正方式，直接对状态预测值进行修正，使滤波算法能及时跟上目标真实状态，以提高滤波精度。使用浮点操作数计算并分析了 CKF 算法、强跟踪 CKF 算法及所提算法的复杂度，同时将 3 种算法应用在建模不准确的目标跟踪中，并进行仿真验证。仿真结果表明：在目标建模不匹配的情况下，低复杂度自适应 CKF 算法和强跟踪 CKF 算法都能保持较好的滤波精度和数值稳定性，同时所提算法在算法复杂度上有明显改善。

关 键 词：容积卡尔曼滤波(CKF)；目标模型不确定性；强跟踪滤波器；自适应修正；算法复杂度
中图分类号：TN953
文献标志码：A **文章编号**：1001-5965(2022)04-0716-09

确定采样型滤波算法估计精度高，实现简单，是近年来非线性系统状态估计领域的一个研究热点，并广泛应用于目标跟踪、金融、信号处理和导航等领域[1]。确定采样型滤波算法包括无迹卡尔曼滤波(unscented Kalman filter，UKF)[2]、中心差分卡尔曼滤波(central difference Kalman filter，CDKF)[3]及容积卡尔曼滤波(cubature Kalman filter，CKF)[4-5]，理论上已经证明，这些算法能至少以二阶泰勒精度逼近各种非线性系统的后验协方差和均值[6]，因此不仅具有比扩展卡尔曼滤波(extended Kalman filter，EKF)更高的滤波精度，还能克服 EKF 需要计算雅可比矩阵、强非线性时较差的数值稳定性等不足。相比于粒子滤波(particle filter，PF)，确定采样型滤波在计算量上具有很大优势，而且不会出现 PF 粒子退化和贫化问题[1]。

Arasaratnam 和 Haykin[4-5]利用球面径向容积准则逼近最优估计的状态后验分布而提出的 CKF 算法，不仅能克服 UKF 在高维度和强非线性系统中的应用局限，同时具备更好的滤波精度及数值稳定性[7]。

但是 CKF 算法不具有自适应性，当目标运动模型不匹配或者目标状态发生突变时就会性能恶化甚至滤波发散。为了解决这一问题，目前主要有 2 种途径：①基于模型的方法，通过提高目标运动模型的准确性和自适应性来改善滤波性能[8-10]；②基于新息的自适应滤波[11]，这种方法中效果较好的是强跟踪滤波器(strong tracking filter，STF)[12]，STF 通过引入渐消因子实时调整增益矩阵，迫使残差序列保持正交，以弥补较大的预测误差[13-18]。但是构建 STF 理论推导复杂，而且渐消因子也存在求解计算量大[17-18]和插入位置

收稿日期：2020-11-17；**录用日期**：2021-01-15；**网络出版时间**：2021-02-01 15:54
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20210201.1422.003.html
基金项目：空军工程大学校长基金（XZJ2020039）
* **通信作者**. E-mail：majiankgd@163.com

引用格式：李春辉，马健，杨永建，等. 低复杂度自适应容积卡尔曼滤波算法[J]. 北京航空航天大学学报，2022，48(4)：716-724.
LI C H，MA J，YANG Y J，et al. Low-complexity adaptive cubature Kalman filter algorithm [J]. Journal of Beijing University of Aeronautics and Astronautics，2022，48(4)：716-724 (in Chinese).

北京航空航天大学学报

北航学报 赠阅

随意[19]等问题。

在降低 CKF 算法复杂度方面,文献[20]提出了一种简化 CKF 算法,但是该算法对克服建模不确定性的提高并不大。文献[17]将 STF 与假设检验理论结合以检测不确定性,在运行时间上有很大优化,但是求解渐消因子的计算量仍然不可避免。

为了能使 CKF 算法具有对目标模型不匹配及状态突变时的鲁棒性,同时能够避免强跟踪 CKF 算法较大的计算量,提出了一种低复杂度自适应 CKF 算法。首先,设立了基于新息的自适应修正判决准则及修正方式,然后,对系统状态预测值进行基于新息的直接补偿修正,使得滤波算法能够及时"跟上"目标真实状态,进而减小滤波误差。本文算法具有良好的滤波性能,同时在算法复杂度和仿真运行时间上有明显改善。

1　容积卡尔曼滤波

对目标建模时考虑如下非线性离散加性噪声系统:

$$\begin{cases} \boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k) + \boldsymbol{w}_k \\ \boldsymbol{y}_{k+1} = h(\boldsymbol{x}_{k+1}) + \boldsymbol{v}_{k+1} \end{cases} \tag{1}$$

式中:f 和 h 分别为非线性的状态转移函数和量测函数;$\boldsymbol{x}_{k+1} \in \mathbf{R}^{n_x}$ 和 $\boldsymbol{y}_{k+1} \in \mathbf{R}^{n_y}$ 分别为系统状态向量和量测向量;$\boldsymbol{w}_k$ 和 $\boldsymbol{v}_{k+1}$ 分别为系统过程噪声和量测噪声,两者是相互独立的零均值白噪声且对应的协方差矩阵分别为 $\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$。系统初始状态为 $\boldsymbol{x}_0 \sim N(0, \boldsymbol{P}_0)$,与 $\boldsymbol{w}_k$、$\boldsymbol{v}_{k+1}$ 互不相关,$\boldsymbol{P}_0$ 为初始协方差矩阵。

CKF 算法采用球面径向容积规则来逼近最优框架中的状态后验均值和后验协方差,根据文献[5,11],将 CKF 算法流程总结如下。

1.1　时间更新

在时间更新中,已知系统在 k 时刻的后验密度函数 $p(\boldsymbol{x}_k \mid \boldsymbol{y}_k) = N(\hat{\boldsymbol{x}}_{k \mid k}, \boldsymbol{P}_{k \mid k})$,则有以下算法流程。

分解预测误差协方差矩阵 $\boldsymbol{P}_{k \mid k}$ 为

$$\boldsymbol{P}_{k \mid k} = \boldsymbol{S}_{k \mid k} (\boldsymbol{S}_{k \mid k})^{\mathrm{T}} \tag{2}$$

式中:$\boldsymbol{S}$ 为协方差矩阵的平方根系数。

计算容积点为

$$\boldsymbol{X}_{i,k \mid k} = \boldsymbol{S}_{k \mid k} \boldsymbol{\xi}_i + \hat{\boldsymbol{x}}_{k \mid k} \qquad i = 1,2,\cdots,m \tag{3}$$

计算非线性状态转移函数传播的容积点为

$$\boldsymbol{X}^*_{i,k+1 \mid k} = f(\boldsymbol{X}_{i,k+1 \mid k}) \qquad i = 1,2,\cdots,m \tag{4}$$

式中:$\boldsymbol{X}_{i,k \mid k}$ 和 $\boldsymbol{X}^*_{i,k+1 \mid k}$ 均为容积点;m 为容积点的个数,且根据三阶容积原则,m 是状态向量维数 n 的 2 倍,即 $m = 2n$;$\boldsymbol{\xi}_i = \sqrt{m/2}\,[1]_i$ 为基本容积点集,$[1]_i$ 为点集[1]中的第 i 列,点集[1]可以表示为

$$[1] = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix} \right\} \tag{5}$$

计算状态预测值和预测协方差矩阵为

$$\hat{\boldsymbol{x}}_{k+1 \mid k} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{X}^*_{i,k+1 \mid k} \tag{6}$$

$$\boldsymbol{P}_{k+1 \mid k} = \frac{1}{2n} \sum_{i=1}^{2n} \boldsymbol{X}^*_{i,k+1 \mid k} (\boldsymbol{X}^*_{i,k+1 \mid k})^{\mathrm{T}} - \hat{\boldsymbol{x}}_{k+1 \mid k} \hat{\boldsymbol{x}}^{\mathrm{T}}_{k+1 \mid k} + \boldsymbol{Q}_k \tag{7}$$

1.2　量测更新

分解预测误差协方差矩阵 $\boldsymbol{P}_{k+1 \mid k}$ 为

$$\boldsymbol{P}_{k+1 \mid k} = \boldsymbol{S}_{k+1 \mid k} (\boldsymbol{S}_{k+1 \mid k})^{\mathrm{T}} \tag{8}$$

计算容积点为

$$\boldsymbol{X}_{i,k+1 \mid k} = \boldsymbol{S}_{k+1 \mid k} \boldsymbol{\xi}_i + \hat{\boldsymbol{x}}_{k+1 \mid k} \qquad i = 1,2,\cdots,m \tag{9}$$

计算非线性量测函数传播的容积点为

$$\boldsymbol{Y}_{i,k+1 \mid k} = h(\boldsymbol{X}_{i,k+1 \mid k}) \qquad i = 1,2,\cdots,m \tag{10}$$

计算量测的预测值为

$$\hat{\boldsymbol{y}}_{k+1 \mid k} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{Y}_{i,k+1 \mid k} \tag{11}$$

计算新息、新息协方差矩阵分别为

$$\boldsymbol{e}_{k+1} = \boldsymbol{y}_{k+1} - \hat{\boldsymbol{y}}_{k+1 \mid k} \tag{12}$$

$$\boldsymbol{P}_{yy,k+1 \mid k} = \frac{1}{2n} \sum_{i=1}^{2n} \boldsymbol{Y}_{i,k+1 \mid k} \boldsymbol{Y}^{\mathrm{T}}_{i,k+1 \mid k} - \hat{\boldsymbol{y}}_{k+1 \mid k} \hat{\boldsymbol{y}}^{\mathrm{T}}_{k+1 \mid k} + \boldsymbol{R}_{k+1} \tag{13}$$

式中:$\boldsymbol{y}_{k+1}$ 为系统 $k+1$ 时刻的量测值。

计算互协方差矩阵为

$$\boldsymbol{P}_{xy,k+1 \mid k} = \frac{1}{2n} \sum_{i=1}^{2n} \boldsymbol{X}_{i,k+1 \mid k} \boldsymbol{Y}^{\mathrm{T}}_{i,k+1 \mid k} - \hat{\boldsymbol{x}}_{k+1 \mid k} \hat{\boldsymbol{y}}^{\mathrm{T}}_{k+1 \mid k} \tag{14}$$

计算滤波增益 $\boldsymbol{K}_{k+1}$、$k+1$ 时刻的状态估计值 $\hat{\boldsymbol{x}}_{k+1 \mid k+1}$ 及估计误差协方差矩阵 $\boldsymbol{P}_{k+1 \mid k+1}$ 分别为

$$\boldsymbol{K}_{k+1} = \boldsymbol{P}_{xy,k+1 \mid k} \boldsymbol{P}^{-1}_{yy,k+1 \mid k} \tag{15}$$

$$\hat{\boldsymbol{x}}_{k+1 \mid k+1} = \hat{\boldsymbol{x}}_{k+1 \mid k} + \boldsymbol{K}_{k+1} (\boldsymbol{y}_{k+1} - \hat{\boldsymbol{y}}_{k+1 \mid k}) \tag{16}$$

$$\boldsymbol{P}_{k+1 \mid k+1} = \boldsymbol{P}_{k+1 \mid k} - \boldsymbol{K}_{k+1} \boldsymbol{P}_{yy,k+1 \mid k} \boldsymbol{K}^{\mathrm{T}}_k \tag{17}$$

2　自适应修正容积卡尔曼滤波

2.1　误差分析

从式(16)可以看出,状态估计值 $\hat{\boldsymbol{x}}_{k+1 \mid k+1}$ 由先验的状态预测值 $\hat{\boldsymbol{x}}_{k+1 \mid k}$ 和后验的量测新息反馈 $\boldsymbol{K}_{k+1} \boldsymbol{e}_{k+1}$ 两部分构成。

当 k 时刻目标运动模型与建立模型不匹配或

北京航空航天大学学报

北航学报 瞰阅

者目标状态发生突变时，系统 $k+1$ 时刻的状态预测值 $\hat{\boldsymbol{x}}_{k+1|k}$ 及经过非线性量测函数传播后的量测预测值 $\hat{\boldsymbol{y}}_{k+1|k}$ 就会发生较大的偏差，进而体现在 $k+1$ 时刻较大的新息 $\boldsymbol{e}_{k+1}$ 里。

为了克服上述情况，STF 通过在预测误差协方差矩阵里引入渐消因子，在线调整滤波增益矩阵，迫使输出残差序列正交，因此具有对不确定模型系统的鲁棒性[13]。由 STF 的构造过程可见，STF 本质上是对式(16)的第 2 部分的增益 $\boldsymbol{K}_{k+1}$ 进行修正。其实对上述分析中第 1 部分的状态预测值 $\hat{\boldsymbol{x}}_{k+1|k}$ 也可以进行修正，即对 $\hat{\boldsymbol{x}}_{k+1|k}$ 进行合理的补偿修正以及时“跟上”目标实际状态，进而避免目标模型不匹配或者状态突变带来的不良影响。

由此可见，这种直接对状态预测值 $\hat{\boldsymbol{x}}_{k+1|k}$ 进行补偿修正的方法更为直观且相比于 STF 明显减小了计算量，但是需要建立自适应修正的判决准则和修正方式。

2.2 基于新息的自适应修正判决准则和修正方式

由 2.1 节的分析可知，当目标模型不匹配或者目标状态发生突变时，新息 $\boldsymbol{e}_{k+1}$ 能反映系统状态预测值和量测预测值出现的较大偏差，同时也能反映出预测值偏离实际状态的“趋势”和“程度”。因此，基于新息来判决是否修正及如何修正。

与量测值息息相关的是量测误差协方差 $\boldsymbol{R}_k$，在无不良量测的情况下，由于系统量测噪声是白噪声，以 $k+1$ 时刻的量测值为中心，以 $\boldsymbol{R}_{k+1}$ 对角线上第 i 个分量的协方差 $\sqrt{\boldsymbol{R}_{k+1,i}}$ 为半径画一个圆，则系统状态真实值及量测的预测值都应该在圆里。

但是如图 1 所示，当 k 时刻之后由于目标模型不匹配或目标状态突变，$k+1$ 时刻时间更新阶段的状态预测值 $\hat{\boldsymbol{x}}_{k+1|k}$ 与状态真实值 $\boldsymbol{x}_{k+1}$ 已经出现了较大偏差，这一偏差经过非线性量测函数 h 传播后使得 $k+1$ 时刻的量测预测值 $\hat{\boldsymbol{y}}_{k+1|k}$ 偏离并超出上述量测圆，进而使得第 i 个量测变量的新息 $e_{k+1,i}$ 变大。

因此，可以建立起基于新息的修正判决准则。当新息第 i 个分量的绝对值大于某一门限时，则需要对状态预测值进行补偿修正，以及时“跟上”目标真实状态。为了形象地将这一门限也等效为一个圆，即图 1 中的蓝色圆，其半径 $r=\lambda\sqrt{\boldsymbol{R}_{k+1,i}}$ 即为修正判决门限，λ 为基于实验得到的缩放因子，与系统量测精度相关，主要作用是调节先验预测值和后验量测值对修正范围的影响，为了避免过度修正，同时考虑到滤波中存在积累误差，λ 的选取区间为[0.6,1)。

修正的主要目的是给系统状态预测值 $\hat{\boldsymbol{x}}_{k+1|k}$ 一个符合实际变化的补偿，使滤波能及时“跟上”目标突变的状态，进而使得修正后的量测预测值 $\hat{\boldsymbol{y}}_{k+1|k}$ 能克服偏差，回到图 1 中的蓝色修正圆内，

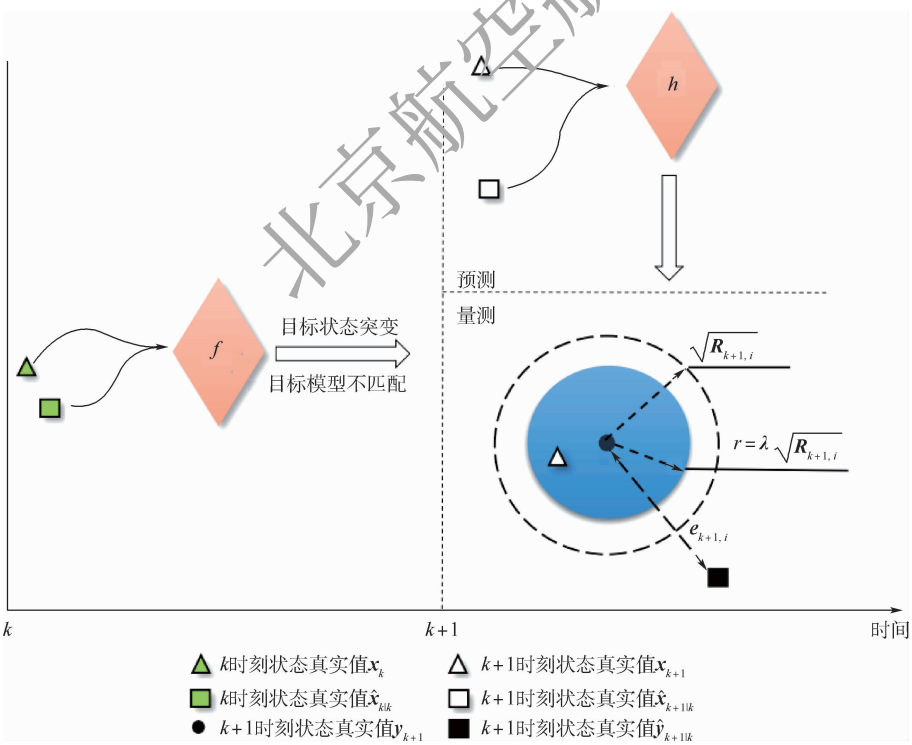


图 1 基于新息的修正判决准则

Fig. 1 Amending judgment criteria based on innovation sequence

北航学报 赠阅

并最终降低滤波估计误差。而新息能够反映出预测值偏离实际状态的“趋势”和“程度”，因此对状态预测值的修正也是基于新息的。假设系统状态向量及量测向量均是 x 轴位置和速度、y 轴位置和速度四维，即

$$\begin{cases}\tilde{\boldsymbol{x}}_{k+1\mid k}(1) = \hat{\boldsymbol{x}}_{k+1\mid k}(1) + \boldsymbol{e}_{k+1,1}c_1\\ \tilde{\boldsymbol{x}}_{k+1\mid k}(2) = \hat{\boldsymbol{x}}_{k+1\mid k}(2) + \boldsymbol{e}_{k+1,2}c_2\\ \tilde{\boldsymbol{x}}_{k+1\mid k}(3) = \hat{\boldsymbol{x}}_{k+1\mid k}(3) + \boldsymbol{e}_{k+1,3}c_3\\ \tilde{\boldsymbol{x}}_{k+1\mid k}(4) = \hat{\boldsymbol{x}}_{k+1\mid k}(4) + \boldsymbol{e}_{k+1,4}c_4\end{cases} \tag{18}$$

式中：$\hat{\boldsymbol{x}}_{k+1\mid k}(i)$ 为状态预测值的第 i 个向量；$\tilde{\boldsymbol{x}}_{k+1\mid k}(i)$ 为对第 i 个向量的修正值；$\boldsymbol{e}_{k+1,i}$ 为 $k+1$ 时刻新息的第 i 个分量；c_1、c_2、c_3 和 c_4 分别为 x 轴位置、x 轴速度、y 轴位置和 y 轴速度的修正系数。

修正系数的选取首先要遵循避免过度修正基本原则，其次主要考虑两方面：一是修正系数与系统量测精度相关；二是新息 $\boldsymbol{e}_{k+1}$ 所提供的信息维度，如假设新息中只含有位置的信息，则在选取时 c_1 或 c_3 可以选取得稍大些，速度的修正可以通过位置变化来传递，但是对应的系数 c_2 或 c_4 选取得稍小些。

对系统状态预测值 $\hat{\boldsymbol{x}}_{k+1\mid k}$ 的修正不可避免地会对预测误差协方差造成影响，因此为了防止滤波过程中这一偏差累积，每次修正后对预测误差协方差矩阵也要进行补偿，即

$$\boldsymbol{P}_{k+1\mid k+1} = \boldsymbol{P}_{k+1\mid k+1} + \left(\frac{1}{4}\sum_{i=1}^{4}c_i\right)\mathrm{sqrt}(\bar{R}_{k+1}) \tag{19}$$

式中：$\bar{R}_{k+1}$ 为 $\boldsymbol{R}_{k+1}$ 对角线元素的均值。

2.3　低复杂度自适应 CKF 算法实现

根据建立的判决准则和修正方式总结了本文算法的流程，如图 2 所示。图中在判别之后的虚线是低复杂度自适应 CKF 算法的路径，实线是标准 CKF 算法的路径，具体步骤总结如下：

步骤 1　通过式(2)～式(12)完成时间更新及量测更新的一部分，并得到新息 $\boldsymbol{e}_{k+1}$。

步骤 2　将新息中对应变量的绝对值与门限进行比较，若满足修正判决条件，则用式(18)进行状态预测值修正，之后用式(9)～式(17)完成滤波，并用式(19)进行预测误差协方差补偿。

步骤 3　若不需要进行修正，则按式(13)～式(17)完成标准 CKF 滤波。

步骤 4　重复步骤 1～步骤 3。

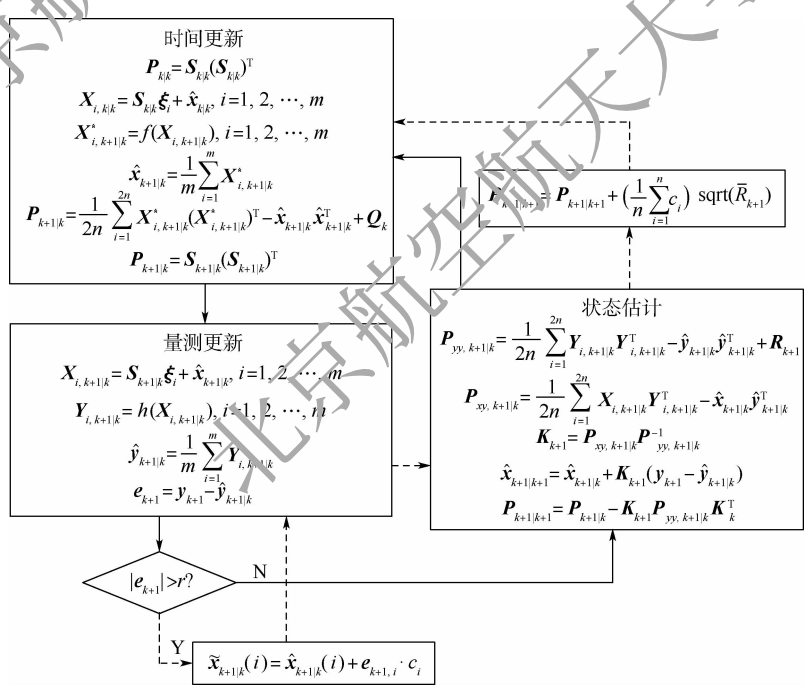


图 2　低复杂度自适应 CKF 算法流程

Fig. 2　Flowchart of low-complexity adaptive CKF algorithm

3　算法复杂度分析

使用浮点操作数(flops)来分析算法复杂度是直观有效的方法，根据文献[21-22]，定义以下 4 种矩阵运算的 flops：

1）矩阵加减法。$\boldsymbol{A}\in\mathbf{R}^{n\times m}$，$\boldsymbol{B}\in\mathbf{R}^{n\times m}$，$\boldsymbol{A}\pm\boldsymbol{B}$ 运算的复杂度为 nm flops。

2）矩阵乘法。$\boldsymbol{A}\in\mathbf{R}^{n\times m}$，$\boldsymbol{B}\in\mathbf{R}^{m\times l}$，$\boldsymbol{AB}$ 运算的复杂度为 $2nml-nl$ flops。

3）矩阵求逆。$\boldsymbol{A}\in\mathbf{R}^{n\times n}$，$\boldsymbol{A}^{-1}$ 运算的复杂度

为 n^3 flops。

4）Cholesky 分解。$\boldsymbol{A}\in\mathbf{R}^{n\times n}$，chol($\boldsymbol{A}$)运算的复杂度为 $n^3/3$ flops。

假设系统状态变量为 n 维，量测变量为 l 维，则利用上述矩阵运算复杂度定义，对标准 CKF 算法、强跟踪 CKF 算法[13]及本文提出的低复杂度自适应 CKF 算法的复杂度进行计算，结果如表 1 所示。

表 1　三种算法复杂度计算结果

Table 1　Computational results of complexity of three algorithms

算法	算法复杂度/flops
CKF	$20n^3/3+10n^2+10n^2l+8nl^2+2nl+l^3+3l^2+l$
强跟踪 CKF	$26n^3/3+10n^2+16n^2l+10nl^2-nl+l^3+8l^2+3l$
低复杂度自适应 CKF	$20n^3/3+11n^2+16n^2l+8nl^2+2nl+l^3+3l^2+3l+n$

CKF 算法的复杂度分析具体过程可见文献[21]，参照该过程得到了其他 2 种算法的复杂度计算结果。

强跟踪 CKF 算法主要考虑文献[13]中式(3)、式(6)、式(10)和式(11)求解渐消因子上，4 个公式对应的算法复杂度依次为 $2l$、$2l^2$、$2n^3+6n^2l+2l^2n+l^2-3nl$ 和 $2l^2$ flops。

低复杂度自适应 CKF 算法主要考虑式(12)新息的计算及式(18)和式(19)的计算，对应步骤的算法复杂度依次为 l、n 和 n^2+l flops。

尽管文献[13]中的强跟踪 CKF 算法已经进行了改进，求解时避免了计算雅可比矩阵，但是相比而言，本文算法在复杂度上仍有明显改善。将强跟踪 CKF 算法与低复杂度自适应 CKF 算法的复杂度求差可得 $\Delta\text{flops}=n(2n^2-n-1)+3nl(2n-1)+2l^2n+5l^2$，在 n 和 l 都大于 1 时，这一差值会随着 n 和 l 的增大而变大，即随着状态向量、量测向量维度的增加，低复杂度自适应 CKF 算法的优势会更大。为了更直观地体现这一差异，根据得到的算法复杂度公式绘制了 3 种算法的复杂度随状态向量、量测向量维度分别变化时的曲线，如图 3和图 4 所示。

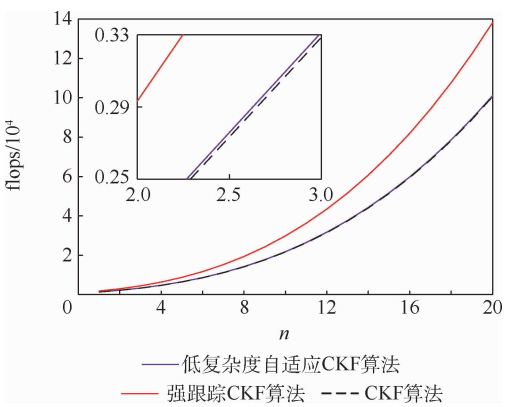


图 3　三种算法的复杂度比较($l=8$)

Fig. 3　Comparison of three algorithms complexity ($l=8$)

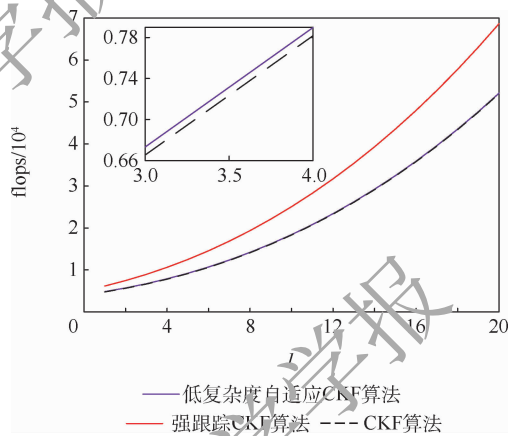


图 4　三种算法的复杂度比较($n=8$)

Fig. 4　Comparison of three algorithms complexity ($n=8$)

4　仿真与分析

4.1　建模与仿真条件设置

仿真中，目标运动模型为匀速运动和匀速转弯运动的混合机动。系统状态向量为 $\boldsymbol{X}=[x,\dot{x},y,\dot{y}]^{\mathrm{T}}$，其中匀速运动模型的状态转移矩阵为 $\boldsymbol{F}_{\mathrm{CV}}$，匀速转弯运动模型的状态转移矩阵为 $\boldsymbol{F}_{\mathrm{CT}}$，分别如下：

$$\boldsymbol{F}_{\mathrm{CV}}=\begin{bmatrix}1&T&0&0\\0&1&0&0\\0&0&1&T\\0&0&0&1\end{bmatrix}\tag{20}$$

$$\boldsymbol{F}_{\mathrm{CT}}=\begin{bmatrix}1&\dfrac{\sin(\omega T)}{\omega}&0&\dfrac{\cos(\omega T)-1}{\omega}\\0&\cos(\omega T)&0&-\sin(\omega T)\\0&\dfrac{1-\cos(\omega T)}{\omega}&1&\dfrac{\sin(\omega T)}{\omega}\\0&\sin(\omega T)&0&\cos(\omega T)\end{bmatrix}\tag{21}$$

系统的量测方程为

$$\boldsymbol{Y}_k=\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1&0&0&0\\0&0&1&0\end{bmatrix}\boldsymbol{X}_k+\boldsymbol{V}_k\tag{22}$$

为了验证低复杂度自适应 CKF 算法在目标建模不匹配时的滤波性能，在跟踪时与 CKF 算

北京航空航天大学学报

北航学报 增阅

法、强跟踪 CKF 算法进行对比，同时 3 种算法在目标建模时都仅使用匀速运动模型 $\boldsymbol{F}_{\mathrm{CV}}$。

在目标运动仿真参数的设置上，转弯率 ω = 5°/s，采样时间 T = 1 s，目标在 0 ~ 250 s 的时间内保持交替变换运动状态，其中匀速直线运动时间段为[0，70] s，(120，145] s，(160，200] s，(225，250] s；匀速转弯运动时间段为(70，120] s，(145，160] s，(200，225] s，共 3 次转弯。

目标初始状态为 $\boldsymbol{x}_0$ = [1 000 m，200 m/s，1 000 m，10 m/s]^{T}，系统初始协方差 $\boldsymbol{P}_0$ = diag(3，1，3，1)；系统过程噪声协方差矩阵 $\boldsymbol{Q}$ = diag(1，1，1，1)，量测噪声协方差矩阵 $\boldsymbol{R}$ = diag(100，100)；强跟踪 CKF 算法的遗忘因子取 0.95。

仿真结果以目标 x 轴和 y 轴位置上的均方根误差(root mean square error，RMSE)来评价，并进行 150 次蒙特卡罗仿真，均方根误差的定义为

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\boldsymbol{X}_i(k) - \hat{\boldsymbol{X}}_i^j(k|k)\right)^2} \tag{23}$$

式中：$\boldsymbol{X}_i(k)$ 和 $\hat{\boldsymbol{X}}_i^j(k|k)$ 分别为第 j 次仿真中状态向量的第 i 个分量在 k 时刻的真实值和状态预测值；N 为蒙特卡罗仿真次数。每次仿真中，目标初始状态预测值和状态初始值取一致。

4.2　仿真结果分析

图 5 为其中一次蒙特卡罗仿真的目标真实运动轨迹及 3 种算法的滤波轨迹。可以看到，低复杂度自适应 CKF 算法及强跟踪 CKF 算法均能很好地跟踪目标，但是由于滤波建模时使用的始终是匀速运动模型，CKF 算法在 3 次机动转弯时跟踪性能就会下降。

图 6 和图 7 分别为滤波阶段 x 轴和 y 轴位置的均方根误差，左上角为对 10 ~ 20 s 阶段放大显示。可以看出，整体上低复杂度自适应 CKF 算法及强跟踪 CKF 算法能够在建模不准确时始终保持较高的滤波精度，但是 CKF 算法只有在目标匀速运动阶段即建模匹配时才能发挥很好的滤波性能，在目标转弯阶段时其滤波精度受到很大影响。但是从图 6 和图 7 中放大显示部分来看，CKF 算法在目标运动建模准确时滤波精度要比另外 2 种算法要高。

表 2 为 3 种算法在滤波时间段里的平均均方根误差。结合其中的数据及图 6 和图 7 可以看出，低复杂度自适应 CKF 算法能够保持和强跟踪 CKF 算法基本一致的滤波精度，可以较好地克服由于目标建模不准确或者目标状态突变带来的不利影响。

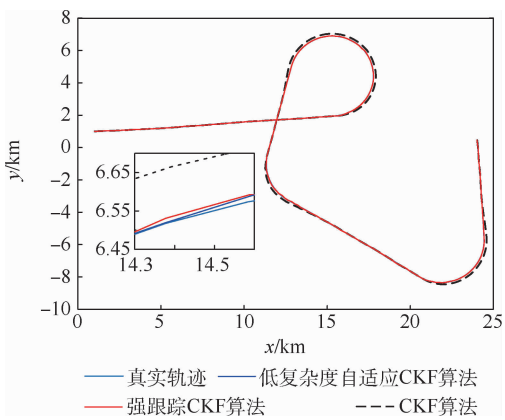


图 5　单次仿真中的目标运动轨迹

Fig. 5　Target motion trajectory in single simulation

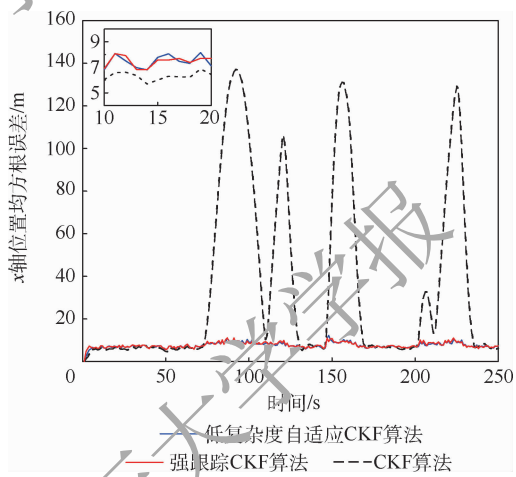


图 6　x 轴位置均方根误差

Fig. 6　RMSE of position in x axis

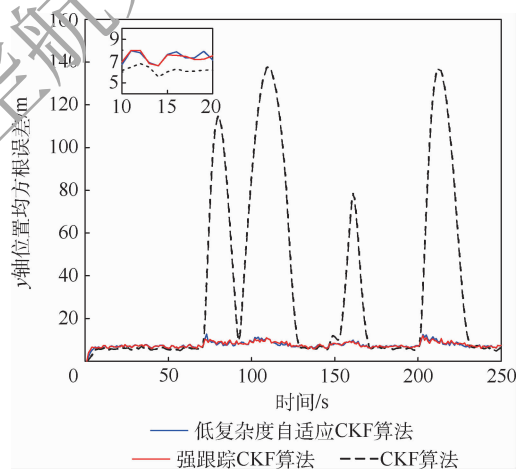


图 7　y 轴位置均方根误差

Fig. 7　RMSE of position in y axis

表 2　不同算法平均均方根误差对比

Table 2　Comparison of mean RMSE among different algorithms

算法	平均 RMSE/m	
	x 轴位置	y 轴位置
CKF	35.732 5	35.062 5
强跟踪 CKF	8.416 2	8.389 1
低复杂度自适应 CKF	8.405 5	8.384 5

北京航空航天大学学报

北航学报 醒阅

在硬件配置为 Intel i7 4710 处理器和 4 GB 内存的笔记本电脑上使用 MATLAB 2017a 进行了试验。3 种算法的单次仿真平均运行时间和运行时间增加百分比如表 3 所示。运行时间增量计算方式为$(T_i - T_0)/T_0$,T_0 为 CKF 算法的运行时间,T_i 为其他 2 种算法的运行时间。

由表 3 可见,CKF 算法耗时最短,低复杂度自适应 CKF 算法在运行时间上略微增加了 1.29%,而强跟踪 CKF 算法由于需要求解渐消因子,耗时最多,相比较 CKF 算法增加了 29.53%。运行时间对比数据和前文算法复杂度计算结果相吻合。

为了验证低复杂度自适应 CKF 算法的滤波数值稳定性,绘制 3 种算法每次蒙特卡罗仿真中 x 轴和 y 轴位置的平均均方根误差,如图 8 和图 9 所示。可以看出,低复杂度自适应 CKF 算法及强跟踪 CKF 算法每次仿真中的位置平均均方根误差都低于 CKF 算法,这是因为 CKF 算法在每次仿真中目标建模不匹配即转弯阶段时的滤波误差会剧增。同时相比于 CKF 算法,另外 2 种算法的位置平均均方根误差没有出现某次仿真发散的情况,而且曲线波动幅度更小,数值稳定性更优。

表 3 不同算法运行时间对比

Table 3 Comparison of running time among different algorithms

算法	平均运行时间/(10^{-2}s)	运行时间增加百分比
CKF	2.677 6	0
强跟踪 CKF	3.468 3	29.53
低复杂度自适应 CKF	2.712 1	1.29

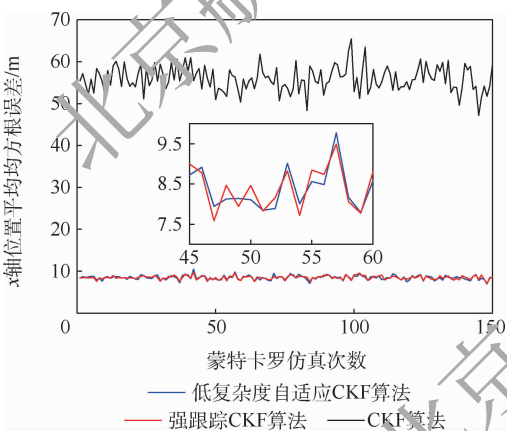


图 8 x 轴位置平均均方根误差

Fig. 8 Mean RMSE of position in x axis

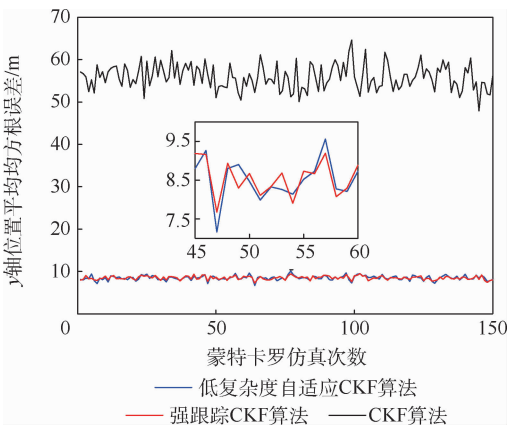


图 9 y 轴位置平均均方根误差

Fig. 9 Mean RMSE of position in y axis

5 结 论

1)仿真结果表明,在目标建模不准确的情况下,低复杂度自适应 CKF 算法能够保持和强跟踪 CKF 算法基本一致的滤波精度和数值稳定性,有效提高了 CKF 算法的自适应性,同时也验证了所提算法的有效性。

2)使用等效浮点操作数进行了算法复杂度计算,低复杂度自适应 CKF 算法相比较强跟踪 CKF 算法具有更低的算法复杂度,且随着状态向量、量测向量维度的增加,所提算法的优势会更大。

3)仿真结果表明,在算法运行时间上,相比较 CKF 算法,强跟踪 CKF 算法运行时间增加了 29.53%,而低复杂度自适应 CKF 算法仅增加了 1.29%,具有明显优势。仿真运行时间结果与复杂度计算结果保持一致。

为了实现相关参数的自动选取,需要进一步考虑与智能寻优算法的结合或者相关理论的推导。

参考文献(References)

[1] 王小旭,潘泉,黄鹤,等. 非线性系统确定采样型滤波算法综述[J]. 控制与决策,2012,27(6):801-812.
WANG X X,PAN Q,HUANG H,et al. Overview of deterministic sampling filtering algorithms for nonlinear system[J]. Control and Decision,2012,27(6):801-812(in Chinese).

[2] JULIER S J,UHLMANN J K,DURRANT-WHYTE H F. A new method for the nonlinear transformation of means and covariances in filters and estimators[J]. IEEE Transactions on Automatic Control,2000,45(3):477-482.

[3] MERWE R V,WAN E A. Sigma-point Kalman filters for integrated navigation[C]// Processing of the 60th Annual Meeting of the Institute of Navigation,2004:641-654.

[4] ARASARATNAM I,HAYKIN S. Cubature Kalman filters[J]. IEEE Transactions on Automatic Control,2009,54(6):1254-1269.

[5] ARASARATNAM I,HAYKIN S. Cubature Kalman filtering for continuous-discrete systems:Theory and simulations[J]. IEEE Transactions on Signal Processing,2010,58(10):4977-4993.

[6] JULIER S J,UHLMANN J K. A general method for approximating nonlinear transformation of probability distributions[EB/

OL]. (1996-11-01)[2020-10-29]. http://www.eng.ox.ac.uk/.

[7] 穆静,蔡远利. 迭代容积卡尔曼滤波算法及其应用[J]. 系统工程与电子技术,2011,33(7):1454-1457.
MU J,CAI L Y. Iterated cubature Kalman filter and its application[J]. Systems Engineering and Electronics,2011,33(7):1454-1457(in Chinese).

[8] ZHANG H W,XIE J W,GE J A,et al. Adaptive strong tracking square-root cubature Kalman filter for maneuvering aircraft tracking[J]. IEEE Access,2018,6:10052-10061.

[9] ZHANG H W,XIE J W,GE J A,et al. Strong tracking SCKF based on adaptive CS model for manoeuvring aircraft tracking[J]. IET Radar,Sonar & Navigation,2018,12(7):742-749.

[10] HAN B,HUANG H Q,LEI L,et al. An improved IMM algorithm based on STSRCKF for maneuvering target tracking[J]. IEEE Access,2019,7:57795-57804.

[11] 徐树生,林孝工. 基于鲁棒 CKF 的多传感器全信息融合算法[J]. 电机与控制学报,2013,17(2):90-97.
XU S S,LIN X G. Multi-sensor all information fusion algorithm based on robust CKF[J]. Machines and Control,2013,17(2):90-97(in Chinese).

[12] ZHOU D H,XI Y G,ZHANG Z J. A suboptimal multiple extended Kalman filters[J]. Chinese Journal of Automation,1992,4(2):145-152.

[13] 孙妍,鲁涤强,陈启军. 一种基于强跟踪的改进容积卡尔曼滤波器[J]华中科技大学学报(自然科学版),2013,41(S1):451-454.
SUN Y,LU D Q,CHEN Q J. An improved cubature Kalman filters based on strong tracking[J]. Journal of Huazhong University of Science and Technology(Nature Science Edition),2013,41(S1):451-454(in Chinese).

[14] ZHAO L Q,WANG J L,YU T,et al. Design of adaptive robust square-root cubature Kalman filter with noise statistic estimator[J]. Applied Mathematics and Computation,2015,256:352-367.

[15] GE Q B,LI W B,WEN C L. SCKF-STF-CN:A universal nonlinear filter for maneuver target tracking[J]. Journal of Zhejiang University Science C,2011,12(8):678-686.

[16] LI N,ZHU R,ZHANG Y. A strong tracking square-root CKF algorithm based on multiple fading factors for target tracking[C]// Processing of the 7th International Joint Conference on Computational Sciences and Optimization. Piscataway:IEEE Press,2014:16-24.

[17] ZHANG A,BAO S D,BI W H,et al. Low-cost adaptive square-root cubature Kalman filter for systems with process model uncertainty[J]. Journal of Systems Engineering and Electronics,2016,27(5):945-953.

[18] ZHANG A,BAO S D,GAO F,et al. A novel strong tracking cubature Kalman filter and its application in maneuvering target tracking[J]. Chinese Journal of Aeronautics,2019,32(11):2489-2502.

[19] 张浩为,谢军伟,葛佳昂,等. 自适应 CS 模型的强跟踪平方根容积卡尔曼滤波算法[J]. 系统工程与电子技术,2019,41(6):19-27.
ZHANG H W,XIE J W,GE J A,et al. Strong tracking square-root cubature Kalman filter over adaptive current statistical model[J]. Systems Engineering and Electronics,2019,41(6):19-27(in Chinese).

[20] 戴定成,蔡宗平,牛创. 基于简化平方根容积卡尔曼滤波的跟踪算法[J]. 电光与控制,2015,22(3):11-14.
DAI D C,CAI Z P,NIU C. Target tracking algorithm based on reduced square-root cubature Kalman filter[J]. Electronics Optics & Control,2015,22(3):11-14(in Chinese).

[21] 张召友,郝燕玲,吴旭. 3 种确定性采样非线性滤波算法的复杂度分析[J]. 哈尔滨工业大学学报,2013,45(12):111-115.
ZHANG Z Y,HAO Y L,WU X. Complexity analysis of three deterministic sampling nonlinear filtering algorithms[J]. Journal of Harbin Institute of Technology,2013,45(12):111-115(in Chinese).

[22] GREWAL M S,ANDREW A P. Kalman filtering,theory and practice using Matlab[M]. 2nd ed. New York:John Wiley & Sons,2008:225-289.

北航学报 赠阅

Low-complexity adaptive cubature Kalman filter algorithm

LI Chunhui, MA Jian*, YANG Yongjian, GAN Yi

(Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: Cubature Kalman filter (CKF) with good filtering performance is one of the deterministic sampling filtering algorithms, but it is not able to overcome the impact caused by the target model uncertainty or the mutation of the target state. Constructing strong tracking CKF can effectively improve the adaptability of the algorithm, but the computation is greatly increased when solving the fading factor. A low-complexity adaptive CKF algorithm is proposed to solve the above problems. By establishing adaptive judgment criteria and amending method based on innovation sequence, the predicted state value is directly amended, so that the filtering algorithm can keep up with the real state of the target in time, and thus improve the filtering accuracy. The complexity of CKF, strong tracking CKF and the proposed algorithm are calculated and analyzed by using floating point operations. At the same time, the above three algorithms are applied to target tracking with inaccurate modeling, and are verified through simulation. The simulation results show that both the proposed algorithm and the strong tracking CKF algorithm can maintain better filtering accuracy and numerical stability in the case of mismatched target modeling, and the proposed algorithm has obvious improvement in algorithm complexity.

Keywords: cubature Kalman filter (CKF); target model uncertainty; strong tracking filter; adaptive amending; algorithm complexity

北京航空航天大学学报

Received: 2020-11-17; **Accepted**: 2021-01-15; **Published online**: 2021-02-01 15:54

URL: kns.cnki.net/kcms/detail/11.2625.V.20210201.1422.003.html

Foundation item: Air Force Engineering University President's Fund (XZJ2020039)

* **Corresponding author**. E-mail: majiankgd@163.com

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码,中文摘要(包括目的、方法、结果及结论),中文关键词(5 ~ 8 个),中图分类号,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词,引言,正文,参考文献。首页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投它刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经刊登,即赠送单行本。

3.4 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 北京航空航天大学学报编辑部

办公地点:北京航空航天大学办公楼 405,407,409 房间

电 话:(010)82315594,82338922,82314839,82315426

E-mail:jbuaa@ buaa. edu. cn

http://bhxb. buaa. edu. cn

http://www. buaa. edu. cn

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

（原《北京航空学院学报》）

（月刊　1956年创刊）

第48卷 第4期　2022年4月

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100083
地　　址　北京市海淀区学院路37号
印　　刷　北京科信印刷有限公司
发　　行　《北京航空航天大学学报》编辑部
发行范围　国内外发行
联系电话　(010) 82315594　82338922　82314839
电子信箱　jbuaa@buaa.edu.cn

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.48 No.4　April 2022

Administrated by　Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by　Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100083, P. R. China)
Chief Editor　ZHAO Qinping
Edited and Published by　Editorial Board of JBUAA
Printed by　Beijing Kexin Printing Co., Ltd.
Distributed by　Editorial Board of JBUAA
Telephone　(010) 82315594　82338922　82314839
E-mail　jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V　　国内定价：50.00元／期

ISSN 1001-5965

9 771001 596229 04