



北航学报
赠阅

ISSN 1001-5965
CODEN BHHDE8

北京航空航天大学 学报

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS



2019-10

Vol.45 No.10

目 次

吸气式高超声速飞行器热气动弹性研究进展	杨超, 赵黄达, 吴志刚 (1911)
民机客舱中太阳辐射对热舒适性的影响	庞丽萍, 李恒, 王天博, 范俊, 邹凌宇 (1924)
基于 SLM 的模拟月壤原位成形技术	李雯, 徐可宁, 黄勇, 胡文颖, 王道宽, 姚思齐 (1931)
面向战斗机云作战的构造型仿真平台架构	田永亮, 王永庆, 熊培森, 郭宇, 武哲 (1938)
调频引信粗糙面目标与干扰信号识别	郝新红, 杜涵宇, 陈齐乐 (1946)
一种电磁定位系统工作空间拓展方法	郑莉芳, 万元宇, 关少亚, 孙凯, 孟偲, 贾佳 (1956)
基于编解码双路卷积神经网络的视觉自定位方法	贾瑞明, 刘圣杰, 李锦涛, 王赞豪, 潘海侠 (1965)
基于原始对偶内点法的 EST 图像重建	薛倩, 刘婧, 马敏, 王化祥 (1973)
基于自适应遗传算法的 MEMS 加速度计快速标定方法	高爽, 张若愚 (1982)
含函数型自变量回归模型中的变量选择	刘科生, 王思洋 (1990)
液体火箭发动机故障诊断器设计及其 HIL 验证	赵万里, 郭迎清, 杨菁, 薛薇, 武小平 (1995)
基于 Gram-Schmidt 变换的有监督变量聚类	刘瑞平, 王惠文, 王珊珊 (2003)
PMMWI 与 VI 优势互补的人体隐蔽违禁物检测与定位	赵国, 秦世引 (2011)
非结构重叠网格显式装配算法	宣传伟, 韩景龙 (2026)
阵列射流冲击复合不同肋化表面的沸腾特性	张添, 张畅, 谢荣建, 董德平 (2035)
一种基于 HXDSP 的移位器查找表技术	叶鸿, 顾乃杰, 林传文, 张孝慈, 陈瑞 (2044)
基于地面基站的定位系统构建和方案	耿珂, 黄智刚, 苏雨, 石培辰, 高强, 熊华钢 (2051)
边条/鸭翼对前掠翼和后掠翼气动特性的影响	张冬, 陈勇, 胡孟权, 付向恒 (2058)
超低信噪比调频连续波引信信号小周期态 Duffing 振子检测	朱志强, 侯健, 闫晓鹏, 栗革, 郝新红 (2069)
一种区域参考大气密度的建模与应用方法	刘一博, 沈作军, 张向宇 (2079)
基于异步卷积分解与分流结构的单阶段检测器	赵蓬辉, 孟春宁, 常胜江 (2089)
自研激光雷达三维点云配准技术	呼延嘉玥, 徐立军, 李小路 (2099)
基于飞行数据的无人机平飞动作质量评价模型	滕怀亮, 李本威, 高永, 杨栋, 张赞 (2108)
基于非等维状态的 IMM 混合估计方法	欧能杰, 汪圣利, 张直 (2115)

CONTENTS

Research progress of aerothermoelasticity of air-breathing hypersonic vehicles
..... YANG Chao, ZHAO Huangda, WU Zhigang (1911)

Effect of solar radiation on thermal comfort in civil aircraft cabin
..... PANG Liping, LI Heng, WANG Tianbo, FAN Jun, ZOU Lingyu (1924)

In-situ forming of lunar regolith simulant via selective laser melting
..... LI Wen, XU Kening, HUANG Yong, HU Wenying, WANG Daokuan, YAO Siqi (1931)

Structured simulation platform architecture for fighter cloud operations
..... TIAN Yongliang, WANG Yongqing, XIONG Peisen, GUO Yu, WU Zhe (1938)

Rough surface target and jamming signal recognition of FM fuze
..... HAO Xinhong, DU Hanyu, CHEN Qile (1946)

A method for expanding workspace of electromagnetic tracking system
..... ZHENG Lifang, WAN Yuanyu, GUAN Shaoya, SUN Kai, MENG Cai, JIA Jia (1956)

A visual localization method based on encoder-decoder dual-stream CNN
..... JIA Ruiming, LIU Shengjie, LI Jintao, WANG Yunhao, PAN Haixia (1965)

EST image reconstruction based on primal dual interior point algorithm
..... XUE Qian, LIU Jing, MA Min, WANG Huaxiang (1973)

Rapid calibration method of MEMS accelerometer based on adaptive GA
..... GAO Shuang, ZHANG Ruoyu (1982)

Variable selection in regression models including functional data predictors
..... LIU Kesheng, WANG Siyang (1990)

Design of liquid rocket engine fault diagnosis device and its HIL verification
..... ZHAO Wanli, GUO Yingqing, YANG Jing, XUE Wei, WU Xiaoping (1995)

Supervised clustering of variables based on Gram-Schmidt transformation
..... LIU Ruiping, WANG Huiwen, WANG Shanshan (2003)

Detection and localization of concealed forbidden objects on human body based on complementary advantages of PMMWI and VI
..... ZHAO Guo, QIN Shiyin (2011)

Explicit assembly algorithm of unstructured overset grid
..... XUAN Chuanwei, HAN Jinglong (2026)

Boiling characteristics of array jet impingement with various pin-finned surfaces
..... ZHANG Tian, ZHANG Chang, XIE Rongjian, DONG Deping (2035)

A shifter look-up table technique based on HXDSP
..... YE Hong, GU Naijie, LIN Chuanwen, ZHANG Xiaoci, CHEN Rui (2044)

Positioning system construction and scheme based on ground base station
..... GENG Ke, HUANG Zhigang, SU Yu, SHI Peichen, GAO Qiang, XIONG Huagang (2051)

Effect of strake and canard on aerodynamic characteristics of forward-swept wing and back-swept wing
..... ZHANG Dong, CHEN Yong, HU Mengquan, FU Xiangheng (2058)

Small-scale periodic state Duffing oscillator FMCW fuze signal detection at ultra-low SNR
..... ZHU Zhiqiang, HOU Jian, YAN Xiaoping, LI Ping, HAO Xinhong (2069)

A method for range reference atmospheric density modeling and application
..... LIU Yibo, SHEN Zuojun, ZHANG Xiangyu (2079)

Single shot multibox detector based on asynchronous convolution factorization and shunt structure
..... ZHAO Penghui, MENG Chunming, CHANG Shengjiang (2089)

Three-dimensional point cloud registration technique for self-designed LiDAR
..... HUYAN Jiayue, XU Lijun, LI Xiaolu (2099)

Quality evaluation model of unmanned aerial vehicle's horizontal flight maneuver based on flight data
..... TENG Huailiang, LI Benwei, GAO Yong, YANG Dong, ZHANG Yun (2108)

IMM mixing estimation method based on unequal dimension states
..... OU Nengjie, WANG Shengli, ZHANG Zhi (2115)

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0120

吸气式高超声速飞行器热气动弹性研究进展



杨超^{1,2}, 赵黄达^{1,2}, 吴志刚^{1,2,*}

(1. 北京航空航天大学 航空科学与工程学院, 北京 100083;

2. 航空器先进设计技术工业和信息化部重点实验室, 北京 100083)

摘 要: 吸气式高超声速飞行器是当前航空航天领域研究的热点, 该类飞行器通常使用超燃冲压发动机作为推进系统, 并采用一体化设计方案, 带来了一系列的气动弹性问题。首先阐述了吸气式高超声速飞行器机体/发动机一体化建模研究进展; 随后介绍了热气动弹性/推进耦合、控制系统耦合以及不确定性分析等方面的热气动弹性动力学研究进展, 并对相关热气动弹性试验研究进行了分析; 最后对吸气式高超声速飞行器的热气动弹性问题提出了若干研究建议。

关 键 词: 气动弹性; 一体化设计; 超燃冲压发动机; 吸气式高超声速飞行器; 控制系统耦合

中图分类号: V211.47

文献标识码: A

文章编号: 1001-5965(2019)10-1911-13

吸气式高超声速飞行器飞行马赫数大于 5, 普遍采用超燃冲压发动机作为推进系统, 由于不需要携带氧化剂, 飞行器的航程增大, 入轨更加灵活^[1]。吸气式高超声速飞行器同其他高超声速飞行器一样, 同样存在着热颤振、气动伺服弹性稳定性、壁板颤振等相关气动弹性问题。

笔者所在团队曾在 2010 年对高超声速飞行器的气动弹性力学问题进行过综述^[2], 并对气动推进/气动弹性耦合问题进行了一定的介绍。但近十年来, 吸气式高超声速飞行器的热气动弹性问题得到了越来越多的关注, 随着研究的推进, 许多新的认识也不断产生。

吸气式高超声速飞行器通常采用机体/发动机一体化构型设计, 这种构型存在着多个学科间的交叉与耦合, 形成了气动-伺服-热-弹性-推力 (ASTEP)^[3] 综合问题。吸气式高超声速飞行器, 作为高超声速飞行器的一种, 除了面对传统的热气动弹性问题, 由于特殊的推进系统的引入, 飞行

器系统产生的特殊而复杂的气动弹性问题需要引起关注。其气动弹性问题的特殊性和挑战性主要体现在以下方面:

1) 机体/发动机一体化构型带来的动力学耦合问题

吸气式高超声速飞行器的前体下表面通常作为发动机的预压缩面, 通过数道激波达到超燃冲压发动机的工作条件; 后体下表面作为发动机喷管的一部分, 产生推力以及附加升力。这样的构型, 使得机体及推进系统产生控制以及飞行力学、气动弹性动力学等方面的耦合问题。刚体与弹性振动引起的飞行迎角的变化会影响发动机入口条件, 从而导致发动机推力的敏感变化, 推力反过来影响飞行器的姿态控制、气动弹性特性, 造成复杂的动力学耦合问题。

2) 刚弹耦合的动力学问题

吸气式高超声速飞行器通常采用细长体构型, 为了满足高超声速巡航要求, 往往需要较大的

收稿日期: 2019-03-22; 录用日期: 2019-05-27; 网络出版时间: 2019-07-01 16:33

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190701.0849.001.html

* 通信作者: E-mail: wuzhigang@buaa.edu.cn

引用格式: 杨超, 赵黄达, 吴志刚. 吸气式高超声速飞行器热气动弹性研究进展[J]. 北京航空航天大学学报, 2019, 45(10): 1911-1923. YANG C, ZHAO H D, WU Z G. Research progress of aerothermoelasticity of air-breathing hypersonic vehicles[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1911-1923 (in Chinese).

燃油质量系数。此外,该类飞行器多使用轻质薄壁结构,并面临高速飞行带来的气动加热效应。这几个方面的因素共同导致了机体固有频率的降低,且具有较大的频率变化范围。机体的低阶弹性模态频率将更加接近其刚体飞行动力学频率,系统刚弹耦合问题越发严重。

3) 气动弹性强不确定性问题

吸气式高超声速飞行器面临的宽马赫数飞行条件、复杂气动/热环境、动力学/控制耦合等问题,给全机系统分析带来了许多困难,对控制系统的设计提出了新的挑战。推进系统涉及复杂流体力学、燃烧动力学问题,本身同样存在不确定性问题,其引入更是增加了问题的复杂度,因此需要有效的飞行控制方法及不确定分析方法保证飞行器的稳定性及飞行品质。

针对上述吸气式高超声速飞行器的新特点,本文着重介绍了该类飞行器的建模及气动弹性动力学问题的研究进展,希望能为相关研究人员提供一定的参考。

1 吸气式高超声速飞行器一体化建模研究进展

1.1 适合全机动力学耦合分析的超燃冲压发动机建模进展

从该类飞行器的特点可以看出,吸气式高超声速飞行器性能的准确估计不仅需要考虑机体的相关动力学特征,推进系统同样是关键部分,有必要对超燃冲压发动机建模进行单独讨论。超燃冲压发动机本身存在一系列气动弹性问题,包括壁板热颤振、激波振荡等问题,受篇幅限制,本节不讨论发动机的气动弹性问题研究进展,而是着重总结气动弹性问题研究中发动机的建模研究工作。

超燃冲压发动机模型可以分为解析模型和数值模型,解析模型通过引入各类假设条件得到简单的参数表达式,在动力学建模时更具有优势。数值模型则保留了更多发动机特性,可以满足工程精度需要。

Chavez 和 Schmidt^[4]先建立了一个低精度超燃冲压发动机模型,见图 1。该模型包括扩散段、燃烧室以及内喷管三部分。建模中假设扩压段和喷管是等熵流动,通过总温的增加来模拟燃烧过程,可以给出推力随马赫数、高度、迎角、燃油当量比的变化关系式。该模型被广泛地应用于各类吸气式高超声速飞行器建模及仿真研究,为后续工

作打下了基础。国内唐硕等^[5]基于一维流理论,建立了类似的适用于飞行动力学分析的刚体推进系统模型,并讨论了飞行参数对于推进系统性能的影响规律。

在一维模型基础上,MASIV (Michigan-AFRL Scramjet In Vehicle)、SRGULL 等二维分析程序考虑了更多影响因素,对超燃冲压发动机进行了更高精度的建模仿真。

MASIV 发动机分析程序^[6-9]考虑了:①隔离段的预燃烧激波串效应;②气体解离和复合反应导致的总压损失;③壁面传热及摩擦效应;④燃料空气混合情况;⑤有限速率燃烧和自点火反应机理。通过求解一系列微分方程组,得到发动机各状态量的分布情况。MASIV 程序建立的超燃冲压发动机模型示意图见图 2。

Torrez 等^[8]为了验证 MASIV 程序的有效性进行了试验研究,试验装置示意图以及计算程序使用的简化模型见图 3,试验结果与计算得到的

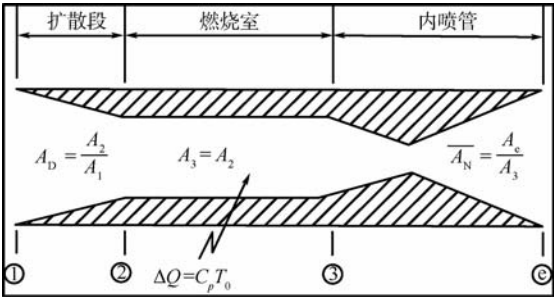


图 1 超燃冲压发动机解析模型^[4]

Fig. 1 Scramjet engine analytical model^[4]

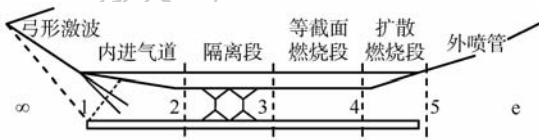


图 2 MASIV 程序使用的发动机模型^[6]

Fig. 2 Engine model used in MASIV^[6]

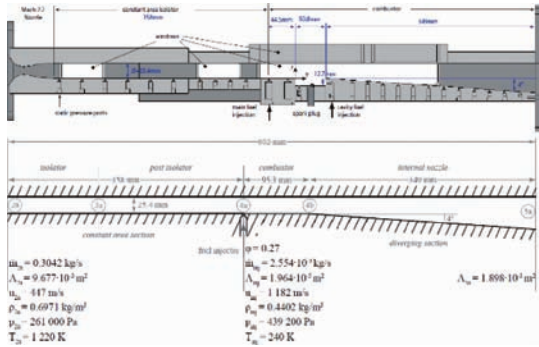


图 3 试验装置示意图及 MASIV 程序计算模型^[8]

Fig. 3 Experimental device schematic and MASIV calculation model^[8]

压力分布的对比曲线见图 4。从结果可以看出,该程序与试验结果趋势相符,但存在一定的误差,需要说明的是,计算中使用的常数项参数并未进行修正,在一定程度上验证了程序的通用性。MASIV 在精度及效率上都可以满足一定的工程需求,已在气动弹性问题的研究上得到了一定的应用^[10-11]。

SRGULL^[12]程序通过求解二维欧拉无黏方程得到前体/进气道流动特性,利用边界层方程预测前体阻力及发动机入口参数,加上一维的燃烧室模型,共同分析整个推进系统流道。该程序可以考虑层流、过渡湍流边界层的影响,得到发动机流道的升力、推力、力矩,并可以模拟发动机不启动现象^[13],同样得到了一定的应用^[14]。

更高精度的三维发动机建模往往需要耗费巨大的计算时间,通过 CFD 方法以及燃烧相关软件 CHEMKIN^[15]等详细分析得到流场情况,并不适用于最初的设计阶段,此处不做过多介绍。

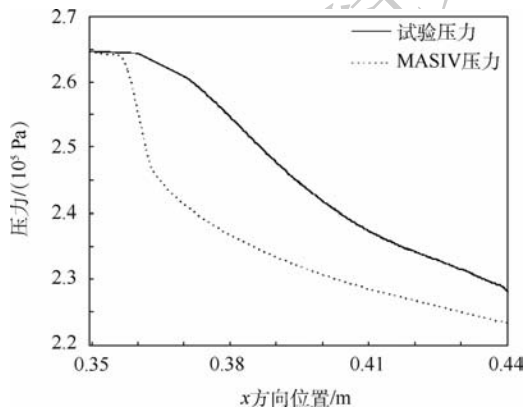


图 4 压力分布结果对比^[8]

Fig. 4 Comparison of pressure distribution results^[8]

1.2 全飞行器一体化动力学建模

为了得到全机的飞行动力学、气动弹性动力学模型,在发动机建模的基础上,还需要考虑气动、气动热、刚性/弹性结构以及飞行动力学建模。

对于高超声速飞行器单纯的气动力及气动热的建模问题,国内外许多学者进行了研究及总结^[16],这部分并非本文的重点。目前高超热气动弹性领域常用的定常/非定常气动力及气动热计算方法,主要包括可与结构动力学耦合的 CFD 方法以及工程计算方法。而气动弹性中常用的工程方法,主要包括细长体理论、活塞理论^[17]、当地活塞流方法^[18]、牛顿冲击理论^[19]等气动力计算方法,以及 Eckert 参考焓^[20]、Spalding-Chi^[21]等气动热计算方法。此外,对于吸气式高超声速飞行器,在气动力/热建模过程中,还需要对发动机的燃烧

动力学问题进行一定的处理。

本节根据研究对象的不同,分别介绍二维及三维全飞行器一体化动力学建模情况。

1) 二维模型

Chavez 和 Schmidt^[4,22]建立了如图 5 所示的二维吸气式高超声速飞行器模型并进行了纵向动力学分析,模型建立过程中使用工程算法得到气动力,引入中间铰接梁模型来模拟弹性振动,如图 6 所示,随后推导得到了控制导数解析表达式,该经典模型至今被许多国内外学者使用。

Bolender 和 Doman^[23]对上述模型进行了拓展,在气动力的计算中,使用斜激波/Prandtl-Meyer 方法代替了原来的牛顿冲击方法。此外,添加了一个可平移的整流罩,用于在非设计条件下保持激波打在唇口状态,其几何模型见图 7。

Williams 等^[24]重点考虑了飞行器质量变化及

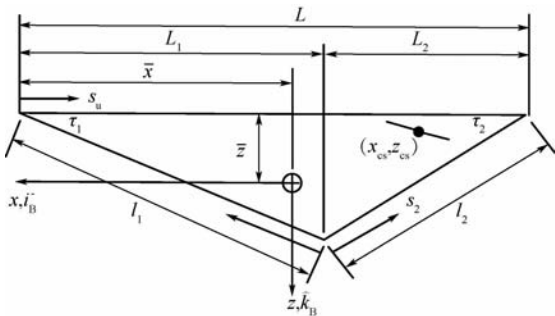


图 5 飞行器几何外形^[4]

Fig. 5 Vehicle geometry^[4]

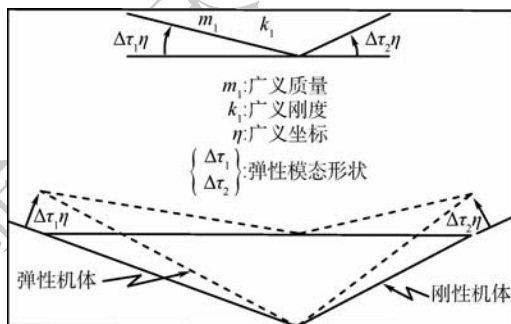


图 6 弹性机体模型^[4]

Fig. 6 Elastic vehicle model^[4]

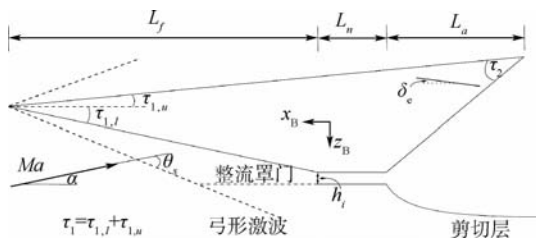


图 7 吸气式高超声速飞行器几何模型^[23]

Fig. 7 Geometric model of air-breathing hypersonic vehicle^[23]

结构温度变化对整个飞行轨迹的影响,模型的内部体积示意图见图8,计算结果表明,飞行器模态的振型及频率主要由质量变化和加热效应主导。图9给出了飞行过程中第一阶模态演变情况,图中 slugs 为英制质量单位,1 ft = 30.48 cm。随后的动图中力学计算发现,飞行器的短周期零极点随着燃料的消耗而移动,对系统的带宽产生了一定的限制,在建模过程中需要考虑上述2个因素的影响。

华如豪和叶正寅^[25]建立了一个与图8类似的吸气式高超声速飞行器多学科耦合动力学模型,其中气动力及气动热分别使用工程方法得到,机体部分使用变截面梁来模拟,并考虑了燃油消耗的影响。随后在此模型基础上进行了静气动弹性配平以及动力学特征分析。

综上所述,二维模型的建模过程中,飞行器的几何外形多基于 Chavez 的构型,并添加不同的改进措施如鸭翼、发动机唇口控制、高精度发动机模型等;弹性的引入主要通过假设模态法,并根据考虑问题的不同进行不同精度的建模。

Sudalagunta 等^[26]建立了面向控制的吸气式高超声速飞行器气动弹性模型,见图11。对机体的刚性截面,使用6个独立位移来表示截面轴向、弯曲、剪切和扭转载荷引起的变形,推导出了弹性飞机的非线性运动方程。

在最近的文献中,Sachan 和 Padhi^[27]对吸气式高超声速飞行器6自由度建模进行了简要的总结,并介绍了针对这些模型不确定性的鲁棒控制设计方法,提出了控制及建模未来可能的发展方向。

国内李惠峰等^[28]使用三维外形参数化建模方法得到了飞行器几何模型,见图12。随后采用工程预估方法得到了气动力分布。发动机燃烧室的建模使用了简化的 LH₂ 燃烧模型。推导了刚体/弹性耦合矢量方程并使用拉格朗日方法建

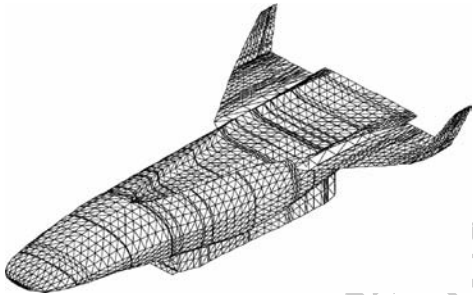


图10 高超声速飞行器等轴视图^[11]

Fig. 10 Isometric view of hypersonic vehicle^[11]

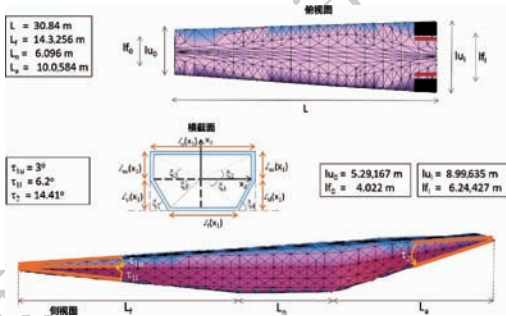


图11 典型的吸气式高超声速飞行器三视图 (非比例绘图)^[26]

Fig. 11 Three-sided view of a typical air-breathing hypersonic vehicle (figures not to scale)^[26]

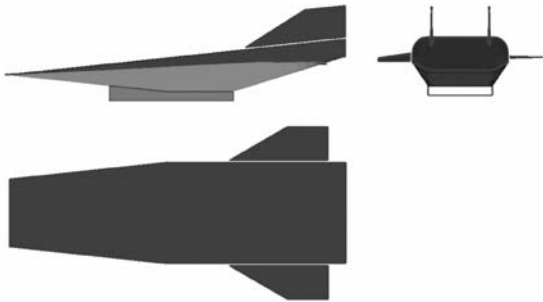


图12 吸气式高超声速飞行器三维视图^[28]

Fig. 12 Air-breathing hypersonic vehicles 3D view^[28]

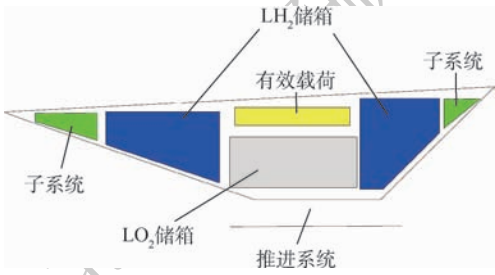


图8 内部体积分布示意图^[24]

Fig. 8 Internal volume layout^[24]

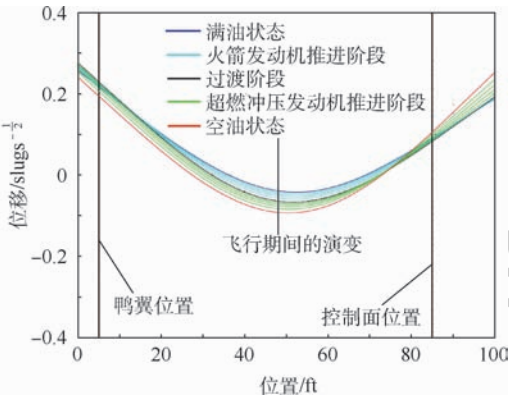


图9 飞行中第一阶模态振型演变情况^[24]

Fig. 9 Evolution of first-order mode shape during flight^[24]

2) 三维模型

Dalle 等^[11]建立了三维6自由度刚体模型,见图10。建模过程分为机体及发动机两部分,在保证计算效率的同时保留了模型的高阶特征,可以很好地反映出机体/发动机的非线性耦合特征。

立了全机动力学模型。肖进等^[29]同样使用参数化建模方法构建了三维模型,不同的是,气动力及气动热通过 CFD 方法得到,并进行了配平分析,具有一定的工程指导价值。

2 吸气式高超声速飞行器热气动弹性动力学问题研究进展

吸气式高超声速飞行器由于涉及 ASTEP 多学科问题,因此不同问题的侧重点有所不同。本节从以下几个方面介绍相关问题的研究进展。

2.1 热气动弹性/推进耦合问题

Mcnamara 和 Friedmann^[3]对高超声速气动弹性和热气动弹性问题进行了总结,从问题特征、建模方法、试验研究、热气动弹性相似律、壁板颤振、机翼和全机数值计算几个方面详细介绍了高超声速飞行器相关研究进展,随后指出了 2 个新的研究方向,其中之一便是热气动伺服弹性/推进耦合问题。文献[3]对吸气式高超声速飞行器部分的关注主要集中在动力学控制问题以及机体推进一体化多学科分析优化问题。

Chavez 和 Schmidt^[4,22]对某二维模型进行了气动推进/气动弹性分析,建模过程参考第 1 节。分析时,首先计算了该模型的稳定性导数及控制导数,数值算例的结果表明,在某些情况下,发动机部分的影响可达到气动力的 50%,且对全机稳定性导数造成不稳定贡献。随后的动力学模态分析表明,该模型在俯仰方向高度不稳定,具有显著的俯仰/弯曲耦合效应。飞行器气动力 10% 的变化将使其短周期频率发生 34% 的变化、弹性模态频率改变 14%、弹性模态阻尼变化 50%。同样,推力特性 10% 的变化将使其短周期频率发生 11% 的变化。证明了气动推进及气动弹性效应对全机动态特性都有显著的影响,因此需要整体建模方法。Schmidt^[30]同样使用上述模型进行了纵向动力学特性分析,值得注意的是,其通过计算表明机身的俯仰运动将显著干扰燃烧室入口条件,且这些扰动的带宽与弹性固有频率一样高,需要对此进行有效的控制。

Raney 等^[14]对三维 X-30 类模型引入了七阶弹性弯曲模态加两阶刚体模态(俯仰及沉浮)进行分析,模型的模态及频率见图 13。气动力计算使用牛顿冲击理论和二阶 Van Dyke 活塞理论的结合,发动机部分使用 SRGULL 代码。文献[14]研究了迎角和机体变形对于推进系统的影响,其中法向力随迎角的变化见图 14,图中 1 lb = 453.592 37 g,

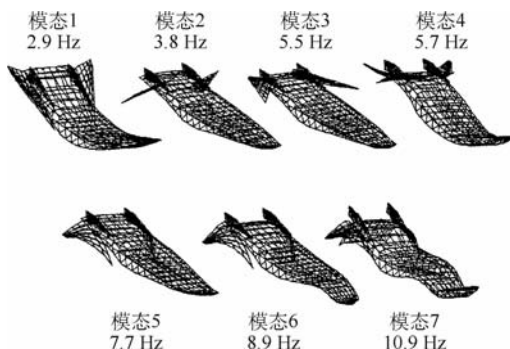


图 13 气动弹性模型中使用的结构模态振型及真空下频率^[14]

Fig. 13 Structural mode shapes and in-vacuo frequencies used in aeroelastic model^[14]

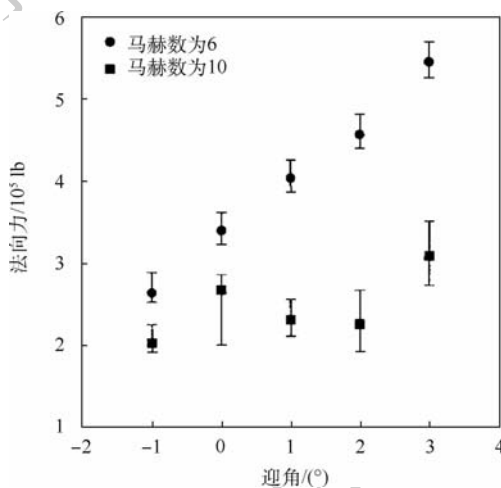


图 14 法向力随迎角变化^[14]

Fig. 14 Normal force variation with angle of attack^[14]

实心符号表示无几何变形条件下的气动力数据,线段表示由湍流引起的几何变形条件下,气动力的变化范围,具有一定的工程参考价值。

Rudd 和 Pines^[31]研究了高超声速乘波体飞行器的动力学特性,气动力使用有限微分方法,并使用参考温度法计算阻力。结果表明,为了建立有效的控制系统,推进系统必须进行精确的建模。Mirmirani 等^[32]利用高精度 CFD 模型研究了一个全尺寸高超声速飞行器。首先在马赫数 10,高度 30 km 条件下对二维模型进行分析,随后基于线性时变模型进行了控制系统研究^[33]。Bolender 和 Doman^[23,34]建立了二维吸气式高超声速飞行的非线性模型,通过拉格朗日方程得到了弹性机体的运动方程,分析了线化系统的长短周期特性。

Culler 等^[35]以二维 X-43 为对象进行了气动及结构热力学分析,给出了燃油消耗及气动加热对于该模型模态以及飞行动力学特征的影响,相关结果分别见图 15 和图 16。

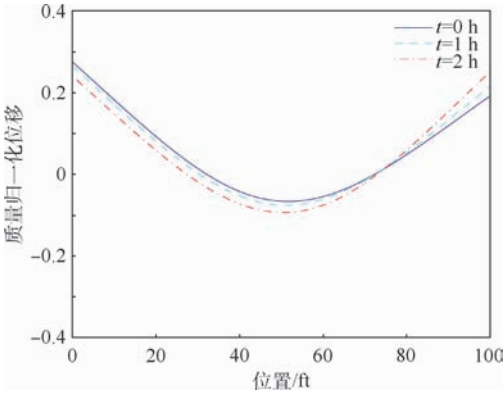


图 15 一阶弯曲模式的演变过程^[35]

Fig. 15 Evolution of first-order bending mode^[35]

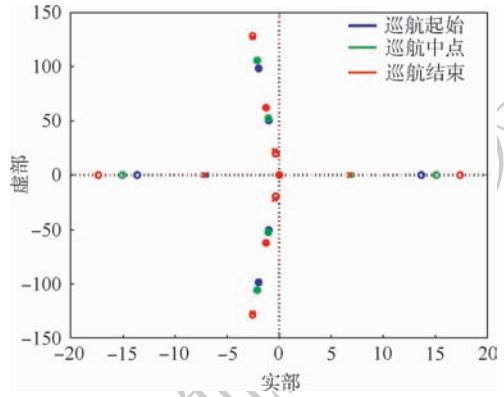


图 16 考虑气动加热和燃料消耗的系统极点/零点演变过程^[35]

Fig. 16 Evolution of system poles/zeros considering aerodynamic heating and fuel consumption^[35]

Klock 和 Cesnik^[10] 研究了吸气式高超声速飞行器热气动弹性仿真方法,建立了用于弹性高超声速飞行器的基于分区的、多物理场、多精度仿真框架,通过数值方法对比了分区求解与整体求解计算结果,两者吻合较好。

国内对热气动弹性/推进耦合问题也进行了一定的调研与研究。作者团队^[2]对高超声速飞行器气动弹性力学研究进行了综述,介绍了气动推进/气动弹性耦合问题,从稳定性分析控制以及不确定性建模 2 个角度指出了需要关注的问题。向锦武等^[36]针对典型吸气式高超声速飞行器构型研究了结构弹性对纵向静态特性和飞行动力学特性的影响。弹性模型采用弹性铰接的两段梁模型,分析了不同机体刚度对飞行配平参数以及动力学特性的影响。结果表明,全飞行包线可配平区域随着刚度下降而减小,见图 17,其中 ω_1 为机体第一阶弯曲频率,用于衡量机体弯曲刚度的大小。图 18 的动态特性分析结果指出,该模型短周期模态稳定性随着机体刚度的降低而下降,指出

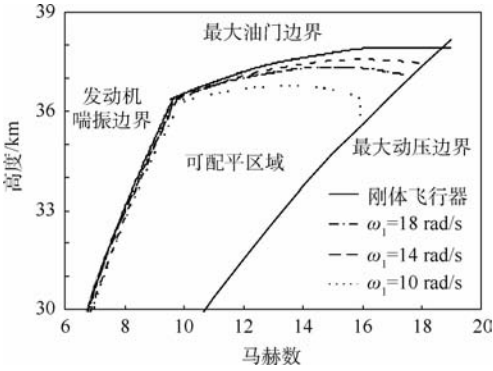


图 17 不同机身弯曲刚度下的飞行包线^[36]

Fig. 17 Flight envelopes under different bending stiffness of fuselage^[36]

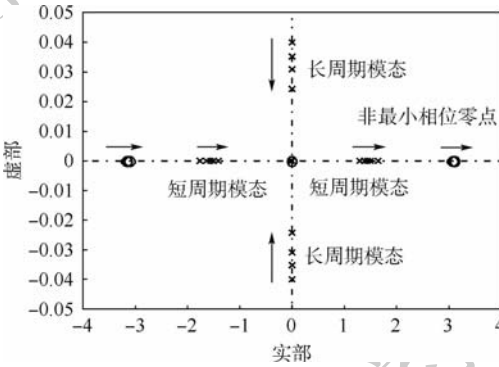


图 18 开环极点随机身刚度减小的变化^[36]

Fig. 18 Open-loop zero-pole variation with stiffness reduction of fuselage^[36]

了将弹性因素引入该类飞行器分析中的必要性。

2.2 气动、结构、控制耦合

飞行控制系统的加入使得吸气式高超声速飞行器同样面临着气动伺服弹性问题,且推进系统的引入导致了全机控制系统设计更加复杂。针对该问题,国内外学者进行了一系列控制方法研究。

Fidan 等^[37]总结对比了相关文献中的吸气式高超声速飞行器控制系统设计方法,包括美国国家航空航天局(NASA)研究中心使用的控制方案。研究表明,吸气式高超声速飞行器控制面临的主要问题包括处理各种耦合效应、非线性、不确定性以及系统参数变化。后续的一个研究^[33]基于时变模型进行了纵向运动学控制。

Parker 等^[38]基于 Bolender 和 Doman^[23]的高精度模型,通过力/力矩拟合,忽略弱耦合项等手段建立了一个面向控制的模型,可用于非线性控制系统的设计。文献[38]随后提出了基于近似反馈线性化方法的控制系统设计方案,仿真结果表明该方法即使在一定的参数扰动情况下,系统仍具有出色的跟踪性能。弹性效应的引入使得控

制系统失效,在前体添加鸭翼可以用来补偿弹性效应的影响。Cui 等^[39]采用相同的建模思路进行研究,采用最小二乘法进行曲线拟合,面向控制的模型保留了原始真实模型的主要特性。

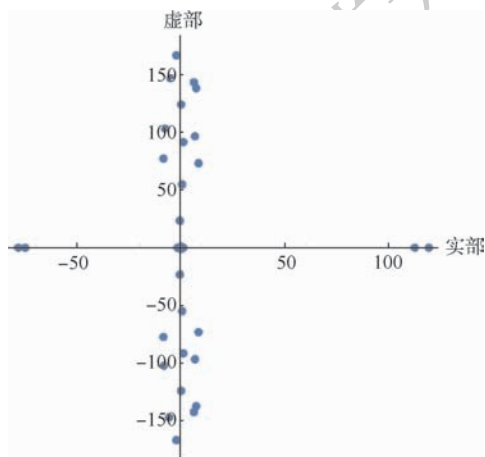
Dickeson 等^[40]开展了针对 X-43 构型的控制相关的建模/分析/设计研究,讨论了发动机及机体参数变化对系统静/动态特性的影响,同时揭示了控制系统设计时这些参数的影响机理。Sudalagunta 等^[26]在建立三维模型基础上,进行了系统开环稳定性分析,随后对线化系统进行了运动模态分析,并设计了线性二次调节器用于补偿初始条件的扰动。图 19 对比了开环和闭环条件下系统特征根分布,计算表明线性二次调节器可以满足系统的稳定性要求,并可通过参数调整使的系统最不稳定特征根更加远离虚轴。

国内 Duan 和 Li^[41]、方洋旺等^[42]分析了高超声速飞行器制导与控制问题的难点,并介绍了典

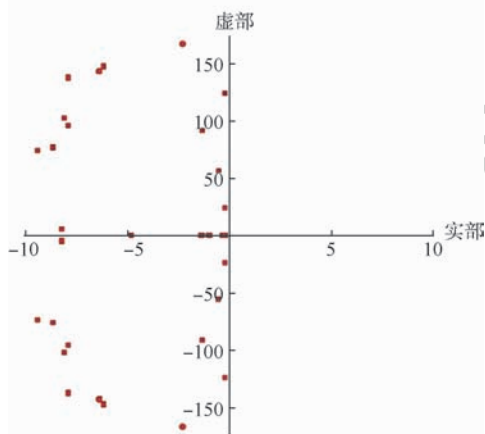
型的线性/非线性控制方法。吴立刚^[43]、王勇^[44]等在此基础上针对吸气式高超声速飞行器的特点总结了控制系统的数学模型。

作者团队^[45]利用二维吸气式高超声速飞行器模型,对机体/推进系统耦合下的气动伺服弹性问题进行了研究。结果表明,推进系统对飞行器的气动伺服弹性特性产生了显著的影响,考虑发动机控制系统的影响时,研究对象的气动伺服弹性稳定裕度下降约 16%,在飞控系统的设计时需要考虑该因素。相关结果见图 20,其中 G 、 F 分别为刚体失稳点以及气动伺服弹性失稳点,升降舵增益 K 取 2.5 时,系统临界失稳。

不同于系统的鲁棒性分析与设计,另一类研究主要针对机体/发动机整体控制^[46-47],Schmidt 和 Velapoldi^[48]将燃油当量比及扩散面积比作为发动机控制输入量,研究表明发动机指令对机体姿态产生了明显的扰动,指出了整体控制的必要性。Hao 等^[49]设计了一种机体/发动机的协调控制方案,该方法在不改变原始机体控制和发动机控制系统结构的基础上,通过引入发动机稳定裕度变量来调整原控制系统的参数,仿真结果表明协调控制方案保持了原系统的稳定性及速度指令跟踪能力,验证了其有效性。



(a) 提出模型的开环特征根



(b) 运用LQR控制后系统的闭环特征根

图 19 复平面中的特征值分布^[26]

Fig. 19 Eigenvalue distribution in complex plane^[26]

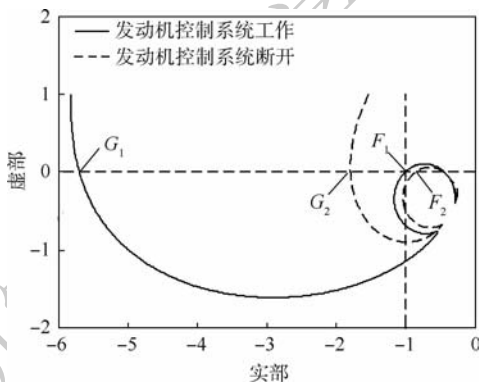


图 20 系统开环传递函数的 Nyquist 曲线对比 ($K = 2.5$)^[45]

Fig. 20 Comparison of system open-loop transfer function Nyquist diagram when $K = 2.5$ ^[45]

2.3 吸气式高超声速飞行器的气动弹性强不确定性

吸气式高超声速飞行器的强不确定性特点,给气动弹性问题的研究带来了很大挑战。对于不确定性分析方法在气动弹性问题中的应用问题,文献^[50]进行了详细的总结。在此基础上,吸气式高超声速飞行器的特殊设计构型带来了许多新的气动弹性不确定性问题,对此,国内外学者开展了一系列研究。

Buschek 和 Calise^[51] 针对一种吸气式高超声速飞行器模型进行了不确定性分析,将推力的变化以及机体的弹性振动分别作为不确定性来源,使用 H_∞ 和 μ 方法分析了控制系统的鲁棒特性。在此基础上对控制系统进行了降阶研究^[52],并将降阶控制系统与全阶情况进行对比,验证了降阶的正确性。

Chavez 和 Schmidt^[53] 建立了非结构不确定性模型以及实参数结构不确定度模型,研究表明结构模态频率和阻尼的估计误差对系统的不确定性贡献较大。是否考虑弹性对系统不确定度产生了明显的影响,因此在控制律的设计时刚体假设可能不再适用。

作者团队^[54] 针对气动伺服弹性系统,采用结构奇异值方法分析了系统的鲁棒稳定性,该方法可以高效地建立弹性飞行器的状态空间方程,具有较强的工程应用价值。曲鑫等^[55] 将吸气式高超声速飞行器不确定性来源归结为飞行器质量变化、压力分布误差等4个方面,进行了不确定性建模并设计了鲁棒协调控制器,仿真结果验证了该控制器的效果。曾开春和向锦武^[56] 在建立动力学模型的基础上,通过基于泰勒展开的区间分析方法(TIAM)和基于多项式逼近的区间分析方法(CIAM)研究了飞行动力学不确定性,并与直接蒙特卡罗(DMC)模拟方法进行对比。图 21 给出了 2 种计算方法得到的特征根不确定范围,结果表明,控制系统改变了不确定边界的形状及大小,同时,由于部分 DMC 模拟点落在了 TIAM 边界之外而仍在 CIAM 内,说明 CIAM 在进行动力学不确定行分析时更加准确保守。

Pu 等^[57] 从 4 个方面分别研究了吸气式高超声速飞行器模型的不确定性:弹性效应、气动参数、外部环境扰动以及建模误差。提出了鲁棒轨迹线性化控制(TLC)方案,仿真研究验证了鲁棒设计方案的良好控制性能和抗不确定性干扰能力。

2.4 相关试验研究进展

对于考虑吸气式超燃冲压发动机的气动弹性风洞试验无法也不大可能实现,目前考虑真实高空条件下的地面无发动机的高超热气弹试验也很难实现。

对于某些特定情况下的高超声速气动弹性风洞试验研究,美国开展的较早并取得了一些的成果^[58-59]。国内目前在高超速风洞进行的气动弹性试验较少。季辰等^[60-61] 开展了针对舵、翼面的高超声速颤振试验研究,试验在中国航天空气动

力技术研究院 FD-07 高超声速风洞中进行。试验装置见图 22。该试验马赫数为 5~6,通过应变片测量部件的振动信号,随后进行结构模态参数辨识以及亚临界颤振边界预测。

南京航空航天大学及合肥工业大学^[62-63] 共同开展了高超声速舵面的颤振试验,试验装置见图 23。试验对象在 29.5 kPa 临界动压下发生了弯扭耦合颤振,颤振频率为 29.3 Hz。数值仿真结果与该试验结果具有很好的一致性,并验证了模型限幅装置的安全性及有效性。

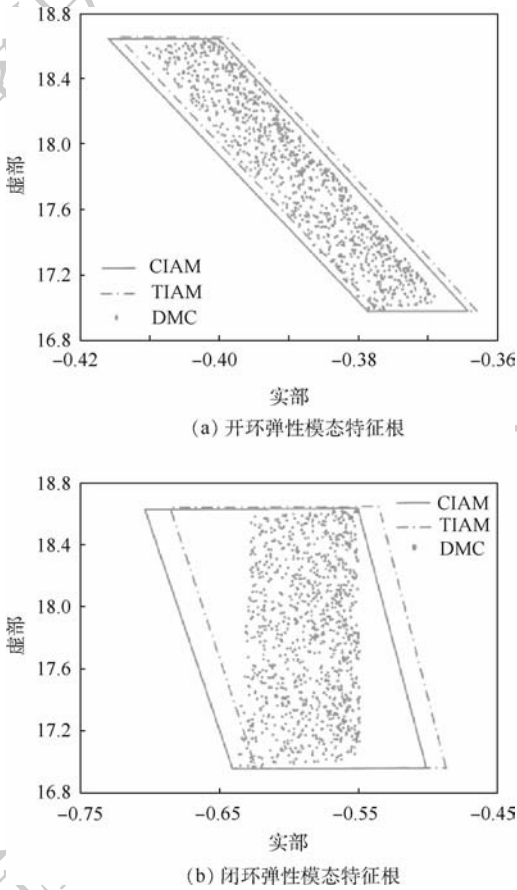


图 21 纵向动力学模态复值特征根不确定范围^[56]
Fig. 21 Complex eigenvalue uncertainty range of longitudinal dynamic modes^[56]



图 22 高超声速风洞中的翼面及颤振试验装置^[60]
Fig. 22 Wing model and flutter testing apparatus in hypersonic wind tunnel^[60]

受限于试验设备及实现难度,目前国内外尚未有针对吸气式高超声速飞行器模型完整的气动弹性试验,现有的试验主要为吸气式高超声速飞行器模型的气动力^[64-67]、气动热试验^[68]以及超燃冲压发动机相关试验^[69]。

为了解决高速风洞气动弹性试验费用高、模型设备要求高等问题,国内外学者开展了许多关于半物理地面模拟加载气动力试验的研究。

目前模拟非定常气动力的加载方式主要为激振器离散点加载。作者团队^[70]首先以细长体导弹为研究对象,对其气动伺服弹性问题进行了半物理地面模拟试验。试验装置见图 24。在该装置基础上进行了系统开环响应以及闭环稳定性试验,数值仿真与实测数据一致性较好。

地面模拟试验的一个重要问题是如何有效使用较少的激振器更准确地模拟真实飞行条件下的气动力,并需要对激振器的位置进行优化选取。西北工业大学胡巍等^[71]对地面试验中的气动力降阶方法进行了研究,使用二次降阶方法有效地减少了操纵面上激振点的个数,但是对于目前的试验能力,文中研究对象的激振器数量仍然偏多。

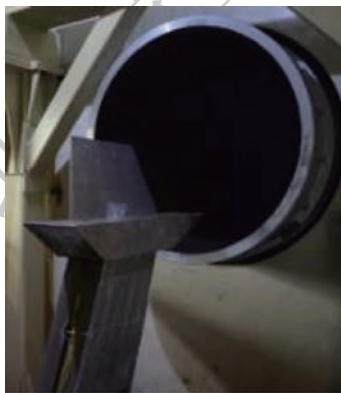


图 23 试验模型与支撑机构^[63]

Fig. 23 Test model and support mechanism^[63]

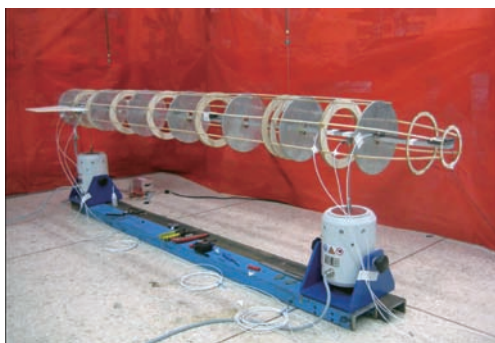


图 24 模拟气动力加载的导弹气动伺服弹性地面试验测试系统^[70]

Fig. 24 Missile aeroservoelastic ground test system with simulated aerodynamic load^[70]

其他的载荷加载方式还包括气囊加载系统^[72]以及非接触式电磁加载^[73]等方式,这些加载方式各有优点,但目前尚未得到广泛的应用。

3 结束语

吸气式高超声速飞行器作为一类特殊类型飞行器,其机体/发动机一体化构型、动力学问题刚弹耦合、强不确定性等特点,给该类飞行器的气动弹性的问题研究带来了新的挑战。本文回顾和总结了超燃冲压发动机及全飞行器动力学的建模方法及应用,在此基础上重点介绍了吸气式高超声速飞行器热气动弹性动力学问题的研究进展。从这些研究工作中得到的建议和结论主要包括:

1) 发动机系统的建模与分析是吸气式高超声速飞行器气动弹性力学研究的重要问题之一。对于目前的发动机分析方法,一维方法精度过低,不能很好地反映发动机特性;三维方法通常需要考虑复杂的燃烧过程,计算量过大,实际应用困难,需要进行降阶处理;二维方法在精度和计算量上适中,比较适用于热气动特弹性力学问题,但应用时需要处理好三维机体模型和二维发动机模型之间不同维度的匹配问题。此外,发动机自身也存在控制系统,可能引入延迟环节,建议在开展热气动伺服弹性研究时考虑该因素的影响。

2) 在研究吸气式高超声速飞行器热气动弹性力学及飞行动力学问题时,采用刚体假设将带来很大的误差,需采用刚弹耦合建模方法。此外,该类飞行器燃料质量系数大,机体频率在整个飞行包线过程中发生较大变化。同时,发动机工作时燃料质量消耗速率大,机体频率具有明显的时变性特点,该特点对热气动弹性问题的影响程度,目前尚不确定,仍需要进一步的研究。机体频率的变化将改变飞行配平参数,并影响系统的运动模态特征根分布,在某些情况下对飞行器的动力学稳定性造成不利影响。吸气式高超声速飞行器机体频率变化范围宽、变化速率大的特点,使得传统的滤波器不再适用,需要重新设计有效的控制系统。

3) 高超声速飞行器面临复杂的飞行环境,不确定性分析将会成为重点问题。对于复杂非线性系统的不确定性问题,需要尽可能分析不确定性来源。在热气动弹性力学问题的分析中,研究人员对于结构和气动载荷的不确定性问题已有一定认识,未来应进一步考虑热不确定性的影响。热问题在时间上是累积过程,在与结构、气动问题共同计算时需要考虑各学科间时间尺度不同的问

题。在空间上,热在不同方向的分布规律通常不同,因此在描述热不确定性问题时,时间和空间上的建模精度,将同时影响结果的准确性。

4) 由于试验周期长、成本高且试验条件有限,目前数值仿真模拟仍然是研究该类飞行器的主要分析方法。常温条件下飞行器的气动特性试验较多,热环境下针对该类构型的试验虽然已经开展,但是仍然存在着许多不足。因此建议继续加强热气弹试验,设计适用于热环境下的测量采集试验装置,开发合理有效的热载荷加载方式,积累试验数据以便更好地理解热气动弹性力学机理;另一方面,地面模拟加载非定常气动力的气动弹性试验由于其成本低、模型尺寸不受限、安全性高等特点,在处理热气动弹性试验问题时具有特有的优势,有望成为一种有效的分析及试飞前验证的重要手段。

5) ASTEP 综合问题是典型的多学科耦合问题,传统的热气动弹性力学处理耦合问题时,多将部分子学科进行解耦处理并/或忽略一些方向的弱耦合,推力通常为多场耦合动力学的外界输入条件。而对于吸气式高超声速飞行器,推进系统与其他学科的耦合将更加严重,往往不能解耦处理,因此需要多学科多场强耦合建模分析与计算方法研究;然而 ASTEP 耦合方法带来计算成本大幅增加,降阶方法可以提高计算效率,并保证一定的精度,因此也是处理多场耦合动力学问题的有效分析途径。

参考文献 (References)

- [1] FLAHERTY K W, ANDREWS K M, LISTON G W. Operability benefits of airbreathing hypersonic propulsion for flexible access to space[J]. *Journal of Spacecraft and Rockets*, 2010, 47(2): 280-287.
- [2] 杨超, 许赞, 谢长川. 高超声速飞行器气动弹性力学研究综述[J]. *航空学报*, 2010, 31(1): 1-11.
YANG C, XU Y, XIE C C. Review of studies on aeroelasticity of hypersonic vehicles[J]. *Acta Aeronautica et Astronautica Sinica*, 2010, 31(1): 1-11 (in Chinese).
- [3] MCNAMARA J J, FRIEDMANN P P. Aeroelastic and aerothermoelastic analysis in hypersonic flow: Past, present, and future[J]. *AIAA Journal*, 2011, 49(6): 1089-1122.
- [4] CHAVEZ F R, SCHMIDT D K. Analytical aeropropulsive-aeroelastic hypersonic-vehicle model with dynamic analysis[J]. *Journal of Guidance, Control, and Dynamics*, 1994, 17(6): 1308-1319.
- [5] 唐硕, 张栋, 祝强军. 吸气式高超声速飞行器推进系统耦合建模与分析[J]. *飞行力学*, 2013, 31(3): 244-249.
TANG S, ZHANG D, ZHU Q J. Coupling model and analysis for air-breathing hypersonic vehicle propulsion system[J]. *Flight Dynamics*, 2013, 31(3): 244-249 (in Chinese).
- [6] TORREZ S, SCHOLTEN N, MICKA D, et al. A scramjet engine model including effects of precombustion shocks and dissociation: AIAA-2008-4619[R]. Reston, VA: AIAA, 2008.
- [7] TORREZ S, DRISCOLL J, BOLENDER M, et al. Effects of improved propulsion modelling on the flight dynamics of hypersonic vehicles: AIAA-2008-6386[R]. Reston, VA: AIAA, 2008.
- [8] TORREZ S, DRISCOLL J, DALLE D, et al. Scramjet engine model MASIV: Role of mixing, chemistry and wave interaction: AIAA-2009-4939[R]. Reston, VA: AIAA, 2009.
- [9] TORREZ S, DRISCOLL J, DALLE D, et al. Hypersonic vehicle thrust sensitivity to angle of attack and mach number: AIAA-2009-6152[R]. Reston, VA: AIAA, 2009.
- [10] KLOCK R, CESNIK C E. Aerothermoelastic simulation of air-breathing hypersonic vehicles[C]//55th AIAA/ASME/ASCE/AHS/SC Structures, Structural Dynamics, and Materials Conference. Reston, VA: AIAA, 2014: 149.
- [11] DALLE D, FRENDEIS S, DRISCOLL J, et al. Hypersonic vehicle flight dynamics with coupled aerodynamic and reduced-order propulsive models: AIAA-2010-7930[R]. Reston, VA: AIAA, 2010.
- [12] PINCKNEY S Z, WALTON J T. Program SRGULL: An advanced engineering model for the prediction of airframe-integrated subsonic/supersonic hydrogen combustion ramjet cycle performance: Nasp Tm-1120[R]. Washington, D. C.: NASA, 1991.
- [13] LOCKWOOD M, PETLEY D, HUNT J, et al. Airbreathing hypersonic vehicle design and analysis methods[C]//34th Aerospace Sciences Meeting and Exhibit. Reston, VA: AIAA, 1996: 381.
- [14] RANEY D L, MCMINN J D, POTOTZKY A S. Impact of aeroelastic-propulsive interactions on flight dynamics of a hypersonic vehicle[J]. *Journal of Aircraft*, 1995, 32(2): 355-362.
- [15] KEE R J, RUPLEY F M, MEEKS E, et al. CHEMKIN-III: A FORTRAN chemical kinetics package for the analysis of gas-phase chemical and plasma kinetics: SAND 96-8216[R]. Livermore, CA: Sandia National Laboratories, 1996.
- [16] 叶正寅, 孟宪宗, 刘成, 等. 高超声速飞行器气动弹性的近期进展与发展展望[J]. *空气动力学学报*, 2018, 36(6): 984-994.
YE Z Y, MENG X Z, LIU C, et al. Progress and prospects on aeroelasticity of hypersonic vehicles[J]. *Acta Aerodynamica Sinica*, 2018, 36(6): 984-994 (in Chinese).
- [17] ASHLEY H. Piston theory—A new aerodynamic tool for the aeroelastician[J]. *Journal of the Aeronautical Sciences*, 1956, 23(12): 1109-1118.
- [18] 张伟伟, 叶正寅. 基于当地流活塞理论的气动弹性计算方法研究[J]. *力学学报*, 2005, 37(5): 632-639.
ZHANG W W, YE Z Y. Numerical method of aeroelasticity based on local piston theory[J]. *Chinese Journal of Theoretical and Applied Mechanics*, 2005, 37(5): 632-639 (in Chinese).
- [19] YATES JR E C, BENNETT R M. Analysis of supersonic-hypersonic flutter of lifting surfaces at angle of attack[J]. *Journal of Aircraft*, 1972, 9(7): 481-489.
- [20] ECKERT E. Engineering relations for heat transfer and friction

- in high-velocity laminar and turbulent boundary-layer flow over surfaces with constant pressure and temperature[J]. Transactions of the ASME, 1956, 78(6): 1273-1283.
- [21] SPALDING D, CHI S W. The drag of a compressible turbulent boundary layer on a smooth flat plate with and without heat transfer[J]. Journal of Fluid Mechanics, 1964, 18(1): 117-143.
- [22] CHAVEZ F, SCHMIDT D. An integrated analytical aeropropulsive/aeroelastic model for the dynamic analysis of hypersonic vehicles; AIAA-92-4567-CP[R]. Reston, VA: AIAA, 1992.
- [23] BOLENDER M A, DOMAN D B. A non-linear model for the longitudinal dynamics of a hypersonic air-breathing vehicle; AIAA-2005-6255[R]. Reston, VA: AIAA, 2005.
- [24] WILLIAMS T, BOLENDER M, DOMAN D, et al. An aerothermal flexible mode analysis of a hypersonic vehicle; AIAA-2006-6647[R]. Reston, VA: AIAA, 2006.
- [25] 华如豪, 叶正寅. 吸气式高超声速飞行器多学科动力学建模[J]. 航空学报, 2015, 36(1): 346-356.
HUA R H, YE Z Y. Multidisciplinary dynamics modeling and analysis of a generic hypersonic vehicle[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(1): 346-356 (in Chinese).
- [26] SUDALAGUNTA P R, SULTAN C, KAPANIA R K, et al. Aeroelastic control-oriented modeling of an airbreathing hypersonic vehicle[J]. Journal of Guidance, Control, and Dynamics, 2018, 41(5): 1136-1149.
- [27] SACHAN K, PADHI R. A brief survey on six-degree-of-freedom modeling for air-breathing hypersonic vehicles[C]//5th IFAC Conference on Advances in Control and Optimization of Dynamical Systems ACODS 2018. Schlossplotz; IFAC, 2018, 51(1): 492-497.
- [28] 李惠峰, 薛松柏, 张冉. 吸气式高超声速飞行器机体推进控制一体化建模方法研究[J]. 宇航学报, 2012, 33(9): 1185-1194.
LI H F, XUE S B, ZHANG R. Research of modeling method based on integrated design for airframe-propulsion-control of AHV[J]. Journal of Astronautics, 2012, 33(9): 1185-1194 (in Chinese).
- [29] 肖进, 完颜振海, 杨亮, 等. 面向控制的高超声速飞行器一体化建模与分析[J]. 计算机仿真, 2017, 34(2): 92-96.
XIAO J, WANYAN Z H, YANG L, et al. Control-oriented integrated modeling and analysis of hypersonic vehicle[J]. Computer Simulation, 2017, 34(2): 92-96 (in Chinese).
- [30] SCHMIDT D. Dynamics and control of hypersonic aeropropulsive/aeroelastic vehicles; AIAA-92-4326-CP[R]. Reston, VA: AIAA, 1992.
- [31] RUDD L, PINES D. Integrated propulsion effects on dynamic stability and control of hypersonic vehicles; AIAA-2000-3826[R]. Reston, VA: AIAA, 2000.
- [32] MIRMIRANI M, WU C, CLARK A, et al. Modeling for control of a generic airbreathing hypersonic vehicle; AIAA-2005-6256[R]. Reston, VA: AIAA, 2005.
- [33] FIDAN B, KUIPERS M, IOANNOU P, et al. Longitudinal motion control of air-breathing hypersonic vehicles based on time-varying models; AIAA-2006-8074[R]. Reston, VA: AIAA, 2006.
- [34] BOLENDER M A, DOMAN D B. Nonlinear longitudinal dynamical model of an air-breathing hypersonic vehicle[J]. Journal of Spacecraft and Rockets, 2007, 44(2): 374-387.
- [35] CULLER A, WILLIAMS T, BOLENDER M. Aerothermal modeling and dynamic analysis of a hypersonic vehicle; AIAA-2007-6395[R]. Reston, VA: AIAA, 2007.
- [36] 向锦武, 曾开春, 聂璐. 考虑弹性影响的乘波体飞行动力学特性[J]. 北京航空航天大学学报, 2012, 38(10): 1306-1310.
XIANG J W, ZENG K C, NIE L. Elastic effects on flight mechanics of waverider[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(10): 1306-1310 (in Chinese).
- [37] FIDAN B, MIRMIRANI M, IOANNOU P. Flight dynamics and control of air-breathing hypersonic vehicles; Review and new directions; AIAA-2003-7081[R]. Reston, VA: AIAA, 2003.
- [38] PARKER J T, SERRANI A, YURKOVICH S, et al. Control-oriented modeling of an air-breathing hypersonic vehicle[J]. Journal of Guidance, Control, and Dynamics, 2007, 30(3): 856-869.
- [39] CUI Y, CAI G, ZHANG S, et al. Control-oriented modeling and characteristic analysis of an airbreathing hypersonic vehicle[C]//2014 IEEE Chinese Guidance, Navigation and Control Conference. Piscataway, NJ: IEEE Press, 2014: 1180-1185.
- [40] DICKESON J, RODRIGUEZ A, SRIDHARAN S, et al. Control-relevant modeling, analysis, and design for scramjet-powered hypersonic vehicles; AIAA-2009-7289[R]. Reston, VA: AIAA, 2009.
- [41] DUAN H, LI P. Progress in control approaches for hypersonic vehicle[J]. Science China: Technological Sciences, 2012, 55(10): 2965-2970.
- [42] 方洋旺, 柴栋, 毛东辉, 等. 吸气式高超声速飞行器制导与控制研究现状及发展趋势[J]. 航空学报, 2014, 35(7): 1776-1786.
FANG Y W, CHAI D, MAO D H, et al. Status and development trend of the guidance and control for air-breathing hypersonic vehicle[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(7): 1776-1786 (in Chinese).
- [43] 吴立刚, 安昊, 刘健行, 等. 吸气式高超声速飞行器控制的最新研究进展[J]. 哈尔滨工业大学学报, 2016, 48(10): 1-16.
WU L G, AN H, LIU J X, et al. Recent progress in control of air-breathing hypersonic vehicles[J]. Journal of Harbin Institute of Technology, 2016, 48(10): 1-16 (in Chinese).
- [44] 王勇, 张艳, 白辰, 等. 吸气式高超声速飞行器制导与控制方法综述[J]. 兵器装备工程学报, 2017, 38(4): 72-76.
WANG Y, ZHANG Y, BAI C, et al. Review of guidance and control approaches for air-breathing hypersonic vehicle[J]. Journal of Ordnance Equipment Engineering, 2017, 38(4): 72-76 (in Chinese).
- [45] 吴志刚, 楚龙飞, 杨超, 等. 推力耦合的高超声速飞行器气动伺服弹性研究[J]. 航空学报, 2012, 33(8): 1355-1363.
WU Z G, CHU L F, YANG C, et al. Study on aeroservoelasticity of hypersonic vehicles with thrust coupling[J]. Acta Aeronautica et Astronautica Sinica, 2012, 33(8): 1355-1363 (in Chinese).

- [46] MCRUER D. Design and modeling issues for integrated airframe/propulsion control of hypersonic flight vehicles [C] // American Control Conference. Piscataway, NJ: IEEE Press, 1991:729-734.
- [47] SERRANI A, ZINNECKER A M, FIORENTINI L, et al. Integrated adaptive guidance and control of constrained nonlinear air-breathing hypersonic vehicle models [C] // American Control Conference. Piscataway, NJ: IEEE Press, 2009:3172-3177.
- [48] SCHMIDT D, VELAPOLDI J. Flight dynamics and feedback guidance issues for hypersonic airbreathing vehicles: AIAA-1999-4122 [R]. Reston, VA: AIAA, 1999.
- [49] HAO X, CHANG J, BAO W, et al. Preliminary design for coordinated control of airbreathing hypersonic vehicle and scramjet engine [C] // Proceedings of the 33rd Chinese Control Conference. Piscataway, NJ: IEEE Press, 2014:909-914.
- [50] PETTIT C L. Uncertainty quantification in aeroelasticity: Recent results and research challenges [J]. Journal of Aircraft, 2004, 41(5):1217-1229.
- [51] BUSCHEK H, CALISE A. Robust control of hypersonic vehicles considering propulsive and aeroelastic effects: AIAA-1993-3762 [R]. Reston, VA: AIAA, 1993.
- [52] BUSCHEK H, CALISE A J. Uncertainty modeling and fixed-order controller design for a hypersonic vehicle model [J]. Journal of Guidance, Control, and Dynamics, 1997, 20(1):42-48.
- [53] CHAVEZ F, SCHMIDT D. Uncertainty modeling for large flexible high-speed aircraft: AIAA-1996-3920 [R]. Reston, VA: AIAA, 1996.
- [54] 吴志刚, 杨超. 气动伺服弹性系统不确定性建模与鲁棒稳定性 [J]. 航空学报, 2003, 24(4):312-316.
WU Z G, YANG C. Modeling and Robust stability for aeroservoelastic systems with uncertainties [J]. Acta Aeronautica et Astronautica Sinica, 2003, 24(4):312-316 (in Chinese).
- [55] 曲鑫, 李菁菁, 宋勋, 等. 考虑推进和气动弹性影响的高超声速飞行器的建模与控制 [J]. 宇航学报, 2011, 32(2):303-309.
QU X, LI J J, SONG X, et al. Modeling and robust coupled control of air-breathing hypersonic vehicle considering propulsion and aeroelastic effects [J]. Journal of Astronautics, 2011, 32(2):303-309 (in Chinese).
- [56] 曾开春, 向锦武. 高超声速飞行器飞行动力学特性不确定分析 [J]. 航空学报, 2013, 34(4):798-808.
ZENG K C, XIANG J W. Uncertainty analysis of flight dynamic character for hypersonic vehicles [J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(4):798-808 (in Chinese).
- [57] PU Z, TAN X, FAN G, et al. Uncertainty analysis and robust trajectory linearization control of a flexible air-breathing hypersonic vehicle [J]. Acta Astronautica, 2014, 101:16-32.
- [58] MORGAN G H, MILLER W R. Flutter tests of some simple models at a Mach number of 7.2 in helium flow: NASA-MEMO-4-8-59L, L-199 [R]. Washington, D. C.: NASA, 1959.
- [59] HARRIS T M, HUTTSELL L J. Aeroelasticity research at wright-patterson air force base (wright field) from 1953-1993 [J]. Journal of Aircraft, 2003, 40(5):813-819.
- [60] 季辰, 李锋, 刘子强. 高超声速风洞颤振试验技术研究 [J]. 实验流体力学, 2015, 29(4):75-80.
JI C, LI F, LIU Z Q. Research on flutter test technique in hypersonic wind tunnel [J]. Journal of Experiments in Fluid Mechanics, 2015, 29(4):75-80 (in Chinese).
- [61] 季辰, 赵玲, 朱剑, 等. 高超声速风洞连续变动压舵面颤振试验 [J]. 实验流体力学, 2017, 31(6):37-44.
JI C, ZHAO L, ZHU J, et al. Hypersonic wind tunnel flutter test research on rudder models by continuously varying dynamic pressure [J]. Journal of Experiments in Fluid Mechanics, 2017, 31(6):37-44 (in Chinese).
- [62] 韩景龙, 宣传伟, 张兵, 等. 高超声速颤振数值与实验研究 [C] // 第十五届全国空气弹性学术交流会, 2017.
HAN J L, XUAN C W, ZHANG B, et al. Numerical and experimental study of hypersonic flutter [C] // 15th National Aeroelastic Academic Conference, 2017 (in Chinese).
- [63] 许辉. 高超声速舵面颤振风洞试验与数值模拟影响因素研究 [D]. 合肥: 合肥工业大学, 2018.
XU H. Study on hypersonic control surface flutter wind tunnel test and its numerical simulation [D]. Heifei: Heifei University of Technology, 2018 (in Chinese).
- [64] 张红英, 孙姝, 程克明, 等. 进气道工作状态对吸气式高超声速飞行器气动特性影响的实验研究 [J]. 宇航学报, 2007, 28(6):1488-1493.
ZHANG H Y, SUN S, CHENG K M, et al. Experimental investigation of inlet start unstart influences on the aerodynamic characteristic of a hypersonic vehicle [J]. Journal of Astronautics, 2007, 28(6):1488-1493 (in Chinese).
- [65] 张红英, 程克明, 伍昭兆. 某高超飞行器流道冷流特征及气动特性研究 [J]. 空气动力学学报, 2009, 27(1):119-123.
ZHANG H Y, CHENG K M, WU Y Z. A study on the flowpath and the aerodynamic characteristic of a hypersonic vehicle [J]. Acta Aerodynamica Sinica, 2009, 27(1):119-123 (in Chinese).
- [66] 姚文秀, 雷麦芳, 杨耀栋, 等. 高超声速乘波飞行器气动实验研究 [J]. 宇航学报, 2002, 23(6):82-85.
YAO W X, LEI M F, YANG Y D, et al. An aerodynamic experiment research of the hypersonic waverider vehicle [J]. Journal of Astronautics, 2002, 23(6):82-85 (in Chinese).
- [67] 吴颖川, 贺元元, 贺伟, 等. 吸气式高超声速飞行器机体推进一体化技术研究进展 [J]. 航空学报, 2015, 36(1):245-260.
WU Y C, HE Y Y, HE W, et al. Progress in airframe-propulsion integration technology of air-breathing hypersonic vehicle [J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(1):245-260 (in Chinese).
- [68] 徐大军, 蔡国飙, 乐川. 吸气式高超声速飞行器气动热试验研究 [J]. 宇航学报, 2006, 27(5):1004-1009.
XU D J, CAI G B, LE C. Aeroheating experiment for airbreathing hypersonic vehicle [J]. Journal of Astronautics, 2006, 27(5):1004-1009 (in Chinese).
- [69] HANK J, MURPHY J, MUTZMAN R. The X-51A scramjet engine flight demonstration program: AIAA-2008-2540 [R]. Reston, VA: AIAA, 2008.
- [70] WU Z G, CHU L F, YUAN R Z, et al. Studies on aeroservoelasticity semi-physical simulation test for missiles [J]. Science China-Technological Sciences, 2012, 55(9):2482-2488.
- [71] 胡巍, 杨智春, 谷迎松. 带操纵面机翼气动弹性地面试验仿真系统中的气动降阶方法 [J]. 西北工业大学学报,

- 2013,31(5):810-815.
- HU W, YANG Z C, GU Y S. A new and effective method for Reducing order of aerodynamics of a wing with control surface for ground flutter test[J]. Journal of Northwestern Polytechnical University, 2013,31(5):810-815(in Chinese).
- [72] 陈帅, 杨智春, 李斌. 结构试验中均布载荷的气囊加载系统设计方法[J]. 工程力学, 2012, 29(6): 146-150.
- CHEN S, YANG Z C, LI B. Design method of uniform loading system using airbags in structure test[J]. Engineering Mechanic, 2012, 29(6): 146-150(in Chinese).
- [73] 侯英昱, 付志超, 朱剑, 等. 气动力模拟非接触式加载方法研究[J]. 空气动力学学报, 2018, 36(2): 357-361.

HOU Y Y, FU Z C, ZHU J, et al. Research on contactless loading method for aerodynamic force test[J]. Acta Aerodynamica Sinica, 2018, 36(2): 357-361(in Chinese).

作者简介:

杨超 男, 博士, 教授, 博士生导师。主要研究方向: 飞行器设计、气动弹性力学、飞行动力学。

赵黄达 男, 博士研究生。主要研究方向: 气动弹性力学。

吴志刚 男, 博士, 副教授, 博士生导师。主要研究方向: 气动弹性力学与主动控制。

Research progress of aerothermoelasticity of air-breathing hypersonic vehicles

YANG Chao^{1,2}, ZHAO Huangda^{1,2}, WU Zhigang^{1,2,*}

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China;

2. Key Laboratory of Aircraft Advanced Design Technology, Ministry of Industry and Information Technology, Beijing 100083, China)

Abstract: Nowadays the air-breathing hypersonic vehicles are focused on by the researchers of aeronautics and astronautics. This type of vehicle usually uses a scramjet engine and an integrated design method, which brings a series of aeroelastic problems. This paper first describes the research progress of integrated air-frame/engine modeling of air-breathing hypersonic vehicles. Then the research status of aerothermoelastic dynamics is introduced from the aspects of aerothermoelastic/propulsion coupling, control system coupling and uncertainty analysis. The related aerothermoelastic experimental research is analyzed. Finally, several suggestions for the study of aerothermoelastic problems of air-breathing hypersonic vehicles are proposed.

Keywords: aeroelasticity; integrated design; scramjet engine; air-breathing hypersonic vehicles; control system coupling

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0056

民机客舱中太阳辐射对热舒适性的影响



庞丽萍^{1,2,*}, 李恒¹, 王天博³, 范俊⁴, 邹凌宇¹

1. 航空航天大学 航空科学与工程学院, 北京 100083; 2. 沈阳航空航天大学 航空发动机学院, 沈阳 110136;

3. 沈阳航空航天大学 经济与管理学院, 沈阳 110136; 4. 陆军航空兵研究所, 北京 101121)

摘 要: 乘客出行过程中对于热舒适度要求不断提高,对民机客舱的整体热舒适性提出了更加严格的要求。通过对南北飞行的航班实际测量发现,由于受到太阳辐射的影响,客舱内部向阳和背阴两侧温度分布极不均匀,特别是窗户周围,平均温差达到 20℃,客舱向阳和背阴两侧的热舒适性相差较大;结合 CFD 动态仿真,基于实际客舱各区实时测量数据,建立等比例客舱仿真模型,以实际测量的温度和压力数据作为仿真边界条件,再现民机客舱内部温度和热舒适度 PMV 的分布情况,为定量分析南北飞行航班客舱的热舒适性提供理论依据。

关 键 词: 热舒适性; 太阳辐射; CFD 动态仿真; 实时测量; 定量分析

中图分类号: V217+.39

文献标识码: A

文章编号: 1001-5965(2019)10-1924-07

民机座舱的热舒适性是目前国内外研究的重要内容,也是民机性能在经济市场中重要的竞争力之一。PMV 指数是以人体热平衡的基本方程及心理学主观热感觉的等级为出发点,考虑人体热舒适感诸多有关因素的全面评价指标^[1-4],是目前国际上公认的衡量环境热舒适性的重要参数。

关于太阳辐射对飞机影响的研究,国外最早开始于 20 世纪中期。如建立于 1947 年的美国麦金利气候实验室,采用发热灯代替太阳光源,并且已经成功地开展了多组飞机实验^[5];俄罗斯 Obninsk 企业为第五代战斗机 T-50 生产复合材料部件,开发了一种独特的涂层,以保护飞机的玻璃座舱免受太阳辐射^[6-7];Beck 等^[5]和 Battistoni^[8]通过具体的物理理论计算,分析了太阳射线中的能量对飞机飞行过程的影响。

国内对于太阳辐射和热舒适性的研究目前还不完善。由于中国无容纳飞机开展气候环境实验

的实验室,相关研究也进行的较少,仅仅在理论和飞机窗户材料方面有所研究^[9]。杨建忠、王振斌等^[10]对飞机座舱的温度扰动进行了实验研究分析;何良、吴长水等^[11]根据雷诺平均的方法,采用射线追踪方法,计算了太阳辐射对汽车乘员的温度场,并且分析了隔热玻璃对舱室内乘员温度场的影响;郝贵和、刘永辉等^[12]基于 ASHRAE 模型分析了客舱表面太阳辐射量。

目前,对太阳辐射和热舒适度的研究很多都是在理想的理论环境下进行的^[13-14],很少有学者做过实际测量^[15-16],也没有提出合理有效的改进措施。本文以实际航班测试与 CFD 仿真为基础,采用实际的动态温度测量数据作为 CFD 动态仿真的边界条件,详细建立了与实际民机客舱相近的物理模型,再现民机客舱内部在太阳辐射下温度和热舒适度的分布情况,为提高民机客舱热舒适性提供详细准确的参考依据。

收稿日期: 2019-02-21; 录用日期: 2019-05-10; 网络出版时间: 2019-05-21 09:49

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190520.0849.001.html

基金项目: 辽宁省“兴辽英才计划”(XLYC1802092)

* 通信作者: E-mail: pangliping@buaa.edu.cn

引用格式: 庞丽萍, 李恒, 王天博, 等. 民机客舱中太阳辐射对热舒适性的影响[J]. 北京航空航天大学学报, 2019, 45(10): 1924-1930. PANG L P, LI H, WANG T B, et al. Effect of solar radiation on thermal comfort in civil aircraft cabin[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1924-1930 (in Chinese).

1 太阳辐射热效应测试与分析

1.1 实际客舱温度数据实时采集实验

1) 测试方法

从飞机地面待机开始,到整个飞行包线结束的过程中,每隔 5 min 测量一次客舱内各壁面的温度和舱内压力。客舱内部区域划分如图 1 所示。

将客舱横向划分为向阳和背阴两侧面,两侧面实验测点分布完全对称,所以取客舱向阳面为例,说明实验测点分布情况。在远离窗户的侧壁上分别取上、中、下 3 点,求平均值得到侧壁的温度,如图 2 所示。

窗户在采取遮光板时,选取 4 个温度测点,分布在窗上,4 组测量数据取平均值得到窗户温度值,如图 3 所示。地板温度测点共 3 个,沿横向分布在乘客脚底,3 组测量数据取平均值得到地板温度值,如图 4 所示。

2) 测试设备

测试设备信息如表 1 所示。

3) 航线信息

为了更好地对比太阳辐射效应对飞机客舱热舒适性的影响,选取春季、夏季和冬季 3 个季节,航线在南北方向且飞行时间较长的 6 组航班进行实验测试,目的是获取更多的测试数据及使测试结果更具对比性。具体的航班信息如表 2 所示。

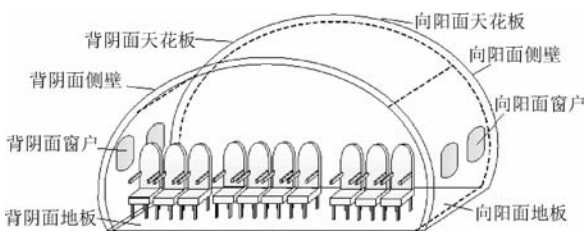


图 1 客舱内部区域划分示意图

Fig. 1 Schematic diagram of cabin interior zoning



图 2 侧壁温度测点分布

Fig. 2 Temperature measurement point distribution of side wall



图 3 窗户温度测点分布

Fig. 3 Temperature measurement point distribution of window



图 4 地板温度测点分布

Fig. 4 Temperature measurement point distribution of floor

表 1 实验测量仪器信息

Table 1 Information of measuring instruments in experiment

测量仪器	测量范围	精度
温湿度压力传感器	30 ~ 110 kPa	0. 15 kPa
Testo 热线风速仪	0 ~ 10 m/s	0. 1 m/s
Testo 红外线温度计	- 30 ~ 400℃	2%

表 2 航班信息

Table 2 Flight information

编号	季节	往返地	航班号	机型
1	春季	北京—厦门	HU7191	B787
2	春季	厦门—北京	HU7192	B787
3	夏季	北京—昆明	HU7211	B787
4	夏季	昆明—北京	HU7212	B787
5	冬季	北京—丽江	CA1469	A319
6	冬季	丽江—北京	CA1470	A319

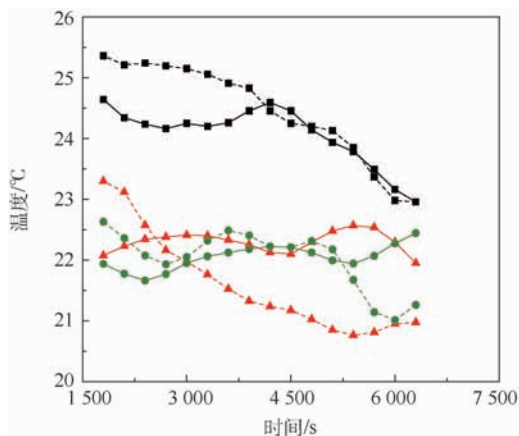
1.2 测试结果分析

根据适航标准和通风卫生标准的规定,乘客区域气流速度不超过 0. 36 m/s,通过整理 6 组航班的测量数据发现,实验测量人体周围的风速约 0. 1 ~ 0. 2 m/s,对热舒适度影响很小;在飞机起飞

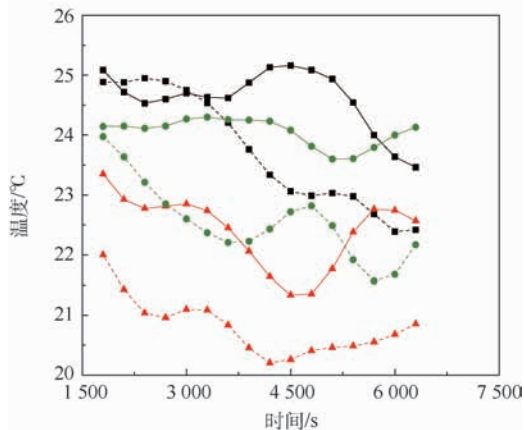
和降落的过程中,向阳面和背阴面随着时间反复变化,数据相差较大。因此,本文主要分析飞机在平飞阶段温度测量数据,结果如图5所示。

根据测量数据结果曲线,可得如下结论:

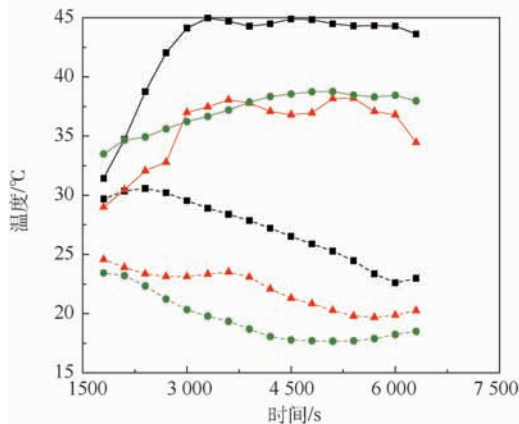
1) 由图5(a)可以得出,春季、夏季和冬季飞机客舱内地板的温度相差很小,同一航班上向阳和背阴两侧地板平均温度最大温差不超过2℃。



(a) 客舱内部地板温度变化曲线



(b) 客舱内部侧壁温度变化曲线



(c) 客舱内部窗户温度变化曲线

—■— 夏季,向阳 —▲— 春季,向阳 —●— 冬季,向阳
—■— 夏季,背阴 —▲— 春季,背阴 —●— 冬季,背阴

图5 不同季节客舱内部温度测试结果曲线

Fig. 5 Temperature measurement curves of cabin interior in different seasons

主要原因是:地板不受太阳光直接辐射,热量经过窗户进入客舱后传递到各个区域,最终太阳辐射的热量只有很少部分传递到地板上。

2) 由图5(b)可以得出,向阳面和背阴面侧壁的温度差值受季节性影响程度比地板较大,夏季最高温度25℃,春季最低温度21℃,两季节客舱侧壁温差可达4℃;但是同一季节里向阳面和背阴面的侧壁温差在2℃,同样说明侧壁温度受太阳辐射影响较小。

3) 由图5(c)可以得出,客舱窗户温度受太阳辐射影响很大,3个季节里向阳面和背阴面的窗户平均温差高达20℃。冬季窗户最高温度36℃,最低温度16℃,局部最大温差达到20℃。主要原因是:向阳面窗户直接受到太阳辐射,导致温度升高。

实验测试结果表明,从飞机客舱内部窗户到侧壁再到地板的垂直方向上,温度场分布极不均匀,从而造成乘客热舒适性较差。特别是靠近向阳侧窗户周围,即使采取遮光措施,但是较高的辐射温度给乘客带来极不舒适的感觉。

2 民机客舱热舒适性仿真

2.1 PMV 评价指标

在ISO 7730—2005^[17]热舒适性标准中,PMV指标的推荐值为-0.5~+0.5,即如果在环境中的90%的人觉得热舒适性情况良好,则该环境可以称为热舒适环境。本文的CFD模块中,PMV公式如下:

$$\begin{aligned} \text{PMV} = & (0.303e^{-0.036M} + 0.0275) \{ (M - W) - \\ & 0.00305[5.733 - 6.99(M - W) - P_a] - \\ & 0.42(M - W - 58.15) - 0.000017M(5867 - \\ & P_a) - 0.0014M(34 - t_a) - \\ & 0.000000396f_{cl}[(t_{cl} + 273)^4 - (t_r + \\ & 273)^4] - f_{cl}h_c(t_{cl} - t_a) \} \end{aligned}$$

式中: M 为人体代谢量, W/m^2 ; W 为人体对外做功, W/m^2 ; P_a 为周围空气中水蒸气分压, Pa; t_a 为人体周围空气温度, $^{\circ}C$; t_{cl} 为服装外表面温度, $^{\circ}C$; t_r 为周围环境平均辐射温度, $^{\circ}C$; f_{cl} 为人体服装面积系数; h_c 为对流换热系数, $W/(m^2 \cdot K)$ 。

周围环境平均辐射温度公式为

$$t_r = \left[\sum_n F_{p-i} (t_i + 273)^4 \right]^{1/4} - 273$$

式中: F_{p-i} 为该表面对其他表面的辐射角系数; t_i 为其他表面的温度。可以看出,环境的辐射温度对热舒适度影响很大。

2.2 动态仿真边界条件

对实体模型等效简化处理后,按照实际尺寸建立双通道飞机客舱模型。图 6 为仿真模型的截面图,客舱乘客编号规则如下:每排从左到右排序,分别为 1~10 号乘客,即 1 号乘客为背阴面窗户乘客,10 号乘客为向阳面靠窗户乘客。

同时,分别选取乘客脚部、胸部和头部 3 个位置计算仿真的 PMV 值。测点具体位置如图 7 所示。

各壁面的压力温度等条件是根据实际测量数据作为仿真边界条件输入。本文根据夏季北京飞往昆明的 HU7211 航班测量数据,利用 MATLAB 软件进行边界条件的函数曲线拟合。

客舱压力的边界条件如图 8 所示。两侧地板、侧壁和窗户的温度拟合曲线如图 9 所示。

根据适航标准和通风卫生标准的规定,民机客舱内的新风量为每人 0.25 kg/min。本文按照 50% 的新风空气和 50% 的舱内再循环空气组成,得出最终送风量为每人 0.5 kg/min,按此边界条件设置送风口边界条件。考虑到 3 h 左右的航班客舱湿度较低,送风空气的湿度设置为 25%;向阳和背阴两侧采用统一的送风温度为 17℃。

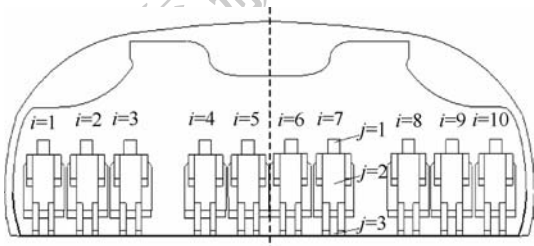


图 6 乘客位置分布
Fig. 6 Passenger location distribution

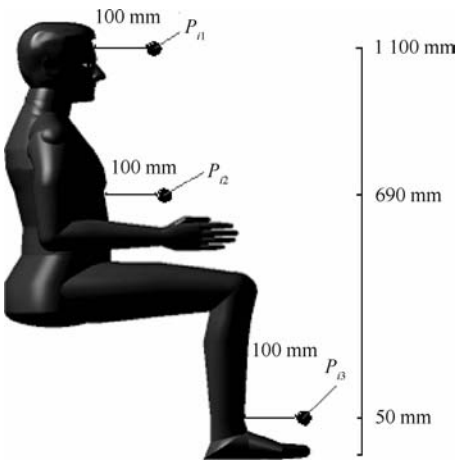


图 7 乘客周围测点分布
Fig. 7 Layout of measuring points around passenger

本文将乘客看成一个整体,计算乘客表面与其他各表面部分的辐射角系数,其他表面包括客舱的侧壁面、天花板、地板、前后表面、座椅表面及其他乘客表面。物理模型如图 10 所示。共计算了 2 个典型位置乘客的辐射角系数,结果如表 3 和表 4 所示。其中,乘客 1 为右侧靠窗乘客,乘客 2 为右侧过道乘客。各乘客对周围环境辐射面角系数之和为 1。

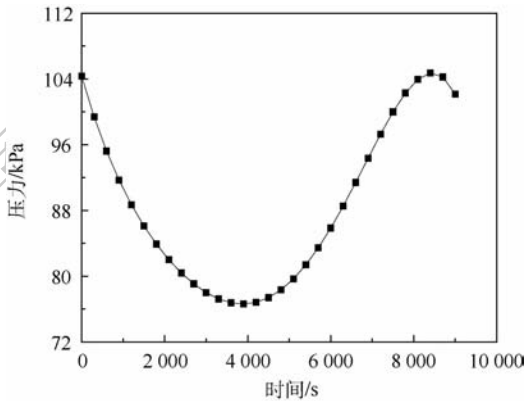
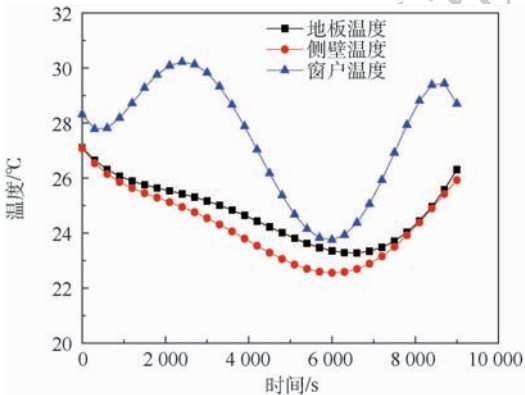
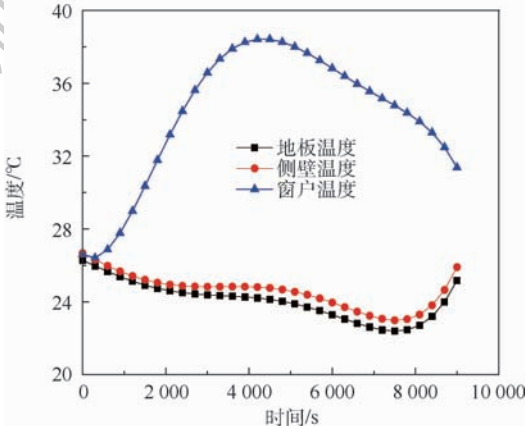


图 8 客舱压力边界条件
Fig. 8 Cabin pressure boundary conditions



(a) 左侧温度边界条件拟合曲线



(b) 右侧温度边界条件拟合曲线

图 9 两侧温度边界条件拟合曲线
Fig. 9 Fitting curves of temperature boundary conditions of two sides

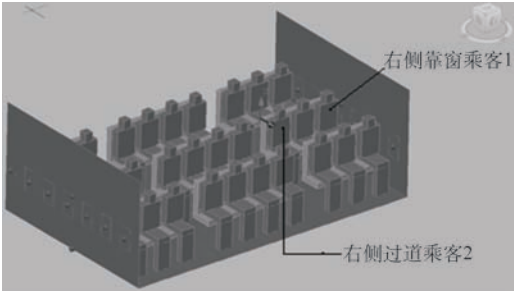


图 10 仿真的物理模型
Fig. 10 Simulation of physical model

表 3 乘客 1 的辐射角系数

Table 3 Radiation angle coefficient of passenger 1			
表面 p	表面 i	辐射类型	辐射角系数 F_{p-i}
乘客 1	座舱顶壁面	表面至表面	0.20
乘客 1	座舱地板	表面至表面	0.13
乘客 1	座舱右侧面	表面至表面	0.29
乘客 1	座舱右窗户	表面至表面	0.14
乘客 1	座椅	表面至表面	0.14
乘客 1	其他乘客	表面至表面	0.10

表 4 乘客 2 的辐射角系数

Table 4 Radiation angle coefficient of passenger 2			
表面 p	表面 i	辐射类型	辐射角系数 F_{p-i}
乘客 2	座舱顶壁面	表面至表面	0.30
乘客 2	座舱地板	表面至表面	0.18
乘客 2	座舱右侧面	表面至表面	0.08
乘客 2	座舱左壁面	表面至表面	0.08
乘客 2	座舱右窗户	表面至表面	0.02
乘客 2	座舱左窗户	表面至表面	0.03
乘客 2	座椅	表面至表面	0.20
乘客 2	其他乘客	表面至表面	0.11

人体作为 35℃ 的恒温源,透射率为 0,发射率为 0.97,反射率由客舱空间布局、人体透射率和发射率根据模型自动计算得出。

2.3 仿真结果分析

在构建乘客与实际客舱模型等效的仿真物理模型后,输入实际测量数据拟合出温度和压力等边界条件,在 CFD 中仿真计算出温度场的分布情况及乘客周围热舒适度 PMV 值。

飞机在平飞阶段舱内温度稳定后的仿真截图如图 11 所示。结果显示,向阳侧乘客周围的空气温度高于背阴侧乘客周围的空气温度,这与实际测量结果大致相符合。但是仿真结果的温差并没有实际测量的温差大,主要原因是仿真过程中忽略了很多实际的辐射热量和客舱内部空气的流场分布情况,而且温度云图显示的是客舱内部空气的温度场分布,并非窗户、侧壁和地板的温度,因此差异性没有实际测量结果明显。

飞机在平飞阶段舱内 PMV 稳定后的仿真截

图如图 12 所示。结果显示,向阳侧乘客周围的 PMV 值明显高于背阴侧,说明在相同的送风温度下,即使采用遮光板向阳侧窗户周围乘客也会有较热的感觉。

根据国际通用的 PMV 计算模式,利用 CFD 软件计算得出整个飞行过程中乘客头部和胸部周围热舒适度在向阳和背阴两侧的平均差值,如表 5 和表 6 所示。分析表 5 和表 6 数据可以看出,采用遮光板以后,向阳和背阴两侧乘客的热舒适度 PMV 仍然存在差异,特别是两侧靠近窗户周围的乘客,PMV 值最高相差 0.64,而人体在热舒适环境下 PMV 的推荐值为 -0.5 ~ +0.5,所以两侧乘客肯定会产生不同的冷热感觉,导致乘客

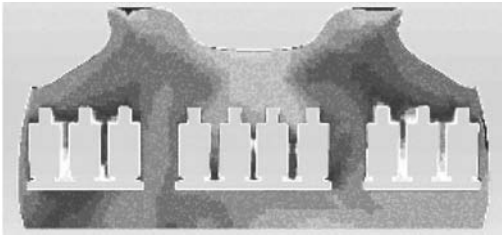


图 11 客舱内部温度场仿真结果
Fig. 11 Simulation results of temperature field in cabin

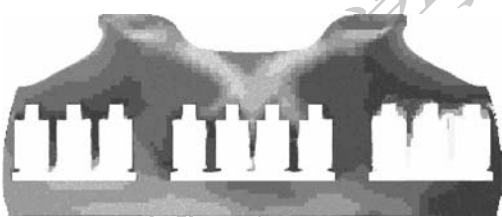


图 12 客舱内部 PMV 仿真结果
Fig. 12 Simulation results of PMV in cabin

表 5 乘客头部 PMV 左右差值仿真结果

Table 5 Simulation results of left-right difference of passenger head PMV		
左右测点	PMV 差值	
10-1	0.64	
9-2	0.56	
8-3	0.40	
7-4	0.1	
6-5	0.02	

表 6 乘客胸部 PMV 左右差值仿真结果

Table 6 Simulation results of left-right difference of passenger chest PMV		
左右测点	PMV 差值	
10-1	0.54	
9-2	0.54	
8-3	0.44	
7-4	0.11	
6-5	0.04	

出行过程中舒适度下降。

3 结 论

本文通过采集分析实际航班飞行过程中的温度实时数据,结合 CFD 动态仿真研究,得出如下结论:

1) 白天的太阳辐射对于飞机客舱内部地板和远离窗户的侧壁温度影响较小,温差不超过 2℃;但是对于窗户周围的温度影响较大,向阳和背阴两侧窗户周围最高温差可达 20℃。

2) 通过 CFD 动态仿真研究发现,现有的送风和控温方式下,采用遮光板时可以将客舱内部的空气温度调节的相对均匀合理,空气温度场分布向阳和背阴两侧没有存在较大的差异性。

3) 仿真结果显示,民机窗户的温度在太阳辐射影响下会急剧升高,在较高的辐射温度影响下,向阳侧乘客的 PMV 值明显高与背阴侧的 PMV 值,因此需要进一步采取新的措施,降低窗户的辐射温度,减小太阳辐射对民机客舱向阳侧窗户周围的热舒适性影响。

参考文献 (References)

- [1] RANCESCA R D A, BORIS I P, GIUSEPPE R. Notes on the calculation of the PMV index by means of Apps[J]. Energy Procedia, 2016, 101: 249-256.
- [2] HASAN M H, ALSALEEM F, RAFAIE M. Sensitivity study for the PMV thermal comfort model and the use of wearable devices biometric data for metabolic rate estimation[J]. Building and Environment, 2016, 110: 173-183.
- [3] SHEN C, YU N. Study of thermal comfort in free-running buildings based on adaptive predicted mean vote[C]// International Conference on E-Product E-Service and E-Entertainment. Piscataway, NJ: IEEE Press, 2010: 1-4.
- [4] MARCEL S, ANDREAS W. A framework for an adaptive thermal heat balance model (ATHB)[J]. Building and Environment, 2015, 94: 252-262.
- [5] BECK P, ROLLET S, LATOCHA M, et al. TEPC reference measurements at aircraft altitudes during a solar storm[J]. Advances in Space Research: The Official Journal of the Committee on Space Research (COSPAR), 2005, 36(9): 1627-1633.
- [6] Paintindia Group. PPG quick-application windshield coating kit available for general aviation[J]. Aviation Maintenance, 2010, 60(10): 66-68.
- [7] LV M Y, YAO Z B, ZHANG L C, et al. Effects of solar array on the thermal performance of stratospheric airship[J]. Applied Thermal Engineering, 2017, 124: 22-33.
- [8] BATTISTONI G. The FLUKA code, galactic cosmic ray and solar energetic particle events: From fundamental physics to space radiation and commercial aircraft doses[C]// IEEE Nuclear Science Symposium Conference Record. Piscataway, NJ: IEEE

Press, 2008: 1609-1615.

- [9] 黑贵罡, 姜曙光, 杨骏, 等. Fanger PMV 热舒适模型发展过程及适用性分析[J]. 低温建筑技术, 2017, 39(10): 125-128.
- HEI S G, JIANG S G, YANG J, et al. Development and applicability analysis of Fanger PMV thermal comfort model[J]. Cryogenic Building Technology, 2017, 39(10): 125-128 (in Chinese).
- [10] 杨建忠, 王振斌, 陈希远, 等. 飞机座舱温度扰动实验研究[J]. 科学技术与工程, 2017, 17(31): 364-368.
- YANG J Z, ZHANG Z B, CHEN X Y, et al. Experimental study on temperature disturbance in aircraft cockpit[J]. Science and Technology and Engineering, 2017, 17(31): 364-368 (in Chinese).
- [11] 何良, 吴长水, 李启杰, 等. 太阳辐射下乘员舱温度场的数值模拟研究[J]. 机械强度, 2017, 39(2): 479-483.
- HE L, WU C S, LI Q J, et al. Numerical simulation of temperature field in passenger compartment under solar radiation[J]. Mechanical Strength, 2017, 39(2): 479-483 (in Chinese).
- [12] 郝贵和, 刘永辉, 伍红英. 基于 ASHRAE 模型的客舱表面太阳辐射量分析[J]. 辽宁工程技术大学学报(自然科学版), 2017, 36(6): 645-650.
- HAO G H, LIU Y H, WU H Y. Analysis of solar radiation on cabin surface based on ASHRAE model[J]. Journal of Liaoning University of Engineering and Technology (Natural Science Edition), 2017, 36(6): 645-650 (in Chinese).
- [13] SHI G L. Discussions on the application of Fanger's thermal comfort theory[C]// The First International Conference on Building Energy and Environment Proceedings, 2008: 8.
- [14] 张昭, 唐虎, 成竹. 军用飞机实验室气候环境试验项目分析[J]. 装备环境工程, 2017, 14(10): 87-91.
- ZHENG Z, TANG H, CHENG Z. Analysis of climate and environment test projects in military aircraft laboratory[J]. Equipment Environment Engineering, 2017, 14(10): 87-91 (in Chinese).
- [15] DUAN P Y, LI H, LIU C C. PMV based hot/cold complaint model for dynamical thermal comfort[C]// International Conference on Manufacturing Science and Technology, 2015, 3: 1185-1188.
- [16] 杜秀媛. 客机座舱喷嘴送风参数优化及热环境评价[D]. 重庆: 重庆大学, 2017.
- DU X Y. Optimization of air supply parameters and thermal environment assessment of nozzles in passenger aircraft cockpit[D] Chongqing: Chongqing University, 2017 (in Chinese).
- [17] ISO/TC 159/SC 5. Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria: ISO 7730—2005[S]. Geneva: ISO, 2005.

作者简介:

庞丽萍 女, 博士, 教授。主要研究方向: 飞行器环境控制。

李恒 男, 硕士研究生。主要研究方向: 飞行器环境控制。

Effect of solar radiation on thermal comfort in civil aircraft cabin

PANG Liping^{1,2,*}, LI Heng¹, WANG Tianbo³, FAN Jun⁴, ZOU Lingyu¹

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China;

2. School of Aero-engine, Shenyang Aerospace University, Shenyang 110136, China;

3. School of Economic and Management, Shenyang Aerospace University, Shenyang 110136, China;

4. Army Air Force Research Institute, Beijing 101121, China)

Abstract: The requirement of thermal comfort for passengers is constantly increasing, which puts forward more urgent requirements for the overall thermal comfort of civil aircraft cabin. Based on the actual measurement of flights flying from north to south, it is found that the temperature distribution on both sides of the cockpit is extremely uneven due to the influence of solar radiation, especially around the windows, the average temperature difference reaches 20℃, and the thermal comfort on both sides of the cockpit is quite different. Combined with the CFD dynamic simulation, based on the actual situation, the thermal comfort of the cockpit on both sides of the cockpit is quite different. Real-time measurements of temperature and thermal comfort PMV in the cabin of civil aircraft are reproduced by setting up a simulation model of equal proportion cabin. The simulation boundary condition is temperature and pressure data measured in practice. The theoretical basis for quantitative analysis of thermal comfort in the cabin of north-south flights is provided.

Keywords: thermal comfort; solar radiation; CFD dynamic simulation; real-time measurement; quantitative analysis

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2018.0690

基于 SLM 的模拟月壤原位成形技术



李雯^{1,*}, 徐可宁¹, 黄勇², 胡文颖¹, 王道宽³, 姚思齐²

1. 中国航空发动机研究院, 北京 101304; 2. 北京航空航天大学 航空科学与工程学院, 北京 100083;

3. 北京科技大学 新材料技术研究院, 北京 100083)

摘 要: 激光选区熔化 (SLM) 技术与原位资源利用 (ISRU) 概念结合, 有望解决地外大规模基地建设的工程难题。利用模拟月壤考察了 SLM 成形技术用于月球原位资源增材制造的可行性。采用高能束激光为热源, 对粉床内模拟月壤颗粒进行逐层照射, 使颗粒熔融固结。以激光体积能量密度为综合评价指标, 开展 SLM 工艺参数研究, 实现模拟月壤的低能耗、高效率、高几何精度成形。研究结果表明: 模拟月壤在激光工作波长吸收率高, 热稳定性好, 利用较低激光能量可实现模拟月壤 SLM 成形, 成形件几何精度高; 激光体积能量密度决定了成形件质量, 增加激光体积能量密度可以提高成形件力学性能, 但过高的激光体积能量密度使成形件发生严重变形; 模拟月壤颗粒形态复杂、粒度分布广、流动性差, 通过优化颗粒粒径分布范围, 可以有效提高粉体的流动性, 从而形成致密且均匀的粉床, 避免成形件缺陷的产生。

关 键 词: 月球探测; 模拟月壤; 原位制造; 原位资源利用 (ISRU); 增材制造

中图分类号: V11; TH16

文献标识码: A

文章编号: 1001-5965(2019)10-1931-07

激光选区熔化 (Selective Laser Melting, SLM) 是随着激光技术发展及高亮度光纤激光器出现, 从激光选区烧结 (Selective Laser Sintering, SLS) 发展起来的一种增材制造技术。SLM 利用高能束激光逐层照射预先铺覆好的粉层颗粒, 使其熔化并固化成形, 最终形成致密的三维零件。近年来, SLM 在航空航天领域获得应用^[1]。

21 世纪伊始, 人类迎来了新一轮的探月高潮。美国国家航空航天局 (NASA) 提出了“重返月球、建立月球永久性基地”计划; 欧洲航天局 (ESA) 制定了“极光”计划, 其主要任务是载人登月, 建立月球基地, 并以月球为跳板实施载人火星探测任务。月球基地建设对人类未来深空探测任务具有重要的军事和经济价值^[2]: ①为长期的空间科学与技术研究提供真实可靠的实验环境, 促

进空间技术发展; ②作为人类深空探测的前哨和补给站, 拓展人类深空探测范围; ③作为太空移民的实验基地, 促进相关工程技术研究和发 展。然而, 月球基地建设尚存在诸多技术挑战, 其中如何克服运载火箭空间和有效载荷的限制并最大限度地节约建设周期和经济成本是亟待解决的关键问题。据估算, 在月球建设一个与国际空间站规模相同的永久性基地 (重约 450 t), 仅“地-月”运输成本就高达 50 亿美元^[3]。

原位资源利用 (In-Situ Resource Utilization, ISRU) 概念最早由 NASA 提出, 旨在通过勘测、获取、利用地外天体的天然或废弃资源, 增强人类在地外空间的自给自足能力, 最大限度地减少对地球供给的依赖, 从而使人类真正走出地球, 迈向深空, 并实现可持续发展过程^[4]。月球 ISRU 技术

收稿日期: 2018-11-23; 录用日期: 2019-05-28; 网络出版时间: 2019-06-11 13:25

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190610.1136.006.html

基金项目: 国家自然科学基金 (51705490, 51876004); 国防科技创新特区项目

* 通信作者。E-mail: mosquato@buaa.edu.cn

引用格式: 李雯, 徐可宁, 黄勇, 等. 基于 SLM 的模拟月壤原位成形技术[J]. 北京航空航天大学学报, 2019, 45(10): 1931-1937.

LI W, XU K N, HUANG Y, et al. In-situ forming of lunar regolith simulant via selective laser melting[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1931-1937 (in Chinese).

是实现月球表面大规模、复杂多样基础设施建设的重要途径,也为实现专用工具和维修零件月面原位制造提供崭新的解决思路,具有重要的经济和军事研究价值。目前,美、德、英等国均有科研机构开展了月面原位成形技术研究,采用不同的成形设备和工艺,以模拟月壤为主要原材料,在地球实验室环境下探索月壤原位成形技术的可行性,并提出了堆积、烧结、熔融、模压、粘结等多种成形方法^[5]。研究发现,以高能束激光作为热源的 SLM 成形技术具有成形所需能量小、自动化程度高、适用于粉体加工、成形几何精度高、力学性能好等优点,因此在利用月球风化层资源实现不同规模和复杂程度的结构件月面原位制造方面具有较好应用前景。利用特殊设计的 SLM 设备(满足环境适应性),使用月球丰富的风化层资源,在月球表面原位制造出不同规模和复杂程度的三维结构(如结构单元、维修工具和零件),可以满足长期无人/载人月面探测任务、月球基地可持续发展的需要^[6-10]。最近 5 年,该领域研究受到国际航空航天研究机构的广泛关注,成为深空探测技术研究的热点问题,是未来“空间制造”的重要发展方向之一。

本文以一种与典型月海风化层物质物理力学属性相近的模拟月壤为研究对象,采用地基实验研究方法,探索 SLM 成形技术与 ISRU 概念相结合实现月壤原位制造的技术可行性,并对成形件几何精度、力学性能进行评估。研究成果可为中国探月工程及未来月球基地建设提供理论指导和技术支撑。

1 模拟月壤特性分析

地基和空基对月遥感信息及 Apollo、Luna 月面探测结果均显示,月球表面普遍覆盖着一层结构松散、颗粒细小、厚度达数米至数十米的月球风化层物质。其中,粒径小于 10 mm 的月球风化层物质被称为月壤。月壤是月岩空间风化作用和陨石高速撞击作用下形成的混合体物质,其颗粒矿物组成与其所处区域的月岩化学成分和矿物组成密切相关。月壤按照起源可分为月海玄武岩起源、月球高地起源、月海和高地混合源 3 种类型^[11-12]。月壤的基本物质类型包括矿物碎屑(橄榄石、斜长石、单斜辉石、钛铁矿等)、原始结晶岩碎屑(玄武岩、斜长岩、橄榄岩、苏长岩等)、角砾岩碎屑、各种玻璃(撞击玻璃、黄色或黑色火成碎屑玻璃)、黏合集块岩、陨石碎片(陨硫铁、橄

石、辉石、锥纹石、镍纹石、合纹石)^[13],化学成分主要是 SiO_2 、 TiO_2 、 Al_2O_3 、 Fe_2O_3 、 FeO 、 MgO 、 CaO 等^[11-12]。

月壤返回样品极为珍贵,即使是拥有 381.7 kg 月壤返回样品的美国也严格控制其使用范围,并规定月壤返回样品仅可用于非破坏性的科学实验研究。为此,国内外研究机构以 Apollo 计划不同登月点月壤返回样品的实验数据为依据,结合不同的科研任务目标,研制了多种模拟月壤。中国多家科研机构以嫦娥探月工程技术发展为牵引,研制了不同类型的模拟月壤。其中,中国空间技术研究院利用吉林省靖宇地区的火山灰为原料,研制出了 TYII-0 工程模拟月壤,其主要矿物相为橄榄石、辉石、长石、钛铁矿,并含有大量的玻璃质。模拟月壤具有与典型月海月壤相近的物理力学属性:颗粒密度 2900 kg/m^3 ,内聚力 0 kPa ,内摩擦角 40° ,中值粒径 $100\text{ }\mu\text{m}$ ^[14]。另外,这种以火山灰为原料研制的模拟月壤具有复杂的颗粒形态,从球形到极端棱角状都有出现,如图 1 所示,4 种最常见的颗粒形态包括长条形、次棱角形、棱角形、次圆形^[15]。

本文将以上述模拟月壤作为原材料,在实验室环境下(暂不考虑月球重力环境),开展基于 SLM 的模拟月壤原位成形技术实验探索研究。目前,国内针对模拟月壤物理力学属性的实验研究主要聚焦在土力学、地面力学等方面,主要用于为月面巡视探测器行走机构设计及性能评估、月面探测设备研制、月面除尘技术发展提供数据支撑,而针对 SLM 成形技术研究的模拟月壤物理力学属性相关实验数据较少。因此,本文将首先开展相关实验研究,分析测量模拟月壤的化学成分、光谱吸收特性(吸收曲线)、质量热稳定性(TGA 曲线)及熔点温度(DSC 曲线)等物理属性,用于

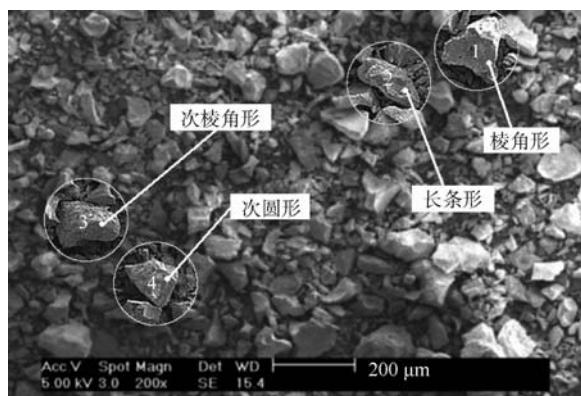


图 1 模拟月壤 SEM 照片^[15]

Fig. 1 SEM image of lunar regolith simulant^[15]

指导 SLM 成形设备及工艺参数选取,并为后续 SLM 铺粉与成形过程热离散元建模与仿真研究、模拟月壤原位成形技术发展提供数据支撑。

1.1 矿物组成及化学成分分析

利用扫描电子显微镜 (SEM) JSM-6510 配备的能谱仪 (EDS) 对模拟月壤化学成分进行实验分析。从 EDS 结果可知,模拟月壤的元素种类包含 O、Si、Fe、Ti、Al、Mg、Ca、Na 等;主要相的成分为 SiO₂、CaO、MgO、Al₂O₃、FeO 等;与 Apollo-12 登月点月壤采集样品 (A12) 的化学成分和矿物组成接近 (见表 1)。

表 1 模拟月壤与 A12 样品的化学成分对比

Table 1 Comparison of composition of chemical elements in lunar regolith simulant and A12 returned sample

化学元素	质量分数/%	
	模拟月壤	A12 月壤 ^[11]
O	36.48	46.31
Si	23.38	21.60
Fe	8.89	11.70
Ti	1.01	1.80
Al	2.78	3.40
Cr		0.12
Mn		0.15
Mg	7.26	5.60
Ca	12.22	7.60
Na	0.85	0.40
K		0.26
P		0.17
S		0.12

1.2 吸收光谱分析

吸收光谱是材料在某些频率上对电磁辐射的吸收所呈现的比率,与发射光谱相对。不同波长光对样品作用不同,吸收强度也不同,通常用一个无量纲标量“吸收率 A”定量描述,定义为吸收的辐射与入射的辐射比值。材料吸收率越高,则发生熔融所需的输入能量越小,因此,吸收率可以作为 SLM 工艺参数选取和优化的参考。本文利用 PerkinElmer Lambda 950 紫外/可见/近红外分光光度计,在惰性气体环境下 (避免地球大气环境中 CO₂、H₂O 对测量结果的影响),测量了模拟月壤实验粉的光谱吸收特性,其吸收曲线如图 2 所示。

从实验获得的吸收曲线可知,模拟月壤在紫外/可见/近红外光区 (波长 λ = 200 ~ 2500 nm) 具有较高的光吸收能力。在紫外光区,模拟月壤吸收率存在 1 个峰值点 (A = 0.98, λ = 249 nm);在可见光区,吸收率随波长单调递减,从 0.72 降低到 0.63;在近红外光区,吸收率随波长非线性

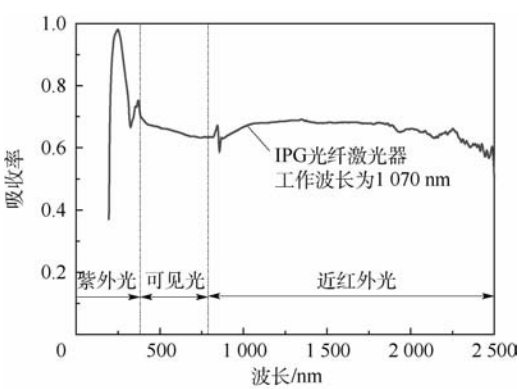


图 2 模拟月壤的吸收曲线

Fig. 2 Absorbance curve of lunar regolith simulant

变化,幅值范围为 0.50 ~ 0.69;SLM 配备的 IPG 光纤激光器工作波长 λ = 1070 nm,模拟月壤吸收率测量值为 0.68,低于 JSC-1A3 模拟月壤测量值^[7]。另外需要特别指出的是,模拟月壤在紫外/可见/近红外光区均表现出较高的光吸收能力,而根据月球表面太阳辐射数据可知,月表太阳辐射强度为 1358 ~ 1421 W/m²^[16],约为地球太阳辐射强度的 15 000 倍,开展以太阳光为热源的 SLM 设备研制将可有效解决地外空间原位制造技术对可持续能源供给的需求,该研究将是未来发展的又一重要研究方向。

1.3 热分析

利用 DSC-TGA 同步热分析仪 SDT Q600,对模拟月壤进行热重分析 (Thermo-Gravimetry Analysis, TGA) 和差示扫描量热分析 (Difference Scanning Calorimetry, DSC)。通过 TGA 判断模拟月壤实验粉中是否存在挥发成分;通过 DSC 分析确定材料的熔点温度。图 3 为模拟月壤 TGA 曲线,图 4 为模拟月壤 DSC 曲线,升温速度为 20℃/min。

如图 3 所示,在 25 ~ 1 200℃ 温度范围,模拟月壤的质量变化很小,最大失重率仅为 0.32%,具有良好的质量热稳定性,有利于在 SLM 成形过程中形成均匀致密的组织结构,降低成形件表

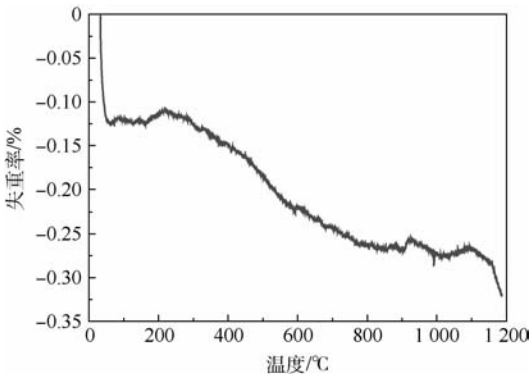


图 3 模拟月壤 TGA 曲线

Fig. 3 TGA curve of lunar regolith simulant

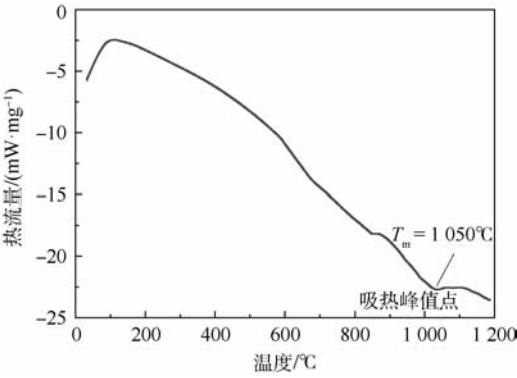


图4 模拟月壤 DSC 曲线

Fig. 4 DSC curve of lunar regolith simulant

面及内部产生缺陷的风险。模拟月壤在宽温域范围保持低失重率的原因可以解释为：实验粉主要矿物成分为橄榄石、辉石、长石、钛铁矿、玻璃质，其在 1200℃ 以下的挥发性小。在 25 ~ 300℃ 温度范围，模拟月壤发生的质量损失主要源于吸附在颗粒表面、颗粒-颗粒之间、颗粒多孔结构内部的液态水的蒸发；此后出现的质量损失主要源于模拟月壤中微量组分发生汽化所致。因此，在 SLM 成形工艺参数选取中，要避免过高的能量输入，否则粉床内颗粒可能发生汽化现象，形成局部缺陷。

如图 4 所示，模拟月壤的一个主要吸热峰值位于 1050℃，可以将其作为实验粉材的熔点温度，其与 JSC-1A3 模拟月壤的熔点温度接近 ($T_m = 1120\text{℃}$ ^[7])。

1.4 粒度分布

利用激光粒度分布仪 BT-9300S 测量了模拟月壤原始粉的粒度分布，并通过粒径级配曲线（见图 5）描述。可见，模拟月壤原始粉的粒度分布范围较广，从亚微米到数百微米的颗粒粒径均存在。其中，等效粒径的最大、最小值分别为 716 μm、0.497 μm；中值粒径 (D_{50}) 为 116.0 μm

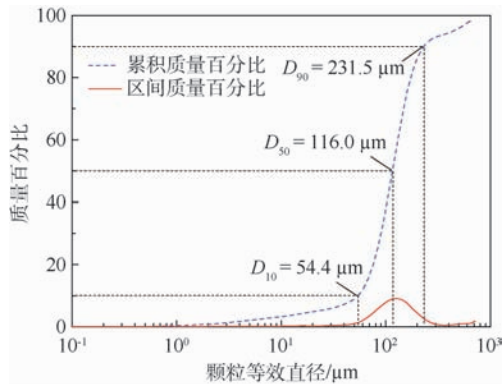


图5 模拟月壤的粒径级配曲线

Fig. 5 Particle size distribution curves of lunar regolith simulant

（与文献[14]略有差别），粒径小于 231.5 μm 的颗粒占总质量比的 90% ($D_{90} = 231.5\text{ μm}$)；粒径小于 54.4 μm 的颗粒占总质量比的 10% ($D_{10} = 54.4\text{ μm}$)。

2 SLM 实验方法

2.1 模拟月壤预处理

大量研究表明，颗粒形态、粒径级配是影响颗粒系统流动性的重要因素。复杂颗粒形态可以增加颗粒间的切向接触力，产生互锁现象；细颗粒间的范德华力、静电力对颗粒接触力学、动力学行为影响显著，将增加颗粒-颗粒、颗粒-壁面之间的粘附强度，形成颗粒团聚、壁面沉积现象。在 SLM 铺粉过程中，粉材流动性对所形成粉床质量存在显著影响，流动性差的粉材会降低粉床内局部区域的颗粒配位数，形成不均匀排列的颗粒系统，在宏观上表现为非均质且各向异性的粉层；粉床质量差会引起 SLM 成形过程中出现不连续的熔池和孔洞，造成成形件表面粗糙、内部存在缺陷、致密度低。因此，本文在开展 SLM 实验研究以前，首先需要制备出具有较好流动性的模拟月壤实验粉。具体思路如下：以 TYH-0 模拟月壤原始粉作为基础材料，通过机械分选处理，获得不同粒径级配的粉体材料，再通过实验和数值仿真相结合的方法，合理选取粒度分布范围，以确保实验粉在 SLM 铺粉过程中具有好的流动性，从而形成结构均匀致密的粉层。

本文选用一种气流磨设备对模拟月壤原始粉进行分级处理，获得不同粒径级配的实验粉，再通过流动性实验与离散元数值仿真相结合的方法，确定具有最佳流动性的粒度分布范围。研究结果表明，当等效粒径控制在 70 ~ 120 μm 范围时，实验粉的流动性最好。因此，在 SLM 成形技术实验研究中，将使用粒度分布在 70 ~ 120 μm 的实验粉。

2.2 SLM 工艺参数确定

选用德国 Solutions SLM125 成形设备，开展模拟月壤的 SLM 实验研究。SLM125 配置的 IPG 光纤激光器工作波长为 1070 nm，最大输出功率为 400 W，可调光斑直径为 70 ~ 100 μm，最大成形尺寸为 125 mm × 125 mm × 125 mm。在激光器的工作波长，实验粉的光吸收率为 0.68。

陶瓷和玻璃材料的 SLM 成形研究表明，若要获得几何精度好的成形件，需要将激光体积能量密度控制在相对低的水平，因为激光体积能量密度较高时，粉床中的颗粒完全熔化，导致熔池周围

出现收缩、球化效应,形成热影响区。但是,若要获得力学性能好的成形件,则需要将激光体积能量密度控制在相对高的水平,因为稳定的熔池将形成致密的成形件,使其密度接近原始材料密度。本文开展 SLM 成形工艺参数研究的目标是:使用最小的激光输出功率,以最快的速度完成 SLM 成形,并保证成形件具有合理的几何精度及可接受的力学性能。前人研究发现,SLM 成形件质量与多个工艺参数相关,这些参数都会对熔池产生影响,但又无法通过串行方式对其进行逐个优化。为此,本文引入激光体积能量密度 E_d ,其常用于描述单位体积材料施加的平均能量,数学表达式为^[17]

$$E_d = \frac{P}{vH_dL_T}$$

(1)

式中: E_d 为激光体积能量密度; P 为激光输出功率; W ; v 为激光扫描速度; H_d 为激光扫描间距; L_T 为铺粉层厚。

基于 Goulas 和 Friel^[18] 利用 JSC-1A3 模拟月壤和 SLM 设备成功制造出 3D 零件所用的工艺参数,初步确定式(1)各个变量值,再利用 SLM125 设备和本文制备的模拟月壤,开展一系列的标准样件(5 mm×5 mm)SLM 成形测试,最终确定本文的最佳工艺参数,如表 2 所示。研究发现,当激光体积能量密度 $E_d=15\text{ J/mm}^3$ 时,成形件几何精度高;但当激光体积能量密度增加到 50 J/mm^3 时,成形件产生了严重的翘曲变形。

表 2 实验采用的工艺参数

Table 2 Process parameters used during experimental procedure		
参数	数值	参考值 ^[17]
激光输出功率 P/W	40	10~50
激光工作波长 $\lambda/\mu\text{m}$	1.07	1.07~1.09
激光光斑直径 $\phi/\mu\text{m}$	70	100~300
激光扫描速度 $v/(\text{mm}\cdot\text{s}^{-1})$	190	20~300
激光扫描间距 $H_d/\mu\text{m}$	140	210~300
铺粉层厚 $L_T/\mu\text{m}$	100	100~350
基板预热温度 $T/^\circ\text{C}$	200	200

3 结果分析

3.1 模拟月壤 SLM 成形技术验证

利用 SLM125 成形设备和制备的模拟月壤粉末材料,在地球重力条件下开展了模拟月壤原位成形技术探索实验研究。结果显示,在较低激光输出功率($P=40\text{ W}$)、较快加工速度下,可以成功 3D 打印出具有复杂几何特征的齿轮和扳手模型

(见图 6)。另外,为了定量评价 SLM 成形件的几何精度,以齿轮的中心圆直径、齿顶圆直径、齿根圆直径为研究对象,选取 5 个测点位置对上述参数进行测量,并取其平均值与齿轮设计值进行对比,结果如表 3 所示。结果显示,利用 SLM 设备及优化的工艺参数和粒度分布的模拟月壤粉末材料,可以成功制造出具有高几何精度的 3D 零件,因此证明了基于 SLM 的模拟月壤原位成形技术的可行性,这将为解决月球基地大规模建设工程问题及载人/无人月面长期探测任务中维修零件和工具现场制造问题提供了有效途径。但需要指出的是,本文实验是在传统 SLM 设备上开展的,没有引入月面低重力、高真空环境因素,后续计划开展考虑月面环境适应性的相关实验及仿真研究,测量模拟月壤在高真空环境下的热物性、低重力环境对粉床形成的作用机制,以及多因素协同对 SLM 成形件质量的影响等。

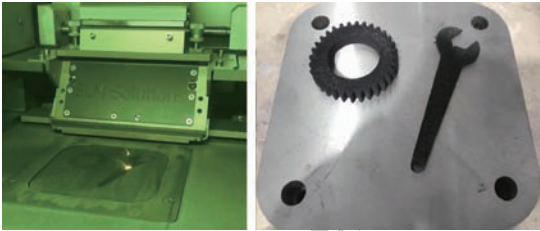


图 6 模拟月壤 SLM 实验结果

Fig. 6 Experimental results of lunar regolith simulant by SLM

表 3 齿轮成形件几何尺寸的设计值与测量值

Table 3 Designed and measured geometric dimension for fabricated gear			
参数	中心圆直径	齿顶圆直径	齿根圆直径
设计值/mm	44.19	39.00	26.00
测量平均值/mm	44.28	39.36	25.75

3.2 压缩强度分析

基于 SLM 的月壤原位成形技术研究目标是:利用月球表面丰富的月壤资源,采用增材制造方法,在月球表面制造出具有高几何精度和良好力学性能的结构单元及功能零件,以实现月球基地和月面探测器的自维持、自修复构想。因此,模拟月壤 SLM 成形件不但要满足高几何精度要求以外,还要满足不同应用场景的力学性能要求。本文通过压缩实验测量了 SLM 成形件的压缩强度,其峰值强度为 10.39 MPa ,比文献^[17]模拟月壤成形件压缩强度提高 3 倍以上。

为了进一步提高 SLM 成形件的力学性能,通过减小粉层厚度的方法来增加激光体积能量密

度,也就是通过减小粉层厚度,使粉床中颗粒通过直接吸收激光能量或颗粒间热辐射和热传导作用间接获取更多的能量后发生相变,在经冷却后固结,形成一层致密的结构体,避免由于颗粒吸收的激光能量不足而引起的层间界面强度低、表面粗糙、致密度低等问题。本文设定粉层厚度为 $70\text{ }\mu\text{m}$ (等于最小等效粒径),保持其他工艺参数不变,在相同条件下进行模拟月壤 SLM 成形,并通过压缩实验测量成形件力学性能。研究结果显示,粉层厚度降低有利于成形件力学性能提高,铺粉厚度为 $70\text{ }\mu\text{m}$ 的模拟月壤 SLM 成形件的压缩强度提高到 35.23 MPa ,约为铺粉厚度 $100\text{ }\mu\text{m}$ 成形件的 3 倍。

4 结 论

本文利用与 A12 月壤样品化学成分近似的模拟月壤和 SLM125 成形设备(配备 IPG 光纤激光器),探索通过 SLM 成形技术与 ISRU 概念相结合实现月壤原位增材制造的可行性。研究结论如下:

1) 使用的模拟月壤在 IPG 光纤激光器工作波长具有较高光吸收率,在较低激光体积能量密度下可以实现 3D 零件 SLM 成形,成形件几何精度和力学性能较好。实验结果证明了基于 SLM 的模拟月壤原位成形技术的可行性,并且该技术有望成为解决月球基地大规模、复杂多样建设任务及月面探测设备维修零件和工具现场制造的一种有效途径。

2) 使用的模拟月壤在紫外/可见/近红外光谱范围都具有较高光吸收率,在波长 249 nm 时吸收率高达 0.98。月球表面太阳辐射强约为地球的 15 000 倍,研制以太阳光为热源的 SLM 成形技术替代现有以高能束激光为热源的 SLM 成形技术,将成为空间风化层原位增材制造技术可持续发展的重要发展方向,具有重要的军事和经济意义。

3) 模拟月壤颗粒形状复杂、粒度分布较广,导致流动性差,无法直接用于 SLM 成形加工,需要对其进行粒度分布优化预处理。本文采用气流磨设备提取等效粒径在 $70\sim 120\text{ }\mu\text{m}$ 范围的模拟月壤颗粒,具有最佳流动性。但从环境适应性角度考虑,气流磨设备不能在月面高真空环境下工作,因此需要研制其他粉材分选处理设备。

4) 铺粉厚度对成形件力学性能影响显著,在其他工艺参数不变的情况下,铺粉厚度为 $70\text{ }\mu\text{m}$ 的模拟月壤成形件,与铺粉厚度为 $100\text{ }\mu\text{m}$ 的模拟

月壤成形件相比,压缩强度显著提高。但需要特别注意的是,铺粉厚度选取要同时考虑颗粒粒径、激光体积能量密度,铺粉厚度过小时,可能出现颗粒破碎和激光体积能量密度过高引起的严重热变形问题,导致成形件几何精度降低甚至成形失败。

本文研究目的是探索基于 SLM 的模拟月壤 3D 打印技术的可行性,所开展的 SLM 成形实验均是在地球实验室环境下进行的,没有考虑月球低重力、高真空等极端环境影响。建议后续开展考虑月球低重力、高真空环境的 SLM 铺粉及成形过程的地面环境模拟实验和多尺度数值仿真研究,定量评价月球环境效应对 SLM 成形件几何精度、力学性能的作用机制,开展相关的月面环境适应性研究。

参考文献 (References)

- [1] 董鹏,陈济轮. 国外选区激光熔化成型技术在航空航天领域应用现状[J]. 航天制造技术,2014(1):1-5.
DONG P, CHEN J L. Current status of selective laser melting for aerospace applications abroad[J]. Aerospace Manufacturing Technology, 2014(1):1-5 (in Chinese).
- [2] SANDERS G B, LARSON W E, PICARD M. Development and demonstration of sustainable surface infrastructure for moon/mars exploration; NASA-20110016205[R]. Washington, D. C.: NASA, 2011.
- [3] FATERI M, GEBHARDT A. Experimental investigation of selective laser melting of lunar regolith for in-situ applications [C] // ASME International Mechanical Engineering Congress and Exposition, 2013: V02AT02A008.
- [4] 李志杰,果琳丽. 月球原位资源利用技术研究[J]. 国际太空, 2017(3):44-50.
LI Z J, GUO L L. Research on the technology of lunar in-situ resource utilization [J]. Space International, 2017(3):44-50 (in Chinese).
- [5] 王志浩,刘宇明,田东波,等. 月壤原位成型技术工程适用性浅析[J]. 航天器环境工程, 2018, 35(3):298-306.
WANG Z H, LIU Y M, TIAN D B, et al. A brief analysis of the engineering applicability of lunar soil in-situ forming technology [J]. Spacecraft Environment Engineering, 2018, 35(3):298-306 (in Chinese).
- [6] CECCANTI F, DINI E, KESTELIER X D, et al. 3D printing technology for a moon outpost exploiting lunar soil [C] // 61st International Astronautical Congress. Paris: International Astronautical Federation, 2010:1-9.
- [7] BALLA V K, ROBERSON L B, OCONNOR G W, et al. First demonstration on direct laser fabrication of lunar regolith parts [J]. Rapid Prototyping Journal, 2012, 18(6):451-457.
- [8] FATERI M, GEBHARDT A. Process parameters development of selective laser melting of lunar regolith for on-site manufacturing applications [J]. International Journal of Applied Ceramic Technology, 2015, 12(1):46-52.
- [9] GOULAS A, BINNER J G P, HARRIS R A, et al. Assessing ex-

- traterrestrial regolith material simulants for in-situ resource utilization based 3D printing[J]. *Applied Materials Today*, 2017, 6:54-61.
- [10] GERDESL N, FOKKEN G, LINKE S, et al. Selective laser melting for processing of regolith in support of a lunar base[J]. *Journal of Laser Applications*, 2018, 30(3):032018.
- [11] MCKAY D S, HEIKEN G H, VANIMAN D T, et al. The lunar regolith in the lunar sourcebook[M]. Cambridge: Cambridge University Press, 1991:285-365.
- [12] 李雯. 轮式月面探测器牵引通过性的细观力学研究[D]. 北京:北京航空航天大学, 2008.
- LI W. Meso-mechanical study on trafficability for wheeled lunar rover vehicle[D]. Beijing: Beihang University, 2008 (in Chinese).
- [13] 欧阳自远. 月球科学概论[M]. 北京:中国宇航出版社, 2005:12-19.
- OUYANG Z Y. Introduction to lunar science[M]. Beijing: China Aerospace Press, 2005:12-19 (in Chinese).
- [14] 贾阳, 申振荣, 党兆龙, 等. 模拟月壤研究及其在月球探测工程中的应用[J]. *航天器环境工程*, 2014, 31(3):241-247.
- JIA Y, SHEN Z R, DANG Z L, et al. Lunar soil simulant and its engineering application in lunar exploration program[J]. *Spacecraft Environment Engineering*, 2014, 31(3):241-247 (in Chinese).
- [15] HUANG Y, ZHAO R, LI W. Radiative characteristics of non-spherical particles based on a particle superposition model[J]. *Journal of Geophysical Research*, 2013, 118(20):11762-11769.
- [16] HEIKEN G, VANIMAN D, FRENCH B. Lunar sourcebook: A user's guide to the moon[M]. Cambridge: Cambridge University Press, 1991.
- [17] GONG H, RAFI K, GU H, et al. Analysis of defect generation in Ti-6Al-4V parts made using powder bed fusion additive manufacturing processes[J]. *Additive Manufacturing*, 2014, 1-4:87-98.
- [18] GOULAS A, FRIEL R J. 3D printing with moon dust[J]. *Rapid Prototyping Journal*, 2016, 22(6):864-870.

作者简介:

李雯 女, 博士, 研究员。主要研究方向: 特种材料增材制造技术及其多尺度数值仿真技术。

黄勇 男, 博士, 教授, 博士生导师。主要研究方向: 空间热辐射理论。

In-situ forming of lunar regolith simulant via selective laser melting

LI Wen^{1,*}, XU Kening¹, HUANG Yong², HU Wenying¹, WANG Daokuan³, YAO Siqi²

(1. Aero Engine Academy of China, Beijing 101304, China; 2. School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China; 3. Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: Selective laser melting (SLM) technique in combination with the in-situ resource utilization (ISRU) concept can be an off-world manufacturing solution to the significant engineering challenge on the large-scale construction for extra-terrestrial bases. The feasibility of SLM process applied to the additive manufacturing of lunar in-situ resource was investigated by utilizing lunar regolith simulant. A laser source was utilized to melt the powder locally in a layer-wise manner. In order to successfully fuse the powder into parts with low laser power, high efficiency and high geometrical accuracy, the SLM process parameters were investigated and evaluated by laser volume energy density. The results show that the simulant can be successfully fused into parts with high geometrical accuracy using SLM process with low laser power due to its high absorbance and low mass loss. The fabricated part quality depends on the laser volume energy density: increase of laser volume energy density input results in better mechanical properties of parts; however, excessive laser volume energy density input leads to high distortion of parts. Poor powder fluidity of the raw simulant is observed due to its complex particulate shape and a wide range of particle size distribution. The powder fluidity of the simulant is improved by optimizing its particle size range, resulting in a denser and more uniform powder bed, which can avoid defects within fabricated parts.

Keywords: lunar exploration; lunar regolith simulant; in-situ manufacturing; in-situ resource utilization (ISRU); additive manufacturing

Received: 2018-11-23; Accepted: 2019-05-28; Published online: 2019-06-11 13:25

URL: kns.cnki.net/kcms/detail/11.2625.V.20190610.1136.006.html

Foundation items: National Natural Science Foundation of China (51705490, 51876004); National Defense Science and Technology Innovation Special Zone Project

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

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0051

面向战斗机云作战的构造型仿真平台架构



田永亮^{1,*}, 王永庆^{1,2}, 熊培森¹, 郭宇², 武哲¹

1. 北京航空航天大学 航空科学与工程学院, 北京 100083; 2. 沈阳飞机设计研究所, 沈阳 110035)

摘

要: 随着大数据、云计算、物联网、移动互联网等信息技术的迅猛发展与广泛应用,

新的作战模式不断涌现,以任务分布式指控流程为核心的云作战成为一种全新的跨域全维作战样式。在分析作战云与云作战特征的基础上,结合传统作战仿真流程提出了云作战体系仿真流程,并提出了云作战构造型仿真平台框架的总体方案设计与系统功能设计。通过云作战构造型仿真示例,对比了传统作战样式与云作战样式的观察—调整—决策—行动(OODA)循环,结果表明,云作战样式能够有效缩短OODA循环时间。

关键词: 云作战; 作战云; 建模; 仿真; 观察—调整—决策—行动(OODA)

中图分类号: V221+.91; TP391.9

文献标识码: A

文章编号: 1001-5965(2019)10-1938-08

对于现代战争来说,争夺制空权、制海权、陆地控制权,都离不开对敌方信息的获取与压制,尤其是在联合体系对抗与复杂体系战场环境下,更需要对传统的作战方式进行改变,以在信息对抗中获得新的优势。然而在实际的体系对抗中,信息主要存在3方面的问题:①信息孤岛问题,同类型武器平台难以实现态势共享;②信息融合问题,不同类型武器平台的数据格式差异大;③信息备份问题,获取信息的存储如何安全可靠。

大数据、云计算、物联网、移动互联网等信息技术与现代军事应用相结合,形成了适用于现代复杂战争环境的云作战体系解决方案,通过作战云能够实现体系对抗中的态势共享,跨域系统能力,解决了信息孤岛、信息融合和信息备份等问题。

2014年8月,美国《航空周刊》发表了云作战的构想图,描述了“作战己方的空中优势空域云”的发展远景:在轨太空侦察/通信/导航卫星、空中预警机、F-15/16等四代机、海面航母战斗群,以及深入敌方纵深空域的F-22/35隐身战机、RQ-

180无人侦察机和新型远程隐身轰炸机等多维作战单元,在作战云的联结下形成一个高度融合的作战体系^[1-3]。

作战云是指综合运用网络通信技术、虚拟化技术、分布式计算技术及负载均衡技术将分散部署的作战资源进行有机重组而形成的一种弹性、动态的作战资源池^[4-5],其采用面向服务的模式,为指挥决策、部队行动、武器打击提供按需、便捷、快速的专业、权威的数据和应用服务。作战云是一种通过战场通信网络互联的弹性作战资源集群,是云计算理念在军事领域的全新运用^[6]。

云作战是指基于作战云提供的各项服及相应保障技术形成的全新作战模式,其依靠作战云提供的已被池化的作战资源、基于云计算技术对资源池中大数据分析处理结果及战场中先进高效的数据链等高效信息传输技术和信息融合技术,使连入作战云中的各终端之间实现信息共享和跨域协同,从而提高作战效率。

云作战体系相比于以往的作战样式,具有动态虚拟资源池、资源融合、分布式云杀伤链等独特

收稿日期: 2019-02-16; 录用日期: 2019-03-29; 网络出版时间: 2019-06-18 17:43

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190617.1409.007.html

*通信作者: E-mail: tianyongliang1@buaa.edu.cn

引用格式: 田永亮, 王永庆, 熊培森, 等. 面向战斗机云作战的构造型仿真平台架构[J]. 北京航空航天大学学报, 2019, 45(10): 1938-1945. TIAN Y L, WANG Y Q, XIONG P S, et al. Structured simulation platform architecture for fighter cloud operations [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1938-1945 (in Chinese).

的特征和优势^[7-9]。云作战采用类似云计算技术,从体系层面实现陆、海、空、天各作战域的战场资源整合,汇聚成云,完成战场数据的网状交互,具备全域性、分布式、网络化特点,将对未来的空战体系产生深远影响。多种作战飞机在执行任务过程中组成的作战云,可以改变作战飞机传统的作战样式,使其能够在作战云的支持下与作战体系中的各种资源结合在一起,真正实现信息融合和态势共享,并在更高的层面上进行作战。具体的作战样式包括云攻击样式、云射击样式、群控制样式、云制导样式^[10-15]。本文在作战云与云作战概念的基础上,结合云作战体系仿真流程,提出了云作战构造型仿真平台框架,为战斗机云作战样式研究与应用提供参考。

1 云作战体系仿真流程

常规仿真通常包含 5 个步骤:构建仿真框架,建立目标分析,创建仿真场景,运行仿真模型,分析仿真结果。本文在此基础上进行扩展,结合云作战体系的具体特点,考虑设计空间的范围、体系的随机性、体系的作战决策智能控制等,分解云作战体系的能力,最终形成能力识别、任务构建、模型构建、策略映射、智能控制、设计实验、样本分析等 7 个步骤。

云作战体系仿真的业务逻辑如图 1 所示。

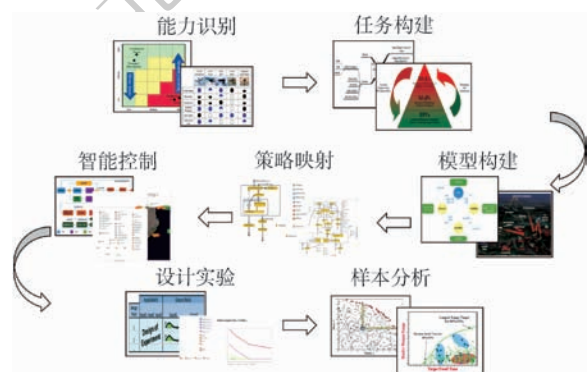


图 1 云作战体系仿真业务逻辑图

Fig. 1 Logic diagram of cloud operation system simulation process

1) 能力识别

能力识别是构建相应体系的前提,也是基于体系研究武器装备性能的基础。针对不同的作战能力,体系任务不同,构成体系的武器装备也有区别。识别将要进行体系仿真的能力,主要是指武器装备和设施的能力,具体包括战斗机的打击能力、行动能力、响应能力,以及导弹的打击能力,雷达的反应能力等。只有准确识别这些能力,才能

良好地进行任务场景的构建,执行相应观察—调整—决策—行动 (OODA) 循环中涉及的武器系统。

2) 任务构建

任务构建步骤不同于模型构建,主要为整个仿真设定一个战争脚本,具体为 6 个方面的脚本设定:①任务使命,是仿真任务背景,包括斩首行动或时效打击等;②相关环境,是仿真的外部环境,包括天气因素、地理因素和地形因素等;③双方环境,包括红方情形、红方行为能力、蓝方情形、蓝方防御能力等;④决策制定,定义时间、地点、方式、行动,为打击的总体过程设定总体方案;⑤红方预测,对比分析红方可能采取的决策方案,为各种替代方案的分析过程;⑥蓝方预测,分析蓝方可能采取的防御方案,预测蓝方可能采取的行动。

3) 模型构建

模型构建步骤先简化现实环境,构建红蓝双方的仿真模型。对仿真来说,并非所有的现实信息对于仿真分析都是有用的,因此,通过分析简化过程,先筛选一部分对仿真分析影响很小的现实信息,再通过解释演绎过程,将实际过程抽象为概念模型,最后通过检验和验证,确定构建的模型能够满足仿真的需要,完成模型构建。

4) 策略映射

策略映射步骤实际为仿真的具体操作步骤。以空军打击策略为例,建模时,设定的逻辑会有“优先攻击敌方的核心目标”,其是空军力量的优势所在,再开始实验,探索仿真结果是否与预想的一样并进行调整,最终关键分析蓝方目标的重要度与威胁度,以及能力评估和仿真任务的关系系数,利用质量功能配置形成打击序列列表,即将宏观的打击能力映射为可操作的打击行为。

5) 智能控制

智能控制步骤为基于智能体 (agent) 对整个任务进行仿真控制,初步想法包括红方控制中心、蓝方控制中心、双方的武器装备,类型为智能体,能够对战场态势进行智能自主控制。也就是说,对各种装备进行智能体模型构建后,就要将其纳入到仿真控制的模块中来,这样才能使得智能体在仿真过程中实现一步步的逻辑。

6) 设计实验

设计实验步骤为从设计空间中选择不同的指标数据,云体系环境下,输入变量并非一个固定的值,而是在一定的上界和下界内变动,由此形成设计空间,从设计空间中随机抽样,进行设计实验,将能够尽量覆盖所有的输入情况,使仿真更为全

面完整。

7) 样本分析

样本分析步骤是对大样本数据点的分析过程,建立输入空间与输出空间的关系,研究同样输入环境下,随机性对体系效能的影响,为指标的优化和技术的评估提供参考。

2 云作战构造型仿真平台框架设计

2.1 总体框架设计

在云作战体系仿真流程研究的基础上,设计了云作战构造型仿真平台的框架(见图2),并研究了不同模块的具体功能及相互关系。在获取大量试验数据的基础上,进行云作战效能评估与分析,为作战体系的改进提供依据。

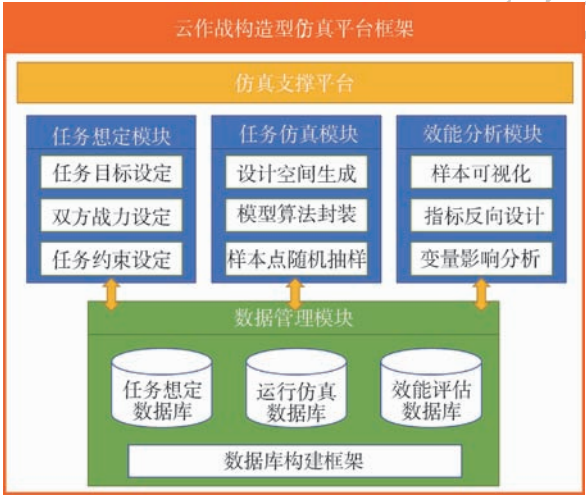


图2 云作战构造型仿真平台设计框架

Fig.2 Structured simulation platform design architecture for clouded operations

云作战构造型仿真平台框架主要由两部分组成:①仿真系统,主要包括任务想定模块、任务仿真模块、效能分析模块、数据管理模块;②仿真支撑平台,主要为仿真系统提供构建的平台,集成常用的建模命令,提供方便的操作界面,并能够支持多种想定的仿真建模与分析工作。

建模与仿真是一系列的活动,进行这一系列的活动需要一个支撑环境或支撑平台。仿真系统与仿真支撑平台有不同的概念,仿真系统具有应用针对性,是一个独立完整的系统,而仿真支撑平台具有通用性、服务性和资源共享性,在仿真支撑平台上可以研究、开发、建立、运行仿真系统。

仿真系统中各个模块之间的关系如下:

1) 任务想定模块主要负责:①任务体系任务目标的设定,包括任务背景、任务要求、打击目标等;②双方战力设定,包括红方蓝方的武器装备

状况、红方蓝方地理位置等;③任务约束的设定,包括任务范围、任务持续时间、任务可消耗的各项最大资源。

2) 任务仿真模块主要为根据任务想定,基于输入变量生成设计空间,执行仿真任务,并输出仿真结果。

3) 效能分析模块主要针对任务仿真模块的数据样本,将数据可视化,对输出数据进行进一步的分析。

4) 数据管理模块负责存储和管理三大任务模块的数据,为任务的设定和分析提供支持。

2.2 仿真系统功能设计

2.2.1 任务想定模块

任务想定模块主要从用户角度出发,为用户提供针对体系任务的各种想定,包括任务背景、任务要求、打击目标、双方战力、双方战场及任务约束等6个方面的内容。

1) 任务背景。需要设定任务的执行原因、任务的具体性质和任务所属类型,如针对敌方的斩首行动或者针对整个敌方设施的时敏打击。

2) 任务要求。需要设定执行任务所必须相关要素和程序,如针对地方地面目标的时敏打击,必须出动远程打击轰炸机还是需要多架战机共同完成。

3) 打击目标。需要设定该完成不同类型任务所必须进行打击的目标,不同类型的任务所需要打击的目标种类和数量有所区别。

4) 双方战力。需要设定一次体系任务中,红方蓝方可能出动的战机类型、数量,携带的导弹类型、数量,防御系统的类型、数量等。

5) 双方战场。需要设定体系任务中,战场的地理地貌、双方基地的地理位置、敌方目标及防御系统的地理位置等。

6) 任务约束。需要设定该次任务持续的时间、任务涉及的地理范围、任务所能够消耗的最大资源等。

2.2.2 任务仿真模块

任务仿真模块的研究内容主要从体系的建模与仿真出发,具体包括模型的封装、算法的封装、设计空间的扩展、样本点的随机抽样、设计实验等5个方面。

1) 模型封装。主要用于封装仿真可能用到的所有元素资源,如战斗机物理模型、导弹的物理模型、敌方地面目标的物理模型等。

2) 算法封装。主要用于封装体系中的智能体的各种规则,如指挥中心的武器选择,战斗机的

目标搜寻、锁定、打击等行为。

3) 设计空间的扩展。主要用于仿真运行前,扩展研究的体系变量的范围,如扩展战斗机起飞重量的上界下界、扩展战斗机的 RCS 范围等。

4) 样本点随机抽样。主要用于从设计空间中抽取输入点,进行仿真,此处选定蒙特卡罗随机抽取设计空间内的样本点,尽可能覆盖整个设计空间。

5) 设计实验。主要为体系的仿真运行,其中由于体系的随机性,每个样本点需要重复运行多次,充分考虑体系环境的各种情况,得到输出空间。

2.2.3 效能分析模块

效能分析模块的研究内容主要从体系效能的数据分析角度出发,具体包括样本筛选、样本可视化、变量影响分析和指标反向设计等 4 个方面。

1) 样本筛选。需要能够根据需求对样本进行筛选,如筛选任务完成项,或者筛选出动一定架次飞机情况下任务完成项。

2) 样本可视化。需要根据需求对样本进行可视化输出,如生成柱状图、曲线图等。

3) 变量影响分析。为根据体系的输入变量及输入变量对应的体系效能,研究两者的影响关系。

4) 指标反向设计。为根据变量与体系效能的影响关系生成的响应面或者包络图,反向优化设计指标。

2.2.4 数据管理模块

数据管理模块的研究内容主要从体系仿真产生的大量数据出发,具体包括数据项维护、实例应用、任务想定数据库、运行仿真数据库和效能评估数据库等 5 个方面。

1) 数据项维护。包含数据分类及数据库建立与维护。

2) 实例应用。查阅浏览所建立及入库的实例数据。

3) 任务想定数据库。对应于任务想定模块数据。

4) 运行仿真数据库。对应于任务仿真模块数据。

5) 效能评估数据库。对应于效能分析模块数据。

3 云作战构造型仿真示例

3.1 任务想定

在云作战构造型仿真平台框架设计的基础

上,考虑到战场复杂性、模型颗粒度、参数实验要求等因素,选择 Anylogic 进行云作战构造型仿真示例,并根据第 1 节所述构造型仿真平台的构建流程,针对战斗机传统作战样式与战斗机云作战样式搭建相应的仿真模型。为了能够清楚地说明问题,采用控制变量的对比方式进行建模,即 2 种作战样式中的基础设施、作战机群和携带的武器均相同,但二者的作战逻辑不同。

示例场景设置(见图 3)如下:①红方出动 1 架有人机和 4 架无人机;②蓝方在 2 个位置部署防空导弹系统(雷达 + SAM);③红方作战机群携带雷达干扰设备和 JDAM 空对地导弹;④红方需要摧毁蓝方 2 处防空导弹系统和 2 处军事建筑。

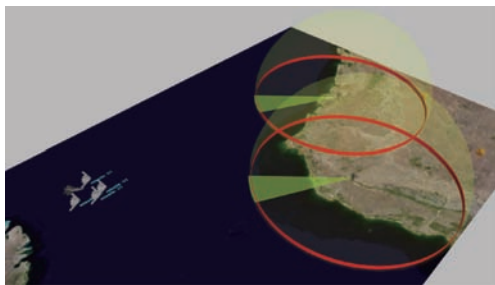


图 3 示例仿真场景

Fig. 3 Simulation scenario demonstration

3.2 传统作战样式仿真

3.2.1 传统作战样式流程

1) 1 架有人机与 4 架无人机在接到作战命令后从我方军用基地起飞,4 架无人机每架携带 1 枚 JDAM 空对地导弹在机群前方飞行,有人机紧随其后。

2) 当发现敌方雷达时,有人机立即减速并对敌方雷达实施电子压制,使雷达失效。

3) 继续飞行,有人机利用自身携带的探测设备对敌方区域进行探测,当发现敌方的 SAM 防空导弹发射车时,立即派出 1 架无人机。

4) 第 1 架无人机进入作战模式,并采用机载探测设备对目标进行探测,当其发现目标时,需向地面指挥中心(CAOC)进行任务的确认。

5) 接到“允许攻击”的指令后,立即对目标进行攻击,导弹由卫星制导,同时有人机对目标进行图像获取,并最终确认其被摧毁。

由于每架无人机在仿真模型中只携带 1 枚 JDAM 空对地导弹,所以当摧毁目标后立即返回基地。

该作战机群对敌方目标的另外 1 个防空导弹系统和 2 个军事建筑均采用这种作战形式对其进

行攻击。即由有人机探测得到目标信息并派出1架无人机执行任务,无人机在利用机载探测设备发现目标后,需向地面指挥中心进行确认,然后对敌方目标进行攻击,有人机实时获取敌方目标的影像,并确认其被摧毁。

3.2.2 传统作战样式 OODA 循环

传统作战样式的 OODA 循环过程如图 4 所示。首先,无人机向有人机提供的敌方目标方向飞行,并利用机载探测设备对敌方目标进行观察 (observe);然后,当探测到敌方目标后,向地面指挥中心请求指令,地面指挥中心协助无人机完成调整 (orient) 和决策 (decide) 过程;最后,无人机接到指令后对目标进行攻击,即为行动 (act) 阶段。

由此可知,在每个 OODA 循环过程中,无人机参与观察和行动阶段,地面指挥中心辅助其完成调整和决策阶段。传统作战样式的总作战时间即为 4 架无人机的 OODA 循环之和。



图 4 传统作战样式的 OODA 循环示意图

Fig. 4 Schematic of OODA circles of traditional combat style

3.3 云作战样式仿真

3.3.1 云作战样式流程

1) 云作战样式的初始阶段与传统作战样式相同,均为 1 架有人机与 4 架无人机从我方基地出发,但不同的是,只有 1 架无人机携带 4 枚 JADM 空对地导弹,2 架无人机负责对战场的探测,1 架无人机负责生成被摧毁敌方目标的影像。

2) 在飞行过程中,2 架侦查无人机以较快的飞行速度飞向战场,并对敌方目标进行探测。当侦查无人机发现敌方雷达时,立即将雷达的信息传回给有人机,有人机随即对雷达实施电子干扰。

3) 在使敌方雷达失效的同时,有人机根据侦查无人机传回的数据对战场进行态势分析,并制定相应的作战方案,指挥携带 JADM 空对地导弹的无人机对目标进行攻击。

4) 无人机接到攻击指令后,立即发射导弹,并在导弹的飞行过程中由有人机对其进行制导,使其准确命中目标。敌方目标被摧毁后,由最后 1 架无人机对其获得图像,确认目标被摧毁。

3.3.2 云作战样式 OODA 循环

云作战样式的 OODA 循环过程如图 5 所示。首先,由 2 架无人机完成观察阶段,快速深入敌方探测敌方目标信息,并将有效信息传回有人机;然后,有人机完成调整阶段和决策阶段,有人机对无人机传回的大量信息进行分析,并根据敌我双方的作战态势形成决策指令,并下达给携带有武器的后方无人机;最后,后方的 2 架无人机完成行动阶段,根据收到的指令及目标的相关信息,发射机载导弹对目标进行攻击。

由此可知,有人机利用其优秀的信息融合能力,真正实现了云作战样式中的跨域共享和多平台协同,有人机不仅可以将威胁锁定,还可以对战场信息做出快速反应,指挥无人机有序地完成任务。云作战的这种优势,可以大大缩短每一个 OODA 循环的过程,而且可以避免战场资源的浪费。

云作战样式的优势在于:对战场进行侦查的无人机由于飞行速度较快,可以对敌方目标实施快速探测,在短时间内获得大量敌方目标的信息,并将这些信息传送给有人机进行处理。这种情况下,有人机对敌方雷达的电子压制距离明显增长。

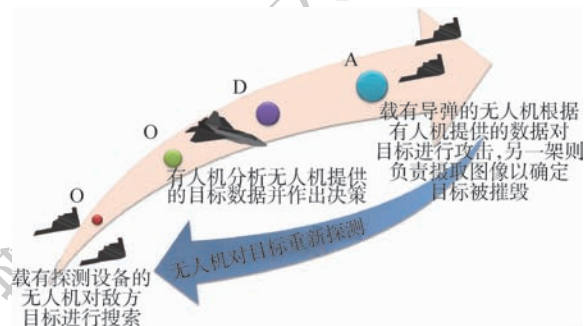


图 5 云作战样式的 OODA 循环示意图

Fig. 5 Schematic of OODA circles of cloud operations style

3.3.3 云作战样式的仿真实现

1) 战场态势共享

有人机作为交战过程中的指挥平台,实时接收其他作战单位的信息,掌握战场中敌我双方的态势,并将态势与相关决定共享给局部网络中的其他作战单位,如图 6 所示。

2) 云攻击样式

云攻击作战样式主要强调无人机的重要作用,即在 OODA 作战循环中无人机为全过程的主要参与者。在案例中,无人机从出发就与有人机

进行实时通信,并且无人机在从发现目标到击毁目标的整个作战过程中,参与程度很高。如高空侦查无人机对敌方目标进行侦查,携带导弹的无人机对敌方目标进行攻击,最后方的无人机实时观测敌方目标并返回敌方目标的状态,如图 7 所示。案例中的这种作战样式体现了云攻击作战样式。

3) 云射击样式

云射击作战样式主要强调机群采用信息共享的方式,依靠群体对目标发动攻击。在案例中,对每一个敌方目标的打击过程均需依靠高空侦察无人机的探测能力、有人机的电子压制能力、携带导弹无人机的攻击能力和后方无人机的观测能力,各个作战单位均能实现数据共享,如图 8 所示。案例中的这种作战样式体现了云射击作战样式。



图 6 基于有人机的云数据链

Fig.6 Cloud data link based on manned fighter

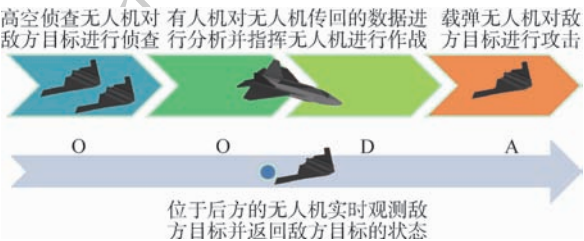


图 7 案例中云攻击样式的体现

Fig.7 Demonstration of "cloud attack" style in example

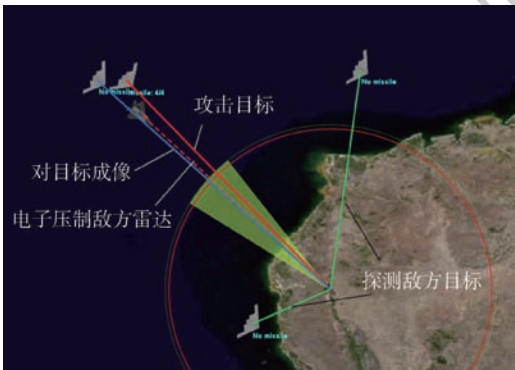


图 8 案例中云射击样式的体现

Fig.8 Demonstration of "cloud shooting" style in example

4) 群控制样式

群控制作战样式主要强调战斗机对多个无人作战单位的指挥。在案例中,有人机在出发时即与 4 个无人机实现通信链接,并指挥具有不同作战能力的无人机执行不同的作战任务,这种控制过程一直持续到结束任务并返回基地,如图 9 所示。案例中的这种作战样式体现了群控制作战样式。

5) 云制导样式

云制导作战样式主要强调武器在攻击过程中其他作战单位对其进行导航控制,使其能够精确打击目标。案例中,在无人机发射导弹直到导弹击毁目标的过程中,作战机群的每一个作战单位就能与导弹实现信息共享,并为导弹提供目标的位置信息,从而精确打击目标,如图 10 所示。案例中的这种作战样式体现了云制导作战样式。



图 9 案例中群控制样式的体现

Fig.9 Demonstration of "group control" style in example

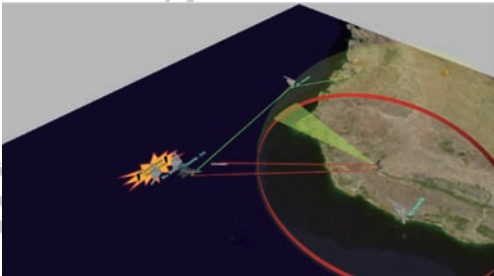


图 10 案例中云制导样式的体现

Fig.10 Demonstration of "cloud guidance" style in example

3.4 效能分析

战斗机作战对抗可以看作是敌我双方互相对量谁能更迅速地完成 OODA 循环。因此,可将 OODA 循环用时作为战斗机作战效能分析的指标。对比分析 2 种作战样式攻击同一个目标的作战时间,当无人机到达起飞地点与目标的中点位置时,即进入 OODA 循环的起点。2 种作战样式的 OODA 循环用时对比如图 11 所示。

传统作战样式中,观察阶段的时间为从循环起点到无人机发现目标的时间,调整阶段的时间

为从发现目标到有人机使敌方雷达失效的时间,决策阶段的时间为雷达失效后到地面指挥中心确认目标并指挥无人机发射导弹的时间,行动阶段的时间为从导弹发射到导弹摧毁目标的时间。

云作战样式中,观察阶段的时间为从循环起点到无人机发现目标的时间,调整阶段的时间为从发现目标到有人机开始对敌方雷达进行电子压制的时间,决策阶段的时间为开始压制敌方雷达到无人机发射导弹的时间,行动阶段的时间为从导弹发射到导弹摧毁目标的时间。

两者对比可以看出,在云作战样式中,当有人机接收到目标信息时,可以立即命令无人机发射导弹,在导弹飞行的过程中形成作战指令,并实时对导弹进行制导,这样先发射再制导的形式可以大大节约导弹的瞄准时间,使得 OODA 循环的时间缩短。

4 结 论

新的作战样式与作战体系一直是军事领域的研究热点,信息技术的迅猛发展与广泛应用也为军事创新提供了强劲动力。云作战是大数据、云计算、物联网、移动互联网等信息技术下的全新作战样式,通过信息共享和跨域协同,能够有效提高作战效率,不仅将改变空天作战样式,也将改变装备的发展路径,对未来的空战体系产生深远的影响。

本文提出的云作战仿真流程与云作战构造型仿真平台框架,仅仅是通过仿真的手段进行云作战体系研究的初步探索。关于云作战的应用还面临着不同来源信息的可靠性、云端系统的安全性、多平台的融合性、信息提取的高效性等诸多方面的挑战,需要进一步深入研究。

参考文献 (References)

[1] FAHRENKRUG D. 21st century warfare: The combat cloud [EB/OL]. (2014-09-15) [2019-01-09]. <http://secure.afa.org/events/Conference/2014/transcripts/Monday-3pm-CloudPanel.asp>.

[2] 胡悦. 美国空军“作战云”发展现状与展望[J]. 现代导航, 2017(1):74-78.

HU Y. Air Force “combat cloud” development and prospect [J]. Modern Navigation, 2017(1):74-78 (in Chinese).

[3] SCHANZ M V. Combat networks that can survive high-threat environments[J]. Air Force Magazine, 2014, 97(7):38-41.

[4] SCHANZ M V. The combat cloud[J]. Air Force Magazine, 2014(7):38-41.

[5] DEPTULA D A. A new era for command and control of aerospace operations[J]. Air & Space Power Journal, 2014(4):5-6.

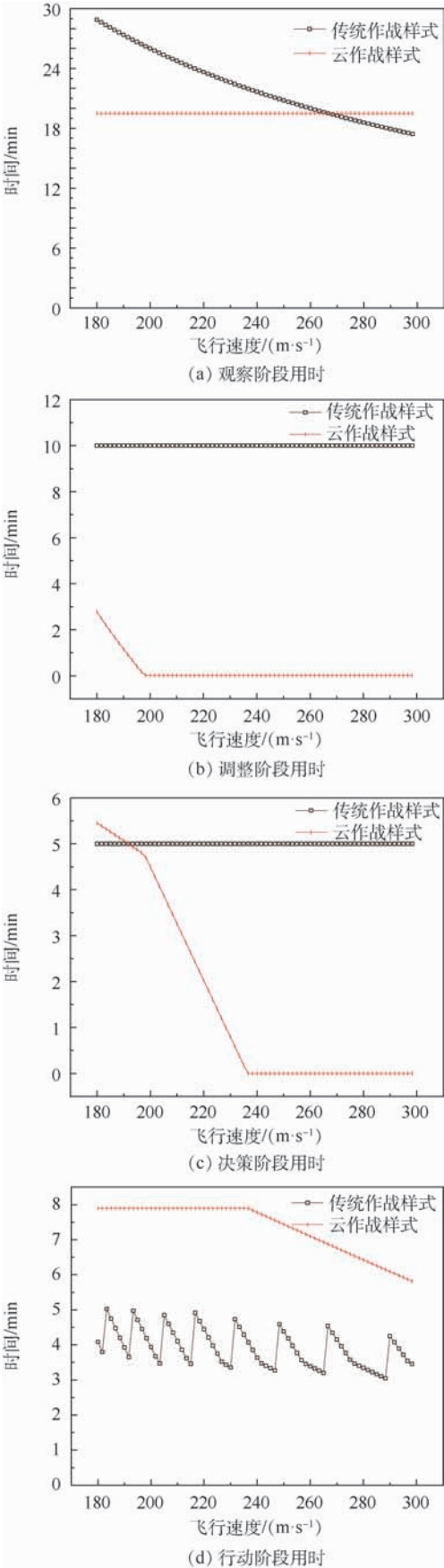


图 11 两种作战样式 OODA 循环用时对比
Fig. 11 OODA circle time comparison between two combat styles

- [6] 罗金亮,宿云波,张恒新. “作战云”体系构建初探[J]. 火控雷达技术,2015,44(3):26-30.
LUO J L, SU Y B, ZHANG H X. Research on system construction of combat cloud[J]. Fire Control Radar Technology, 2015, 44(3):26-30 (in Chinese).
- [7] GOZTEPE K, CEHRELI I, SENSOY S E. A decision framework for combat cloud computing strategy[C]//6th International Information Security & Cryptology Conference, 2013, 9:28-31.
- [8] 李飞,胡荣,胡剑波,等. 空天云作战指挥控制效能评估[C]//第六届中国指挥控制大会. 北京:电子工业出版社, 2018:904-909.
LI F, HU R, HU J B, et al. Research on command and control effectiveness evaluation of aerospace cloud operations[C]//6th China Command and Control Conference Proceedings. Beijing: Publishing House of Electronics Industry, 2018:904-909 (in Chinese).
- [9] 梁维泰,毛晓彬,黄松华. 面向空中战斗云的协同任务规划框架研究[C]//第六届中国指挥控制大会. 北京:电子工业出版社, 2018:37-41.
LIANG W T, MAO X B, HUANG S H. Framework of cooperative mission planning for air combat cloud[C]//6th China Command and Control Conference Proceedings. Beijing: Publishing House of Electronics Industry, 2018:37-41 (in Chinese).
- [10] 时东飞,蔡疆,黄松华,等. 美国空军“战斗云”作战理念及启示[J]. 指挥信息系统与技术,2017,8(3):27-32.
SHI D F, CAI J, HUANG S H, et al. Operational concept and enlightenment of United States Air Force “combat cloud”[J]. Command Information System and Technology, 2017, 8(3):27-32 (in Chinese).
- [11] 刘鹏,戴锋,闫坤. 基于复杂网络的“云作战”体系模型及仿真[J]. 指挥控制与仿真,2016,38(6):6-11.
LIU P, DAI F, YAN K. System model and simulation of “cloud operations” based on complex network[J]. Command Control & Simulation, 2016, 38(6):6-11 (in Chinese).
- [12] WAN K, GAO X, LIU X, et al. A cloud cooperative attack system for networking anti-stealth combat[C]//2013 4th IEEE International Conference on Software Engineering and Service Science (ICSESS). Piscataway, NJ: IEEE Press, 2013:515-520.
- [13] MA S D, ZHANG H Z, YANG G Q. Target threat level assessment based on cloud model under fuzzy and uncertain conditions in air combat simulation[J]. Aerospace Science and Technology, 2017, 67(8):49-53.
- [14] HUANG Y Y. Modeling and simulation method of the emergency response systems based on OODA[J]. Knowledge-Based Systems, 2015, 89:527-540.
- [15] 赵国宏. 作战云体系结构研究[J]. 指挥与控制学报, 2015, 1(3):292-295.
ZHAO G H. Architecture of combat clouds[J]. Journal of Command and Control, 2015, 1(3):292-295 (in Chinese).

作者简介:

田永亮 男,博士。主要研究方向:飞行器总体设计、隐身设计、体系化设计与运用支持技术。

王永庆 男,博士,研究员。主要研究方向:飞行器总体设计、隐身设计。

Structured simulation platform architecture for fighter cloud operations

TIAN Yongliang^{1,*}, WANG Yongqing^{1,2}, XIONG Peisen¹, GUO Yu², WU Zhe¹

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China;

2. Shenyang Aircraft Design and Research Institute, Shenyang 110035, China)

Abstract: With the rapid development and wide application of information technology such as big data, cloud computing, Internet of Things, and mobile Internet, new combat modes continue to emerge. The “cloud operation” with the core of the task distributed command & control process becomes a brand-new cross-domain full-dimensional combat style. Based on the analysis of the feature of combat cloud and cloud operation, combined with conventional combat simulation process, cloud operation system simulation process is proposed, and the cloud operations structured simulation platform architecture design and system function design are proposed. Through cloud operation structured simulation examples, it compares the observe-orient-decide-act (OODA) cycle of traditional combat styles with cloud operation styles. The results show that the cloud operation style can effectively shorten the OODA cycle time.

Keywords: cloud operation; combat cloud; modeling; simulation; observe-orient-decide-act (OODA)

Received: 2019-02-16; Accepted: 2019-03-29; Published online: 2019-06-18 17:43

URL: kns.cnki.net/kcms/detail/11.2625.V.20190617.1409.007.html

* Corresponding author. E-mail: tianyongliang1@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0036

调频引信粗糙面目标与干扰信号识别



郝新红*, 杜涵宇, 陈齐乐

(北京理工大学 机电动态控制重点实验室, 北京 100081)

摘 要: 为有效识别调频(FM)引信地面目标回波信号和数字射频存储(DRFM)转发式干扰信号,提出了一种对地调频引信粗糙面目标与干扰信号识别方法。建立了对地调频引信粗糙面差频信号模型,利用二维距离-速度提取方法提取其差频频率和多普勒频率,采用差频频率峰值带宽和多普勒频率峰值带宽2个特征量识别地面目标回波信号和DRFM转发式干扰信号,并使用蒙特卡罗仿真和非参数假设统计检验对其进行了验证。结果表明:差频频率峰值带宽和多普勒频率峰值带宽的展宽幅度与引信天线照射范围内粗糙面的大小呈正相关,多普勒频率峰值带宽的展宽与载频成正比,利用峰值带宽特性可有效区分粗糙面目标回波信号和DRFM转发式干扰信号。

关 键 词: 调频(FM)引信;粗糙面目标;二维距离-速度提取;数字射频存储(DRFM)干扰;非参数假设检验

中图分类号: TJ43+4.1

文献标识码: A **文章编号:** 1001-5965(2019)10-1946-10

调频(FM)无线电引信具有定距精度高、算法相对简单、硬件容易实现等优点,目前已经在常规弹药中得到了大规模应用^[1-3]。在调频引信信号处理算法设计过程中,通常认为引信目标是点目标,但实际对地引信的目标大地表面是一种粗糙的面目标。随着无线电引信干扰技术的不断发展,以数字射频存储(DRFM)干扰为代表的转发式干扰已经可以完全模拟点目标回波^[4-5],这给调频引信造成了致命威胁。

在无线电引信和雷达领域已经有大量对于面目标特性的研究,对面目标特性的研究主要有2种方法。一种是使用多点散射叠加模拟面目标或体目标回波信号^[6-9],如文献[6]使用3个点模拟面目标,指出多点目标回波造成引信多普勒信号频谱展宽,而且弹目越接近,频谱展宽越大;文献[7]建立了引信远场单点目标和近场多点目标

模型,将信号的调幅带宽和调频带宽作为特征参量,对干扰信号和体目标信号进行分类识别。但是多点散射叠加建模方法比较简单,DRFM干扰系统可以模拟^[10-11]。另一种方法是通过近似法研究粗糙面散射特性^[12-15],在雷达领域较为常见,如文献[12]用随机粗糙面模拟实际地面,利用基尔霍夫近似(KA)法和微扰近似法(SPM)计算了粗糙面的雷达后向和双站散射系数;文献[13]分析了连续波调频雷达高度表随机粗糙面回波和差频信号的统计特性,对差频信号的二维功率谱密度进行了计算和仿真。基于近似法的粗糙面目标建模方法更贴近实际情况,可以用于引信面目标建模。

针对对地调频引信粗糙面目标与干扰信号识别的问题,本文对调频引信随机粗糙面回波信号和差频信号进行了建模,利用二维距离-速度提取

收稿日期: 2019-01-23; 录用日期: 2019-05-18; 网络出版时间: 2019-06-11 15:15

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190610.0900.001.html

基金项目: 国家自然科学基金(61871414)

* 通信作者: E-mail: haoxinhong@bit.edu.cn

引用格式: 郝新红, 杜涵宇, 陈齐乐. 调频引信粗糙面目标与干扰信号识别[J]. 北京航空航天大学学报, 2019, 45(10): 1946-1955.
HAO X H, DU H Y, CHEN Q L. Rough surface target and jamming signal recognition of FM fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1946-1955 (in Chinese).

方法提取其差频频率和多普勒频率,对点目标和随机粗糙面目标作用下调频引信差频信号的二维频率特性进行了仿真。

1 调频引信粗糙面差频信号建模

根据瑞利准则,对于工作在 S 波段的调频引信,稍有起伏的土地或覆盖植被的地面即可被认为是粗糙面。当粗糙面的表面相关长度和高度起伏均方差大于波长时,可以使用基尔霍夫近似法进行粗糙面回波信号建模。本文使用一维随机粗糙面代替粗糙地面进行建模,即在 xOy 平面内,只有 x 维度存在粗糙度,并假设其粗糙度参数和引信工作波长符合基尔霍夫近似法要求。

设一维随机粗糙面高度起伏为 $z=f(x)$,其参考平面为 $z=0$,引信在 $t=0$ 时刻高度为 H ,横坐标为 $x=0$,运动速度为 v ,运动方向与垂直方向夹角为 α ,在引信天线主瓣辐射范围内一维粗糙面上点 $P(x,f(x))$,引信到 P 点距离为 ρ ,引信与 P 点连线和垂直方向夹角为 θ ,引信与粗糙面的位置关系如图 1 所示。

若引信发射频率为 f_0 的点频信号,则根据基尔霍夫近似法计算,粗糙面散射后引信接收到回波信号为^[13]

$$s_{f_0}(t) = -\frac{\sqrt{k}C}{8\pi^{3/2}}e^{j2\pi f_0 t}e^{j\frac{\pi}{4}}\int_{-\infty}^{+\infty}\frac{F^2(\theta)}{\rho^{3/2}}e^{-j2k\rho}e^{j2kf(x+vt\sin\alpha)\cos\theta}dx \quad (1)$$

式中: $k=2\pi f_0/c$ 为发射信号的波数, c 为光速; $\rho=\sqrt{(x-vt\sin\alpha)^2+(H-vt\cos\alpha)^2}$; C 为与天线辐射功率、发射和接收天线增益有关的常数; $F(\theta)$ 为天线方向性函数。

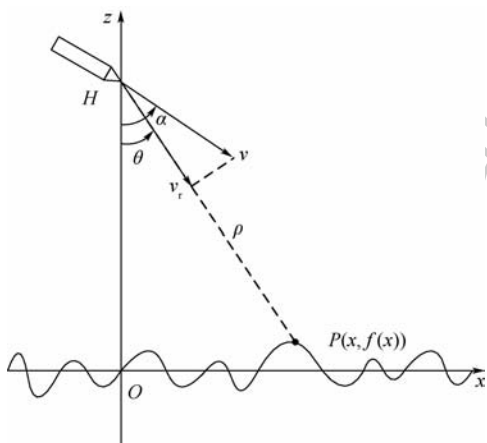


图 1 引信与粗糙面的位置关系

Fig. 1 Positional relationship between fuze and rough surface

设引信发射信号为锯齿波调频信号 $s_i(t)$,将 $s_i(t)$ 表示成傅里叶级数的形式:

$$s_i(t) = \sum_{i=1}^{+\infty} a_i e^{j2\pi i f_m t} \quad (2)$$

式中: i 为谐波次数 ($i=1,2,\dots,+\infty$); a_i 为发射信号的傅里叶系数; f_m 为信号调频。

根据式(1)、式(2)可得出调频引信粗糙面回波信号为

$$s_r(t) = -\sum_{i=1}^{+\infty} a_i e^{j2\pi i f_m t} \frac{\sqrt{k_i}C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} \int_{-\infty}^{+\infty} \frac{F^2(\theta)}{\rho^{3/2}} e^{-j2k_i\rho} e^{j2k_i f(x+vt\sin\alpha)\cos\theta} dx \quad (3)$$

式中: k_i 为第 i 次谐波的波数, $k_i=2\pi i f_m/c$ 。

假设锯齿波调频信号调制频偏 $\Delta F=50$ MHz,根据差频谐波次数与距离对应关系,10 次调频谐波的包络峰值对应 30 m 的弹目距离,因此可以认为在发射信号带宽内 $\sqrt{k_i}$ 的值变化幅度不大,可近似取得 $\sqrt{k_i} \approx \sqrt{k_1}$,其中 $k_1=2\pi f_m/c$ 。则式(3)可以改写为

$$s_r(t) = -\frac{\sqrt{k_1}C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} \int_{-\infty}^{+\infty} \frac{F^2(\theta)}{\rho^{3/2}} \sum_{i=1}^{+\infty} a_i e^{j2\pi i f_m [t - \frac{2\rho - 2f(x+vt\sin\alpha)\cos\theta}{c}]} dx \quad (4)$$

把调频引信粗糙面回波信号与本地参考信号进行混频,得到其差频信号。根据式(2)、式(4)可以得到粗糙面作用下调频引信差频信号表达式为

$$s_{if}(t) = s_r(t) \cdot s_i^*(t) = -\frac{\sqrt{k_1}C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} \int_{-\infty}^{+\infty} \frac{F^2(\theta)}{\rho^{3/2}} s_i\left(t - \frac{2\rho - 2f(x+vt\sin\alpha)\cos\theta}{c}\right) s_i^*(t) dx \quad (5)$$

式中:“ $*$ ”表示共轭。

假设发射信号为锯齿波调频信号 $s_i(t) = Ae^{j2\pi(f_c t + \frac{1}{2}\beta t^2)}$, $t \in [0, T]$, A 为视频包络的幅度, f_c 为载频, $\beta = \Delta F/T$ 为调制斜率, T 为锯齿波调频周期。将其代入式(5),设 $\tau = 2[\rho - f(x+vt\sin\alpha)\cos\theta]/c$ 为引信发射信号经粗糙面上 P 点反射回引信接收端的传输时延,可得第 1 个调制周期内差频信号为

$$s_{if}(t) = -\frac{\sqrt{k_1}C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} \int_{-\infty}^{+\infty} \frac{F^2(\theta)}{\rho^{3/2}} A e^{-j2\pi(f_c \tau + \beta t \tau - \frac{1}{2}\beta \tau^2)} dx \quad t \in [\tau, T + \tau) \quad (6)$$

对于工作在 S 波段的调频引信, $\frac{1}{2}\beta \tau^2$ 要远小于 $f_c \tau$,因此可忽略 $\frac{1}{2}\beta \tau^2$,由此可得粗糙面作用

下第 $n+1$ 个调制周期调频引信差频信号数学模型为

$$s_{\text{IF}}(t, n) = -\frac{\sqrt{k_1} C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} A \int_{-\infty}^{+\infty} \frac{F^2(\theta)}{\rho^{3/2}} \cdot e^{-j2\pi[f_c + \beta(t-nT)]\tau} dx = -\frac{\sqrt{k_1} C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} A \int_{-\infty}^{+\infty} \frac{F^2(\theta)}{\rho^{3/2}} \cdot e^{-j4\pi[f_c + \beta(t-nT)][\rho - f(x + vt\sin\alpha)\cos\theta]/c} dx \quad (7)$$

式中: $n=0, 1, \dots, +\infty$ 。

2 调频引信二维距离-速度提取方法

传统的调频引信大多采用调频谐波定距方法,其信号处理方法是将对差频信号进行二次混频和多普勒检波,提取出特定次谐波的多普勒检波信号^[2]。调频谐波定距方法本质是从频域提取与距离信息一一对应的差频频率,因此难以对抗能够模拟点目标回波的 DRFM 干扰。因此,本文提出一种基于粗糙面目标的调频引信距离-速度二维特征提取方法,直接对基于粗糙面目标特性的差频信号进行距离-多普勒处理,通过第 1 维傅里叶变换提取一个调制周期内与距离相对应的差频频率,再通过第 2 维傅里叶变换提取与速度相对应的多普勒频率,进而提取出其距离和速度信息。调频引信二维距离-速度提取方法示意图如图 2 所示,其中, f_B 为差频频率, f_D 为多普勒频率。

根据式(7),对 t 和 $f(x)$ 进行采样,由于引信在每个调制周期内移动距离远小于弹目距离,因此可以假设引信在每个调制周期内弹目距离不变,不同调制周期之间弹目距离发生改变。所以,第 $n+1$ 个调制周期初始采样时刻引信与粗糙面上的点之间距离可表示为 $\rho_{n,q}$ =

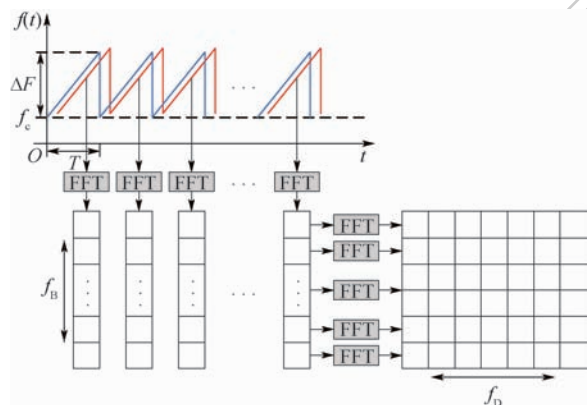


图 2 调频引信二维距离-速度提取方法

Fig. 2 Two-dimensional distance-speed

extraction method for FM fuze

$\sqrt{(x_q - vnT\sin\alpha)^2 + (H - vnT\cos\alpha)^2}$, 信号传输时延可表示为 $\tau_{n,q} = 2[\rho_{n,q} - f(x_q + vnT\sin\alpha) \cdot \cos\theta]/c$, x_q 为 x 轴上的采样点, $q=0, 1, \dots, Q-1, Q$ 为引信天线照射范围内对粗糙面采样的点数。随着弹目不断接近,天线辐射范围也在逐渐变小, Q 的值也在不断变小, Q 的变化满足:

$$Q = \left\lfloor (H - vnT\cos\alpha) \left(\tan\left(\frac{\theta_B}{2} - \theta_0\right) + \tan\left(\frac{\theta_B}{2} + \theta_0\right) \right) / \Delta x \right\rfloor \quad (8)$$

式中: θ_B 为天线 3 dB 波束宽度; θ_0 为天线波束中心与垂直方向夹角; Δx 为 x 轴采样间隔; $\lfloor \cdot \rfloor$ 表示向下取整。

则第 $n+1$ 个调制周期的差频信号为

$$s_{\text{IF}}(m, n, q) = -\frac{\sqrt{k_1} C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} A \cdot \sum_{q=0}^{Q-1} \frac{F^2(\theta)}{\rho_{n,q}^{3/2}} e^{-j2\pi(f_c + \beta\frac{T}{M}m)\tau_{n,q}} \Delta x \quad n=0, 1, \dots, N-1; m=0, 1, \dots, M-1; q=0, 1, \dots, Q-1 \quad (9)$$

式中: M 为一个调制周期内采样的点数。

把每个周期的差频信号 $s_{\text{IF}}(m, n, q)$ 作为列向量,取 N 个周期的差频信号依次排列,即得到一个 $M \times N$ 的矩阵 $s_{\text{IF}}(m, n, q)$ 。对此矩阵做二维离散傅里叶变换:

$$S_{\text{IF}}(p, l, q) = \sum_{m=1}^M \sum_{n=1}^N s_{\text{IF}}(m, n, q) e^{-j2\pi\frac{m}{M}p} e^{-j2\pi\frac{n}{N}l} = -\frac{\sqrt{k_1} C}{8\pi^{3/2}} e^{j\frac{\pi}{4}} A \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{q=0}^{Q-1} \frac{F^2(\theta)}{\rho_{n,q}^{3/2}} \Delta x \cdot e^{-j2\pi(2\beta T \frac{\rho_{n,q} - f(x + vnT\sin\alpha)\cos\theta}{c} + p) \frac{m}{M}} \cdot e^{-j2\pi(2f_c \frac{\rho_{n,q} - f(x + vnT\sin\alpha)\cos\theta}{c} + l) \frac{n}{N}} \quad (10)$$

由式(10)可以看出,粗糙面散射使得距离对应的差频频率和速度对应的多普勒频率存在更多分量,造成差频频率和多普勒频率都存在展宽,同时也造成了距离和速度的耦合。为了加快运算速度,二维距离-速度提取方法可以采用二维快速傅里叶变换(FFT)实现,而为了进一步细化频谱,更好地提取目标特性,可以采用补零二维 FFT。

3 仿真与讨论

3.1 模拟点目标 DRFM 干扰建模与仿真

考虑对地调频引信的理想情况,若地面目标为理想导体平面,引信天线波束中心与地面垂直,则地面反射为镜面反射,目标可等效为位于引信正下方的点目标,其到引信的距离为引信实际高

度的 2 倍。在此条件下,对引信目标特性进行仿真,设引信在 $t=0$ 时刻弹目距离为 $R=20\text{ m}$,弹目相对速度为 $v_r=500\text{ m/s}$;引信发射信号为锯齿波调频信号,载频为 3 GHz ,调制频偏为 50 MHz ,调频为 100 kHz ,仿真 200 个调频周期,在这一段时间内弹目接近了 1 m ,在炸高较高时对引战配合影响不大。使用 MATLAB 对点目标作用下引信差频信号进行仿真,并对差频信号进行二维补零 FFT,提取目标二维频率特性,仿真结果如图 3 所示。

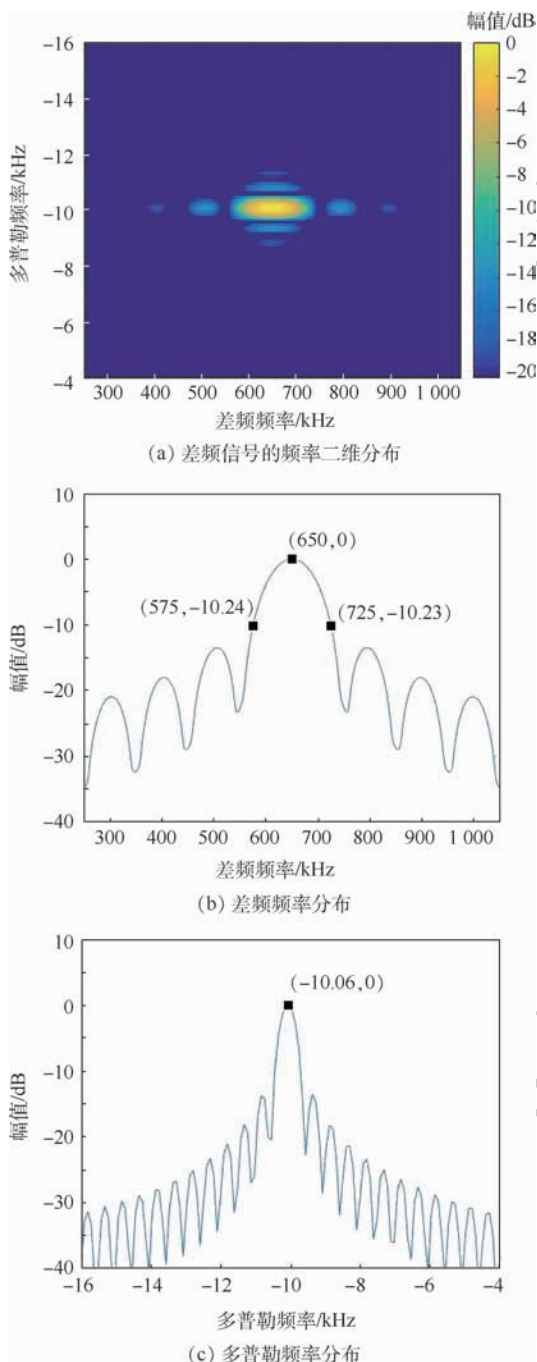


图 3 点目标作用下调频引信差频频率与多普勒频率分布

Fig. 3 Beat frequency and Doppler frequency distribution of FM fuze under action of point target

对于点目标或完全模拟点目标的 DRFM 干扰,调频引信的差频频率和多普勒频率都近似为点频,因此对差频信号进行采样相当于乘以宽度为 T 的矩形窗,对其进行 FFT 后其频谱有近似的 sinc 包络,包络的主瓣宽度为 $2/T$,本文选取差频频谱和多普勒频谱主瓣 -10 dB 带宽作为差频频率和多普勒频率峰值带宽。

从仿真结果可以看出,点目标作用下调频引信差频二维频率分布近似为一个点,差频频率分布是二维频率分布的横坐标投影,多普勒频率分布是二维频率分布的纵坐标投影;差频频率峰值带宽为 150 kHz ,多普勒频率峰值带宽不足 1 kHz ;差频频率在 650 kHz 处出现最大值,多普勒频率在 10.06 kHz 处出现最大值。弹目距离 20 m 对应差频频率理论值为 666.7 kHz ,相对速度 500 m/s 对应多普勒频率理论值为 10 kHz ,由于实际差频频率等于距离差频减去多普勒频率,差频频率出现了 10 kHz 的偏移。因此,仿真频率最大值基本对应仿真设置的距离和速度,仿真结果表明,二维距离-速度提取方法可以提取目标的距离和速度信息。

对于完全模拟点目标回波的 DRFM 干扰信号,调频引信二维频率分布等效于点目标,因此可以使用点目标仿真代替模拟点目标 DRFM 干扰。

3.2 粗糙面目标建模与仿真

根据瑞利准则,只有地面起伏远小于引信工作波长时,地面反射才可近似为镜面反射。而对于工作在 S 波段的调频引信来说,地面起伏通常可与波长比拟,甚至大于波长。因此,粗糙面目标作用下调频引信目标特性更接近实际战场的情况。

本文采用高斯粗糙面模拟粗糙地面,这里选取一维高斯粗糙面。高斯分布的功率谱密度为

$$G(s) = \delta L \sqrt{\pi} \exp\left(-\frac{s^2 L^2}{4}\right) \quad (11)$$

式中: δ 为高斯分布的标准差; L 为表面相关长度。

仿真生成的高斯粗糙面局部如图 4 所示。

以高斯粗糙面作为模拟地面目标,建立引信弹目交会模型。为了更符合实际情况,假设引信天线 3 dB 波束宽度为 $\theta_B=60^\circ$,天线波束中心与垂直方向夹角为 15° ,天线方向性函数为高斯函数:

$$\begin{cases} F(\theta) = \exp(-a^2 \tan^2 \theta) \\ a = \frac{1}{2} \ln 2 / \tan^2 \frac{\theta_B}{2} \end{cases} \quad (12)$$

其他仿真参数与点目标一致,对调频引信粗糙面目标特性进行仿真,对差频信号进行二维补

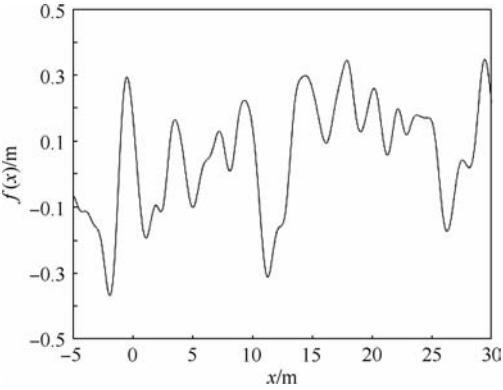


图 4 高斯粗糙面局部示意图

Fig. 4 Partial schematic of Gaussian rough surface

零 FFT, 单次仿真的结果如图 5 所示。

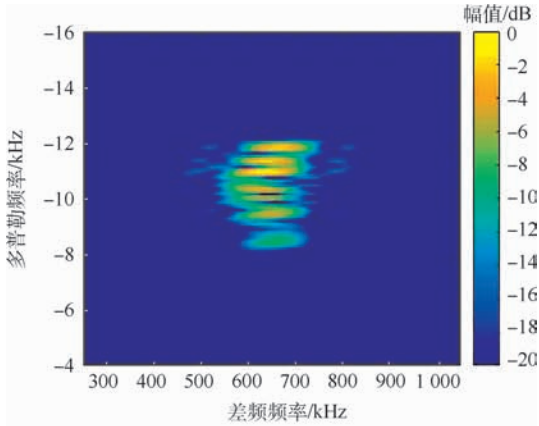
从单次仿真的结果可以看出, 与点目标作用下二维频率分布不同, 粗糙面目标作用下调频引信二维频率分布存在一定随机性; 差频频率峰值带宽为 182.8 kHz, 多普勒频率峰值带宽约为 4 kHz; 差频频率和多普勒频率都出现了明显的展宽, 峰值点也都出现了一定程度的偏移。由于每次仿真程序生成的随机粗糙面都不一样, 所以每次仿真得到的结果也不太一样, 对粗糙面目标特性进行 100 次仿真, 对其结果取平均, 得到仿真结果如图 6 所示。

经过 100 次仿真对其结果取平均, 可以得到粗糙面作用下调频引信差频信号的二维频率分布范围, 可以看出, 粗糙面目标作用下差频频率和多普勒频率均出现展宽, 多普勒频率展宽更为明显, 其高频处能量相对较高; 同时出现了距离和速度的耦合, 造成差频频率和多普勒频率峰值点偏移。

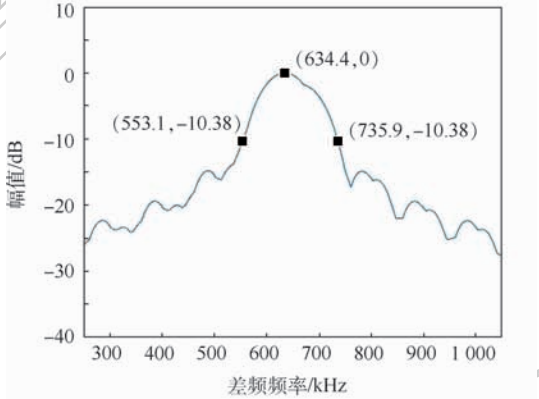
3.3 讨论

为了探讨调频引信的载频对粗糙面作用下调频的二维频率分布的影响, 对粗糙面作用下不同频段的调频引信差频频率与多普勒频率分布进行了仿真, 结果如图 7 和图 8 所示。

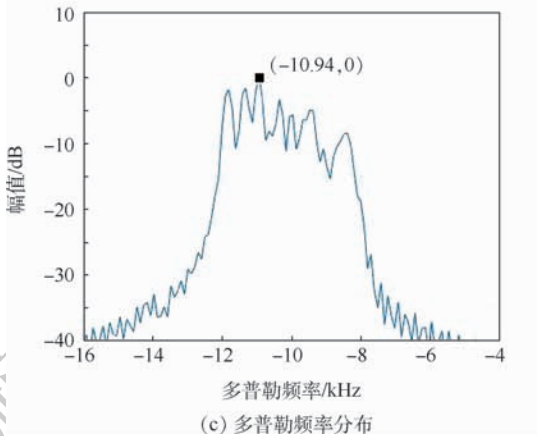
图 7 和图 8 中, 仿真的引信载频分别为 10、24 GHz, 差频频率峰值带宽分别为 217.2、875 kHz, 多普勒频率峰值带宽分别为 12、30 kHz。仿真生成的粗糙面起伏度相同, 由仿真结果可以看出, 随着引信载频的增加, 粗糙面目标作用下引信多普勒频率峰值带宽也随之增加, 带宽大致与载频成正比。差频频率峰值带宽的展宽不明显, 对比不同波段仿真结果可知, 载频对差频频率峰值带宽影响较小。K 波段由于多普勒频率较高, 将仿真的调频提高到了 400 kHz, 可以看出, 差频频率峰值带宽也相应成倍数增加, 与 3.1 节分析一致。



(a) 差频信号的频率二维分布



(b) 差频频率分布



(c) 多普勒频率分布

图 5 随机粗糙面目标作用下调频引信差频频率与多普勒频率分布 (单次仿真)

Fig. 5 Beat frequency and Doppler frequency distribution of FM fuze under action of random rough surface target (single simulation)

为了研究引信天线照射范围内粗糙面区域大小对差频频率峰值带宽和多普勒频率峰值带宽的影响, 对不同高度处调频引信的差频二维频率分布进行多组仿真, 得到差频频率峰值带宽和多普勒频率峰值带宽的平均值随高度变化的规律, 如图 9 所示。可以看出, 差频频率峰值带宽和多普勒频率峰值带宽均随着高度的降低和天线照射范围内粗糙面区域变小而降低, 同时由于引信照射

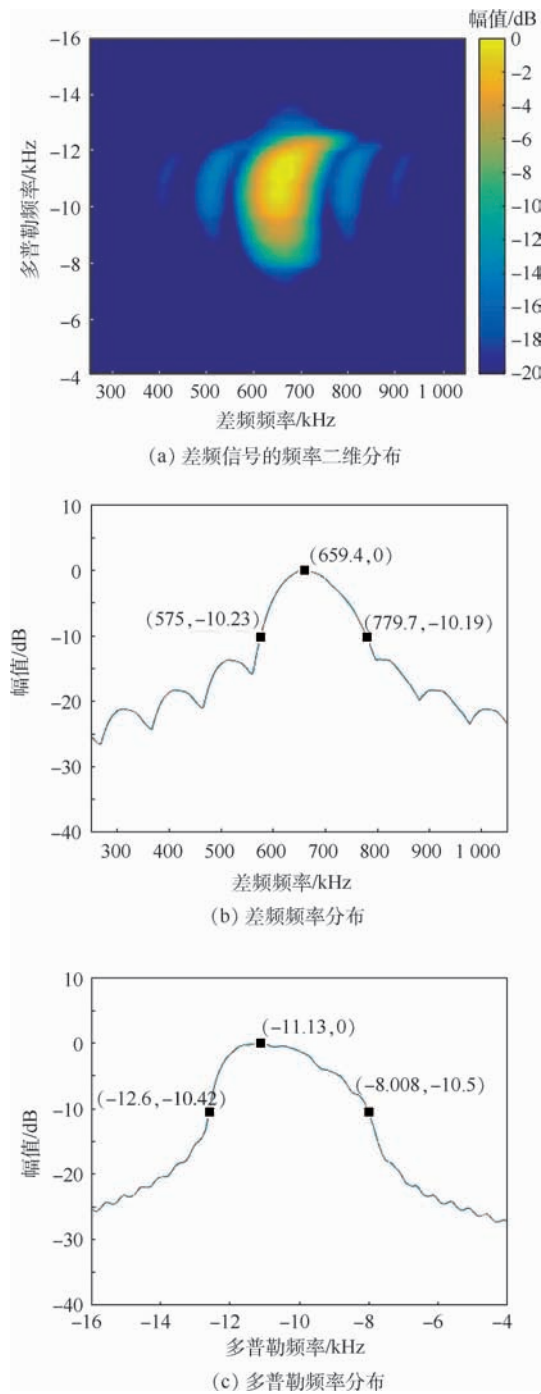


图 6 随机粗糙面目标作用下调频引信差频频率与多普勒频率分布(100 次仿真)

Fig.6 Beat frequency and Doppler frequency distribution of FM fuze under action of random rough surface target (100 simulations)

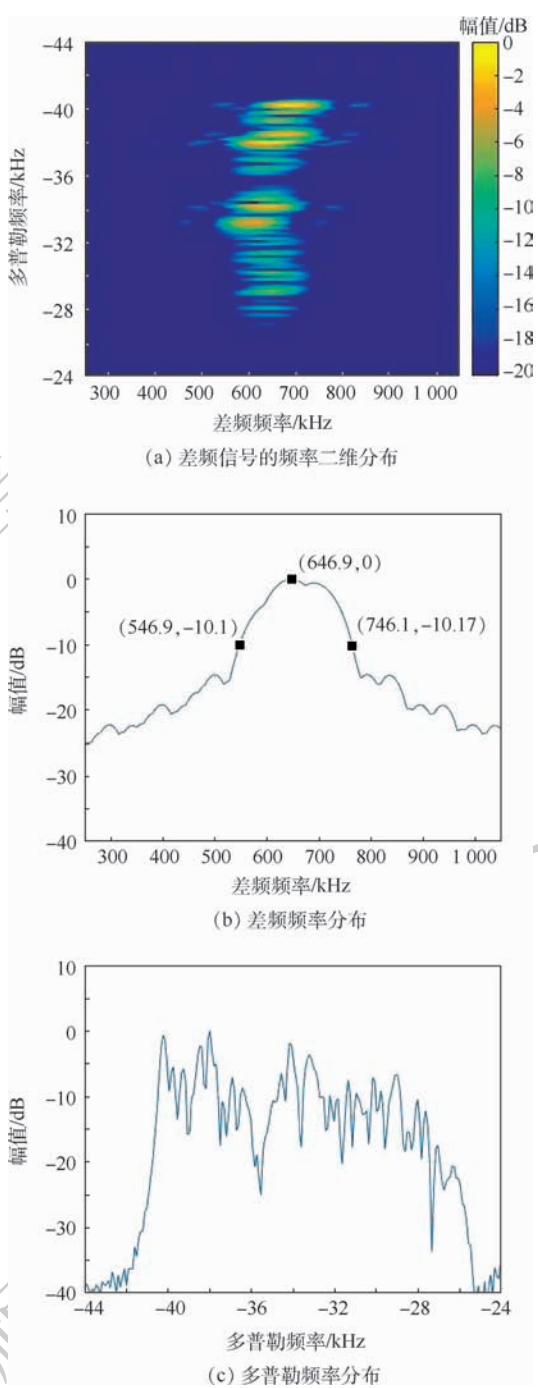


图 7 随机粗糙面目标作用下 X 波段调频引信差频频率与多普勒频率分布

Fig.7 Beat frequency and Doppler frequency distribution of X-band FM fuze under action of random rough surface target

到的区域只是高斯粗糙面的局部,其高度起伏也存在差异,导致带宽随高度的变化出现波动。由仿真可以看出,引信天线照射范围内粗糙面的高度起伏造成调频引信差频频带和多普勒频带展宽,天线照射范围内粗糙面的大小影响了差频频率峰值带宽和多普勒频率峰值带宽,同时,多普勒频率峰值带宽还随着引信载频变化而变化,其大

致与载频成正比。

为了验证地面目标的多普勒展宽特性,基于收发共用天线的 X 波段调频多普勒引信样机,分别测试了针对点目标与土地目标的滑弹试验,滑弹高度 60 m,入射角 45°,弹速 15 ~ 30 m/s。则调频谐波定距的多普勒包络信号输出波形分别如图 10 与图 11 所示。对比分析图 10 与图 11 可知,点

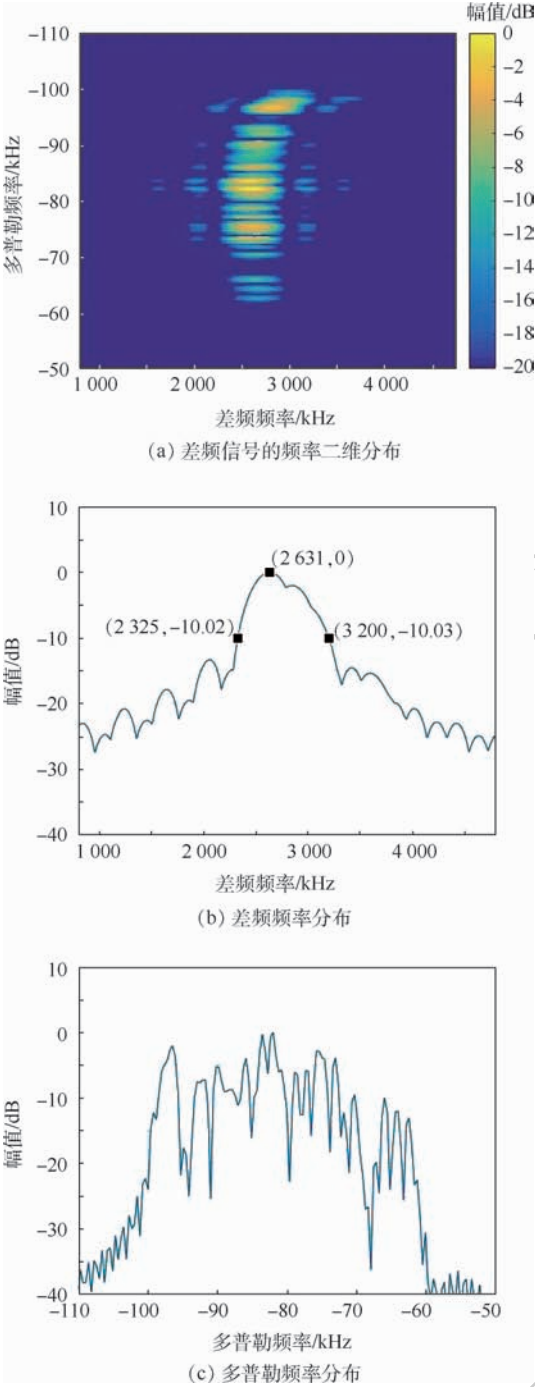


图8 随机粗糙面目标作用下K波段调频引信差频频率与多普勒频率分布

Fig.8 Beat frequency and Doppler frequency distribution of K-band FM fuze under action of random rough surface target

目标作用下多普勒频率近似为点频情况,粗糙面作用下调频引信多普勒输出信号频域存在明显展宽,且频带内高频处能量较高,实测多普勒输出频域波形与仿真得到的频率分布一致,验证了粗糙面目标建模的正确性。

而后进行了基于双通道调频谐波定距的调频多普勒引信多延时叠加DRFM干扰对抗试验,试

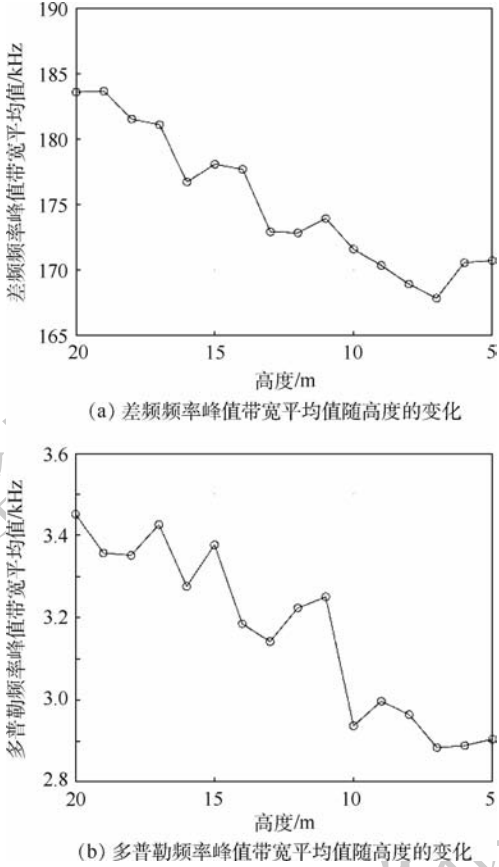


图9 差频频率峰值带宽和多普勒频率峰值带宽的平均值随高度的变化

Fig.9 Variation of average value of beat frequency peak bandwidth and Doppler frequency peak bandwidth with height

验测试结果如图12所示。可知,如果仅采用双通道谐波多普勒定距的方式,基于点目标模拟的DRFM干扰可有效干扰引信。

针对以调频引信差频频率峰值带宽和多普勒频率峰值带宽2个特征参量区分干扰与目标信号的验证,使用Wilcoxon秩和非参数假设检验来验证。Wilcoxon秩和检验又被称为Mann-Whitney秩和检验,是一种适用于两样本的非参数假设检验。Wilcoxon秩和检验计算后返回 P 值, P 值越小,表示两样本分布差异越显著。使用蒙特卡罗方法仿真100组调频引信在随机粗糙面和模拟点目标DRFM干扰作用下的二维频率分布,Wilcoxon秩和检验得到统计箱型图如图13所示。可以明显看出,随机粗糙面作用下调频引信差频频率峰值带宽和多普勒频率峰值带宽都分布在一定区间内,而模拟点目标DRFM干扰作用下引信差频频率峰值带宽和多普勒频率峰值带宽分布非常集中,几乎是固定值。

使用Wilcoxon方法计算得到差频频率峰值带宽 $P_b = 1.8265 \times 10^{-26}$,多普勒频率峰值带宽

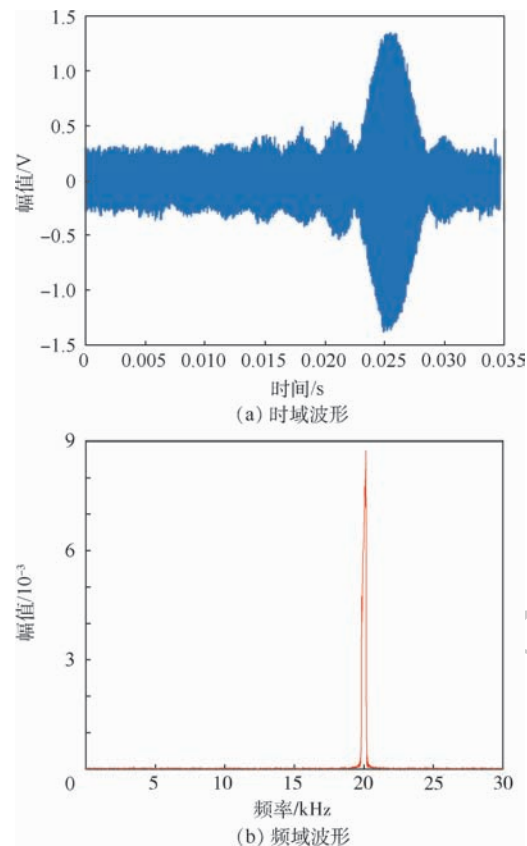


图 10 点目标作用下调频引信多普勒输出波形
Fig. 10 FM fuze Doppler output waveform under action of point target

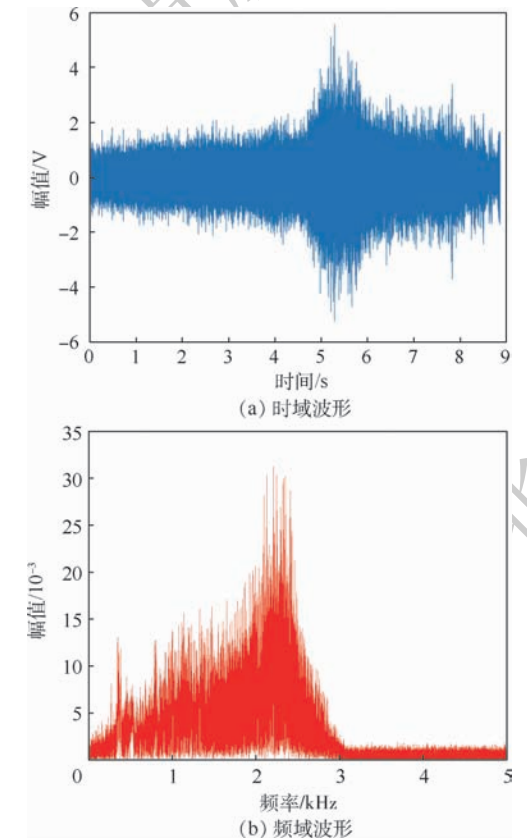


图 11 土地目标作用下调频引信多普勒输出波形
Fig. 11 FM fuze Doppler output waveform under action of earth target

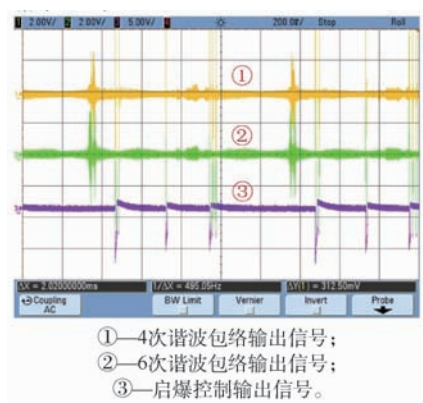


图 12 DRFM 干扰下引信对抗试验测试结果
Fig. 12 Fuze countermeasure test results under DRFM jamming

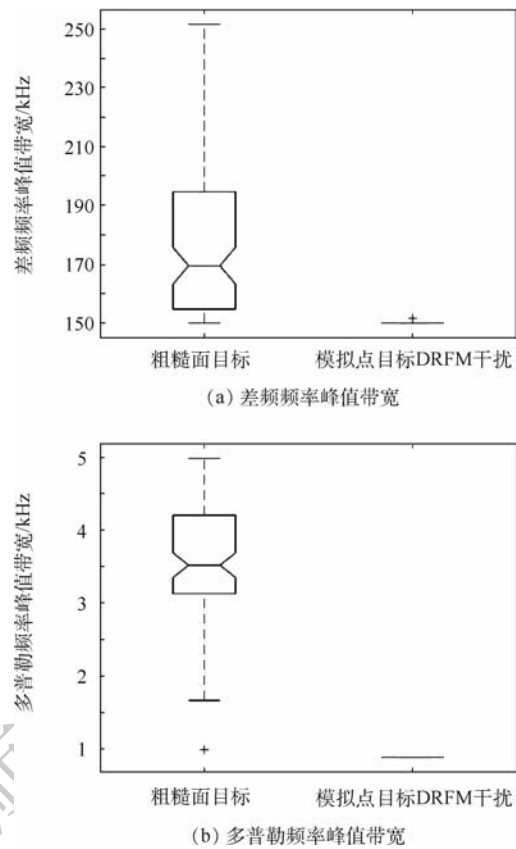


图 13 粗糙面目标和模拟点目标的 DRFM 干扰作用下
调频引信差频频率峰值带宽和多普勒频率
峰值带宽箱型图
Fig. 13 Box diagram of beat frequency peak bandwidth and
Doppler frequency peak bandwidth of FM fuze under action of
rough surface targets and DRFM jamming signals
simulating point targets

$P_D = 5.4838 \times 10^{-39}$ 。计算结果表明,粗糙面目标和模拟点目标 DRFM 干扰作用下引信差频频率峰值带宽和多普勒频率峰值带宽差异极为显著,可有效区分粗糙面目标和模拟点目标的 DRFM 干扰信号。

4 结 论

1) 粗糙面目标作用下,调频引信差频频率峰值带宽和多普勒频率峰值带宽展宽的幅度与引信天线照射范围内粗糙面的大小呈正相关,粗糙面高度起伏造成展宽量出现随机的波动。

2) 粗糙面目标作用下,调频引信多普勒频率展宽更为明显,其高频处能量相对较高,峰值带宽展宽幅度与载频成正比,实测引信多普勒频域波形与仿真多普勒频率分布一致。

3) 蒙特卡罗仿真和非参数假设检验计算表明,利用差频频率峰值带宽和多普勒频率峰值带宽特征可有效区分粗糙面目标和模拟点目标的DRFM干扰信号。

参考文献 (References)

- [1] 崔占忠,宋世和,徐立新.近炸引信原理[M].3版.北京:北京理工大学出版社,2009:76-103.
CUI Z Z, SONG S H, XU L X. Principle of proximity fuze[M]. 3rd ed. Beijing: Beijing Institute of Technology Press, 2009: 76-103 (in Chinese).
- [2] 赵惠昌.无线电引信设计原理与方法[M].北京:国防工业出版社,2012:35-93.
ZHAO H C. Fundamentals and methodology of radio fuze[M]. Beijing: National Defense Industry Press, 2012: 35-93 (in Chinese).
- [3] 孔志杰,郝新红,栗苹,等.调频引信谐波时序检测抗干扰方法及实现[J].北京航空航天大学学报,2018,44(3):549-555.
KONG Z J, HAO X H, LI P, et al. Harmonic timing sequence detection anti-jamming method and its implementation for FM fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(3): 549-555 (in Chinese).
- [4] ADAMY D. DRFMs-Part2[J]. Journal of Electronic Defence, 2013, 36(11): 44-46.
- [5] HANBALI S B, KASTANTIN R. A review of self-protection deceptive jamming against chirp radars[J]. International Journal of Microwave and Wireless Technologies, 2017, 9(9): 1853-1861.
- [6] 潘曦,崔占忠.无线电引信近场目标特性研究[J].兵工学报,2008,29(3):277-281.
PAN X, CUI Z Z. Near-field characteristic of target for radio fuze[J]. Acta Armamentarii, 2008, 29(3): 277-281 (in Chinese).
- [7] LI Z Q, HAO X H, CHEN H L, et al. Target signal recognition for CW Doppler proximity radio detector based on SVM[C]// International Conference on Mechatronic Sciences, Electric Engineering and Computer. Piscataway, NJ: IEEE Press, 2013: 1160-1163.
- [8] HUANG H, PAN M, LU Z. Hardware-in-the-loop simulation technology of wide-band radar targets based on scattering center model[J]. Chinese Journal of Aeronautics, 2015, 28(5): 1476-1484.
- [9] 肖泽龙,张恒,董浩,等.多普勒对空引信回波分析及碰炸优先判决准则研究[J].兵工学报,2016,37(10):1820-1827.
XIAO Z L, ZHANG H, DONG H, et al. Research on modeling and simulation of echo signal of pulse Doppler fuze and judgment criterion of its impact[J]. Acta Armamentarii, 2016, 37(10): 1820-1827 (in Chinese).
- [10] 杨旭.毫米波引信体目标模拟器的实现[J].航空兵器,2009(3):55-58.
YANG X. Implementation of a MMW fuze body target simulator[J]. Aero Weaponry, 2009(3): 55-58 (in Chinese).
- [11] 张珂,王震,舒建涛,等.基于数字射频存储的引信面目标回波模拟器[J].探测与控制学报,2016,38(5):15-21.
ZHANG K, WANG Z, SHU J T, et al. Echo simulator of area target based on DRFM[J]. Journal of Detection & Control, 2016, 38(5): 15-21 (in Chinese).
- [12] 宋运龙.实际粗糙地面的电磁散射特征研究[D].西安:西安电子科技大学,2008:21-32.
SONG Y L. Study on the characteristic of electromagnetic scattering from the actual rough land surface[D]. Xi'an: Xidian University, 2008: 21-32 (in Chinese).
- [13] MONAKOV A, NESTEROV M. Statistical properties of FMCW radar altimeter signals scattered from a rough cylindrical surface[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(1): 323-333.
- [14] 任新成,朱小敏,刘鹏.大地土壤表面与浅埋多目标宽带复合电磁散射研究[J].物理学报,2016,65(20):51-58.
REN X C, ZHU X M, LIU P. Wide-band composite electromagnetic scattering from the earth soil surface and multiple targets shallowly buried[J]. Acta Physica Sinica, 2016, 65(20): 51-58 (in Chinese).
- [15] 彭鹏,童创明,鲍峻松,等.基于二维粗糙面模型的大地土壤表面散射特性研究[J].微波学报,2013,29(4):38-42.
PENG P, TONG C M, BAO J S, et al. Study on the EM scattering from the earth surface based on the 2D rough surface model[J]. Journal of Microwaves, 2013, 29(4): 38-42 (in Chinese).

作者简介:

郝新红 女,博士,副教授,博士生导师。主要研究方向:中近程探测及控制技术、引信抗干扰技术。

杜涵宇 男,硕士研究生。主要研究方向:无线电引信信号处理及抗干扰技术。

Rough surface target and jamming signal recognition of FM fuze

HAO Xinhong^{*}, DU Hanyu, CHEN Qile

(Science and Technology on Electromechanical Dynamic Control Laboratory, Beijing Institute of Technology, Beijing 100081, China)

Abstract: A rough surface target and jamming signal recognition method for the ground frequency modulation (FM) fuze is proposed in order to effectively identify the ground target echo signal of the FM fuze and the digital radio frequency memory (DRFM) transmissive jamming. A rough surface beat frequency signal model of the ground FM fuze is established, and a two-dimensional distance-speed extraction method is used to extract the beat frequency and Doppler frequency. The beat frequency peak bandwidth and the Doppler peak bandwidth are used to identify ground target and the DRFM transmissive jamming, and their utility is verified by Monte Carlo simulation and non-parametric hypothesis statistical test. The results show that the spread of the beat frequency peak bandwidth and the Doppler frequency peak bandwidth is positively correlated with the size of the rough surface under the illumination of the fuze antenna. The Doppler frequency peak bandwidth spread is proportional to the carrier frequency. The peak bandwidth characteristics can be used to distinguish between rough surface target echo and DRFM transmissive jamming.

Keywords: frequency modulation (FM) fuze; rough surface target; two-dimensional distance-speed extraction; digital radio frequency memory (DRFM) jamming; nonparametric hypothesis test

Received: 2019-01-23; **Accepted:** 2019-05-18; **Published online:** 2019-06-11 15:15

URL: kns.cnki.net/kcms/detail/11.2625.V.20190610.0900.001.html

Foundation item: National Natural Science Foundation of China (61871414)

*** Corresponding author.** E-mail: haoxinhong@bit.edu.cn

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0037



一种电磁定位系统工作空间拓展方法

郑莉芳¹, 万元宇¹, 关少亚², 孙凯³, 孟偲^{3,4,*}, 贾佳¹

(1. 北京科技大学 机械工程学院, 北京 100083; 2. 北京航空航天大学 机械及自动化学院, 北京 100083;

3. 北京航空航天大学 宇航学院, 北京 100083; 4. 北京航空航天大学 生物医学工程高精尖创新中心, 北京 100083)

摘 要: 针对 NDI 电磁定位跟踪设备工作空间有限且在工作空间内定位精度不一致的问题, 提出了一种利用机械臂移动磁场发生器从而拓展电磁定位系统工作空间且保证定位精度的方法。利用 NDI 系统返回的误差指示值衡量定位精度, 当误差指示值超出设定的阈值时, 利用机械臂移动磁场发生器使传感器重新位于 NDI 系统的最佳测量工作区, 并将电磁定位系统测量的位姿通过空间变换方式统一到机械臂基座坐标系, 从而在保证定位精度的同时也起到扩展工作空间的作用。为验证所提方法的有效性, 通过实验验证定位误差与误差指标值及传感器到磁场发生器中心的距离成正相关关系; 通过拓展前后的误差分析表明, 所提方法能有效降低定位误差, 平均位置误差从 2.61 mm 降低到 1.34 mm, 平均姿态误差从 2.42° 降低到 1.37°。所提方法可应用于类似血管介入手术导管在大范围移动的器械定位与跟踪。

关 键 词: 无遮挡定位; 电磁定位系统; 工作空间拓展; 矩阵变换; Aurora

中图分类号: TP241.2

文献标识码: A

文章编号: 1001-5965(2019)10-1956-09

电磁定位系统 (Electromagnetic Tracking System, EM) 利用电磁感应原理进行位姿测量, 具有实时定位、精度高、不惧遮挡的优点^[1-2], 因而被广泛应用于医学手术中器械的跟踪定位^[3-4]。如 Wallace 等^[5]将 EM 应用于肾上腺、肝脏、肺等位置病变活检手术中的穿刺针导航; Krücker 等^[6]将 EM 应用于消融手术引导; Wood 等^[7]将 EM 应用于血管介入手术。

EM 应用于手术导航的缺点之一是定位精度易受环境磁场及导磁手术器械的影响。为了测定环境磁场对定位的影响, Gergel 等^[8]将一个三脚架并行机械臂作为标定机器人, 将 EM 的传感器固定在机械臂末端, 并设计程序分别采集机械臂及 EM 测量的位姿, 通过比较二者的差异来测定 EM 的定位精度。为了提高 EM 在干扰环境下的

定位精度, Boutaleb 等^[9]设计出一种 60 mm × 60 mm × 15 mm 的可移动标定块, 对器械在工作空间中的位置和姿态误差进行定量分析, 进而提高器械的定位精度。Kwartowitz 等^[10]设计出一种用于测量定位系统空间定位误差的平面模型, 并用该模型改进 EM 的定位精度。文献[9-10]的方法仅适用于磁场干扰固定的情形, 针对移动导磁器械对测量磁场的干扰, 笔者在前期工作中, 提出一种基于薄板样条函数的方法来动态矫正传感器受到干扰后的误差^[11]。Reichl 等^[12]为了提高电磁定位精度, 提出利用电磁伺服跟踪, 通过移动磁场发生器, 使传感器一直处于系统工作磁场的中心, 在提高定位精度的同时拓展了 EM 的定位空间。

EM 应用于手术导航的另外一个缺点是工作

收稿日期: 2019-01-23; 录用日期: 2019-05-10; 网络出版时间: 2019-06-11 09:32

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190610.0914.003.html

基金项目: 国家自然科学基金 (61873010, 61533016)

* 通信作者: E-mail: Tsai@buaa.edu.cn

引用格式: 郑莉芳, 万元宇, 关少亚, 等. 一种电磁定位系统工作空间拓展方法[J]. 北京航空航天大学学报, 2019, 45(10): 1956-1964. ZHENG L F, WAN Y Y, GUAN S Y, et al. A method for expanding workspace of electromagnetic tracking system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1956-1964 (in Chinese).

空间有限,且定位精度与传感器的距离相关。对于特定脏器的穿刺手术,因脏器尺寸有限,影响不大。但对于血管介入等手术,由于器械的运动范围较大,超出了 EM 的工作空间,因而只能用于局部区域的导管跟踪,且难以保证定位精度的可靠性。受 Reichl 等^[12]研究的启发,本文提出了一种利用机械臂移动 EM 中磁场发生器的位置来扩展其工作空间,并通过使传感器位于最佳工作区域的方法来保证定位的精确度。与 Reichl 等^[12]方法不同,本文方法并不试图使传感器始终位于磁场中心周围区域,而是通过置信度确定最佳工作区域,允许根据不同的应用环境设置不同的置信度阈值,更适合临床应用。

1 EM 工作原理及局限

1.1 EM 工作原理

EM 是一种基于法拉第电磁感应定律的空间位姿测量装置,可实现目标的位姿测量及动态跟踪。图 1 为 NDI 公司的交流磁场三轴正交式电磁定位系统 Aurora,主要包括磁场发生器、系统控制器、工具联结器及传感线圈。工作时,磁场发生器生成磁场,待测目标上的传感线圈在磁场内接收磁场信号,传感器联结器利用电磁感应将磁场信号转换为电信号,由系统控制器根据信号强度计算得到待测目标的 6 自由度坐标值,实现对目标的定位跟踪^[13-14]。



图 1 NDI 电磁定位系统 Aurora 的组成

Fig. 1 Composition of NDI Aurora electromagnetic tracking system

1.2 EM 的局限性

EM 是基于电磁感应定律工作的,在应用中,当磁场周围存在铁磁性物质时,EM 将会受到电磁涡流产生的磁场干扰而出现定位误差,所以应用 EM 需要考虑不同环境对定位精度的影响。此外,EM 的工作空间有限,当传感器超出其磁场范围时,便无法进行定位跟踪。表 1 为几种比较成

熟的电磁定位产品指标^[15]。可以看出,产品的测量范围有限,应用到较大的介入范围的手术(如结肠镜检查、神经内窥镜检查、支气管镜检查 and 血管介入导管导航等)无法满足工作空间的要求。

另外,EM 在其工作空间范围内跟踪精度并不一致。一般来说,越远离磁场发生器、靠近磁场范围外侧时,定位精度越低^[16]。如果忽略这种定位精度的不均匀性,就会导致应用中整个工作空间内的平均定位精度较低,定位效果较差。所以,在实际手术过程中,为了保证较高的定位精度,需要找出定位范围内定位精度较高的区域,在此区域内对手术器械进行跟踪。

表 1 几种电磁定位产品主要性能指标

Table 1 Main performance indicators of several electromagnetic tracing products

产品	测量距离/mm	位置精度/mm	方位精度/(°)
Aurora	500	1.6	1.1
Fastrak	762	0.762	0.15
Isotrak	762	2.54	0.75

2 EM 工作空间拓展

本文提出了一种基于机械臂移动的拓展 EM 工作空间的方法,使得术中应用 EM 跟踪大范围运动的手术器械不再受原系统生成磁场范围的限制。如图 2 所示,采用通用机械臂(SG-MOTOMAN NX100-HP3,该机械臂有 6 个自由度,重复定位精度可达 $\pm 0.03\text{ mm}$)及电磁定位系统 Aurora(定位精度为 0.6 mm ,可测量 6 自由度位姿坐标)进行 EM 工作空间拓展。将 Aurora 的磁场发生器与机械臂腕部刚性连接,由机械臂带动磁场发生器移动,并融合机械臂位姿将传感器位姿参数统一到机械臂基座坐标系上。如图 3 所示,手术过程中系统不断地检测反映测量置信度的误差指示

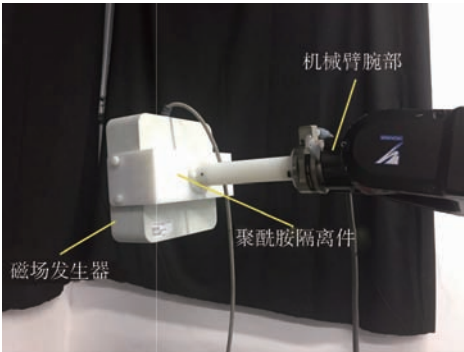


图 2 电磁定位系统空间拓展装置

Fig. 2 Space expansion device of electromagnetic tracking system

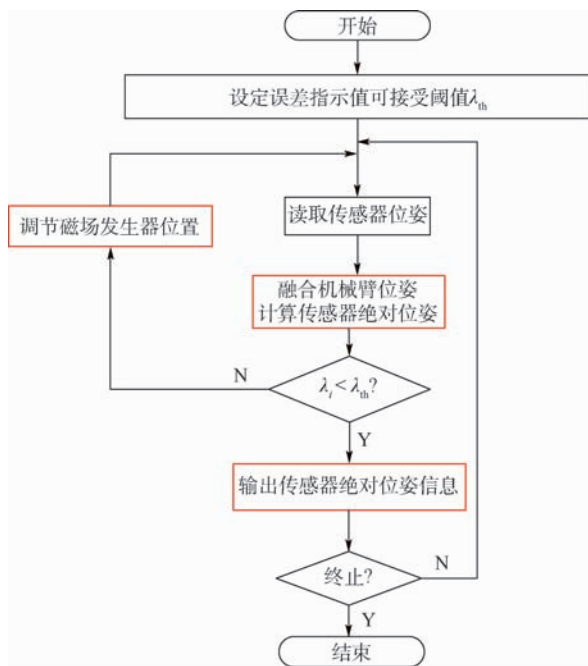


图3 EM工作空间拓展流程

Fig. 3 Process of EM workspace expansion

值(indicator value),当判断误差指示值超过预设阈值时,调整磁场发生器的位置,使得传感器线圈一直处在最佳工作区域内。可以由理论推导出,当金属材料到磁场发生器的距离大于磁场发生器到被测传感器的距离的2倍时,由金属材料引起的二次磁场对被测磁场强度的影响不足1%^[17]。由于最佳工作区域靠近磁场发生器,传感器距离磁场发生器大致在50~120 mm之间,采用长度为250 mm的聚酰胺材料对磁场发生器与机械臂隔离连接,以忽略机械臂上的金属物质对EM磁场的干扰。此外,拓展要保持传感器在非金属桌子上进行,隔离其他电子设备(如电脑、手机、电灯等),尽可能减少环境对EM定位进行干扰。

2.1 EM工作空间拓展原理

2.1.1 传感器绝对位姿的计算

EM测得的位姿为磁场发生器坐标系中传感器线圈的位姿。通过机械臂移动磁场发生器后,由于磁场发生器的位置发生变化,不能用来描述传感器线圈的绝对位姿,需要将移动后用磁场发生器坐标系描述的传感器线圈的位姿统一转化为相对于世界坐标系的绝对位姿。在整个EM拓展流程中,机械臂基座一直处于静止状态,将机械臂基座坐标系视为世界坐标系。对于2个不同的坐标系,其关系可以用6个参数来描述,分别为3个姿态参数(R_x, R_y, R_z)及3个位置参数(x, y, z)。对于空间中任意一点 P ,其在磁场发生器坐标系下的坐标为 $P_s = [x_s, y_s, z_s]^T$,在机械臂基座坐标

系下的坐标为 $P_b = [x_b, y_b, z_b]^T$,二者的转换由式(1)确定:

$$[x_b, y_b, z_b]^T = R \cdot [x_s, y_s, z_s]^T + T \quad (1)$$

式中: R 为 3×3 旋转矩阵,由3个姿态参数决定; T 为 3×1 平移向量,由3个位置参数决定。

根据式(1),可以将磁场发生器测得点的坐标统一到机械臂基座坐标系下。对于求解旋转矩阵 R 和平移向量 T ,一般将式(1)改写为

$$[x_b, y_b, z_b, 1]^T = {}^B T_S \cdot [x_s, y_s, z_s, 1]^T \quad (2)$$

式中: ${}^B T_S = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$ 为 4×4 传感器线圈到机械臂基座的转换矩阵。

图4为EM工作空间拓展坐标系,从中可以得出

$${}^B T_S = {}^B T_H \cdot {}^H T_F \cdot {}^F T_S \quad (3)$$

式中: ${}^B T_H$ 为 4×4 机械臂腕部到机械臂基座的转换矩阵; ${}^H T_F$ 为 4×4 磁场发生器到机械臂腕部的转换矩阵; ${}^F T_S$ 为 4×4 传感器线圈到磁场发生器的转换矩阵。

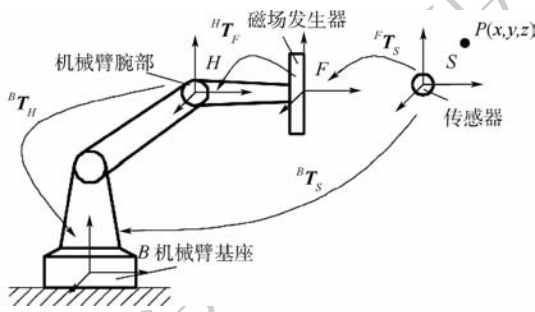


图4 EM工作空间拓展坐标系

Fig. 4 Coordinate system of EM workspace expansion

2.1.2 矩阵标定

根据式(1)~式(3),可以将传感器线圈坐标系下点的坐标转换到机械臂基座坐标系下,完成点从传感器线圈坐标系到世界坐标系的转换。其中, ${}^B T_H$ 、 ${}^F T_S$ 可以在实际定位过程中获得, ${}^H T_F$ 是未知的。在实际应用中,磁场发生器与机械臂腕部刚性连接, ${}^H T_F$ 保持不变,可以通过标定获得。如图5所示,为了获得 ${}^H T_F$,固定传感器线圈及机械臂基座的位置,使得机械臂基座到传感器线圈的 ${}^B T_S$ 不变,通过机械臂腕部运动带动磁场发生器,改变 ${}^B T_H$ 、 ${}^F T_S$ 来计算得出 ${}^H T_F$:

$${}^B T_{H_i} \cdot {}^H T_F \cdot {}^F T_{S_i} = {}^B T_{H_j} \cdot {}^H T_F \cdot {}^F T_{S_j} \quad (4)$$

式中: ${}^B T_{H_i}$ 为当前机械臂腕部到机械臂基座的转换矩阵; ${}^F T_{S_i}$ 为当前传感器线圈到磁场发生器的转换矩阵。

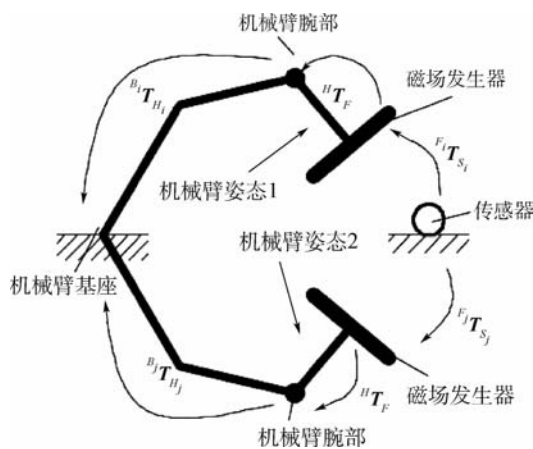


图 5 磁场发生器与机器人末端的转换矩阵标定

Fig. 5 Calibration of conversion matrix between magnetic field generator and robot wrist

$$\begin{cases} A = ({}^B T_{H_j})^{-1} \cdot {}^B T_{H_i} \\ B = {}^F T_{H_j} \cdot ({}^F T_{H_i})^{-1} \\ X = {}^H T_F \end{cases} \quad (5)$$

通过式(5),将式(4)转化为 $AX = XB$ 形式,利用 Park 和 Martin^[18] 的求解方法即可标定出磁场发生器与机器人末端的坐标变换关系。标定过程可在手术前一次性完成,简单迅速。

2.2 基于误差指示值调整磁场发生器

由于 EM 在其工作空间内的定位精度并不一致,应用过程中需要在具有较高精度的定位区域跟踪传感器。不同的 EM 会有不同的评估测量置信度的方法,如 Aurora 在每次完成对传感器位姿的计算后,都会返回一个无单位误差指示值(范围为 0~9.9),可以利用误差指示值反映定位精度。6 自由度传感器内含 2 个相对位置已知且正交放置的传感器线圈,Aurora 系统控制器得到传感器中传感器线圈 1 的测量值后,通过传感器预定义文件的信息计算得出传感器线圈 2 的位姿,将传感器线圈 2 位姿的计算值与测量值进行比较计算得出误差指示值。因此,磁场受到外界的干扰越大,测量值越不准确,误差指示值也越大^[19]。

为保证 EM 定位传感器能保持在最佳工作区域内,通过设定一个误差指示值的阈值 λ_{th} 作为传感器线圈是否处于最佳工作区域的评判标准,当误差指示值超出此阈值时视为定位精度不满足手术要求。在手术中,系统随时检测传感器的位姿,对与系统返回的误差指示值进行判断,一旦误差指示值超出设定的阈值,机器人就会对磁场发生器的位置进行调整:

$${}^B T_{H_{i+1}} = {}^B T_{H_i} \cdot {}^{H_i} T_{F_i} \cdot {}^{F_i} T_{F_{i+1}} \cdot {}^{F_{i+1}} T_{H_{i+1}} \quad (6)$$

式中: ${}^{H_i} T_{F_i}$ 为磁场发生器到机械臂腕部的转换矩

阵; ${}^{F_i} T_{F_{i+1}}$ 为磁场发生器当前位姿到磁场发生器下一次位姿的转换矩阵。

3 实验分析

3.1 误差指示值验证实验

为了验证 Aurora 的误差指示值能否反映 EM 的定位误差的变化趋势,设计了实验进行验证。

3.1.1 实验模型

如图 6 所示,实验用磁场发生器为 Aurora 方盒型磁场发生器,传感器为直径 25 mm 的 6 自由度盘状传感器。标定板为激光切割的亚克力板,标定板上均布间隔距离一致的 D 型定位孔。传感器与带有定位销的放置架固定可以插入标定板上的定位孔中。标定板上的 D 型定位孔朝向一致,当保持磁场发生器与标定板的相对位置不变时,可以保证放置架插入任一定位孔时传感器相对于磁场发生器姿态不变。可以通过可调支柱来调节标定板相对磁场发生器在 Z 方向上的高度。

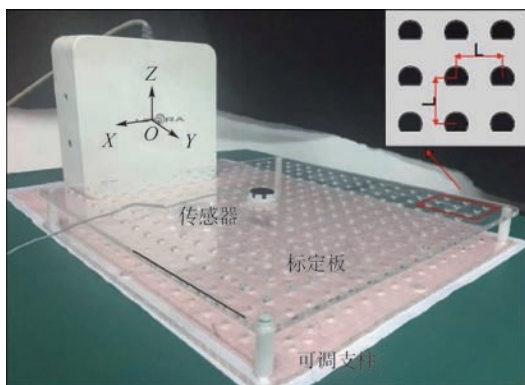


图 6 误差指示值验证实验模型

Fig. 6 Experimental model for error indicator value validation

3.1.2 实验过程与评价指标

保证磁场发生器与标定板位置固定,可调支柱高度固定,在标定板上选择 8×8 个定位孔进行采样,记录传感器在每个采样点的 6 自由度位姿 (x, y, z, R_x, R_y, R_z) 和误差指示值,进行位置误差和姿态误差的计算。

对于位置误差,设标定板上一点 i 的位置坐标为 (x_i, y_i, z_i) ,其相邻点 j 的位置坐标为 (x_j, y_j, z_j) ,则点 i 与相邻点 j 的距离为

$$L_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

设标定板平面内相邻两点的距离为 L_{CT} ,则点 i 与相邻点 j 的距离误差为 $\Delta L_{i,j} = |L_{CT} - L_{i,j}|$ 。对与点 i 相邻的四邻域点分别计算距离误差,定义点 i 位置误差为

$$u_i = \frac{1}{4} \sum_{j=1}^4 \Delta L_{i,j}$$

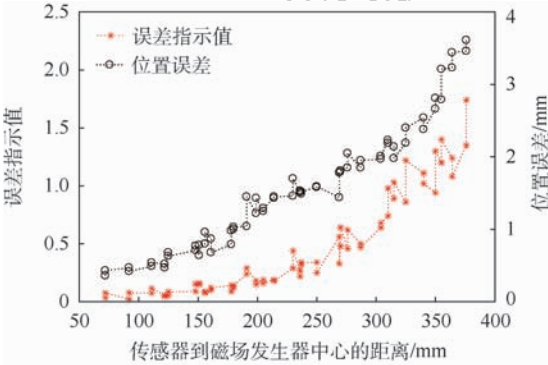
对于姿态误差,选择距离磁场发生器最近的 3×3 个点的姿态参数的平均值(R_x, R_y, R_z)作为整个标定板上所有采样的姿态真值。测得任意一点*i*的姿态为($R_{x_i}, R_{y_i}, R_{z_i}$),则点*i*的姿态误差为

$$v_i = \sqrt{(R_{x_i} - R_x)^2 + (R_{y_i} - R_y)^2 + (R_{z_i} - R_z)^2}$$

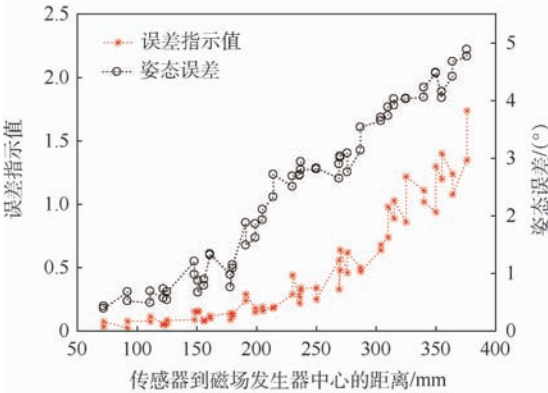
3.1.3 实验结果与分析

采集 8×8 个采样点,记录每个采样点的6自由度位姿和误差指示值后,计算每个采样点的位置误差及姿态误差,并以传感器到磁场发生器中心的距离为横坐标,以误差指示值及位置误差或姿态误差为纵坐标绘制图7。

随着传感器与磁场发生器中心距离的增大,误差指示值呈现增大的趋势,同时位置误差和姿态误差也同样呈现增大的趋势,说明误差指示值可以反映定位误差的变化趋势。从图7也可以看出,距离磁场发生器中心较近的区域,如距离区间为70~120 mm内,传感器在此区间内逐渐远离磁



(a) 位置误差与误差指示值随传感器到磁场发生器中心距离的变化趋势



(b) 姿态误差与误差指示值随传感器到磁场发生器中心距离的变化趋势

图7 位置误差、姿态误差与误差指示值随传感器到磁场发生器中心距离的变化趋势

Fig.7 Variation trend of position error, orientation error and indicator value with distance from sensor to center of magnetic field generator

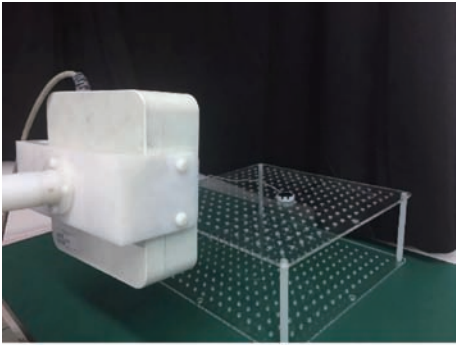
场发生器中心,虽然位置误差、姿态误差和误差指示值都在增加,但是增加的幅度相对较慢。在进行EM工作空间拓展时,可以在此区间内选定误差指示值阈值来移动磁场发生器保证定位精度。

3.2 EM工作空间拓展验证实验

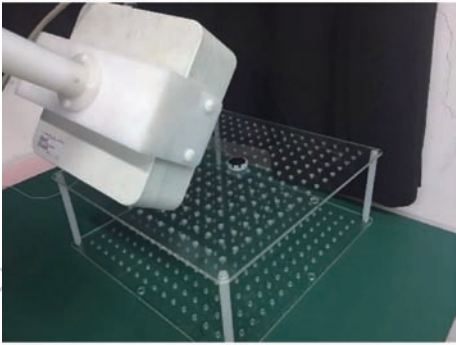
为了验证本文提出的根据误差指示值阈值调整磁场发生器的方法对于提升EM的定位精度是否有效,设计了EM工作空间拓展验证实验进行验证。

3.2.1 实验模型

如图8所示,采用机械臂通过聚酰胺连接件夹持Aurora磁场发生器,对比EM工作空间拓展前后的位置误差及姿态误差。



(a) 拓展前(磁场发生器不动)



(b) 拓展后(磁场发生器运动)

图8 EM工作空间拓展验证实验模型

Fig.8 Experimental model of EM workspace expansion verification

3.2.2 实验过程与评价指标

如图8(a)所示,固定标定板在非金属实验台上不动,在未进行EM空间拓展前,保持机械臂不动,磁场发生器与标定板的相对位姿不变,以实验台平面为 $Z=0$ 高度平面,通过调节可调支柱,分别测量了靠近磁场发生器底部($Z=0$)、靠近磁场发生器中心($Z=90$ mm)、靠近磁场发生器顶部($Z=150$ mm)3个平面上 8×8 个点的位置与姿态,计算位置误差及姿态误差。

如图8(b)所示,同样保持标定板位置不变,

根据本文的 EM 工作空间拓展方法分别测量 $Z = 0, 90, 150 \text{ mm}$ 3 个平面上 8×8 个点的位置与姿态,误差指示值阈值 λ_{th} 设置为 0.3,一旦 EM 在定位传感器时误差指示值超过 0.3,则利用机械臂调整磁场发生器的位置,并重新计算传感器的位姿。通过 3.1.2 节的评价方法计算位置误差及姿态误差。

3.2.3 实验结果与分析

对于 $Z = 0, 90, 150 \text{ mm}$ 上的 8×8 个采样点,采集其位置参数和姿态参数,计算位置误差和姿态误差,将位置误差和姿态误差反映在 Z 轴上,

X 轴和 Y 轴如图 6 中的布置反映 8×8 个采样点的分布,对比拓展前后的位置误差与姿态误差,如图 9 和图 10 所示。图中:左列为拓展前位置误差或姿态误差,右列为拓展后位置误差或姿态误差。

通过图 9、图 10 可以看出,当保持磁场发生器固定不动时,测得的位姿误差依旧随着传感器到磁场发生器中心的距离增大而增大,同样符合误差指示值的变化趋势。表 2 统计了不同平面拓展前后的位置误差与姿态误差的平均值 (Mean) 与均方差 (SD),可以看出,当磁场发生器固定不动时,不同平面采样点的位置误差和姿态误差也

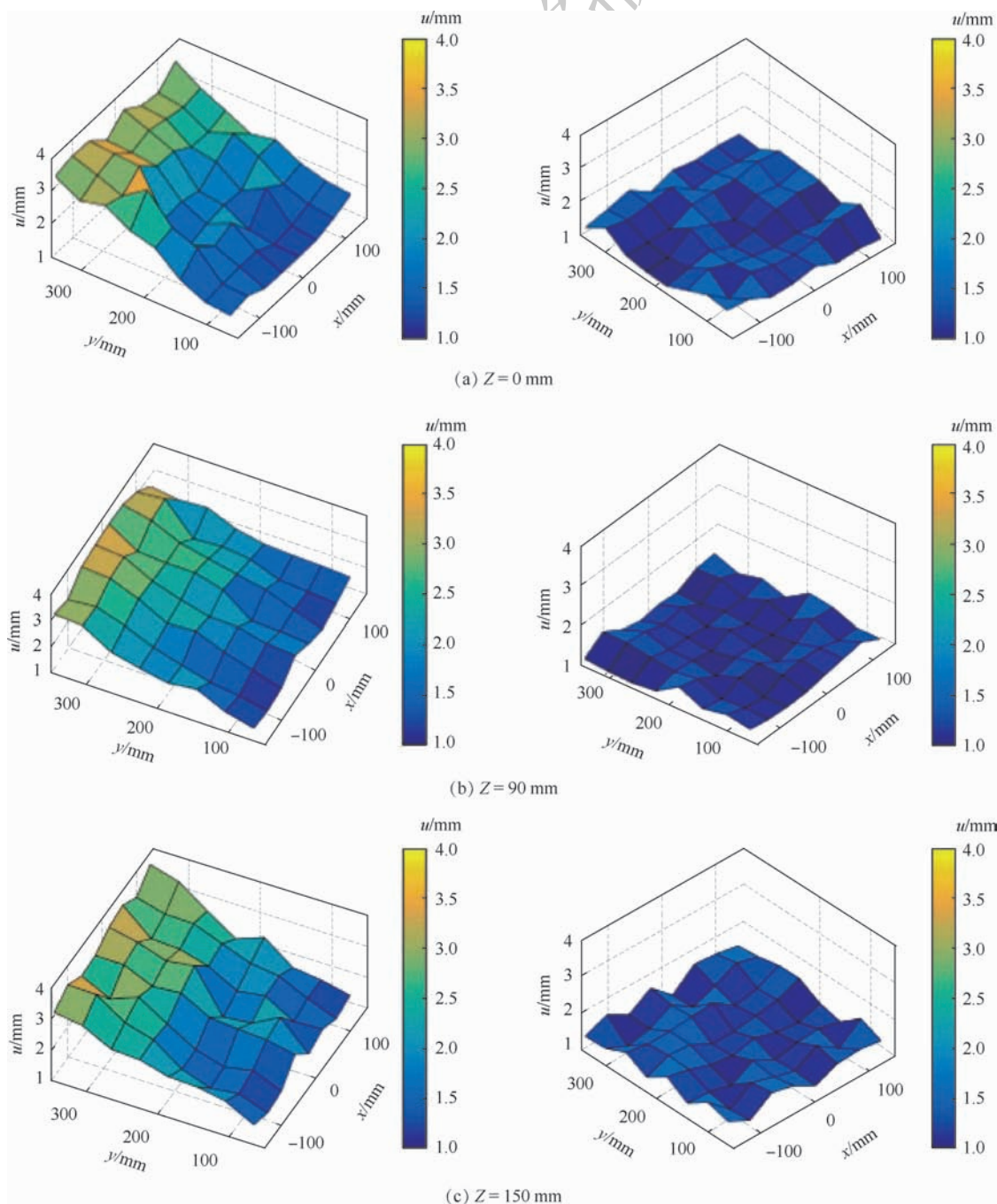


图 9 EM 工作空间拓展前后的位置误差

Fig. 9 Position error before and after EM workspace expansion

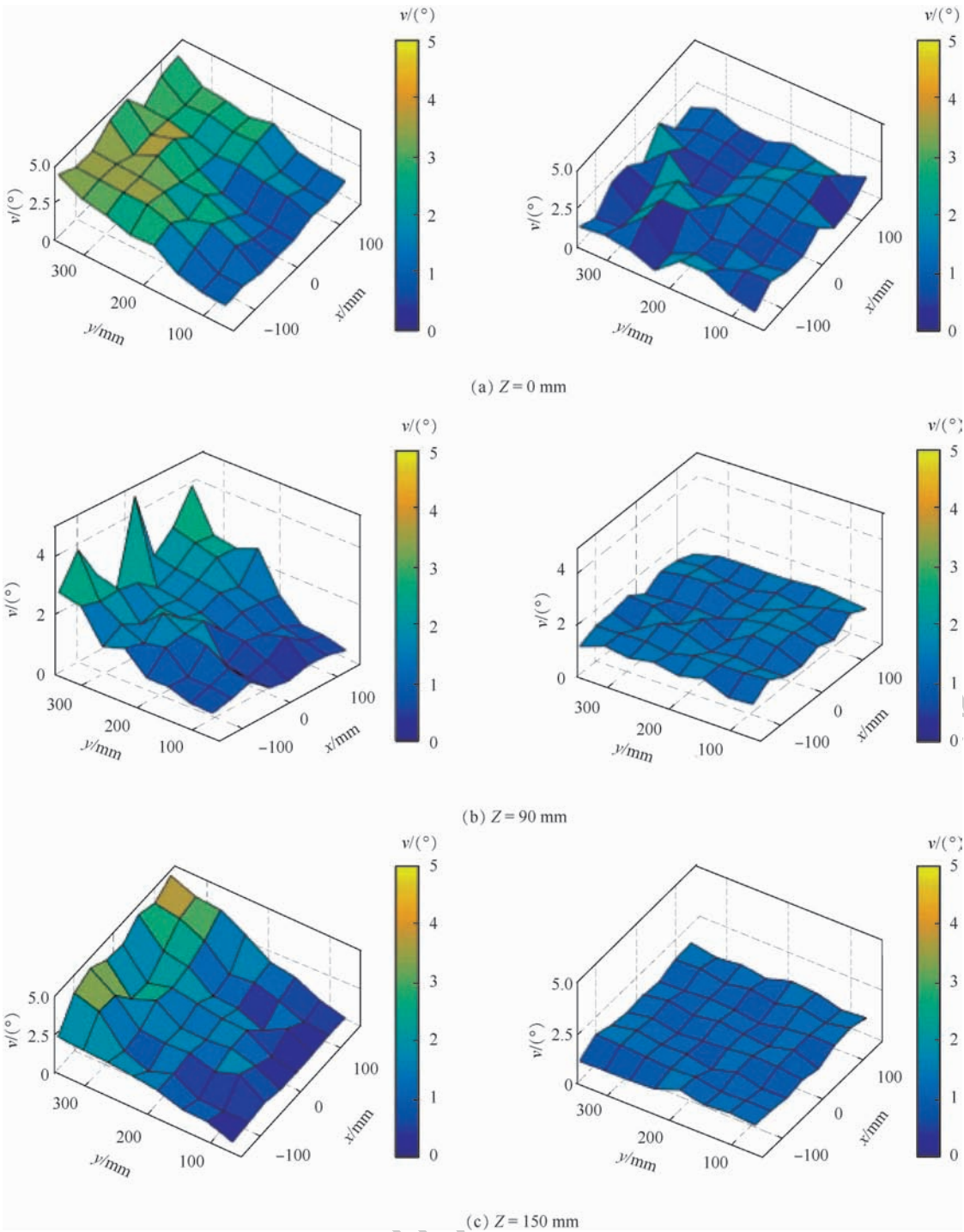


图 10 EM 工作空间拓展前后的姿态误差

Fig. 10 Orientation error before and after EM workspace expansion

表 2 不同平面拓展前后位置误差与姿态误差

Table 2 Position error and orientation error of different planes before and after expansion

Z/mm	拓展前位置误差/mm		拓展后位置误差/mm		拓展前姿态误差/(°)		拓展后姿态误差/(°)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	2.74	0.71	1.34	0.28	2.36	1.13	1.41	0.41
90	2.42	0.64	1.28	0.19	2.29	1.25	1.32	0.28
150	2.68	0.65	1.39	0.22	2.62	1.38	1.37	0.32
平均值	2.61	0.67	1.34	0.23	2.42	1.25	1.37	0.34

不同,磁场发生器中心平面上的采样点误差较小,其余 2 个平面采样点的误差相对较大。拓展后 3 个平面采样点的平均位置误差从 2.61 mm 降低到 1.34 mm,平均姿态误差由 2.42°降低到 1.37°,且 SD 较拓展前也有显著减小,即拓展后同一平面的误差分布更为均匀,验证了根据误差指示值移动磁场发生器在拓展 EM 工作空间的同时可以有效提升电磁定位的精度。

4 结 论

本文针对血管介入等大范围手术中应用 EM 跟踪手术器械时存在的定位范围有限等问题,提出了一种拓展 EM 工作空间的方法。根据误差指示值调整磁场发生器,使得传感器一直处于最佳工作区域内,提升了 EM 的定位精度。

下一步将考虑在临床应用中,磁场发生器移动对手术设备、医生、病人的影响,结合手术室环境对磁场发生器移动的限制条件,得到误差指示值超出预设阈值时,磁场发生器具体的移动路径。

参考文献 (References)

- [1] FRANZ A M, HAIDEGGER T, BIRKFELLNER W, et al. Electromagnetic tracking in medicine—A review of technology, validation, and applications[J]. IEEE Transactions on Medical Imaging, 2014, 33(8): 1702-1725.
- [2] BIRKFELLNER W. Calibration of tracking systems in surgical environment[J]. IEEE Transactions on Medical Imaging, 1998, 17(5): 737-742.
- [3] PÉRIÉ D, TATE A J, CHENG P L, et al. Evaluation and calibration of an electromagnetic tracking device for biomechanical analysis of lifting tasks[J]. Journal of Biomechanics, 2002, 35(2): 293-297.
- [4] FEUERSTEIN M, REICHL T, VOGEL J, et al. Magneto-optical tracking of flexible laparoscopic ultrasound: Model-based online detection and correction of magnetic tracking errors[J]. IEEE Transactions on Medical Imaging, 2009, 28(6): 951-967.
- [5] WALLACE M J, GUPTA S, HICKS M E. Out-of-plane computed-tomography-guided biopsy using a magnetic-field-based navigation system[J]. Cardio Vascular and Interventional Radiology, 2006, 29(1): 108-113.
- [6] KRÜCKER J, XU S, GLOSSOP N, et al. Electromagnetic tracking for thermal ablation and biopsy guidance: Clinical evaluation of spatial accuracy[J]. Journal of Vascular and Interventional Radiology, 2007, 18(9): 1141-1150.
- [7] WOOD B J, ZHANG H, DURRANI A, et al. Navigation with electromagnetic tracking for interventional radiology procedures: A feasibility study[J]. Journal of Vascular and Interventional Radiology, 2005, 16(4): 493-505.
- [8] GERGELI, GAA J, MÜLLER M, et al. A novel fully automatic system for the evaluation of electromagnetic tracker[C] // Conference on Medical Imaging-Image-Guided Procedures, Robotic Interventions and Modeling. Bellingham: SPIE, 2012, 8316: 831608.
- [9] BOUTALEB S, RACINE E, FILLION O, et al. Performance and suitability assessment of a real-time 3D electromagnetic needle tracking system for interstitial brachytherapy[J]. Journal of Contemporary Brachytherapy, 2015, 7(4): 280-289.
- [10] KWARTOWITZ D M, RETTMANN M E, HOLMES D R, et al. A novel technique for analysis of accuracy of magnetic tracking systems used in image guided surgery[C] // Conference on Medical Imaging 2010-Visualization, Image-Guided Procedures, and Modeling. Bellingham: SPIE, 2010, 7625: 76251L.
- [11] 关少亚, 孟德, 万元宇, 等. 基于薄板样条函数的电磁定位系统位姿校正方法[J]. 北京航空航天大学学报, 2018, 44(11): 2350-2355.
GUAN S Y, MENG C, WAN Y Y, et al. A thin plate spline based method for correction of position and posture of electromagnetic tracking system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(11): 2350-2355 (in Chinese).
- [12] REICHL T, GARDIAZABAL J, NAVAB N. Electromagnetic servoing—A new tracking paradigm[J]. IEEE Transactions on Medical Imaging, 2013, 32(8): 1526-1535.
- [13] LUND K. Electromagnetic navigation vs fluoroscopy in aortic endovascular procedures—A phantom study[J]. International Journal of Computer Assisted Radiology & Surgery, 2016, 12(1): 1-7.
- [14] VILLAGRAN C R T, IKEDA S, FUKUDA T, et al. Catheter insertion path reconstruction with autonomous system for endovascular surgery[C] // International Symposium on Computational Intelligence in Robotics & Automation. Piscataway, NJ: IEEE Press, 2007: 47.
- [15] 秦成. 电磁跟踪系统的研究[D]. 武汉: 华中科技大学, 2013.
QIN C. The research of electromagnetic tracking system[D]. Wuhan: Huazhong University of Science and Technology, 2013 (in Chinese).
- [16] 罗伟, 张庆, 李珊珊, 等. 新一代 Aurora 电磁跟踪系统在医学手术导航中的应用[J]. 中国医疗器械杂志, 2013, 37(2): 126-128.
LUO W, ZHANG Q, LI S S, et al. New generation Aurora electromagnetic tracking system in the medical surgical navigation[J]. Chinese Journal Medical instrumentation, 2013, 37(2): 126-128 (in Chinese).
- [17] RAAB F, BLOOD E, STEINER T, et al. Magnetic position and orientation tracking system[J]. IEEE Transactions on Aerospace and Electronic Systems, 1979, 15(5): 709-718.
- [18] PARK F C, MARTIN B J. Robot sensor calibration: Solving $AX = XB$ on the Euclidean group[J]. IEEE Transactions on Robotics and Automation, 1994, 10(5): 717-721.
- [19] Northern Digital Inc. Aurora user guide[Z]. Revision 4. Waterloo: Northern Digital Inc., 2008: 38.

作者简介:
郑莉芳 女,博士,副教授。主要研究方向:先进机电液系统。
万元宇 男,硕士研究生。主要研究方向:医疗机器人、电磁

定位。
孟德 男,博士,副教授。主要研究方向:机器视觉、机器人智能系统。

A method for expanding workspace of electromagnetic tracking system

ZHENG Lifang¹, WAN Yuanyu¹, GUAN Shaoya², SUN Kai³, MENG Cai^{3,4,*}, JIA Jia¹

- (1. School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China;
2. School of Mechanical Engineering and Automation, Beihang University, Beijing 100083, China;
3. School of Astronautics, Beihang University, Beijing 100083, China;
4. Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China)

Abstract: Aimed at the problem of limited workspace and inconsistent measurement accuracy of NDI electromagnetic tracking equipment, a method for expanding the workspace of electromagnetic tracking system and guaranteeing measurement accuracy by moving magnetic field generator is proposed. This method uses the indicator value returned by NDI system as the measurement of accuracy. When the indicator value exceeds the set threshold, the magnetic field generator connected with the manipulator is moved to relocate the sensor in the optimum working area, and the position and attitude measured by the system are unified into the coordinate system of the manipulator base through spatial transformation. In order to verify the effectiveness of the proposed method, experiments are conducted to verify that the measurement error is positively correlated with the indicator value and the distance between the sensor and the center of the magnetic field generator. Then, by comparing the errors before and after the expansion, it is shown that the mean position error can be reduced from 2.61 mm to 1.34 mm, and the mean orientation error can be reduced from 2.42° to 1.37°. This method can be used to locate and track large-scale moving instruments such as vascular interventional catheters.

Keywords: non-occlusion positioning; electromagnetic tracking system; workspace expansion; matrix transformation; Aurora

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0046

基于编解码双路卷积神经网络的视觉自定位方法



贾瑞明^{1,*}, 刘圣杰¹, 李锦涛², 王赞豪², 潘海侠²

(1. 北方工业大学 信息学院, 北京 100144; 2. 北京航空航天大学 软件学院, 北京 100083)

摘 要: 为了从单张 RGB 图像估计出相机的位姿信息, 提出了一种深度编解码双路卷积神经网络(CNN), 提升了视觉自定位的精度。首先, 使用编码器从输入图像中提取高维特征; 然后, 使用解码器提升特征的空间分辨率; 最后, 通过多尺度位姿预测器输出位姿参数。由于位置和姿态的特性不同, 网络从解码器开始采用双路结构, 对位置和姿态分别进行处理, 并且在编解码之间增加跳跃连接以保持空间信息。实验结果表明: 所提网络的精度与目前同类型算法相比有明显提升, 其中相机姿态角度精度有较大提升。

关 键 词: 视觉自定位; 编解码结构; 卷积神经网络(CNN); 跳跃连接; 双路网络

中图分类号: V249.32⁺9; TP391

文献标识码: A

文章编号: 1001-5965(2019)10-1965-08

视觉自定位(Visual Localization, VL)是一种通过待查询的视觉数据来获取相机位姿的方法, 其在计算机视觉中扮演着重要的角色, 在航天器交会对接、轨道飞行器回收空间碎片、火星漫游车等应用场景中, 均需要通过视觉自定位技术获取精确的自身位姿参数。不同于从运动恢复结构(Structure from Motion, SFM)及即时定位与地图构建(Simultaneous Localization and Mapping, SLAM)提供图像序列间的相对位置关系, 视觉自定位直接给出待查询的视觉数据对应的相机绝对位姿。视觉自定位可以分为直接法和间接法。间接法通常的做法是将待查询图像与一系列相似的候选图像比较以获得位置, 这成为一种图像检索问题^[1-3]。直接法从待查询图像中直接恢复相机位姿, 大致可以分为如下3种方法: ①假设已知待查询图像对应的相机位姿的先验知识, 可以得到一个粗糙的位姿预估计。先验知识可以从定位传感器(如GPS^[1,4-5])中获得。②利用特征点匹配^[6-9]

来实现, 通过计算从二维待查询图像中提取到的特征在三维点云模型中的对应来获得相机位姿。这通常需要使用从SFM或其他方法中重建的三维模型。③直接从输入视觉数据中估计出对应的位姿信息, 有标准预测技术^[10]和利用卷积神经网络(Convolutional Neural Network, CNN)的方法。

随着深度学习的发展, 神经网络在实现室内或者室外场景的坐标定位问题上体现出很好的精度。相比于传统算法, 深度学习方法在处理有复杂纹理的场景、不同光照的环境、有运动模糊等挑战性因素的数据集时具有良好的鲁棒性。本文讨论的算法属于直接法视觉自定位中的第3种方法, 即运用卷积神经网络直接从待查询视觉数据中估计出对应相机位姿。

2015年, Kendall等^[11]提出的网络PoseNet尝试直接从输入图像恢复六自由度的相机位姿, 作为使用深度学习处理视觉自定位问题的开拓性工作, 使用了第1版GoogleNet作为PoseNet的基础

收稿日期: 2019-02-13; 录用日期: 2019-05-18; 网络出版时间: 2019-06-12 16:30

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190612.0854.001.html

基金项目: 国家重点研发计划(2017YFB0802300); 北京市教委面上项目(KM201510009005); 北方工业大学学生科技活动项目(110051360007)

*通信作者. E-mail: jiaruiming@ncut.edu.cn

引用格式: 贾瑞明, 刘圣杰, 李锦涛, 等. 基于编解码双路卷积神经网络的视觉自定位方法[J]. 北京航空航天大学学报, 2019, 45(10): 1965-1972. JIA R M, LIU S J, LI J T, et al. A visual localization method based on encoder-decoder dual-stream CNN[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1965-1972 (in Chinese).

网络,将该网络的 3 个 softmax 分类器替换为 1 个全连接层,之后输出位置坐标和姿态角四元数。2016 年, Walch 等^[12]对 PoseNet 提取到的特征进行调整,将这些特征不直接输出到预测器,而是作为空域 LSTM 的输入序列,得到了更精确的结果。科研人员还研究了如下代表性方法: Clark 等^[13]在时域上运用了 LSTM 结构,利用视频帧间的时域信息来辅助预测相机位置,但仅有相机位置信息而缺少姿态信息; Kendall 和 Cipolla^[14]提出了运用模型不确定性的贝叶斯 PoseNet,在多个随机裁剪的输入图像和 PoseNet 之间插入了一个 dropout 层,并为每个生成的位置建模,提升了定位精度; Melekhov 等^[15]提出了一种沙漏网络结构,使用 ResNet34 网络结构代替 PoseNet 的 GoogLeNet,并使用反卷积及跳跃连接,定位精度超过了其他成果; Li 等^[16]提出利用深度图进行研究,相比 PoseNet 提升了精度,但引入了附加深度信息,由于真实场景中可能无法有效获取深度信息(如使用数码相机),不利于实时定位。其他方法如使用附加的语义信息及视觉里程计进行约束^[17-18]、估计像素级坐标之后利用 RANSAC 进行匹配^[19],虽然提升了算法精度,但是大幅增加了算法复杂度及需要提供的信息。

本文提出了一种编解码双路视觉自定位网络(Bifurcate Localization Net, BiLocNet)结构,可以从单张 RGB 图像预测相机六自由度位姿参数;该网络使用改进的 Inception-Resnet-V2 作编码器,用双路结构解码输出位置和姿态参数,提出一种多尺度位姿预测器结构,从解码后的特征图中预测出相机位姿;由于数据集较小,其他方法^[11,13,15,18,20]均利用在 ImageNet 上进行预训练的模型权重来执行权重初始化,该网络无需预训练即可达到优于预训练模型的收敛速度和精度;此外,损失函数加入同方差不确定性^[21]作为目标函数,提升了网络可训练性。

1 编解码双路视觉自定位网络结构

本文提出了一种编解码双路视觉自定位网络结构,如图 1 所示,其由编码器(encoder)、解码器(decoder)和预测器(estimator)组成。首先, BiLocNet 用编码器对输入图像进行编码以提取高级语义特征;然后,采用双路解码器、预测器分别解码输出位置 p 和姿态 q 。其中,在编、解码之间加入跳跃连接使编码部分的空间信息可以传递给解码器,对解码过程中的空间信息缺失进行一定程度的补偿。

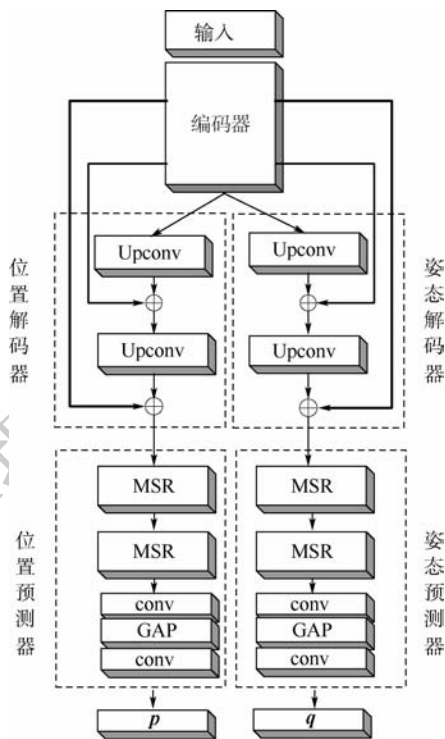


图 1 BiLocNet 结构

Fig. 1 Architecture of BiLocNet

1.1 编码器

编码器由改进的 Inception-Resnet-V2^[22]组成, Inception-Resnet-V2 如图 2 所示。编码器中, stem 结构用于维持基础网络层结构的稳定。Inception-Resnet 有 A、B、C 共 3 种模块,区别在于卷积核的尺寸、通道数和子分支。Inception-Resnet 模块使用最大值池化层及多尺度卷积层获得不同尺度下的特征,同时结合 ResNet 的思想,将残差模块中的残差结构替换为上述多尺度结构,既保留了 Inception 一贯的拓宽及加深网络的思想,又改善了过深的网络可能出现的梯度消失及梯度爆炸问题。残差模块还使网络梯度回传更加有效,从而显著提高网络收敛速度。Reduction 模块用于降低分辨率、提升通道数,达到甚至超越传统卷积神经网络中利用最大值池化层接卷积层得到的效果。

由于本文网络在编、解码之间增加了跳跃连接,为了保证跳跃连接节点的特征图的分辨率相同,调整了池化层及部分卷积层的填充方式以适应解码器输出的特征图尺寸,从而使编码器中间层输出特征与解码器恢复的特征可直接相加,还调整了某些输出通道数以避免出现原网络某些层之间张量维度不匹配的情况,同时去除了图 2 中的 dropout 层、全连接层(FC)和 softmax 层,使输出的特征向量不经过分类任务的预测器而直接作为解码器的输入。

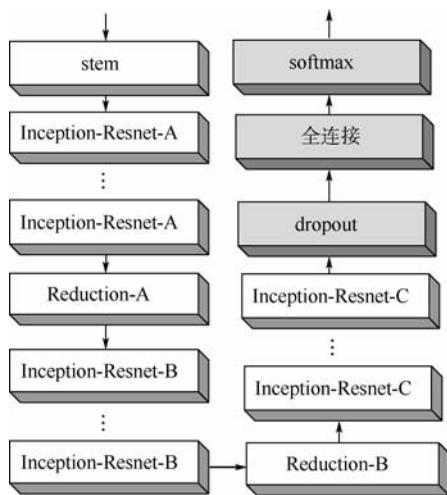


图 2 Inception-Resnet-V2 编码器结构

Fig. 2 Architecture of encoder Inception-Resnet-V2

1.2 解 码 器

由于编码过程中特征图的空间分辨率会逐步降低,导致了神经元间的空间信息被压缩,而这种空间依赖关系对于图像处理任务至关重要,故本文在编码之后加入解码器。

解码器采用 2 个上卷积层,卷积核大小为 4×4 ,步长为 2,每次上卷积之后特征图分辨率变为原来的 2 倍,通道数也相应降低。解码器尝试从高维特征中恢复特征间的空间信息,且增加了网络的深度,对预测精度有显著提升。同时,为了更好地保持编码过程中的图像细节信息,编解码之间增加了 2 个跳跃连接(解码器有 2 个上卷积)。引出编码器 Inception-Resnet-A、Inception-Resnet-B 后的 2 个特征图,与上卷积后的特征图分别相加,实现跳跃连接以弥补解码后恢复的空间信息损失。经过实验验证,增加解码器提升空间分辨率对于提升自定位精度有明显作用。

1.3 多尺度位姿预测器

预测器是深度学习回归任务中不可或缺的结构,解码后的特征图如要转换为需要的参数,需要使用预测器。

本文提出的多尺度位姿预测器结构如图 1 所示,主要分为多尺度降维(Multi-scale Reduction, MSR)模块和全局均值池化模块(Global Average Pooling, GAP)。如图 3 所示,MSR 模块由 1×1 和 3×3 两种尺度的卷积层组成,一共包含 5 个卷积层,conv 代表卷积层, k 代表卷积核大小, s 代表步长。解码出的信息经过第 1 个 MSR 模块,输出特征图大小将变为输入特征图的一半,将得到的输出经过第 2 个 MSR 模块,输出特征图分辨率再次减半。GAP 模块包含一个用于降分辨率升通道

的卷积层、一个全局均值池化层及一个 1×1 卷积层,最终输出一个特征矢量。

MSR 模块用于将解码器输出的特征进行重组,使其通过不同尺度的卷积核获得不同的感受野,同时将这些特征进行降维升通道为接下来的预测做准备,且 MSR 模块拓展了网络的深度和宽度,有效提升了网络性能。单尺度降维模块(Single-Scale Reduction, SSR)结构如图 4 所示。

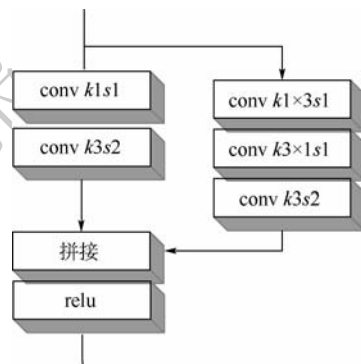


图 3 MSR 模块结构

Fig. 3 Module architecture of MSR

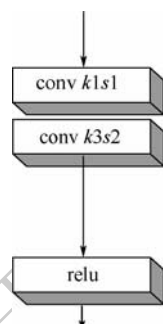


图 4 SSR 模块结构

Fig. 4 Module architecture of SSR

1.4 双路结构

目前,视觉自定位网络中位置与姿态参数均从单路网络输出,未考虑位置和姿态的区别,本文根据位置与姿态的不同特性,采用双路结构分别处理位置参数和姿态参数。

视觉自定位输出六自由度参数,包括 3 个位置和 3 个姿态角度。位置与姿态有以下不同:①相机的位置和姿态具有不同的量纲;②相机位置和姿态的改变反映在二维图像中是不同的,位置的变化将导致图像中物体间的遮挡关系及像素位置发生变化,而姿态的变化导致了视角的变化。真实世界在图像中的投影对于位置和姿态变化的表现是不同的,根据投影关系可知,视角变化会使物体在图像中的大小和形状发生变化。

综上可知,不分离位置和姿态而直接单路输出会导致两者间互相干扰,使网络无法精确解算

位姿信息,故本文将位置和姿态考虑为2种任务,设计了双路网络来解决此问题。多任务学习通过从共享的特征表示中同时学习多个任务的特定表示,提高了学习效率及预测精度。多任务学习是一种任务之间的知识共享和迁移,从一个任务中学习得到的表示通常可以帮助改善其他任务。BiLocNet将编码器输出的特征矢量分裂为位置和姿态两路子任务网络。

2 损失函数

相机位姿中,位置 p 由相机在世界坐标系中的坐标 $[x, y, z]$ 表示,姿态 q 由一个四元数 $q = [v, a, b, c]^T$ 表示。

四元数可以表示三维空间中任意一个旋转,避免了用欧拉角表示出现的万向锁现象。通过归一化为单位长度,任意的四维矢量可以很容易地映射到合理的旋转。使用四元数作为姿态角的优化目标比求取旋转矩阵要容易得多,避免了旋转矩阵增加的计算量及计算标准正交化矩阵。给定一个单位长度的旋转轴 $[i, j, k]$ 和角度 θ ,对应的四元数为

$$q = \left[\cos \frac{\theta}{2}, [i, j, k] \sin \frac{\theta}{2} \right] \quad (1)$$

给定一个欧拉角 $[\varphi, \rho, \omega]$, $[\varphi, \rho, \omega]$ 为沿着 x - y - z 坐标轴依次旋转的角度,将其转换为四元数的方法如下:

$$q = \begin{bmatrix} v \\ a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \cos \frac{\varphi}{2} \cos \frac{\rho}{2} \cos \frac{\omega}{2} + \sin \frac{\varphi}{2} \sin \frac{\rho}{2} \sin \frac{\omega}{2} \\ \sin \frac{\varphi}{2} \cos \frac{\rho}{2} \cos \frac{\omega}{2} - \cos \frac{\varphi}{2} \sin \frac{\rho}{2} \sin \frac{\omega}{2} \\ \cos \frac{\varphi}{2} \sin \frac{\rho}{2} \cos \frac{\omega}{2} + \sin \frac{\varphi}{2} \cos \frac{\rho}{2} \sin \frac{\omega}{2} \\ \cos \frac{\varphi}{2} \cos \frac{\rho}{2} \sin \frac{\omega}{2} + \sin \frac{\varphi}{2} \sin \frac{\rho}{2} \cos \frac{\omega}{2} \end{bmatrix} \quad (2)$$

预测器分别输出一个维度为3的预测矢量 $\hat{p}$ 和一个维度为4的预测四元数 $\hat{q}$ 。预测误差如下:

$$L_p = \|p - \hat{p}\|, \quad (3)$$

$$L_q = \left\| q - \frac{\hat{q}}{\|\hat{q}\|} \right\|, \quad (4)$$

式中: τ 为距离范数。单位化的四元数存在于四维空间的球面上,当预测四元数 $\hat{q}$ 足够接近 q 时,球面距离和欧氏距离的差异可以忽略^[11],因此本文选择欧氏距离作为误差函数。在范数的选择上,由于 L_1 范数对于所有样本点误差都以相同权

重优化,而 L_2 范数则给予不同的误差以不同的优化权重,且由于 L_1 范数的一阶导数在极小值点附近也保持固定,损失函数值震荡较大。本文实验亦证明 L_2 范数较优,故选用 $\tau = 2$ 作为损失函数的范数。

单任务损失函数 g 定义为

$$g(L, \sigma) = Le^{-\ln(1+\sigma^2)} + \ln(1 + \sigma^2) \quad (5)$$

式中: L 为式(3)或式(4)所示欧氏预测误差; σ 作为模型的同方差不确定性^[23]被加入损失函数中,设置为可训练参数。同方差不确定性是描述系统固有噪声的指标,如传感器噪声,特点是增大数据量也无法减小这种系统噪声。加入 σ 有助于增加网络可训练性。

为避免损失函数出现负值,未采用初始化 σ^2 方法^[21],而改为初始化 σ ,使正则项落在正实数区间,从而使网络收敛更加稳定。由于将位置和姿态看成2个任务,总损失函数定义为

$$L_{\text{total}} = g(L_p, \sigma_p) + g(L_q, \sigma_q) \quad (6)$$

由于多任务之间的损失函数尺度大小不同,需要有不同的权重来进行平衡以利于网络训练,所以本文的损失函数加入了不确定性。在训练过程中,网络自动学习平衡位置和姿态损失项的最佳权重 σ_p, σ_q ,既可以保证稳定优化 L_p 和 L_q ,同时作为正则项被加入损失函数又不会因为 σ 太大而使数据项权重太小导致网络无法收敛。

3 实验

3.1 数据集

为便于与其他算法进行比较,使用 Microsoft 提供的 7-scenes 室内场景数据集^[10]来评价本文 BioLocNet 结构。7-scenes 数据集广泛运用于相机位姿估计任务,其包含 7 个场景,同时提供 RGB 及原始深度图。图像由 KinectRGB-D 传感器采集,真实位姿标注由 KinectFusion^[24]计算得出。每个场景的图像由分辨率为 640 × 480 的图片组成,被划分为训练集及测试集。场景具有复杂的纹理及光照信息,同时视角的改变和运动模糊也为相机位姿估计带来极大挑战。图 5 展示了 7 个场景的数据集图片,每个场景随机采样一张展示。

3.2 实验场景

按照官方推荐划分的训练集及测试集来训练和测试每个场景。本文使用 GPU 型号为 NVIDIA1080 Ti,在内存为 32GB、CPU 为 i7-6700 的计算机上进行实验。程序在 tensorflow 框架下运行,初始学习率设为 10^{-3} ,衰减率设为 0.955,优化器使用 Adam。在预处理阶段,每幅图像先缩

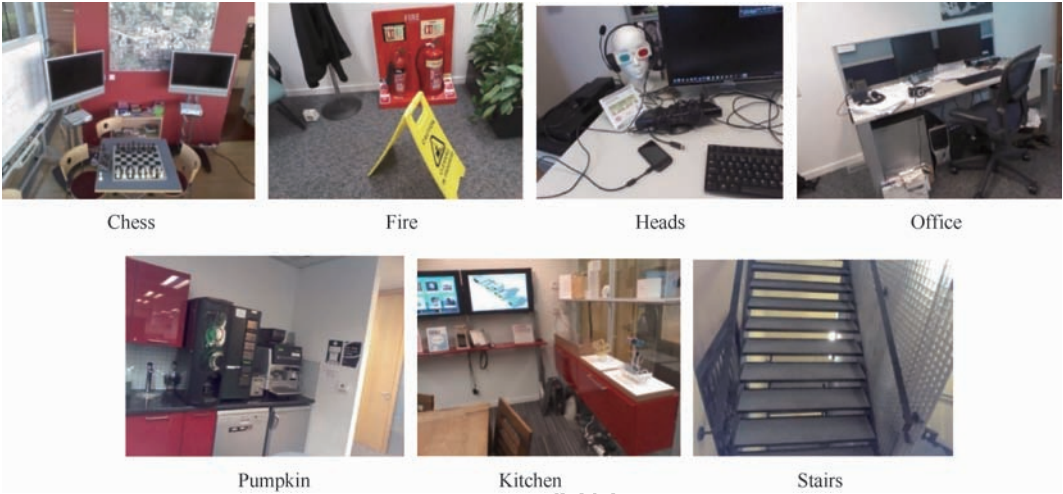


图 5 7-scenes 数据集样例
Fig. 5 Samples of 7-scenes dataset

放到 343×343 , 随后减去每个场景图像均值再除以标准差, 最后在线随机裁剪到 300×300 。在测试阶段, 使用图像的中心裁剪进行测试。训练时, 每个 epoch 将数据集顺序打乱。损失函数权重初始化为 $\sigma_p = 0, \sigma_q = -1$, batchsize 设为 16, 迭代 $20 \times 10^3 \sim 100 \times 10^3$ 步, 训练网络直至收敛。

3.3 位姿估计精度

选用与其他算法相同的评价标准来评估本文的实验结果。通常的做法是分别求出测试集预测结果中位置和姿态的中位数进行比较。表 1 展示了本文结果与其他目前最先进算法结果的对比, 所使用的网络为本文提出的 BiLocNet。

表 1 给出了 7 个场景的测试误差, 表中粗黑体数据表示该数据在横向对比中最优。实验结果表明, 本文网络的表现超过了目前最先进的输入为 RGB 图像的同类型算法, 有 3 个场景的位置精度达到了最优, 6 个场景的姿态精度超越其他算法, 平均精度值也超越了其他算法。由实验结果可以看出, 得益于将位置和姿态分别进行预测的双路网络, 预测出的姿态精度大大超越其他算法; 且实验发现, 只使用单路解码器及预测器的结构 (Hourglass) 在 Chess 场景下位置误差为 0.15 m, 姿态误差为 6.02° , 比使用双路结构位置精度下降 15%, 姿态精度下降 17%。实验结果证明了本文算法双路结构的必要性和有效性。

表 1 不同场景下不同算法的位置误差和姿态误差

Table 1 Position error and orientation error with different scenes for various algorithms

场景	位置误差/m, 姿态误差/($^\circ$)					
	PoseNet ^[11]	Bayesian DS ^[15]	LSTM-Pose ^[13]	VidLoc ^[12]	Hourglass ^[14]	BiLocNet
Chess	0.32, 8.12	0.28, 7.05	0.24, 5.77	0.18, N/A	0.15, 6.17	0.13, 5.13
Fire	0.47, 14.4	0.43, 12.52	0.34, 11.9	0.26, N/A	0.27, 10.84	0.29, 10.48
Heads	0.29, 12.0	0.25, 12.72	0.21, 13.7	0.14, N/A	0.19, 11.63	0.16, 12.67
Office	0.48, 7.68	0.30, 8.92	0.30, 8.08	0.26, N/A	0.21, 8.48	0.25, 6.82
Pupkin	0.47, 8.42	0.36, 7.53	0.33, 7.00	0.36, N/A	0.25, 8.12	0.25, 5.23
Kitchen	0.59, 8.64	0.45, 9.80	0.37, 8.83	0.32, N/A	0.27, 10.15	0.26, 6.95
Stairs	0.47, 13.8	0.42, 13.06	0.40, 13.7	0.26, N/A	0.29, 12.46	0.33, 9.86
均值	0.44, 10.44	0.35, 10.22	0.31, 9.85	0.25, N/A	0.23, 9.69	0.23, 8.16

3.4 预测器结构分析

预测器的结构直接决定了解码后的特征如何映射为位姿矢量。

本文在网络解码部分尝试过全连接预测器, 结构如图 6 所示, 即将解码器的输出直接通过一个特征向量大小为 2 048 的全连接层做预测, 发

现其精度较差。这种情况的出现是因为全连接层将特征矩阵展开为一维特征向量, 会损失特征矩阵的空间结构, 更适用于分类任务, 所以应避免预测器出现全连接层。本文进一步尝试使用 GAP 预测器, 结构如图 7 所示, 即将解码器输出使用卷积层降分辨率升通道, 之后的全连接层替换为全

局均值池化层,由于 GAP 避免了空间信息的损失及不必要的特征映射,并将解码器输出特征矩阵进行了调整,所以模型效果大幅提升。

为了比较不同预测器模型之间的差异,选择数据集中的一个 Chess 场景进行对比实验,结果见表 2,其中多尺度位姿预测器指本文使用带有 MSR 模块的多尺度位姿预测器。

对于位置和姿态估计任务来说,使用 MSR 增加了网络的深度及宽度,其多尺度卷积核可以重组不同空间尺度的特征信息,保证了较高分辨率的特征图经过有效调整之后合理地映射为位姿矢量。经实验表明,将 MSR (见图 3) 替换为 SSR (见图 4),精度下降 8%,证明了 MSR 多尺度结构的有效性。

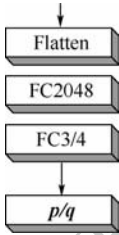


图 6 全连接预测器
Fig. 6 FC estimator



图 7 GAP 预测器
Fig. 7 GAP estimator

表 2 不同预测器的网络精度对比

Table 2 Comparison of network precision of different estimators		
预测器	位置误差/m	姿态误差/(°)
全连接预测器	0.26	8.03
GAP 预测器	0.14	5.21
多尺度位姿预测器	0.13	5.13

3.5 损失函数分析

神经网络可以看为一个复杂的高维函数,而损失函数的选择是优化此高维函数至关重要的步骤。PoseNet 的 α 权重损失函数定义为^[11]

$$L_{total} = L_p + \alpha L_q \tag{7}$$

式中: α 为平衡位置项与姿态项的权重。 α 权重损失函数的缺点是需要手动设置,欲达到最优效

果,通常会使用网格搜索法来确定。在没有先验知识支持的条件下,耗时巨大且表现往往并非最佳。

本文引入了可训练的权重(σ 权重)以自动地学习不同任务间的平衡(式(7)),效果优于 α 权重损失函数,在无 MSR 的 BiLocNet 上位置精度提升了 10.5%,姿态精度提升了 13.9%。数据见表 3,同样使用 Chess 子数据集作为说明。

表 3 α 权重损失函数与 σ 权重损失函数的网络精度对比
Table 3 Comparison of network precision between α weights loss function and σ weights loss function

损失函数	位置误差/m	姿态误差/(°)
α 权重	0.19	6.55
σ 权重	0.17	5.64

3.6 预训练分析

迁移学习指通过预训练从大数据集的分类任务中学习到特征的表示,将保存的权重模型加载到其他任务的网络作为初始化权重,从而加速网络收敛的一种方式。多数研究都借助于迁移学习的预训练模型初始化网络权重,但在本文网络结构下,无预训练的方式优于有预训练的方式,这是因为:①分类数据集与位姿估计数据集样本分布及内容相差较大,迁移学习得到的特征表示由于是从分类数据集中学得,用其作为初始化权重反而对位姿估计问题是有害的。②进行预训练需要花费较长的时间,使用数据集增强及增加迭代轮数等手段实现方便快速且在本文讨论问题上精度更优。何凯明等^[25-26]也在研究成果中提到对于非分类任务,迁移学习可能并不会取得更好的效果。本文在数量为 100 000 的部分 ImageNet 数据集上进行编码器的预训练,并将训练好的权重作为网络训练前的权重初始化进行迁移学习。有无预训练的对比结果如图 8 所示。

可以看出,无预训练具有收敛速度快、收敛精度高的特点,可以快速使网络收敛到最小值附

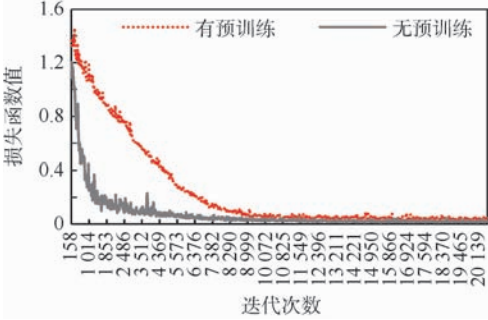


图 8 有无预训练的 BiLocNet 损失函数曲线
Fig. 8 Loss function curves of pre-trained and non-pre-trained BiLocNet

近。表 4 展示了有无迁移学习的结果对比。表中使用 Fire 场景作为测试场景,示例图片见图 5。定性分析和定量分析结果均表明,在本文网络上使用预训练模型对结果没有提升,甚至有所下降。

表 4 有无迁移学习的网络精度对比

Table 4 Comparison of network precision with and without transfer learning		
有无迁移学习	位置误差/m	姿态误差/(°)
有	0.32	10.64
无	0.29	10.48

4 结 论

1) 本文算法可实现较为优异的预测性能,在室内数据集上可将平均位置误差缩小至 0.23 m,平均姿态误差缩小至 8.16°。

2) 跳跃连接和解码器可以补充编码过程中损失的空间信息,与多尺度位姿预测模块相互配合,可以大幅提升预测精度。

3) 实验表明,由于数据集差异较大,迁移学习对于本文算法效果不佳。损失函数中加入可训练权重使网络的位姿精度平均提升了 12.2%,且可以避免花费大量时间手动选择超参数。

4) 研究待解决问题的特性有助于提升实验结果,本文算法挖掘位置和姿态的特性和区别,针对性地设计双路网络,使位姿精度平均提升了 14%。

此外,依然有几个值得继续研究的问题:多任务之间通过训练得到的损失函数权重是否是最优解;解码器恢复特征图到什么程度的空间分辨率是最佳效果;2 个子任务之间是否可以通过交互提升性能? 下一步将继续研究这些问题。

参考文献 (References)

[1] CHEN D M,BAATZ G,KOSER K,et al. City-scale landmark identification on mobile devices[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE Press,2011:12258110.

[2] TORII A,SIVIC J,PAJDLA T,et al. Visual place recognition with repetitive structures[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE Press,2013:883-890.

[3] SCHINDLER G,BROWN M,SZELISKI R. City-scale location recognition[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE Press,2007:1-7.

[4] ARTH C,PIRCHHEIM C,VENTURA J,et al. Instant outdoor localization and SLAM initialization from 2.5 D maps[J]. IEEE Transactions on Visualization and Computer Graphics,2015,21(11):1309-1318.

[5] POGLITSCH C,ARTH C,SCHMALSTIEG D,et al. A particle filter approach to outdoor localization using image-based rendering[C] // IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Piscataway, NJ: IEEE Press,2015:132-135.

[6] SATTLER T,LEIBE B,KOBELT L. Improving image-based localization by active correspondence search[C] // Proceedings of European Conference on Computer Vision. Berlin:Springer,2012:752-765.

[7] LI Y, SNAVELY N, HUTTENLOCHER D, et al. Worldwide pose estimation using 3D point clouds[C] // Proceedings of European Conference on Computer Vision. Berlin:Springer,2012:15-29.

[8] CHOUDHARY S, NARAYANAN P J. Visibility probability structure from SFM datasets and applications[C] // Proceedings of European Conference on Computer Vision. Berlin:Springer,2012:130-143.

[9] SVARM L,ENQVIST O,OSKARSSON M,et al. Accurate localization and pose estimation for large 3D models[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE Press,2014:532-539.

[10] SHOTTON J,GLOCKER B,ZACH C,et al. Scene coordinate regression forests for camera relocation in RGB-D images[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press,2013:2930-2937.

[11] KENDALL A,GRIMES M,CIPOLLA R. PoseNet: A convolutional network for real-time 6-DOF camera relocation[C] // Proceedings of IEEE International Conference on Computer Vision. Piscataway,NJ:IEEE Press,2015:2938-2946.

[12] WALCH F,HAZIRBAS C,LEAL-TAIXÉ L,et al. Image-based localization with spatial LSTMs[EB/OL]. (2016-11-23) [2018-12-25]. https://arxiv.org/pdf/1611.07890v1.

[13] CLARK R,WANG S,MARKHAM A,et al. VidLoc: A deep spatio-temporal model for 6-DOF video-clip relocation[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press,2017:6856-6864.

[14] KENDALL A, CIPOLLA R. Modelling uncertainty in deep learning for camera relocation[C] // Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Piscataway,NJ:IEEE Press,2016:4762-4769.

[15] MELEKHOV I,YLIOINAS J,KANNALA J,et al. Image-based localization using hourglass networks[EB/OL]. (2017-08-24) [2018-12-25]. https://arxiv.org/abs/1703.07971.

[16] LI R,LIU Q,GUI J,et al. Indoor relocation in challenging environments with dual-stream convolutional neural networks[J]. IEEE Transactions on Automation Science and Engineering,2018,15(2):651-662.

[17] RADWAN N,VALADA A,BURGARD W. Vlocnet + + : Deep multitask learning for semantic visual localization and odometry[EB/OL]. (2016-10-11) [2018-12-25]. https://arxiv.org/abs/1804.08366.

[18] BRAHMBHATT S,GU J,KIM K,et al. Geometry-aware learn-

ing of maps for camera localization[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 2616-2625.

[19] LI X, YLIOINAS J, KANNALA J. Full-frame scene coordinate regression for image-based localization [EB/OL]. (2018-01-25) [2018-12-25]. <https://arxiv.org/abs/1802.03237>.

[20] LASKAR Z, MELEKHOV I, KALIA S, et al. Camera relocation by computing pairwise relative poses using convolutional neural network[C] // Proceedings of IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 929-938.

[21] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 7482-7491.

[22] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C] // Thirty-First AAAI Conference on Artificial Intelligence, 2017: 4-12.

[23] KENDALL A, GAL Y. What uncertainties do we need in Bayesian deep learning for computer vision? [EB/OL]. (2017-10-05) [2018-12-26]. <https://arxiv.org/abs/1703.04977>.

[24] IZADI S, KIM D, HILLIGES O, et al. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera [C] // Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. New York: ACM, 2011: 559-568.

[25] HE K M, ROSS G, PIOTR D. Rethinking imagenet pre-training [EB/OL]. (2015-11-21) [2018-12-25]. <https://arxiv.org/abs/1811.08883>.

[26] WU Y X, HE K M. Group normalization[C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018, 3-19.

作者简介:
贾瑞明 男, 博士, 助理研究员。主要研究方向: 计算机视觉、深度学习、模式识别。
刘圣杰 男, 硕士研究生。主要研究方向: 计算机视觉、深度学习。

A visual localization method based on encoder-decoder dual-stream CNN

JIA Ruiming^{1,*}, LIU Shengjie¹, LI Jintao², WANG Yunhao², PAN Haixia²

(1. School of Information Science and Technology, North China University of Technology, Beijing 100144, China;
2. College of Software, Beihang University, Beijing 100083, China)

Abstract: In order to calculate the camera pose from a single RGB image, a deep encoder-decoder dual-stream convolutional neural network (CNN) is proposed, which can improve the accuracy of visual localization. The network first uses an encoder to extract advanced features from input images. Second, the spacial-resolution is enhanced by a pose decoder. Finally, a multi-scale estimator is used to output pose parameters. Because of the different performance of position and orientation, the network adopts a dual-stream structure from the decoder to process the position and orientation separately. To restore the spatial information, several skip connections are added to encoder-decoder architecture. The experimental results show that the accuracy of the network is obviously improved compared with the congener state-of-the-art algorithms, and the orientation accuracy of camera pose is improved dramatically.

Keywords: visual localization; encoder-decoder architecture; convolutional neural network (CNN); skip connection; dual-stream network

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0013

基于原始对偶内点法的 EST 图像重建



薛倩^{1,*}, 刘婧¹, 马敏¹, 王化祥²

¹民航大学 电子信息与自动化学院, 天津 300300; ²天津大学 电气自动化与信息工程学院, 天津 300072)

摘

要: 静电层析成像(EST)被动感应电荷的机理决定了其独立测量值数等于电极

数目,远小于电容层析成像(ECT)等相对成熟的电学成像(ET)技术的测量值数,导致逆问题的欠定性更加严重。为此,对基于压缩感知理论的 EST 图像重建算法进行了研究。利用奇异值分解(SVD)处理灵敏度矩阵使其满足有限等距性质(RIP),采用 l_1 范数正则化模型和原始对偶内点法(PDIPA)实现图像重建,并在迭代过程中针对荷电磨粒稀疏分布的特点,对图像向量中非零元素个数施加约束。仿真实验表明:该算法相对于基于“Circle of Apollonius”的反投影(BP)算法和 Landweber 迭代算法,明显改进了成像质量,对不同位置的单个电荷可准确重建;2 个电荷距离不小于 1 mm 时可正确分辨电荷数目与位置;对 10 组随机分布的 3 个电荷模型进行测试,荷电磨粒数目监测的准确率约为 80%。

关键词: 静电层析成像(EST); 油液监测; 压缩感知; 图像重建; 正则化

中图分类号: V221+.3; TH89

文献标识码: A

文章编号: 1001-5965(2019)10-1973-09

航空发动机高温、高压、高载荷的工况条件使其零部件异常磨损几率相对较高,由发动机齿轮、轴承等零部件磨损造成的故障占提前换发故障的 60% 和空中停车故障的 37.5%。航空发动机零部件异常磨损不易被及时察觉,研究在线油液监测技术,通过实时监测滑油回路中的异常大磨粒(粒径 $\geq 40 \mu\text{m}$)监测发动机的磨损状况,对故障预警具有重要意义^[1-2]。

基于静电感应原理的静电监测技术是一种有效的在线油液磨粒监测技术。发动机零部件磨损过程中产生的荷电磨粒随油液进入润滑系统中,并受油液的绝缘保护而保持荷电特性。荷电磨粒流经静电传感器时,环状或棒状电极感应区域中油液的净电荷水平变化,由此反映各个时刻流经传感器的油液中磨粒的含量^[2]。但静电传感器测的是区域内所有荷电磨粒的静电场叠加值,无

法获取异常大磨粒的数量、位置等信息^[3]。在此基础上,静电层析成像(Electrostatic Tomography, EST)技术利用特殊的阵列式传感器获取静电信号,通过理论分析或数值计算建立电荷分布与测量数据之间的映射模型,然后求解逆问题以重建管道中电荷分布图像,由此可直观、实时地反映荷电磨粒数量、位置等信息。相对于电阻抗层析成像(EIT),EST 是被动感应式的成像,无需激励,响应速度更快,但测量值数等于电极数目,欠定性较 EIT 更加严重。在 EST 逆问题求解方面,1997 年,英国的 Green 等^[4]针对监测重力输送的双截面 16 电极 EST 系统,采用 EIT 中常用的线性反投影(LBP)算法重建图像并获得物料浓度。1999 年,日本的 Machida 和美国的 Scarlett 设计了一套 EST 系统,后于 2005 年利用仿真实验比较了基于“Circle of Apollonius”的反投影(BP)算法

收稿日期: 2019-01-16; 录用日期: 2019-02-02; 网络出版时间: 2019-02-22 08:45

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190221.1541.005.html

基金项目: 国家自然科学基金(61401466); 中国民航大学科研启动基金(2013QD01S)

* 通信作者。E-mail: xueqian@tju.edu.cn

引用格式: 薛倩, 刘婧, 马敏, 等. 基于原始对偶内点法的 EST 图像重建[J]. 北京航空航天大学学报, 2019, 45(10): 1973-1981.
XUE Q, LIU J, MA M, et al. EST image reconstruction based on primal dual interior point algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1973-1981 (in Chinese).

和最小方差(LS)算法^[5],与直接采用 EIT 成像算法相比,针对 EST 设计的基于等势线交叠定位电荷的 BP 算法可准确重建不同位置的单个电荷,但难以分辨场中的 2 个荷电磨粒;LS 算法可区分 2 个荷电磨粒,但空间分辨率很低。2009 年,马来西亚的 Rahmat 等^[6]将带正则化的 LS(LSR)算法用于 EST 图像重建,并与 CCD 摄像机提取的颗粒分布特征比较,验证了 LSR 算法的准确性;2013 年,他们又利用有限元法(FEM)建模,相对于 LBP 算法和滤波反投影(FBP)算法提高了成像质量,但空间分辨率同样受到限制^[7]。国内,东南大学的研究人员将 Tikhonov 迭代算法用于 EST 图像重建^[8],后又针对浓相气固两相流提出以电容层析成像(ECT)重建图像为先验知识,排除浓度低的像素点,以提高 EST 的成像质量,并通过仿真实验验证了该方法的有效性^[9]。2014 年,Thuku 和 Rahmat^[10]比较了 4 阵列传感器与 16 阵列传感器在电荷层析成像中的适用性,其图像重建过程采用了 Landweber 迭代算法。2015 年,Qian 和 Wang 等^[11-12]分别研究了静电传感器在测量气力输送系统中颗粒平均流速方面的应用。浓相气固两相流中颗粒形成连续分布, l_2 范数正则化可取得较好的成像效果,而对于发动机气路或油路磨粒监测,监测对象本身具有稀疏性。根据 NAS-1638 污染度等级标准,污染度 8 级(即认为滑油很脏)时,100 mL 油液中含 50 ~ 100 μm 颗粒的数目约为 360^[13]。本文设计的 EST 传感器模型中,滑油管道直径为 10 mm,阵列电极轴向长度为 10 mm,敏感场空间体积为 0.785 mL,对应污染度 8 级时的大颗粒数目约为 3。因此,适于选择图像向量的稀疏性为约束,并用 l_1 范数正则化算法求解图像重建问题。2016 年,Tang 等^[14]建立了一种半球形阵列式静电传感器模型,并分析了其空间灵敏独特性;后在此基础上,将静电传感器阵列和 l_1 范数正则化模型用于发动机气路异常磨粒监测,并用基追踪算法重建图像,相对于 l_2 范数正则化算法明显提高了空间分辨率^[3]。

本文通过设计小尺度阵列电极传感器,并针对逆问题的严重欠定性研究了基于压缩感知理论的 EST 图像重建算法,以探索 EST 在荷电磨粒在线监测领域的应用。

1 EST 数学模型

根据静电感应原理,结合滑油管道的实际尺寸,采用与小尺度 8 电极 ECT 传感器^[15]类似的结构,利用有限元分析软件 COMSOL 建立小尺度

EST 传感器 3D 模型,如图 1 所示。管道内径 10 mm,为安装方便,将电极设置在管道外壁。考虑空间限制和电荷值范围,取阵列电极数目为 8,增加电极数目可增加边界测量值从而提高重建精度,但会减小电极面积从而降低信噪比。可利用有限元仿真对电极尺寸进行优化。本文取电极占空比为 0.9,轴向长度为 10 mm。

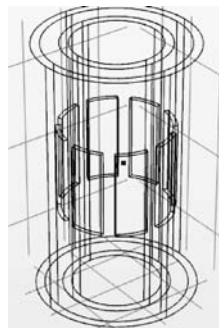


图 1 EST 传感器仿真模型

Fig. 1 EST sensor model used for simulation

当点电荷经过测量电极时,由于静电感应将在电极内外表面产生大小相等、符号相反的电荷。点电荷所产生的静电场与电极上感应电荷相互作用,导体达到静电平衡,此过程在极短时间内完成,因此移动点电荷与静电传感器之间的相互作用可以用静电场的 Poisson 方程及其边界条件来描述^[8]:

$$\begin{cases} \nabla \cdot [\varepsilon_0 \varepsilon(x, y, z) \nabla \phi(x, y, z)] = -\rho(x, y, z) \\ \phi(x, y, z) |_{(x, y, z) \cdot \tau_p} = 0 \\ \phi(x, y, z) |_{(x, y, z) \cdot \tau_{ie}} = \text{Const} \\ E_x = 0 \end{cases} \quad (1)$$

式中: $\phi(x, y, z)$ 为场域内的电势分布; τ_p 为接地屏蔽罩构成的边界,电势为 0; τ_{ie} 为与检测电路相接的阵列式电极 i 构成的边界,其为一等势体; $\rho(x, y, z)$ 为场域内的体电荷密度函数; ε_0 为真空介电常数; $\varepsilon(x, y, z)$ 为材料的相对介电常数分布; E_x 为无穷远处的电场强度。

由于式(1)难以求得解析解,可用有限元法求其数值解,得到场强分布后可求电极上的感应电量。

EST 各电极上的感应电荷可表示为

$$q_i = \int_D g(x, y) s_i(x, y) dx dy \quad (2)$$

式中: q_i 为电极 i 上的感应电荷值; $g(x, y)$ 为管道截面上 (x, y) 处的电荷值,对应重建图像中该位置的灰度值; D 表示管道截面; $s_i(x, y)$ 为电极 i 在 (x, y) 处的灵敏度,有限元仿真中按下式计算:

$$s_i(x, y) = \left| \frac{q_i}{q(x, y)} \right| \quad (3)$$

其中: $q(x, y)$ 为设置于 (x, y) 处的点电荷电量^[8]。

将式(2)离散化处理后,可得如下矩阵形式:

$$\mathbf{Q} = \mathbf{S}\mathbf{G} \quad (4)$$

式中: $\mathbf{Q}$ 为 $M \times 1$ 维的感应电荷值向量; $\mathbf{G}$ 为 $N \times 1$ 维的电荷分布向量,即重建图像的灰度向量; $\mathbf{S}$ 为 $M \times N$ 维的灵敏度矩阵,用于描述感应电荷值与电荷磨粒之间的映射。

为保证一定的空间分辨率,将测量截面用 20×20 的网格离散化,得到管道内像素单元数 $N = 317$,而测量值数等于电极数目 $M = 8$,因此,由灵敏度矩阵和感应电荷测量值求解电荷分布的逆问题具有严重欠定性。

2 EST 图像重建

2.1 灵敏度矩阵预处理

2.1.1 归一化

图 2(a) 展示了一个测量电极对场域内全部像素点的灵敏度。可以发现,电极附近区域的灵敏度较高,即靠近管壁的区域若存在电荷,较容易被检测到。而场域中心附近灵敏度偏低,可以认为管道中心处若出现电荷,对传感器的输出几

乎没有影响。这将导致基于灵敏度矩阵的图像重建算法对管道中央区域的分辨率低,严重影响成像质量^[16]。对灵敏度系数进行归一化处理将有所改善,经仿真分析对比,对 EST 灵敏度矩阵设计如下归一化公式:

$$\bar{S}(i, j) = \frac{S(i, j)}{\max(\mathbf{S}_j)} \quad (5)$$

式中: $S(i, j)$ 为灵敏度矩阵中第 i 行第 j 列的元素; $\bar{S}(i, j)$ 为归一化后的灵敏度系数; $\mathbf{S}_j$ 为灵敏度矩阵中第 j 列向量。

归一化后灵敏度分布均匀性有所改善,效果如图 2(b) 所示。相对于文献[8]中提出的“非等网格划分法”(即对中央区域采用较粗的剖分网格以提高灵敏度值的方法),理论上更合理,改善效果也更好。

2.1.2 奇异值分解

压缩感知理论指出,若信号是可压缩的或在某个正交基空间中是稀疏的,就可用远低于奈奎斯特采样定理所要求的采样频率对信号进行采样并准确恢复原始信号,仅要求压缩感知矩阵满足有限等距性质(RIP)。在电荷磨粒检测中,重建目标 $\mathbf{G}$ 本身具有稀疏性,而灵敏度矩阵 $\mathbf{S}$ 是否满足 RIP 难以判定。为此,利用奇异值分解(SVD)对归一化后的灵敏度矩阵进行处理,将其变换为一个部分正交矩阵^[3,17-18](部分正交矩阵满足 RIP 性质),以用于逆问题求解。

灵敏度矩阵 $\mathbf{S}$ 行满秩,可分解为

$$\mathbf{S} = \mathbf{U}[\mathbf{\Delta} \quad \mathbf{0}]\mathbf{V}^T \quad (6)$$

式中: $\mathbf{U} \in \mathbf{R}^{M \times M}$ 和 $\mathbf{V} \in \mathbf{R}^{N \times N}$ 为正交矩阵; $\mathbf{\Delta} = \text{diag}(\delta_1, \delta_2, \dots, \delta_M)$, $\delta_1 \geq \delta_2 \geq \dots \geq \delta_M > 0$ 为 $\mathbf{S}$ 的奇异值。

令 $\mathbf{\Delta}^* = \text{diag}(1/\delta_1, 1/\delta_2, \dots, 1/\delta_M)$, 并定义 $\mathbf{Z} = \mathbf{\Delta}^* \mathbf{U}^T \mathbf{S} = [\mathbf{I}^{M \times M} \quad \mathbf{0}]\mathbf{V}^T$, 则可按下式处理灵敏度矩阵^[16]:

$$\mathbf{S}_{\text{SVD}} = \mathbf{Z} \begin{bmatrix} 1/\|\mathbf{z}_1\|_2 & 0 & \cdots & 0 \\ 0 & 1/\|\mathbf{z}_2\|_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\|\mathbf{z}_N\|_2 \end{bmatrix} \quad (7)$$

式中: $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ 为 $\mathbf{Z}$ 的列向量。

相应地,感应电荷值向量变换为

$$\mathbf{Q}_{\text{SVD}} = \mathbf{\Delta}^* \mathbf{U}^T \mathbf{Q} \quad (8)$$

于是,EST 图像重建模型变换为

$$\mathbf{Q}_{\text{SVD}} = \mathbf{S}_{\text{SVD}} \mathbf{G}_{\text{SVD}} \quad (9)$$

式中: $\mathbf{G}_{\text{SVD}}$ 为对灰度向量 $\mathbf{G}$ 缩放后的向量。求解 $\mathbf{G}_{\text{SVD}}$ 后可由下式计算 $\mathbf{G}$:

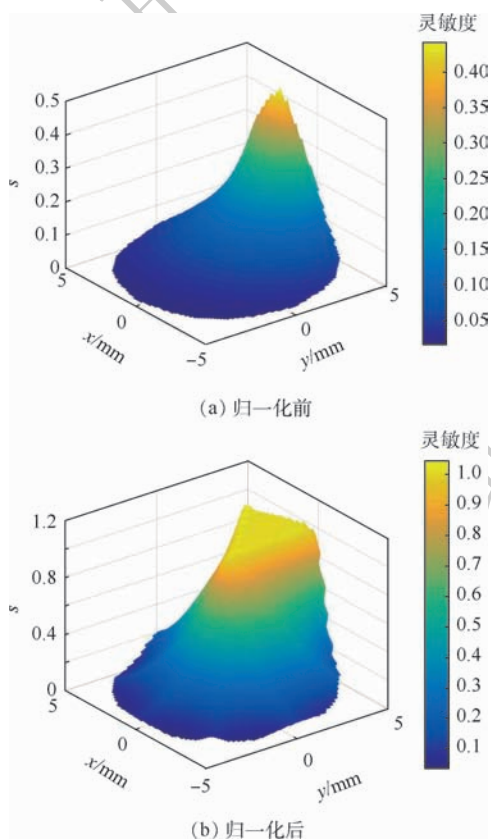


图 2 EST 传感器敏感场分布

Fig. 2 Sensitivity distribution of EST sensor

$$\mathbf{G} = \mathbf{\Delta}^* \mathbf{G}_{\text{SVD}} \quad (10)$$

2.2 原始对偶内点法

针对式(9)的严重欠定性,在图像重建过程中加入稀疏性约束。根据压缩感知理论,可将 EST 图像重建问题转化为求式(9)的最小 l_1 范数解问题:

$$\begin{cases} \min \|\mathbf{G}_{\text{SVD}}\|_1 \\ \text{s. t. } \mathbf{Q}_{\text{SVD}} = \mathbf{S}_{\text{SVD}} \mathbf{G}_{\text{SVD}} \end{cases} \quad (11)$$

本文采用原始对偶内点法(PDIPA)^[19-20]求解该问题。为简单起见,对 $\mathbf{S}$ 和 $\mathbf{Q}$ 分别取灵敏度系数和感应电荷的绝对值,令 $\mathbf{G}_{\text{SVD}}$ 是非负的。由此式(11)可转化为如下线性规划问题:

$$\begin{aligned} & \text{Primal(P)} \\ & \begin{cases} \min \mathbf{1}^T \mathbf{G}_{\text{SVD}} \\ \text{s. t. } \mathbf{S}_{\text{SVD}} \mathbf{G}_{\text{SVD}} = \mathbf{Q}_{\text{SVD}} \\ \mathbf{G}_{\text{SVD}} \geq 0 \end{cases} \quad (12a) \\ & \text{Dual(D)} \\ & \begin{cases} \max \mathbf{Q}_{\text{SVD}}^T \mathbf{F} \\ \text{s. t. } \mathbf{S}_{\text{SVD}}^T \mathbf{F} + \mathbf{H} = \mathbf{1} \\ \mathbf{H} \geq 0 \end{cases} \quad (12b) \end{aligned}$$

PDIPA 法的基本思想是迭代地将不等式约束问题(11)表示为等式约束问题。首先,将原始问题(P)转化为对偶问题(D),然后,求解原始对偶间隙,得到互补条件。问题转化为求互补条件的最优解,可通过牛顿法求解;传统的牛顿法由于计算量大不适用于大型线性规划问题,为此,研究人员开发了获取近似解的快速算法^[17],可以相对较小的计算成本获得高精度解。在迭代求互补条件最优解的过程中不断更新原始变量和对偶变量,直到满足迭代停止条件。下文图像重建中基于 PDIPA 法的 $\mathbf{G}_{\text{SVD}}$ 的求解均在 MATLAB 平台上采用 L-1 Benchmark Package 实现^[21-22]。

在 8 电极 EST 系统中,由于边界测量值有限,上述方法重建的电荷分布图像仍不可避免地含有伪影。例如, (x, y) 处的一个点电荷在 8 个电极上感应的电荷值,可能与点 (x, y) 附近若干点电荷共同作用产生的感应电荷值相近,或者临近的 2 个点电荷产生的感应电荷值,与两点之间一个点电荷作用的结果相近。在不增加电极数目的前提下,为降低重建误差,将一般情况下荷电磨粒数不超过 3 个的先验信息融入迭代过程,即在 PDIPA 法每次更新变量之后,补充如下约束:

$$\mathbf{G}_{\text{SVD}}^t(k) = \begin{cases} 0 & \mathbf{G}_{\text{SVD}}^t(k) < \mathbf{g}_{p_{\max}}^t \\ \mathbf{G}_{\text{SVD}}^t(k) & \text{其他} \end{cases} \quad (13)$$

式中: $P_{\max}$ 为荷电磨粒数最大值; $k = 1, 2, \dots, N$; $\mathbf{g}_{p_{\max}}^t$ 为第 t 次迭代中 $\mathbf{G}_{\text{SVD}}^t$ 向量中第 $P_{\max}$ 大的元素值。

3 仿真实验

为验证 PDIPA 法的有效性,利用图 1 所示 8 电极 EST 传感器模型进行仿真实验。管道内径为 10 mm,管内填充滑油(相对介电常数取为 3),油液中分布点电荷以表示荷电磨粒,电荷值设为 10^{-9} C。在 COMSOL 5.3 和 MATLAB 2016b 平台上进行数值仿真,获取灵敏度矩阵和不同电荷分布情况下 8 个电极上的感应电荷值,用第 2 节方法进行图像重建,并对重建结果进行分析。计算机配置为 i7-6500U CPU 2.5 GHz CPU、12 GB 内存。

3.1 单个荷电磨粒的重建结果

由于磨损初期异常磨粒很少,多数情况下在传感器敏感空间内可能只有一个荷电磨粒,为测试本文算法对不同位置处单个点电荷的重建效果,在测量截面上沿圆心与电极中心连线、圆心与电极空隙中心连线这 2 个方向上,每间隔 1 mm 选一个点,即在图 3 所示的 $a \sim i$ 共 9 个位置依次放置单个点电荷进行仿真测试。分别用基于“Circle of Apollonius”的 BP 算法^[5]、Landweber 迭代算法^[23]和 PDIPA 重建电荷分布图像,并比较成像效果。

成像实验中,对 3 种算法采用相同的灵敏度矩阵归一化方法,且在 BP 算法中也加入了图像向量中非零元素不超过 3 的约束,以减少伪影。由图 4 可见,BP 算法对不同位置的单个点电荷可以较准确地重建,但相对于 PDIPA 重建的伪影较多。此外,BP 算法需要设定合适的阈值,不同

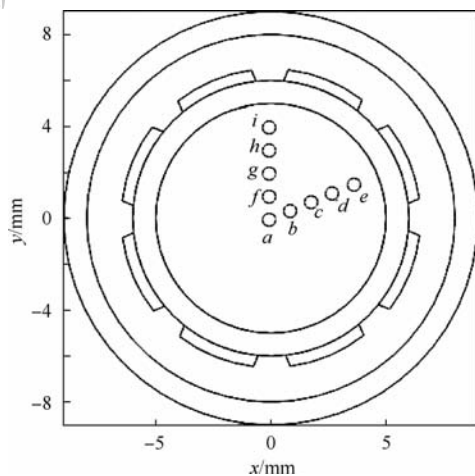


图 3 单个电荷仿真中设置的位置

Fig. 3 Positions used for single point charge simulation

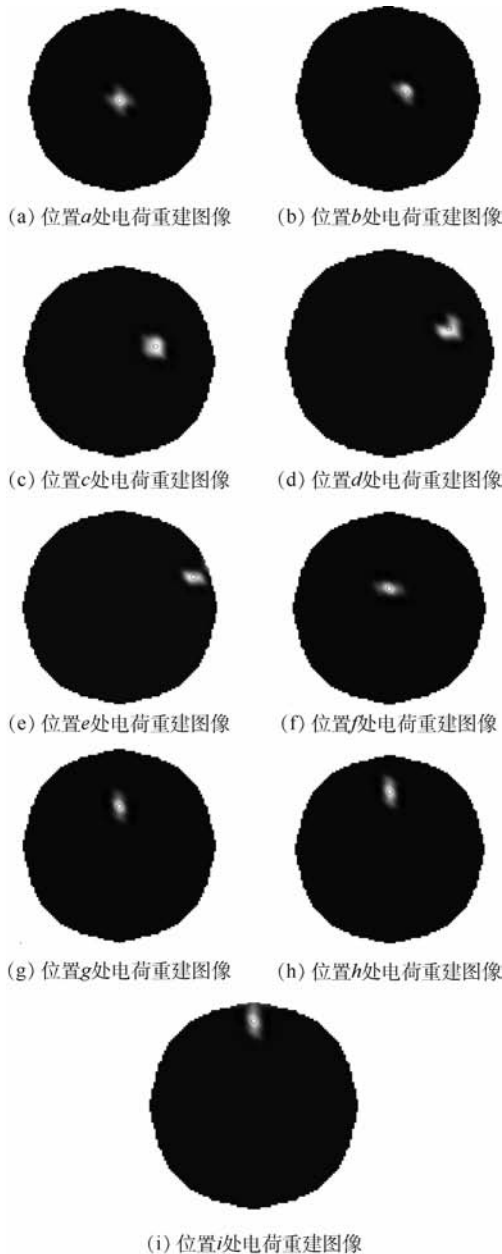


图 4 基于 BP 算法的单个电荷图像重建结果
Fig. 4 Image reconstruction results for single point charge based on BP algorithm

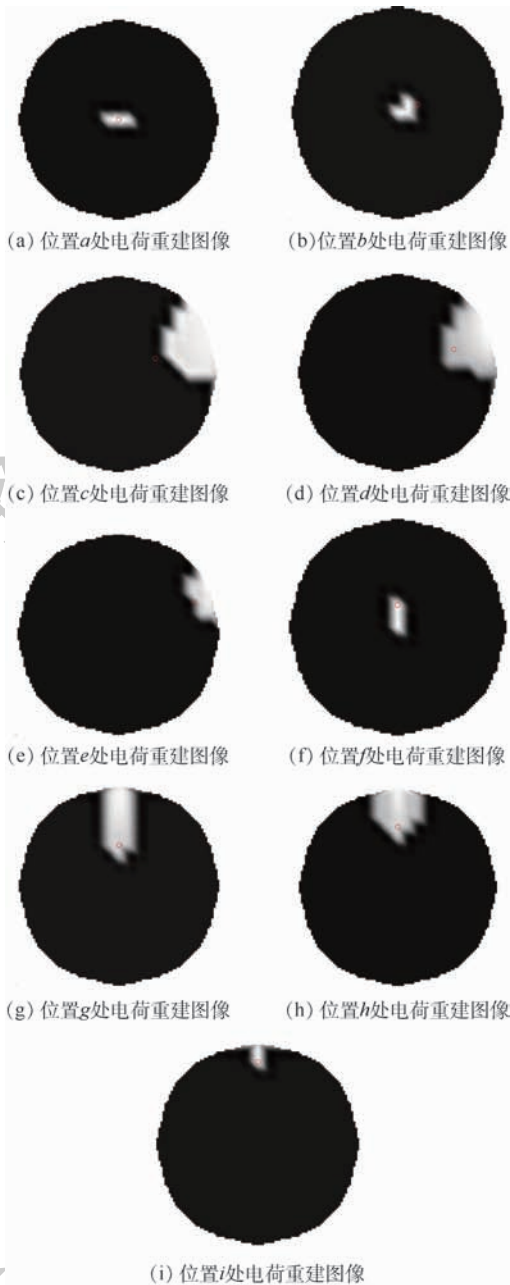


图 5 基于 Landweber 迭代算法的单个电荷图像重建结果
Fig. 5 Image reconstruction results for single point charge based on Landweber iteration algorithm

的检测模型需要不同的阈值,否则严重影响成像效果。

在图 5 所示的 Landweber 迭代算法重建图像中,电荷距离管道中心或边缘区域较近时,通过加入 $N_{\max}=3$ 的约束可减少伪影,而电荷位于 c 、 d 、 g 、 h 等位置时,由于电极附近伪影像素值较大,需调大 $N_{\max}$,否则无法重建出相应位置的电荷,因此图 5(c)、(d)、(g)、(h) 中伪影较严重。

由图 6 所示的 PDIPA 图像重建结果可见,基于 PDIPA 的图像重建可准确获得敏感场内不同位置处的单个荷电磨粒数目与坐标,相对于图 4 中 BP 算法和图 5 中 Landweber 迭代算法重建的

图 5 基于 Landweber 迭代算法的单个电荷图像重建结果
Fig. 5 Image reconstruction results for single point charge based on Landweber iteration algorithm

图像明显提高了成像质量。由于本文实验中剖分网格的尺寸为 0.5 mm,故按网格边长的整数倍选取了实验电荷坐标,若电荷位置坐标不在剖分网格点上,将出现一定的误差,若有必要可通过进一步细化网格来改进。实时性方面,PDIPA 对偶间隙的容差设为 10^{-5} ,最大原始-对偶迭代次数设为 50,耗时 0.3 s, BP 算法耗时 0.003 s, Landweber 迭代算法(迭代 100 次)耗时 0.01 s,均可满足在线监测的要求。

3.2 多个荷电磨粒的重建结果

当敏感场内出现 2 个异常荷电磨粒时,为验

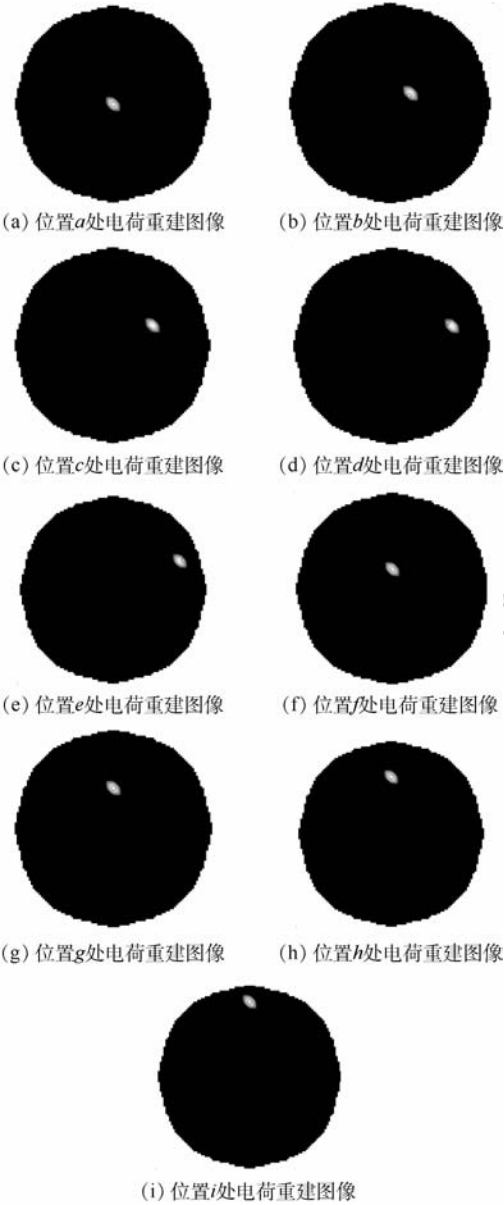


图 6 基于 PDIPA 的单个电荷图像重建结果
Fig. 6 Image reconstruction results for single point charge based on PDIPA

证本文算法的重建效果,设计一对电荷的 6 组分布位置,如图 7 所示。6 组位置分别用 $a'-a''$ 、 $b'-b''$ 、 $c'-c''$ 、 $d'-d''$ 、 $e'-e''$ 、 $f'-f''$ 表示,其坐标分别为: $a'(1,0)$ mm, $a''(0,1)$ mm, $b'(2,0)$ mm, $b''(0,2)$ mm, $c'(3,0)$ mm, $c''(0,3)$ mm, $d'(4,0)$ mm, $d''(0,4)$ mm, $e'(-0.5,0)$ mm, $e''(0.5,0)$ mm, $f'(-0.25,0)$ mm, $f''(0.25,0)$ mm, 以上坐标位置用圆圈标记在重建截面图像上。

依次在 $a'-a''$ 、 $b'-b''$ 、 $c'-c''$ 、 $d'-d''$ 四组位置处放置 2 个点电荷,并分别用 BP 算法、Landweber 迭代算法、PDIPA 进行图像重建,结果如图 8 ~ 图 10 所示。图 8 所示的 BP 算法在 2 个及以上电荷同时存在的情况下会产生严重的伪影,无法

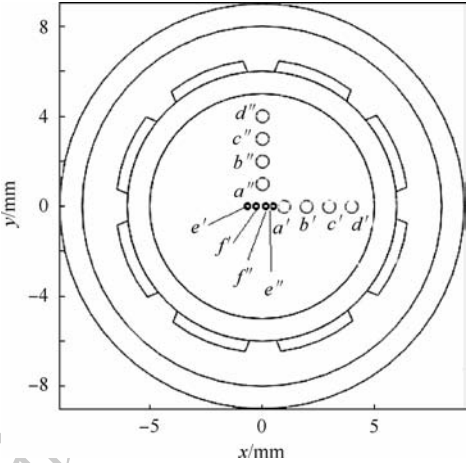


图 7 两个电荷仿真中设置的位置
Fig. 7 Positions used for two point charges simulation

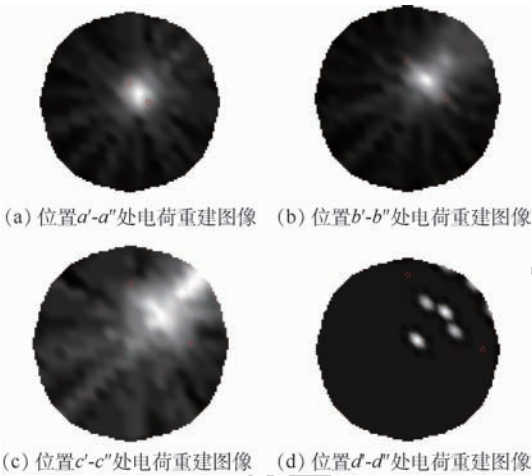


图 8 基于 BP 算法的 2 个电荷图像重建结果
Fig. 8 Image reconstruction results for two point charges based on BP algorithm

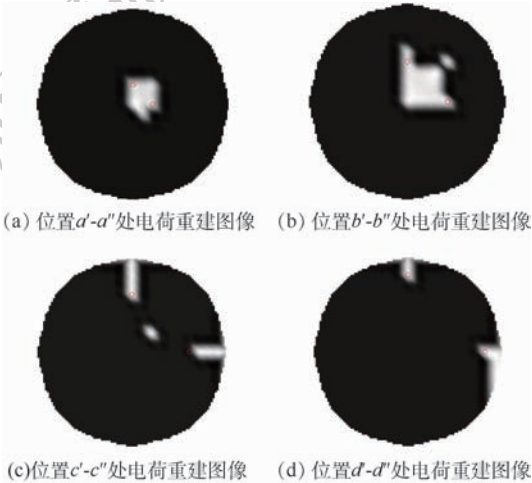


图 9 基于 Landweber 迭代算法的 2 个电荷图像重建结果
Fig. 9 Image reconstruction results for two point charges based on Landweber iteration algorithm

正确判断电荷数目。由图 9 (c)、(d) 可见, Landweber 迭代算法对相距较远的 2 个电荷可以由重

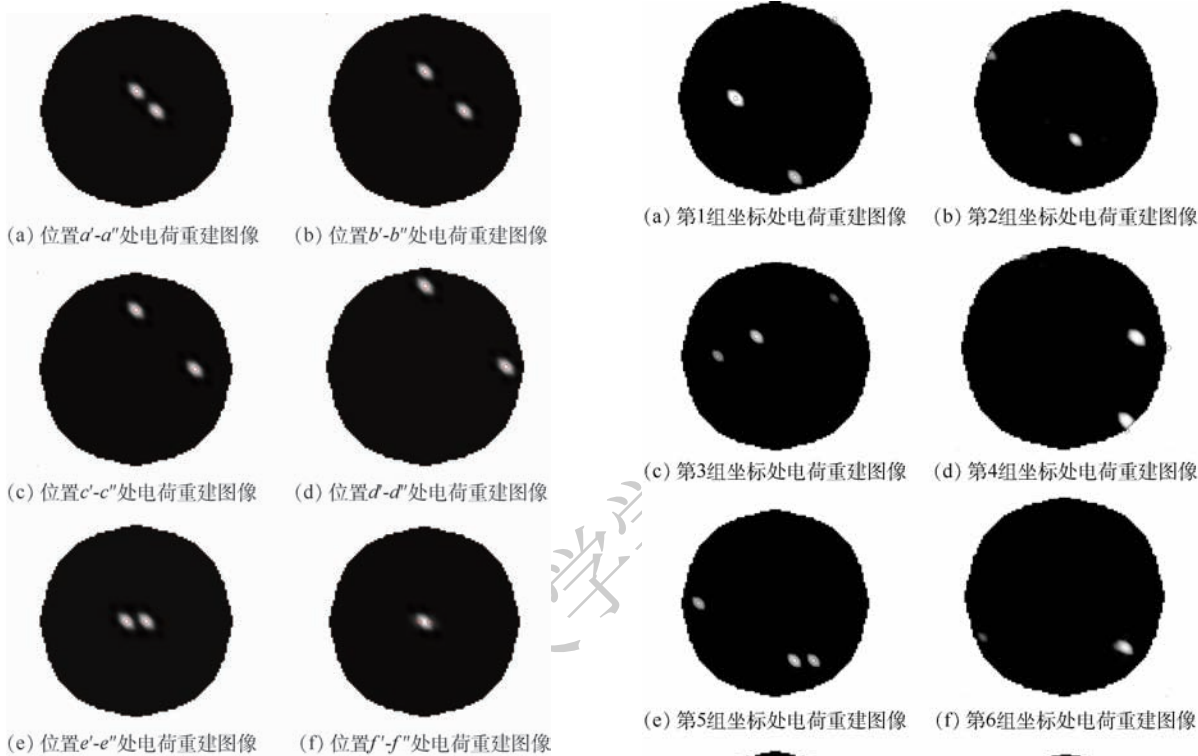


图 10 基于 PDIPA 的 2 个电荷图像重建结果

Fig. 10 Image reconstruction results for two point charges based on PDIPA

建图像区分,当电荷位于 $c'-c''$ 位置时,中心出现伪影。2 个电荷进一步靠近时,如图 9(a)、(b) 所示,由重建图像无法正确识别电荷数目。

由图 10 可见,对位于径向的不同位置、不同间距的 2 个荷电磨粒,采用 PDIPA 均可准确分辨。为进一步验证 PDIPA 的空间分辨能力,在间距 1 mm 的 $e'-e''$ 位置和间距 0.5 mm 的 $f'-f''$ 位置分别放置电荷进行实验,结果表明,对于场域内含 2 个电荷的情况,在 0.5 mm 边长的网格剖分下,PDIPA 可分辨的最小间距达到 1 mm。

在发动机异常磨损的情况下,可能出现更多荷电磨粒同时流经静电传感器的情况。随着荷电磨粒数增多,8 电极传感器由于测量信息有限,将难以保证重建电荷数目的准确率。下面通过测试 PDIPA 对 3 个随机分布的点电荷的重建效果,以进一步分析其可行性与局限性。用 MATLAB 产生 10 组随机整数坐标,坐标位置用圆圈标记在图 11 中的管道截面图像上,以方便对比重建结果。

由图 11 可见,对于图 11(a)、(c)、(e)、(g)、(h) 5 种电荷分布情况,PDIPA 对电荷数目与位置均实现了准确重建;对图 11(b)、(d)、(i) 3 种电荷分布模型,重建电荷位置出现偏差,但仍能正确显示电荷数目;而图 11(f)、(j) 的重建图像中,2 个相距 1 mm 的电荷图像融为一体,导致无法准

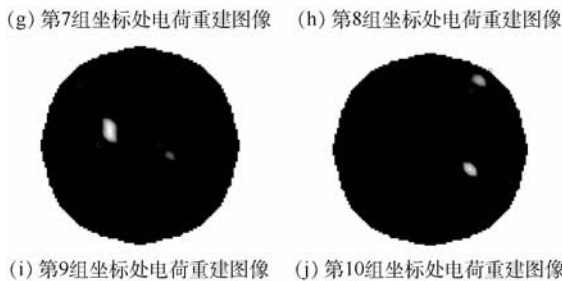


图 11 基于 PDIPA 的 3 个随机分布电荷图像重建结果

Fig. 11 Image reconstruction results for three randomly distributed point charges based on PDIPA

确判断电荷数目。可见,随着电荷数目增加,算法的空间分辨率较 2 个电荷分布的情况(见图 11)有所下降。要实现对 3 个及以上荷电磨粒的可靠监测,需要增加测量信息。

4 结 论

在基于静电感应的荷电磨粒在线监测中,荷电磨粒数目是一个重要指标和检测难点,EST 技术为实时、直观地监测荷电磨粒数提供了可行途径。EST 图像重建的难点在于其测量值数目等于电极数目,用于求解逆问题的独立测量值非常有限。仿真实验表明,基于压缩感知的图像重建算法相对于传统的 BP 算法和 EIT 中常用的迭代算

法可明显改善 EST 成像质量。

1) 基于 PDIPA 的图像重建可准确检测测量截面内不同位置处的单个和 2 个电荷,实时性满足应用要求。

2) 对于 3 个及以上电荷的检测,需要进一步增加测量值,以提高检测的可靠性。

3) 在 3 个电荷同时存在的情况下,对距离 1 mm 的两电荷仍实现了准确区分,说明图像重建的空间分辨率与电荷相对于测量电极的位置有关,在不增加电极数目的情况下,传感器结构仍存在优化空间,还需进一步实验研究。

参考文献 (References)

- [1] 陈志雄,左洪福,詹志娟,等. 滑油系统全流量在线磨粒静电监测技术研究[J]. 航空学报,2012,33(3):446-452.
CHEN Z X, ZUO H F, ZHAN Z J, et al. Study of oil system oil-line debris electrostatic monitoring technology[J]. Acta Aeronautica et Astronautica Sinica, 2012, 33(3):446-452 (in Chinese).
- [2] 李绍成. 基于静电感应和显微图像的油液磨粒监测技术研究[D]. 南京:南京航空航天大学,2009:4-8.
LI S C. Research on technologies of oil wear particle monitoring based on electrostatic induction and microscopical image[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2009:4-8 (in Chinese).
- [3] TANG X, CHEN Z S, LI Y, et al. Compressive sensing-based electrostatic sensor array signal processing and exhausted abnormal debris detecting[J]. Mechanical Systems and Signal Processing, 2018, 105:404-426.
- [4] GREEN R G, RAHMAT M F, EVANS K, et al. Concentration profiles of dry powders in a gravity conveyor using an electrodynamic tomography system[J]. Measurement Science and Technology, 1997, 8(2):192-197.
- [5] MACHIDA M, SCARLETT B. Process tomography system by electrostatic charge carried by particles[J]. IEEE Sensors Journal, 2005, 5(2):251-259.
- [6] RAHMAT M F, ISA M D, RAHM R A, et al. Electrodynamics sensor for the image reconstruction process in an electrical charge tomography system[J]. Sensors, 2009, 9(12):10291-10308.
- [7] THUKU I T, RAHMAT M F, WAHAB N A, et al. 2-D finite-element modeling of electrostatic sensor for tomography system[J]. Sensor Review, 2013, 33(2):104-113.
- [8] 高鹤明. 管内气固两相流的静电层析成像技术[D]. 南京:东南大学,2011:56-61.
GAO H M. Electrostatic tomography in gas/solid two-phase flow within pipe[D]. Nanjing: Southeast University, 2011:56-61 (in Chinese).
- [9] ZHOU B, ZHANG J Y, XU C L, et al. Image reconstruction in electrostatic tomography using a priori knowledge from ECT[J]. Nuclear Engineering and Design, 2011, 241(6):1952-1958.
- [10] THUKU I T, RAHMAT M F A. Finite-element method modeling in 4 and 16 sensors electric-charge tomography systems for particles moving in pipeline[J]. Flow Measurement and Instrumentation, 2014, 38:9-20.
- [11] QIAN X, YAN Y, WANG L, et al. An integrated multi-channel electrostatic sensing and digital imaging system for the on-line measurement of biomass coal particles in fuel injection pipelines[J]. Fuel, 2015, 151:2-10.
- [12] WANG S, LI J, XU C, et al. Local particle mean velocity measurement in pneumatic conveying pipelines using electrostatic sensor arrays[J]. Particulate Science and Technology, 2015, 33(1):81-90.
- [13] 曹宏庆. 某型发动机滑油颗粒监测试验研究[D]. 南京:南京航空航天大学,2012:11-12.
CAO H Q. Research on experiments of aero-engine lubricating oil debris monitoring[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012:11-12 (in Chinese).
- [14] TANG X, CHEN Z S, LI Y, et al. Analysis of the dynamic sensitivity of hemisphere-shaped electrostatic sensors' circular array for charged particle monitoring[J]. Sensors, 2016, 16(9):1403.
- [15] XUE Q, SUN B Y, CUI Z Q, et al. Online monitoring of oil film using electrical capacitance tomography and level set method[J]. Review of Scientific Instruments, 2015, 86(8):085106.
- [16] DING M L, YUE S H, LI J, et al. Second-order sensitivity coefficient based electrical tomography imaging[J]. Chemical Engineering Science, 2019, 199:40-49.
- [17] YU Y Y, HONG M J, LIU F, et al. Compressed-sensing MRI using singular value decomposition based sparsity basis[C]// Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ: IEEE Press, 2011:5734-5737.
- [18] ZHANG Q B, LI X N, CHENG H, et al. SVD-based compressive sensing[J]. AISS, 2013, 5(3):580-588.
- [19] 张立峰,刘昭麟,田沛. 基于压缩感知的电容层析成像图像重建算法[J]. 电子学报,2017,45(2):353-358.
ZHANG L F, LIU Z L, TIAN P. Image reconstruction algorithm for electrical capacitance tomography based on compressed sensing[J]. Acta Electronica Sinica, 2017, 45(2):353-358 (in Chinese).
- [20] WANG G Q, BAI Y Q. Primal-dual interior-point algorithm for convex quadratic semi-definite optimization[J]. Nonlinear Analysis: Theory, Methods & Applications, 2009, 71(7-8):3389-3402.
- [21] YANG A Y, ZHOU Z H, GANESH A, et al. Fast l_1 -minimization algorithms for robust face recognition[J]. IEEE Transactions on Image Processing, 2013, 22(8):3234-3246.
- [22] KIM S J, KOH K, LUSTIG M, et al. An interior-point method for large-scale l_1 -regularized least squares[J]. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(4):606-617.
- [23] LI Y, YANG W. Image reconstruction by nonlinear Landweber iteration for complicated distributions[J]. Measurement Science and Technology, 2008, 19(9):094014.

作者简介:

薛倩 女,博士,讲师。主要研究方向:电学层析成像。

EST image reconstruction based on primal dual interior point algorithm

XUE Qian^{1,*}, LIU Jing¹, MA Min¹, WANG Huaxiang²

(1. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;

2. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: The passive induction mechanism of electrostatic tomography (EST) determines that the number of independent measurements is equal to the number of electrodes, which is much less than the number of independent measurements of relatively mature electrical tomography (ET) technologies such as electrical capacitance tomography (ECT), resulting in a more severe underdetermined inverse problem. In order to address this problem, compressed sensing-based EST image reconstruction algorithm is studied. The sensitivity matrix is processed by singular value decomposition (SVD) to satisfy the restricted isometry property (RIP), and thereafter the l_1 norm regularization model and primal dual interior point algorithm (PDIPA) are utilized to reconstruct the image. Besides, constraint on the number of non-zero elements in the image vector is imposed in the iteration process according to the sparsity of debris distributed in oil. Simulation experiment demonstrates that compared to the “Circle of Apollonius” based back-projection (BP) algorithm and Landweber iteration algorithm, the aforementioned algorithm has obviously improved the imaging quality: accurate reconstruction can be obtained for single charge distributed at different positions; for two point charges whose distance is more than or equal to 1 mm, both the number and positions of the point charges can be correctly observed; for 10 groups of three randomly distributed point charge models, the accuracy rate of charged debris number monitoring is about 80%.

Keywords: electrostatic tomography (EST); oil monitoring; compressed sensing; image reconstruction; regularization

Received: 2019-01-16; Accepted: 2019-02-02; Published online: 2019-02-22 08:45

URL: kns.cnki.net/kcms/detail/11.2625.V.20190221.1541.005.html

Foundation items: National Natural Science Foundation of China (61401466); Startup Scientific Research Foundation of Civil Aviation University of China (2013QD01S)

* Corresponding author. E-mail: xueqian@tju.edu.cn

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0040

基于自适应遗传算法的 MEMS 加速度计快速标定方法



高爽*, 张若愚

(北京航空航天大学 仪器科学与光电工程学院, 北京 100083)

摘 要: 微惯性测量单元(MIMU)的标定技术是低精度惯性导航领域中的重要研究方向,传统标定方法操作复杂,标定精度严重依赖转台精度。为解决大批量 MIMU 快速标定的问题,提出了一种基于自适应遗传算法(GA)的微机电系统(MEMS)加速度计快速标定方法,将加速度计标定问题转化为参数优化问题。首先,利用模观测原理构造目标优化函数;然后,分析系统可观测度确定最优标定编排方案;最后,采用全局搜索的自适应遗传算法优化标定参数。实验结果表明:与牛顿迭代法相比,标定精度提升 1~3 个数量级,运算速度提高 61%。标定后解算的水平姿态角误差小于 0.1° ,可实现与传统标定方法相同量级的姿态精度,验证了所提方法的优越性和实用性。

关键词: 微机电系统(MEMS); 加速度计; 遗传算法(GA); 模观测; 标定; 可观测度分析

中图分类号: V241.62

文献标识码: A

文章编号: 1001-5965(2019)10-1982-08

作为惯性测量系统的核心元件,微惯性测量单元(MIMU)是微机电系统(MEMS)的一个重要分支,具有体积小、成本低、质量轻、功耗低和耐冲击性等优点,在普通民用和战略军用导航等领域具有广阔的应用前景^[1]。然而,在实际应用中,由于 MEMS 器件误差及系统误差受惯性测量单元的结构设计、外部环境等因素影响较大,MEMS 惯性器件的精度成为制约 MIMU 导航性能的主要技术瓶颈之一。因此,MIMU 的标定补偿成为低精度惯性导航领域的一大研究热点^[2]。

分立式标定方法操作复杂,步骤繁琐,标定精度主要由转台精度决定。由于 MIMU 标定精度要求低、大批量生产等特点,需寻求简单易行的快速标定方法。为克服传统标定方法的弊端,Lötters 等^[3]首次提出基于模观测的加速度计标定方法,将标定问题转化为优化问题。在模观测标定中,

研究关键在于非线性方程组的求解,其中,最常用的方法是迭代算法。Frosio 等^[4]采用牛顿迭代法求解非线性方程组,实现 MEMS 加速度计的外场自动标定,标定精度比传统标定方法提升一个数量级;Won 和 Golnaraghi^[5]应用非线性最小二乘法进行迭代计算,在获得待标定参数的前提下降低了运算时间;Fong 等^[6]将下山单纯形算法应用于最小化目标函数,为加速度计的免转台标定提供了理论基础。然而,迭代算法存在的问题是求解过程复杂,初始值选取的优劣直接影响其收敛性和迭代结果。针对迭代算法存在的不足,基于迭代思想的智能算法以其在优化求解中的适用性、鲁棒性等突出优势,为标定优化问题提供了一种新思路。戴邵武等^[7]采用粒子群优化(PSO)算法对加速度计进行快速标定,但未进行标定方法的现场试验验证;杨管金子等^[8]利用遗传算法

收稿日期: 2019-01-29; 录用日期: 2019-04-26; 网络出版时间: 2019-06-17 15:35

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190617.1117.005.html

* 通信作者。E-mail: gaoshuang@buaa.edu.cn

引用格式: 高爽, 张若愚. 基于自适应遗传算法的 MEMS 加速度计快速标定方法[J]. 北京航空航天大学学报, 2019, 45(10): 1982-1989. GAO S, ZHANG R Y. Rapid calibration method of MEMS accelerometer based on adaptive GA[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1982-1989 (in Chinese).

(GA)完成加速度计的免转台标定,但未实现对安装误差参数的标定。

针对以上问题,在不依赖转台的前提下,本文提出了一种基于自适应遗传算法的 MEMS 加速度计快速标定方法。实验结果表明,该方法能快速准确标定出全部误差参数,缩短标定时间,提高标定精度,且标定后能达到与传统标定方法相同量级的姿态精度。

1 加速度计标定模型

1.1 加速度计输出模型

在实际应用中,由于工作原理、制造工艺及使用环境等因素的影响,MIMU 测量的输出信息通常会含有误差,测量输出应根据输入加速度信息和误差项建立数学模型。MIMU 中加速度计的输出误差模型为

$$\begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} = \begin{bmatrix} S_{xx} & E_{xy} & E_{xz} \\ E_{yx} & S_{yy} & E_{yz} \\ E_{zx} & E_{zy} & S_{zz} \end{bmatrix} \times \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} + \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} \quad (1)$$

式中: $[N_x \ N_y \ N_z]^T$ 为加速度计的测量输出; $[f_x \ f_y \ f_z]^T$ 为加速度计的实际输入; $[B_x \ B_y \ B_z]^T$ 为加速度计的常值零偏; $S_{ii} (i = x, y, z)$ 为加速度计的标度因数; $E_{ij} = S_{ii} \cdot M_{ij} (i, j = x, y, z)$, M_{ij} 为加速度计的安装误差。

根据式(1)加速度计的输入输出关系,将加速度计的零偏、标度因数和安装误差进行补偿后,可得载体的实际加速度。补偿后的加速度计输出模型为

$$\begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} = \begin{bmatrix} S_{xx} & E_{xy} & E_{xz} \\ E_{yx} & S_{yy} & E_{yz} \\ E_{zx} & E_{zy} & S_{zz} \end{bmatrix}^{-1} \times \left(\begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} - \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} \right) \quad (2)$$

1.2 模观测标定原理

在静止状态下,加速度计的测量输出满足:

$$C_b^n f^b = -g^n \quad (3)$$

式中: C_b^n 为载体坐标系到导航坐标系的方向余弦阵; f^b 为载体坐标系下的加速度计输出; g^n 为导航坐标系下的当地重力加速度。

对式(3)两侧分别取模值:

$$|f^b| = |C_b^n f^b| = |-g^n| = g \quad (4)$$

在静止状态下,加速度计测量输出的模总是等于当地重力加速度的模,与加速度计的姿态无关^[9]。因此,以加速度计测量输出的模为观测量,可实现对加速度计误差参数的标定。

根据模观测原理,将补偿后的加速度计输出模型代入式(4)后取平方得

$$|f^b|^2 = (K_a^{-1}(N_a - B_a))^T \cdot (K_a^{-1}(N_a - B_a)) = |g|^2 \quad (5)$$

式中: K_a 、 N_a 、 B_a 分别与式(1)中元素相对应, K_a 为标度因数和安装误差矩阵, N_a 为测量输出矩阵, B_a 为零偏矩阵。

式(5)为单位置下加速度计的非线性输出方程。为求解标定参数,需对多位置下的测量输出进行观测,即联立非线性方程组。然而,由于器件固有的噪声和外界环境干扰等因素,加速度计实际输出的模与重力加速度之间存在误差,非线性方程组可改写为误差函数的形式:

$$E_k(\theta) = \sum_{k=1}^N ((f^b)_k^2 - g^2)^2 \quad (6)$$

式中: $\theta = [K_a \ B_a]$ 为待标定参数; N 为测试位置数。

通过构造目标函数,以标定参数为状态量,可将非线性方程组的求解问题转化为非线性函数的优化问题,通过最小化目标函数实现对误差参数的标定。

2 基于自适应遗传算法的标定参数求解

遗传算法是一种借鉴生物界自然选择和遗传机制发展起来的随机搜索优化算法,基于适者生存的原则,通过模拟自然进化在种群内的优胜劣汰实现对最优解的搜索^[10]。遗传算法以群体搜索和个体间的信息交换为准则,采用概率法随机寻优,以其并行性、鲁棒性及自适应性等突出优势被广泛应用于系统辨识、故障诊断、自动控制等领域中复杂优化问题的求解^[11-13]。

基于遗传算法的全局搜索能力,本文将遗传算法应用于 MEMS 加速度计的快速标定,建立目标函数与遗传算法适应度函数的关系,在全局范围内搜索标定参数的最优解。随机产生初始种群开始最优解迭代搜索,对种群中的个体进行适应度计算,以最大适应度为标准,更新最优适应度和最优染色体,通过选择、交叉和变异等遗传操作生成下一代种群,进行当前种群的个体适应度计算。如此循环迭代,直至满足优化终止条件,获得最优个体。算法流程如图 1 所示。

2.1 初始种群与编码

以标定参数作为种群中的个体,根据加速度计输出模型反推出最优解的取值范围,形成初始

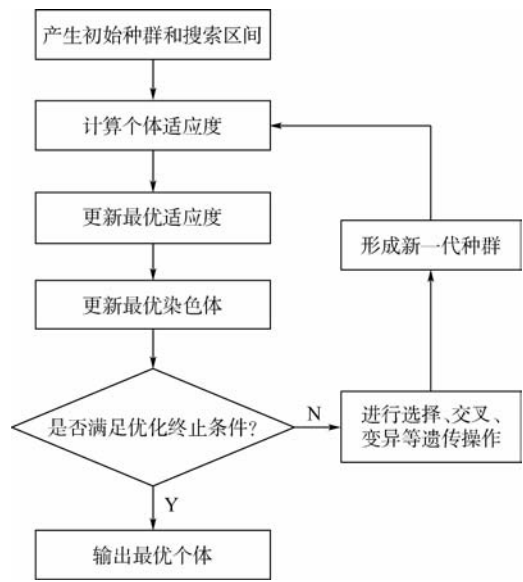


图 1 遗传算法流程

Fig. 1 Flowchart of genetic algorithm

搜索区间。为避免连续函数离散化的映射误差,减少编码解码的复杂运算,本文采用实数编码代替传统的二进制编码方法,在搜索区间内随机产生实数串数据生成初始种群,以该数据作为初始种群开始迭代^[14]。

2.2 适应度计算

适应度函数又称为评价函数,是根据目标函数确定的判别种群中个体优劣性的评价标准^[15]。式(6)是加速度计测量输出模值与当地重力加速度模值的误差函数,通过适当变换可得遗传算法的适应度函数为

$$F^{(ij)}(\theta) = \frac{1}{E^{(ij)}(\theta)} \quad (7)$$

式中: $E^{(ij)}(\theta)$ 为第*i*代种群中第*j*个个体对应的误差函数值。适应度函数 $F^{(ij)}(\theta)$ 为该误差函数的倒数。最大适应度对应的个体作为当前种群的最优个体被保存下来,进入下一代种群继续寻优。

2.3 选择、交叉与变异

选择、交叉与变异等遗传操作能够有效增加种群多样性,避免算法陷入局部最优,提升全局搜索能力。选择操作根据个体的适应度从上一代种群中择优选择个体复制到下一代种群,通过选择过程,种群中的最优个体不断更新,体现了优胜劣汰的进化原则。本文采用均匀排序的选择方法,对种群中的个体适应度大小进行排序,分配各个个体被选择的概率,使适应度高的个体具有更高的存活概率。

交叉操作是遗传算法中的核心运算,通过模拟遗传学的杂交原理,将种群中的2个个体随机搭配,以交叉概率交换部分染色体形成新的个体,

体现出信息交换的思想。变异操作将种群的每个个体以变异概率任意改变基因值,维持种群的多样性,抑制算法过早收敛。在传统遗传算法中,通常采用固定的交叉概率($P_c = 0.4 \sim 0.99$)和变异概率($P_m = 0.000\ 1 \sim 0.1$),在复杂优化问题中存在收敛速度慢及早熟现象等问题。本文采用自适应交叉和变异概率代替固定概率^[16-17],根据个体适应度自动调整 P_c 和 P_m 以实现最优参数的选择,保证算法较强的全局搜索能力和收敛性能。

$$P_c = \begin{cases} k_1 \frac{(f_{\max} - f')}{f_{\max} - f_{\text{avg}}} & f' \geq f_{\text{avg}} \\ k_3 & f' < f_{\text{avg}} \end{cases} \quad (8a)$$

$$P_m = \begin{cases} k_2 \frac{(f_{\text{avg}} - f)}{f_{\max} - f_{\text{avg}}} & f \geq f_{\text{avg}} \\ k_4 & f < f_{\text{avg}} \end{cases} \quad (8b)$$

式中: $f_{\max}$ 为当前种群的最大适应度值; f_{avg} 为当前种群的平均适应度值; f' 为交叉个体中较大的适应度值; f 为变异个体的适应度值;常值参数取为 $k_1 = k_3 = 1, k_2 = k_4 = 0.5$ 。

3 最优标定编排设计

由模观测原理可知,加速度计标定需通过观测多位置下的测量输出实现,测试位置的选择通常根据标定参数的数量和特点来确定。针对多位置标定方法,不同的路径编排导致加速度计的输入激励不同,合理的标定路径要求充分激励所有误差参数,且需要保证标定结果的唯一性。将标定参数作为状态变量,分析不同位置下加速度计测量输出对状态变量的可观测度,确定出最优的静态位置编排,保证标定路径的合理性^[18]。

由于加速度计在静态位置下的测量输出不随时间变化而改变,则离散化时不变系统模型为

$$\begin{cases} X(k+1) = AX(k) \\ Y(k) = HX(k) \end{cases} \quad (9)$$

式中: X 为状态变量; Y 为观测输出; A 和 H 分别为系统状态矩阵和观测矩阵。

线性时不变系统的可观测性判别矩阵为

$$Q = [H \quad AH \quad \cdots \quad AH^{n-1}]^T \quad (10)$$

若满足 $\text{rank}(Q) = n$,则系统完全可观测,否则系统存在不可观测量。

针对加速度计的静态误差模型,由于模观测标定方法无转台坐标系基准,安装误差矩阵只包含3项误差参数,本文以 M_{xy} 、 M_{xz} 、 M_{yz} 作为待标定的安装误差。选取9项标定参数为状态变量:

$$\mathbf{X}(k) = \begin{bmatrix} B_x & B_y & B_z & S_{xx} & S_{yy} & S_{zz} \\ M_{xy} & M_{xz} & M_{yz} \end{bmatrix}$$

(11)

系统的状态变量为不随时间变化的确定值,其状态矩阵为

$$\mathbf{A} = \text{diag}[1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

(12)

系统的观测输出能够反映不同位置下的输入激励,维数与标定方案的位置数目有关,其观测矩阵为

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{G}_{12} \\ \vdots & \vdots \\ \mathbf{I}_{N1} & \mathbf{G}_{N2} \end{bmatrix}$$

(13)

式中: $\mathbf{I}_{11}, \mathbf{I}_{21}, \dots, \mathbf{I}_{N1}$ 为 3 维单位阵; $\mathbf{G}_{12}, \mathbf{G}_{22}, \dots, \mathbf{G}_{N2}$ 为对应位置下重力加速度的输入激励矩阵。

根据奇异值分解理论,对系统可观测性判别矩阵 $\mathbf{Q}$ 进行奇异值分解:

$$\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

(14)

式中: $\mathbf{U}$ 和 $\mathbf{V}$ 分别为由 m 维列向量 $\mathbf{u}_i$ 和 n 维列向量 $\mathbf{v}_j$ 构成的正交阵, $\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m]^T$,

$\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n]^T$; $\mathbf{S} = \begin{bmatrix} \mathbf{A}_{r \times r} & 0 \\ 0 & 0 \end{bmatrix}$ 为 $\mathbf{Q}$ 的奇异值构成的对角阵, r 为 $\mathbf{Q}$ 的秩。

为了分析矩阵 $\mathbf{Q}$ 奇异值和状态变量之间的关系,设正定对称矩阵 $\mathbf{M} = \mathbf{Q}^T \mathbf{Q}$,其特征值和特征向量分别为 λ_i 和 $\mathbf{w}_i$,则特征值和特征向量之间的关系为

$$\mathbf{Q}^T \mathbf{Q} \mathbf{w}_i = \lambda_i \mathbf{w}_i$$

(15)

由式(15)可知,矩阵 $\mathbf{M}$ 的特征值越大,其特征向量对应状态变量的可观测度越高,通过多位置下的系统可观测度分析,可得最大奇异值对应的位置对状态变量具有最高的可观测度,由此确定出最优位置编排。根据标定参数的数量和特点设计静态位置编排方案,如图 2 所示,多位置下可观测性判别矩阵的秩和奇异值如表 1 所示。

由表 1 可知,当 $N \geq 4$ 时,矩阵满秩,系统完全可观测;当 $N \geq 7$ 时,各项奇异值明显增大,这是由于与加速度计敏感轴成 45° 的位置使得系统

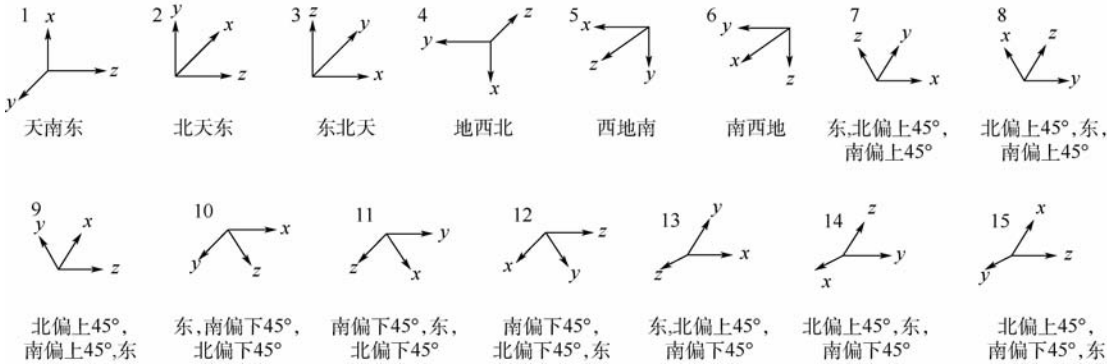


图 2 静态多位置标定编排方案

Fig. 2 Static multi-position calibration scheme

表 1 多位置下可观测性判别矩阵的秩和奇异值

Table 1 Rank and singular value of observable matrix under multi-position

位置	秩	奇异值								
3	8	29.861	29.713	29.562	29.405	29.405	29.405	4.220	2.969	1.715×10^{-7}
4	9	41.585	41.585	41.585	29.716	29.562	29.405	6	5.168	4.198
5	9	41.585	41.585	41.585	41.585	41.585	29.564	6.708	6.708	5.968
6	9	41.805	41.585	41.585	41.585	41.585	41.585	7.3484	7.3484	5.9683
7	9	51.106	46.513	41.693	41.585	41.585	41.585	7.9372	7.8196	6.0063
8	9	55.280	46.533	46.513	46.493	46.493	41.657	8.3749	8.2633	6.0331
9	9	59.063	55.107	50.992	46.596	46.493	46.493	8.6499	8.3946	6.4519
10	9	64.408	57.427	50.991	48.495	48.261	46.572	9.1542	9.1223	8.2048
11	9	68.476	58.829	55.023	50.931	50.931	47.303	9.8828	9.8328	9.4636
12	9	72.041	65.751	58.810	51.012	50.931	50.931	10.392	10.392	9.8817
13	9	72.041	67.576	58.813	58.810	52.764	50.951	10.816	10.765	10.605
14	9	72.041	68.973	62.385	62.377	55.011	55.011	11.178	11.149	11.132
15	9	72.041	68.973	65.779	62.394	62.394	62.377	11.458	11.456	11.439

可观测度增强;当 $N \geq 12$ 时,最大奇异值不再增大,且其他各项奇异值的增幅较小。综合分析标定精度和标定复杂度等因素,本文选取12位置的静态多位置标定方案。

4 实验结果分析

本文以Xsens公司惯性导航系统MTi-1系列为测试对象,其结构如图3所示。根据第3节所述标定路径,在大理石平台上对MTi-1进行位置翻转实验,每个位置静止1 min,采集传感器数据,再翻转至下一位置。

完成数据采集后,采用本文方法对加速度计的标定参数进行最优化求解,遗传算法的适应度函数变化曲线及加速度计误差参数标定结果如图4和图5所示。

由图4可知,最优适应度能够准确跟踪平均适应度的变化。进化初期保持较小的适应度,以提高种群的全局搜索能力;进化后期适应度增大,快速收敛至全局最优解。由图5可知,通过自适应遗传算法的迭代搜索,标定参数全部实现收敛。由此可见,自适应遗传算法能够较好地应用于MEMS加速度计的标定问题。

为了验证本文方法的优越性,以传统多位置法的标定结果作为标定真值,将本文方法与牛顿

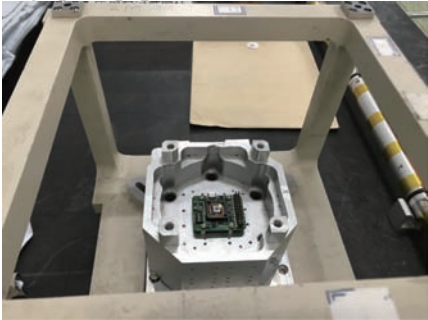


图3 MTi-1系列惯性测量组合

Fig.3 MTi-1 series inertial measurement unit

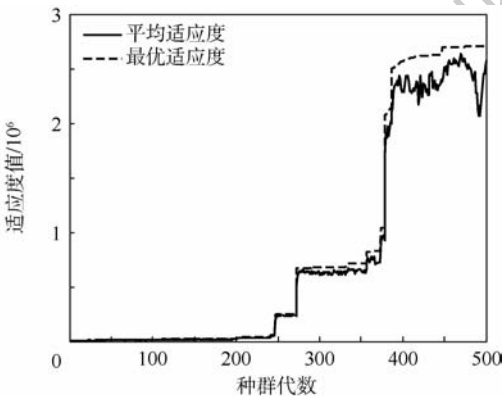
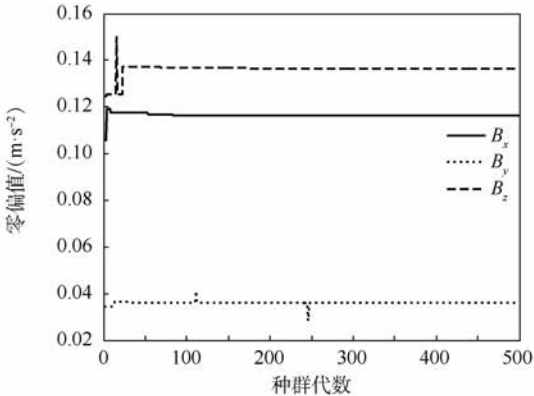
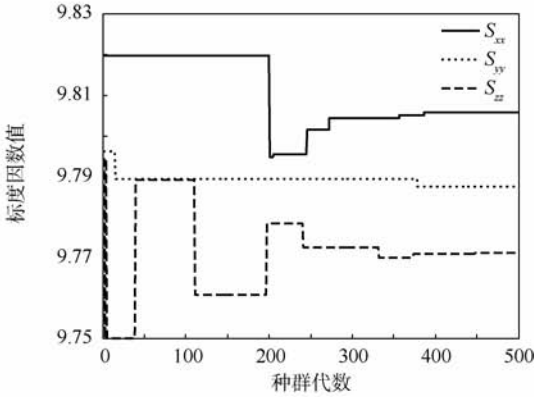


图4 适应度函数变化曲线

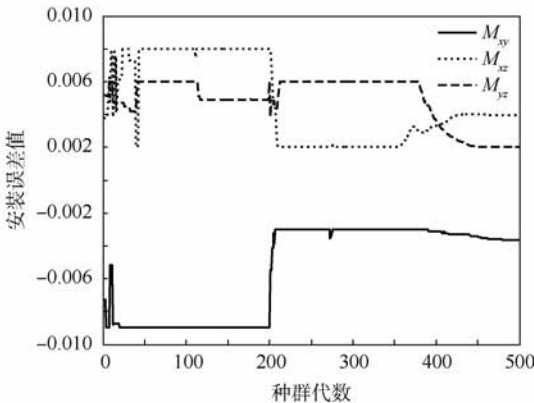
Fig.4 Curves of fitness function variation



(a) 零偏标定结果



(b) 标度因数标定结果



(c) 安装误差标定结果

图5 误差参数标定结果

Fig.5 Calibration results of error parameters

迭代法进行对比,2种方法的标定结果如表2和表3所示。

由表2可知,在牛顿迭代法的标定结果中,三轴零偏的相对误差分别为-0.146%、-0.554%、1.451%,三轴标度因数相对误差分别为0.000204%、0.0104%、0.414%,未能有效标定安装误差;在本文方法的标定结果中,三轴零偏的相对误差分别为-0.017%、0.019%、-0.051%,三轴标度因数相对误差分别为0.007%、0.004%、0.0003%,三轴安装误差的相对误差分别为10.69%、53.64%、21.37%,标定精度比牛顿迭代法提升1~3个数量级。由表3可知,本文方

表 2 本文方法与牛顿迭代法的标定结果对比

误差参数		牛顿迭代法		本文方法	
		标定结果	相对误差/%	标定结果	相对误差/%
$B_x/(m \cdot s^{-2})$	0.116 46	0.116 29	-0.146	0.116 44	-0.017
$B_y/(m \cdot s^{-2})$	0.036 1	0.035 9	-0.554	0.036 107	0.019
$B_z/(m \cdot s^{-2})$	0.136 44	0.138 42	1.451	0.136 37	-0.051
S_{xx}	9.805 18	9.805 2	0.000 204	9.805 9	0.007
S_{yy}	9.787 18	9.788 2	0.010 4	9.787 6	0.004
S_{zz}	9.771 07	9.811 5	0.414	9.771 1	0.000 3
M_{xy}	-0.005 8	0.276 8	×	-0.006 42	10.69
M_{xz}	0.005 22	0.405 1	×	0.002 42	-53.64
M_{yx}	0.006 07	—	—	—	—
M_{yz}	0.003 93	-0.768 9	×	0.003 09	-21.37
M_{zx}	-0.003 6	—	—	—	—
M_{zy}	-0.006 2	—	—	—	—

注：“×”表示无相对误差结果；“—”表示无此项安装误差。

表 3 本文方法与牛顿迭代法的运算时间对比

Table 3 Comparison of operational time between proposed method and Newton's iteration	
方法	运算时间/s
本文方法	164.28
牛顿迭代法	421.92

法的运算速度比牛顿迭代法提升 61%。

在静态条件下,将传统标定方法、牛顿迭代法与本文方法分别应用于 MTi-1 惯性导航系统,开展纯惯性导航验证实验,不同方法下的水平姿态角误差曲线及误差均值如图 6 和表 4 所示。

由图 6 和表 4 可知,本文方法标定后的俯仰

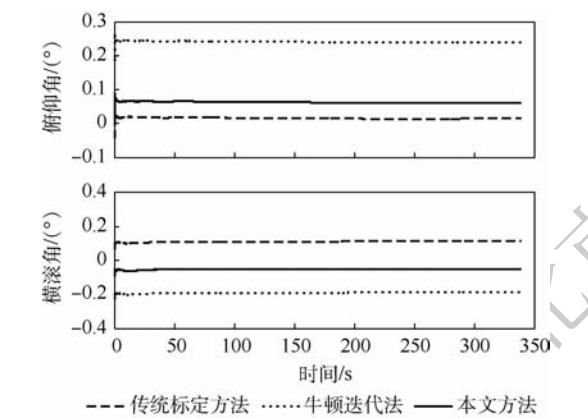


图 6 水平姿态角误差曲线

Fig. 6 Curves of horizontal attitude errors

表 4 水平姿态角误差均值

Table 4 Mean of horizontal attitude errors		
方法	俯仰角误差/(°)	横滚角误差/(°)
传统标定方法	0.015	0.109
牛顿迭代法	0.24	-0.191
本文方法	0.062	-0.051

角误差约为 0.062°,相较于牛顿迭代法减小 74%,与传统标定方法精度相当;本文方法标定后的横滚角误差约为 -0.051°,精度优于牛顿迭代法和传统标定方法。

结合不同方法的标定结果与导航实验结果,可以得出本文方法的优势如下:

- 1) 能够准确标定出 MEMS 加速度计的全部误差参数,与牛顿迭代法相比,标定精度提升 1 ~ 3 个数量级,运算速度提升 61%,本文方法可有效应用于 MEMS 加速度计的快速标定。
- 2) 标定后解算的水平姿态角误差小于 0.1°,能达到与传统标定方法相同量级的精度,验证了本文方法在实际导航中的应用价值。

5 结 论

- 1) 本文针对传统标定方法标定时间长、标定精度依赖转台精度等问题,提出了一种基于自适应遗传算法的 MEMS 加速度计快速标定方法,实现对全部误差参数的快速准确标定。根据模观测原理,将标定问题转化为非线性方程组的优化求解问题;以系统可观测性分析为依据,设计最优标定路径。采用自适应交叉和变异概率,提升遗传算法的全局搜索和收敛性能。
- 2) 实际测试结果表明,与牛顿迭代法相比,本文方法具有标定精度高、标定速度快等优点,标定后能达到与传统标定方法相同量级的姿态精度,验证了本文方法的优越性和有效性。
- 3) 本文方法缩短标定时间,降低标定成本,具有重要的理论研究和工程应用价值。同时,由于遗传算法的参数选取直接影响标定精度及运算

速度,如何通过适应度函数的动态变化及对选择、交叉、变异算子的改进以提升遗传算法的寻优精度和收敛速度是未来工作中值得进一步研究的问题。

参考文献 (References)

- [1] 尹文. MIMU 微惯性测量单元误差建模与补偿技术[D]. 长沙:国防科学技术大学,2007:1-2.
YIN W. Error modeling and compensation of micro inertial measurement unit[D]. Changsha: National University of Defense Technology,2007:1-2(in Chinese).
- [2] 王如胜. MEMS 陀螺捷联惯导系统标定方法研究[D]. 哈尔滨:哈尔滨工业大学,2015:1-6.
WANG R S. Research on the calibration method of MEMS gyroscope strapdown inertial navigation system[D]. Harbin: Harbin Institute of Technology,2015:1-6(in Chinese).
- [3] LÖTTERS J C, SCHIPPER J, VELTINK P H, et al. Procedure for in-use calibration of triaxial accelerometers in medical applications[J]. Journal of Applied Behavior Analysis, 1998, 68(1-3): 221-228.
- [4] FROSIO I, PEDERSINI F, BORGHESE N A. Autocalibration of MEMS accelerometers[J]. IEEE Transactions on Instrumentation and Measurement, 2009, 58(6): 2034-2041.
- [5] WON S P, GOLNARAGHI F. A triaxial accelerometer calibration method using a mathematical model[J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(8): 2144-2153.
- [6] FONG W T, ONG S K, NEE A Y C. Methods for in-field user calibration of an inertial measurement unit without external equipment[J]. Measurement Science & Technology, 2008, 19(8): 817-822.
- [7] 戴邵武, 王克红, 戴洪德. 基于 PSO 算法的加速度计快速标定方法[J]. 光电与控制, 2014, 21(12): 57-60.
DAI S W, WANG K H, DAI H D. A rapid calibration method for accelerometer based on PSO algorithm[J]. Electronics Optics & Control, 2014, 21(12): 57-60(in Chinese).
- [8] 杨管金子, 李建辰, 黄海. 基于遗传算法的加速度计免转台标定方法[J]. 中国惯性技术学报, 2017, 25(1): 119-123.
YANG G J Z, LI J C, HUANG H. Non-turntable calibration method for three-accelerometer based on genetic algorithm[J]. Journal of Chinese Inertial Technology, 2017, 25(1): 119-123(in Chinese).
- [9] 张红良. 陆用高精度激光陀螺捷联惯导系统误差参数估计方法研究[D]. 长沙:国防科学技术大学,2010:39-40.
ZHANG H L. Research on error parameter estimation of land high-precision ring laser gyroscope strapdown inertial navigation system[D]. Changsha: National University of Defense Technology, 2010:39-40(in Chinese).
- [10] 葛继科, 邱玉辉, 吴春明, 等. 遗传算法研究综述[J]. 计算机应用研究, 2008, 25(10): 2911-2916.
GE J K, QIU Y H, WU C M, et al. Summary of genetic algorithms research[J]. Application Research of Computers, 2008, 25(10): 2911-2916(in Chinese).
- [11] DU Y, FANG J, MIAO C. Frequency-domain system identification of an unmanned helicopter based on an adaptive genetic algorithm[J]. IEEE Transactions on Industrial Electronics, 2013, 61(2): 870-881.
- [12] MATIAS T, SECO R, BELCHIOR C. Automatic extraction of the fuzzy control system by a hierarchical genetic algorithm[J]. Engineering Applications of Artificial Intelligence, 2014, 29(3): 70-78.
- [13] SU L M, HOU C Z, DAI Z J, et al. Application of improved genetic algorithm in network fault diagnosis expert system[J]. Journal of Beijing Institute of Technology, 2003, 12(3): 225-229.
- [14] 管小艳. 实数编码下遗传算法的改进及其应用[D]. 重庆:重庆大学,2012:4-5.
GUAN X Y. Improvement and application of genetic algorithm under real number coding[D]. Chongqing: Chongqing University, 2012:4-5(in Chinese).
- [15] 张思才, 张方晓. 一种遗传算法适应度函数的改进方法[J]. 计算机应用与软件, 2006, 23(2): 108-110.
ZHANG S C, ZHANG F X. A modified method to fitness function of genetic algorithms[J]. Computer Applications and Software, 2006, 23(2): 108-110(in Chinese).
- [16] SRINIVAS M, PATNAIK L M. Adaptive probabilities of crossover and mutation in genetic algorithms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1994, 24(4): 656-667.
- [17] 卢兆兴, 张金生, 王仕成, 等. 基于自适应遗传算法的三轴磁强计误差校正方法[J]. 传感技术学报, 2014, 27(3): 331-336.
LU Z X, ZHANG J S, WANG S C, et al. Calibration for tri-axial magnetometer based on adaptive genetic algorithm[J]. Chinese Journal of Sensors and Actuators, 2014, 27(3): 331-336(in Chinese).
- [18] 丁智坚. 三轴惯性/地磁组合测量系统标定方法研究[D]. 长沙:国防科学技术大学,2012:36-40.
DING Z J. Research on the calibration method of inertial/geomagnetic integrated measurement system[D]. Changsha: National University of Defense Technology, 2012:36-40(in Chinese).

作者简介:

高爽 女,博士,讲师。主要研究方向:惯性导航。

张若愚 女,硕士研究生。主要研究方向:惯性导航。

Rapid calibration method of MEMS accelerometer based on adaptive GA

GAO Shuang*, ZHANG Ruoyu

(School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100083, China)

Abstract: MEMS inertial measurement unit (MIMU) calibration is one of an important research direction in low-precision inertial navigation. Traditional calibration method has complex operating procedures and depends most on turntable accuracy. In order to overcome the problem of MIMU calibration in batch production, this paper presents a rapid micro-electro-mechanical system (MEMS) accelerometer calibration method based on adaptive genetic algorithm (GA), which converts calibration task to parameter optimization. Firstly, the principle of norm observation is adopted to establish the objective optimization function. Secondly, the optimal calibration scheme is designed on the basis of system observability analysis. Finally, calibration parameters can be optimized through adaptive GA with global search capability. Experimental results demonstrate that, compared with Newton's iteration, the proposed method can improve calibration accuracy by 1 – 3 orders of magnitude and increase operational speed by 61%. After the proposed calibration, the horizontal attitude error is less than 0.1° and its accuracy can reach the same order of magnitude as that in traditional method, which verifies its superiority and practicability.

Keywords: micro-electro-mechanical system (MEMS); accelerometer; genetic algorithm (GA); norm observation; calibration; observability analysis

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0157

含函数型自变量回归模型中的变量选择



刘科生¹, 王思洋^{2,*}

1. 北京航空航天大学 学生大数据中心, 北京 100083; 2. 中央财经大学 统计与数学学院, 北京 100081)

摘 要: 针对含有函数型和多元向量数据的回归模型中变量选择和参数估计问题进行研究, 扩展了函数型数据分析和变量选择方法的应用范围。首先, 函数型自变量基于函数型主成分基函数空间进行投影; 然后, 对投影后的函数型自变量(按组)及多元向量自变量采用惩罚变量选择方法, 同时估计相应的系数。惩罚项调节参数采用自适应调节参数, 损失函数采用中位绝对损失函数, 以此为例, 通过引入松弛变量将估计算法转化为求解线性规划问题, 算法复杂度低。数值模拟结果表明, 所提方法对于含函数型自变量回归模型的变量选择和参数估计均具有良好效果。

关 键 词: 函数型数据; 变量选择; 参数估计; 分位数; 函数型主成分

中图分类号: O212

文献标识码: A

文章编号: 1001-5965(2019)10-1990-05

在目前的大数据时代, 数据采集的途径越来越多样化, 数据量越来越大, 采集到的数据类型也日益丰富。在对这些数据进行分析的过程中, 不可避免地会遇到混合类型的数据, 无法直接使用已有方法进行分析处理。在已有方法的基础上, 构建新方法对混合类型的数据进行统计分析具有理论和现实意义。例如, 图像数据、音频数据和矩阵数据混合的数据分析问题, 在图像处理、语音识别、推荐系统构建等领域中并不少见, 且已引起广泛的关注。

事实上, 音频数据和图像数据分别可以采用一元和二元函数型数据进行表示, 矩阵数据可以采用多元向量加以描述。因此, 对函数型数据和多元向量数据混合的模型及其估计方法进行研究并加以推广, 可解决一系列实际问题。

对于函数型数据的研究成果众多^[1], 集中于函数型线性回归模型的参数估计和假设检验^[2]、函数型数据的聚类分类等诸多方面^[3-4]。多元统

计分析的成果更是丰富, 近年来围绕着高维情形下的多元统计分析, 如变量选择、假设检验等也有一系列理论与实际结果。但对于函数型数据和多元向量数据混合的情形, 研究成果相对较少。因为在处理该类混合数据时, 需要考虑不同类型数据之间的相关性度量及对模型估计结果的影响, 情况比较复杂。此外, 由于函数型系数的存在, 理论上研究估计量的渐近性质也具有难度。因此, 通常在尽可能不过多损失信息的前提下, 对混合数据进行转化, 基于转化后的数据改进已有方法进行处理^[5]。

基于此, 本文对含有函数型和多元向量自变量的回归模型中变量选择和参数估计问题进行探讨。首先, 对函数型自变量利用函数型主成分基函数进行投影, 对模型加以转化。然后, 采用 L_1 损失函数并考虑组变量选择方法, 同时进行变量选择和参数估计, 其中调节参数的选择采用了自适应算法, 目标函数的最优化借助于线性规划相

收稿日期: 2019-04-11; 录用日期: 2019-04-26; 网络出版时间: 2019-06-19 11:19

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190618.1535.001.html

基金项目: 国家自然科学基金(11501586, 71420107025); 中央财经大学科研创新团队支持计划

* 通信作者: E-mail: siyangw@163.com

引用格式: 刘科生, 王思洋. 含函数型自变量回归模型中的变量选择[J]. 北京航空航天大学学报, 2019, 45(10): 1990-1994.

LIU K S, WANG S Y. Variable selection in regression models including functional data predictors[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1990-1994 (in Chinese).

关算法。最后,通过数值模拟验证了本文方法在变量选择和参数估计上的有效性。

1 函数型和多元向量混合回归模型

本节引入函数型和多元向量混合回归模型,并给出对模型进行变量选择和参数估计的方法。

假设 $X_1(t), X_2(t), \dots, X_p(t)$ 为 p 个函数型自变量,满足 $E(X_j(t)) = 0, E(X_j^2(t)) < \infty, n$ 个样本的取值分别为 $X_{ij}(t), 1 \leq i \leq n, 1 \leq j \leq p$ 。

考虑如下模型:

$$Y_i = \sum_{j=1}^p \int_F X_{ij}(t) \beta_j(t) dt + \mathbf{Z}_i^T \boldsymbol{\gamma} + \epsilon_i \quad (1)$$

式中: F 为积分区域; Y_i 为响应变量; $X_{ij}(1 \leq j \leq p)$ 为函数型自变量; $\mathbf{Z}_i \in \mathbf{R}^q$ 为多元向量自变量; ϵ_i 为随机误差项,满足 $E(\epsilon_i) = 0, E(\epsilon_i^2) < \infty; \beta_j(t) (1 \leq j \leq p)$ 和 $\boldsymbol{\gamma} \in \mathbf{R}^q$ 为待估参数。

在模型中,如果 $\beta_j = 0 (1 \leq j \leq p)$, 模型即为常见的多元线性回归模型;如果 $\boldsymbol{\gamma} = \mathbf{0}$, 模型退化为通常的多元函数型线性模型;进一步,如果 $p = 1, \boldsymbol{\gamma} = \mathbf{0}$, 模型则成为函数型线性模型。因此,该模型具有较强的泛化能力。

2 模型变量选择和参数估计

本节先对函数型自变量在主成分基函数所张成的函数空间进行投影,再采用 L_1 损失函数和组 LASSO (Least Absolute Shrinkage and Selection Operator) [6] 惩罚方法进行变量选择。

2.1 函数型主成分及模型转化

假设任意函数型自变量 $X_j(t) (1 \leq j \leq p)$, 定义 $X_j(t)$ 的协方差函数为 $K_j(s, t) = \text{cov}(X_j(s), X_j(t))$, 并进行如下谱分解 [7]:

$$K_j(s, t) = \sum_{k=1}^{\infty} \kappa_{jk} \phi_{jk}(s) \phi_{jk}(t)$$

式中: $\kappa_{j1} \geq \kappa_{j2} \geq \dots \geq 0$ 为算子 K_j 的各个特征值; ϕ_{jk} 为特征值对应的特征函数。 $\{\phi_{jk}\}$ 构成 $L_2(F)$ 空间的一组规范正交基,从而有

$$\beta_j(t) = \sum_{k=1}^{\infty} b_{jk} \phi_{jk}(t) \quad 1 \leq j \leq p$$

$$X_{ij}(t) = \sum_{k=1}^{\infty} \xi_{ijk} \phi_{jk}(t) \quad 1 \leq j \leq p$$

式中: b_{jk} 和 ξ_{ijk} 为系数。

进一步,模型(1)可转化为

$$Y_i = \sum_{j=1}^p \sum_{k=1}^{\infty} \xi_{ijk} b_{jk} + \mathbf{Z}_i^T \boldsymbol{\gamma} + \epsilon_i$$

以上是关于理论的探讨,当面对样本时,需要

对 $K_j(s, t)$ 进行估计,可使用样本协方差函数进行估计。类似地,可定义函数型自变量的样本协方差函数为

$$\hat{K}_j(s, t) = \frac{1}{n} \sum_{i=1}^n [X_{ij}(s) - \bar{X}_j(s)] \cdot [X_{ij}(t) - \bar{X}_j(t)]$$

$$\text{式中: } \bar{X}_j(s) = \sum_{i=1}^n X_{ij}(s)/n; \bar{X}_j(t) = \sum_{i=1}^n X_{ij}(t)/n。$$

对 $\hat{K}_j(s, t)$ 进行类似谱分解可得到, $\hat{K}_j(s, t) = \sum_{k=1}^{\infty} \hat{\theta}_{jk} \hat{\phi}_{jk}(s) \hat{\phi}_{jk}(t), \forall (s, t) \in F^2$ 。 $\hat{\theta}_{jk}$ 为函数型谱分解中的特征根。注意, $\{X_{ij}(t), 1 \leq i \leq n\}$ 张成的空间中任意一组基函数的维数不超过样本量 n , 所以若 $\{\hat{\theta}_{jk}\}$ 以降序排列,则有

$$\hat{K}_j(s, t) = \sum_{k=1}^n \hat{\theta}_{jk} \hat{\phi}_{jk}(s) \hat{\phi}_{jk}(t) \quad \forall (s, t) \in F^2$$

在实际数据分析中,对样本协方差估计时通常会进行截断处理,最常用的做法是依据方差占比进行基函数个数选择。这里,函数型主成分基函数的个数可以通过累计方差占比 CPV 进行选择 [8-10],如设定选取主成分基函数的个数后能保留 $\text{CPV} = 85\%$ 的方差信息,根据

$$L_j = \min \left\{ l: \sum_{i=1}^l \hat{\theta}_{jk} / \sum_{i=1}^n \hat{\theta}_{jk} \geq 85\% \right\}$$

来对 $X_j(t)$ 选择 L_j 个函数型主成分基函数。

模型(1)可转化为

$$Y_i = \sum_{j=1}^p \hat{\boldsymbol{\xi}}_{ij}^T \mathbf{b}_j + \mathbf{Z}_i^T \boldsymbol{\gamma} + \tilde{\epsilon}_i \quad (2)$$

式中: $\hat{\boldsymbol{\xi}}_{ij} = [\hat{\xi}_{ij1}, \hat{\xi}_{ij2}, \dots, \hat{\xi}_{ijL_j}]$; $\mathbf{b}_j = [b_{j1}, b_{j2}, \dots, b_{jL_j}]^T$ 。

2.2 参数估计

本节基于模型(2)进行变量选择及参数估计的讨论。考虑到将每个函数型自变量展开为几个函数型主成分方向上的投影,若对原始的函数型自变量进行选择,自然会考虑组变量选择方法 [11-13]。组变量选择方法不对单个变量的系数添加惩罚,而是对变量组的系数向量整体添加惩罚,从而达到变量选择的效果。构造如下目标函数:

$$Q(\mathbf{b}, \boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n \rho \left(Y_i - \sum_{j=1}^p \hat{\boldsymbol{\xi}}_{ij}^T \mathbf{b}_j - \mathbf{Z}_i^T \boldsymbol{\gamma} \right) + \sum_{j=1}^p P_{\lambda_{1j}}(\|\mathbf{b}_j\|_{\infty}) + \sum_{j=1}^q \lambda_{2j} |\gamma_j| \quad (3)$$

式中: $\rho(\cdot)$ 为损失函数; $P_{\lambda}(t) = \lambda t$ 为惩罚函数, $\lambda_{1j} (1 \leq j \leq p), \lambda_{2j} (1 \leq j \leq q)$ 为惩罚项的调节参

数; $\|\cdot\|_\infty$ 为针对 $\mathbf{b}_j$ 的范数定义, $\|\mathbf{b}_j\|_\infty = \max\{|b_{j1}|, |b_{j2}|, \dots, |b_{jp}| \}$ 。通过最小化目标函数 $Q(\mathbf{b}, \boldsymbol{\gamma})$ 可得到对应系数的估计量 $(\hat{\mathbf{b}}, \hat{\boldsymbol{\gamma}})$ 。

损失函数 $\rho(\cdot)$ 可以选择为任意常见损失函数或根据需要构造, 如最小二乘损失、分位数损失函数等, 或负的对数似然函数等。综合考虑效率及稳健性质, 选择分位数损失函数^[14], 并以 50% 分位数为例, 则 $\rho(t) = |t|$ 。

2.3 调节参数选择及目标函数优化

注意到, $\|\mathbf{b}_j\|_\infty < C$ 等价于 $|b_{jk}| < C, 1 \leq k \leq L_j$ 。因此, 最小化目标函数式(3)中的 $Q(\mathbf{b}, \boldsymbol{\gamma})$, 可通过引入松弛变量转化为线性规划问题^[15]。

通过如下定义引入松弛变量 $(u_i, v_i)_{i=1}^n$:

$$Y_i - \sum_{j=1}^p \hat{\xi}_{ij} \mathbf{b}_j - \mathbf{Z}_i^T \boldsymbol{\gamma} = u_i - v_i$$

同样, 待估参数向量 $(\mathbf{b}, \boldsymbol{\gamma})$ 的各个元素都可以表示成正部、负部相减的形式, 即 $b_{jk} = b_{jk}^+ - b_{jk}^-$, $1 \leq k \leq L_j, 1 \leq j \leq p, \gamma_l = \gamma_l^+ - \gamma_l^-, 1 \leq l \leq q$ 。最小化目标函数式(3)中的 $Q(\mathbf{b}, \boldsymbol{\gamma})$, 即转化为在如下约束条件下:

$$M_{1j} \geq b_{jk}^+ + b_{jk}^-, b_{jk}^+ \geq 0, b_{jk}^- \geq 0, 1 \leq j \leq p$$

$$M_{2l} \geq \gamma_l^+ + \gamma_l^-, \gamma_l^+ \geq 0, \gamma_l^- \geq 0, 1 \leq l \leq q$$

$$Y_i - \sum_{j=1}^p \hat{\xi}_{ij} \mathbf{b}_j - \mathbf{Z}_i^T \boldsymbol{\gamma} = u_i - v_i, u_i \geq 0, v_i \geq 0,$$

$$1 \leq i \leq n$$

最小化如下目标函数:

$$\sum_{i=1}^n (u_i + v_i)/2 + \sum_{j=1}^p \lambda_{1j} M_{1j} + \sum_{l=1}^q \lambda_{2l} M_{2l}$$

该优化问题转化为线性优化问题, 简化了目标函数的优化过程。充分利用了损失函数和惩罚函数的具体形式, 对于其他损失函数和惩罚函数需要另行考虑。

关于调节参数, 主要涉及到损失函数和惩罚函数之间的权重选择。调节参数过小, 损失函数权重较大, 模型复杂度惩罚不足; 调节参数过大, 模型复杂度惩罚过重, 模型过于简单, 无法很好地拟合数据。调节参数选择有诸多准则供参考^[16-17], 本节采用 SIC 准则。

引入不加惩罚项时的估计量 $(\tilde{\mathbf{b}}, \tilde{\boldsymbol{\gamma}})$, 令 $\lambda_{1j} = \lambda \|\tilde{\mathbf{b}}_j\|_\infty^{-1} (1 \leq j \leq p), \lambda_{2j} = \lambda \|\tilde{\boldsymbol{\gamma}}_j\|^{-1} (1 \leq j \leq q)$, 只需对 λ 依据 SIC 准则进行选择即可。

$$\text{SIC}(\lambda) = \ln \left(\frac{1}{2n} \sum_{i=1}^n \left| Y_i - \sum_{j=1}^p \hat{\xi}_{ij} \mathbf{b}_j - \mathbf{Z}_i^T \boldsymbol{\gamma} \right| \right) +$$

$$\frac{\ln n}{2n} |\text{df}_\lambda|$$

式中: df_λ 表示调节参数为 λ 时 $Y_i = \sum_{j=1}^p \hat{\xi}_{ij} \mathbf{b}_j + \mathbf{Z}_i^T \boldsymbol{\gamma}$ 的样本数量。

以上针对含有函数型和多元向量自变量的混合回归模型, 从模型本身出发, 利用函数型主成分分析、组变量选择方法、线性规划等, 对模型实现了变量选择和参数估计。

3 数值模拟

在不同误差分布、样本量和信噪比下, 对函数型和多元向量混合回归模型进行变量选择和参数估计, 考查其有限样本性质。

关于函数型自变量及其函数型系数、多元向量自变量及其系数, 参考文献[10]进行如下设计: $p = q = 3, r_{ijm} = \max_m(\mu_{ij}(t_m^{(j)})) - \min_m(\mu_{ij}(t_m^{(j)}))$,

$$X_{ijm} = \mu_{ij}(t_m^{(j)}) + \epsilon_{ijm}, \epsilon_{ijm} \sim N(0, 0.025r_{ij}^2), \mu_{i1}(t) = \cos(2\pi(t - a_1)) + a_2 t, t \in [0, 1], \mu_{i2}(t) = b_1 \sin(2t) + b_2, t \in (0, \pi/3), \mu_{i3}(t) = c_1 t^3 + c_2 t^2 + c_3 t + c_4, t \in [-1, 1], a_1 \sim N(-5, 3^2), a_2 \sim N(7, 1), b_1 \sim U(3, 7), b_2 \sim N(0, 1), c_1 \sim N(-3, 1.2^2), c_2 \sim N(2, 0.5^2), c_3 \sim N(-2, 1), c_4 \sim N(2, 1.5^2),$$

$$\mathbf{Z} \sim N \left(0, \begin{pmatrix} 1 & 0.2 & -0.4 \\ 0.2 & 1 & 0.3 \\ -0.4 & 0.3 & 1 \end{pmatrix} \right)$$

按照如下公式生成因变量:

$$Y_i = \int_0^1 \mu_{i1}(t) \beta_1(t) dt + \mathbf{Z}_i^T \boldsymbol{\gamma} + \sigma \epsilon_i$$

式中: σ 为标准差; $\beta_1(t) = \sin(2\pi t), \boldsymbol{\gamma} = [0.3, 0, 0]$ 。为对比不同样本量、信噪比下的模型估计效果, 分别设定 $n = 100, 300, \sigma = 0.05, 0.2$ 。

关于效果评价指标, 分别使用将非零参数估计为非零的参数个数(TP)、将为零的参数估计为非零的参数个数(FP)来衡量变量选择的效果, $\text{TP} = 2, \text{FP} = 0$ 是最理想的结果。参数估计效果分别使用均方误差根(RMSE)和偏差(Bias)加以衡量:

$$\text{RMSE}(\hat{\beta}_1(t)) = \sqrt{\int_0^1 (\hat{\beta}_1(t) - \beta_1(t))^2 dt}$$

$$\text{Bias}(\hat{\gamma}_1) = |\hat{\gamma}_1 - \gamma_1|$$

为保证结果的稳定性, 将各情况均随机模拟 200 次, 表 1 和表 2 给出了各指标的均值(Mean)和标准差(Sd)。

由表 1 可知, 在正态误差项下能将非零参数都估计为非零, 但某些情况下会将为零的参数估计为非零, 在模型中引入了无关变量。在样本量大、信噪比低情形($n = 300, \sigma = 0.05$), 模型变量

表 1 正态误差下的数据模拟结果

Table 1 Data simulation results with normal error

(n, σ)	统计指标	TP	FP	RMSE	Bias
(100, 0.05)	Mean	2	0.22	0.028 2	0.005 8
	Sd	0	0.52	0.007 6	0.004 4
(100, 0.2)	Mean	2	0.34	0.084 4	0.022 9
	Sd	0	0.61	0.033 0	0.017 9
(300, 0.05)	Mean	2	0.09	0.016 8	0.002 7
	Sd	0	0.30	0.004 8	0.002 0
(300, 0.2)	Mean	2	0.18	0.049 1	0.012 0
	Sd	0	0.42	0.019 5	0.009 8

表 2 柯西误差下的数据模拟结果

Table 2 Data simulation results with Cauchy error

(n, σ)	统计指标	TP	FP	RMSE	Bias
(100, 0.05)	Mean	2	0.01	0.036 0	0.008 3
	Sd	0	0.07	0.007 6	0.004 4
(100, 0.2)	Mean	2	0.03	0.116 8	0.035 5
	Sd	0	0.16	0.054 7	0.030 1
(300, 0.05)	Mean	2	0	0.019 5	0.003 8
	Sd	0	0	0.006 5	0.002 8
(300, 0.2)	Mean	2	0.12	0.062 1	0.014 0
	Sd	0	0.32	0.026 6	0.011 6

选择效果最好。根据 RMSE 和 Bias 可以看出,该方法对于函数型自变量的参数和多元向量自变量的参数均具有良好效果。表 2 结果类似。由表 1 和表 2 的结果对比可知,误差分布为厚尾分布时,变量选择的结果受到影响较小,而参数估计精度受到较大影响。

4 结 论

- 1) 本文同时考虑了函数型自变量和多元向量自变量,拓展了函数型数据分析的应用领域,给出了一种新的数据混合回归模型。
- 2) 引入惩罚函数同时进行变量选择和参数估计,对函数型自变量引入了组变量选择方法,对经过函数型主成分分析投影后的函数型自变量具有选择效果。
- 3) 在变量选择过程中,将目标函数优化问题转化为线性优化问题,降低了参数估计的复杂性。
- 4) 在参数估计过程中考虑了异常值的影响,采用了稳健变量选择方法,扩大了适用性。

参考文献 (References)

[1] FERRATY F. Recent advances in functional data analysis and related topics[M]. Berlin: Springer, 2011.

[2] CHEN S T, XIAO L, STAICU A M. A smoothing-based goodness-of-fit test of covariance for functional data[J]. Biometrics, 2018, 75(2): 562-571.

[3] CUEVAS A. A partial overview of the theory of statistics with functional data[J]. Journal of Statistical Planning and Inference, 2014, 147: 1-23.

[4] PARK J, AHN J. Clustering multivariate functional data with phase variation[J]. Biometrics, 2017, 73(1): 324-333.

[5] KATO K. Estimation in functional linear quantile regression[J]. Annals of Statistics, 2012, 40(6): 3108-3136.

[6] TIBSHIRANI R. Regression shrinkage and selection via the Lasso[J]. Journal of the Royal Statistical Society. Series B(Statistical Methodology), 1996, 58(1): 267-288.

[7] HALL P, HOROWITZ J L. Methodology and convergence rates for functional linear regression[J]. Annals of Statistics, 2007, 35(1): 70-91.

[8] HALL P, HOSSEINI-NASAB M. On properties of functional principal components analysis[J]. Journal of the Royal Statistical Society. Series B(Statistical Methodology), 2005, 68(1): 109-126.

[9] LIN X, LU T, YAN F, et al. Mean residual life regression with functional principal component analysis on longitudinal data for dynamic prediction[J]. Biometrics, 2018, 74(4): 1482-1491.

[10] HUANG L, ZHAO J, WANG H, et al. Robust shrinkage estimation and selection for functional multiple linear model through LAD loss[J]. Computational Statistics & Data Analysis, 2016, 103: 384-400.

[11] QIAN J, SU L. Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso[J]. Journal of Econometrics, 2016, 191(1): 86-109.

[12] VINCENT M, HANSEN N R. Sparse group lasso and high dimensional multinomial classification[J]. Computational Statistics & Data Analysis, 2014, 71: 771-786.

[13] LIU X, LIN Y, WANG Z. Group variable selection for relative error regression[J]. Journal of Statistical Planning and Inference, 2016, 175: 40-50.

[14] WANG H J, LI D, HE X. Estimation of high conditional quantiles for heavy-tailed distributions[J]. Journal of the American Statistical Association, 2012, 107(500): 1453-1464.

[15] BANG S, JHUN M. Simultaneous estimation and factor selection in quantile regression via adaptive sup-norm regularization[J]. Computational Statistics & Data Analysis, 2012, 56(4): 813-826.

[16] WANG T, ZHU L. Consistent tuning parameter selection in high dimensional sparse linear regression[J]. Journal of Multivariate Analysis, 2011, 102(7): 1141-1151.

[17] HIROSE K, TATEISHI S, KONISHI S. Tuning parameter selection in sparse regression modeling[J]. Computational Statistics & Data Analysis, 2013, 59: 28-40.

作者简介:

刘科生 男,博士研究生,助理研究员。主要研究方向:教育大数据、复杂数据分析。

王思洋 女,博士,副教授,硕士生导师。主要研究方向:高维数据分析、信用风险建模。

Variable selection in regression models including functional data predictors

LIU Kesheng¹, WANG Siyang^{2,*}

(1. Big Data Center of Student Affairs Department, Beihang University, Beijing 100083, China;
2. School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China)

Abstract: The variable selection and parameter estimation problem is researched in the framework of mixed-type regression model with both functional and multivariate predictors, which broadens the scope of functional data analysis and the application fields of variable selection methodology. First the functional predictors are projected into spaces spanned by functional principal component basis functions. Then variable selection and parameter estimation are implemented simultaneously for the multivariate predictors and derived projection predictors in the form of grouping, where the tuning parameter of the penalized term is adaptively selected and the loss function is based on absolute median loss function. As to the optimization procedure, by introducing slack variables, it is transformed into a linear programming problem with several constraint conditions, which simplifies the computation. The simulation results illustrate that the proposed method performs quite well in variable selection and parameter estimation in the mixed-type regression model.

Keywords: functional data; variable selection; parameter estimation; quantile; functional principal component

Received: 2019-04-11; Accepted: 2019-04-26; Published online: 2019-06-19 11:19
URL: kns.cnki.net/kcms/detail/11.2625.V.20190618.1535.001.html
Foundation items: National Natural Science Foundation of China (11501586,71420107025); Program for Innovation Research in Central University of Finance and Economics
* Corresponding author. E-mail: siyangw@163.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0044

液体火箭发动机故障诊断器设计及其 HIL 验证



赵万里¹, 郭迎清^{1,*}, 杨菁¹, 薛薇², 武小平²

(1. 西北工业大学 动力与能源学院, 西安 710129; 2. 北京航天动力研究所, 北京 100076)

摘

要: 为了实现某型液体火箭发动机机载实时故障诊断, 采用 FPGA 与 DSP 相结合

的方式作为硬件架构设计了故障诊断器, 其中 FPGA 控制高精度 A/D 转换器进行传感器数据采集, DSP 运行故障诊断算法并将结果输出, 对故障诊断器的硬件和软件分别进行了设计。提出了一种递归结构识别 (RESID) 算法用于液体火箭发动机故障诊断, 该算法可在 6 ms 内诊断出流量衰减故障。搭建基于故障诊断器和工控机的硬件在环 (HIL) 试验平台, 采用自动代码生成技术与手写代码结合的方式对 RESID 算法进行了试验验证, 通过上位机界面进行观察。结果表明: RESID 算法能准确地诊断出发动机常见的故障并在故障诊断器上实现, 算法运行时间为 3.9 ms, 故障诊断器可以实现实时数据监测和故障诊断, 相对于传统平台更加小型化和经济化, 既可以作为机载装置使用, 也可作为通用平台来开发新算法。

关键词: 液体火箭发动机; 故障诊断; 故障诊断器; FPGA + DSP; 代码生成; HIL

平台

中图分类号: V433.9

文献标识码: A

文章编号: 1001-5965(2019)10-1995-08

液体火箭发动机是在极端物理条件下运行的复杂热力学系统, 其故障的发生和发展具有极端的快速性和极大的破坏性。液体火箭对发动机的安全性和可靠性提出了更高的要求, 发展可靠有效的发动机故障诊断系统, 可以对发动机工作过程中出现的故障予以预警和判断, 并能及时采取有效措施, 保护液体火箭和载荷安全, 有效避免由于发动机故障而引起的灾难性事故^[1-2]。

目前, 国内外针对液体火箭发动机故障诊断算法的研究还停留于试验验证阶段, 主要是基于硬件在环 (HIL) 仿真平台和实际试车台进行验证^[3-4]。以上研究都是基于地面试验开展的, 这些设备体积较大, 很难应用于机载设备上, 仅能用于地面仿真验证, 而直接应用于机载系统故障诊断器的研究较少。将故障诊断器应用于液体火箭发动机机载系统, 对飞行过程中的故障进行监测

和处理显得尤为重要, 因此有必要开发一种可以用于机载故障诊断的硬件装置。

近年来, 随着硬件技术的发展, 对单处理器结构 (如单片机、ARM、DSP 和 FPGA 等) 用于数据采集和算法处理的研究越来越多^[5]。但液体火箭发动机故障诊断器是一项极其复杂的工程, 并且其综合性极强, 有必要采用一种能应用于复杂、性能高的场合的多处理器架构, 将以上单一架构相互结合起来, 进行优势互补。液体火箭发动机故障诊断器主要完成高精度数据采集和故障诊断算法的验证。采用 FPGA 灵活的配置能力和 DSP 运算能力构建双处理器进行故障诊断器设计是一种非常合理高效的选择, 其体积小, 价格低廉, 并且通用性高, 易于扩展。

因此, 本文设计了以 FPGA 和 DSP 为硬件的故障诊断器用于液体火箭发动机机载在线故障诊

收稿日期: 2019-01-31; 录用日期: 2019-04-26; 网络出版时间: 2019-06-12 14:57

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190611.1601.004.html

* 通信作者。E-mail: yqguo@nwpu.edu.cn

引用格式: 赵万里, 郭迎清, 杨菁, 等. 液体火箭发动机故障诊断器设计及其 HIL 验证[J]. 北京航空航天大学学报, 2019, 45(10): 1995-2002. ZHAO W L, GUO Y Q, YANG J, et al. Design of liquid rocket engine fault diagnosis device and its HIL verification [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1995-2002 (in Chinese).

断。首先,对故障诊断器的整体架构进行了介绍;其次,分别设计了故障诊断器的硬件和软件;然后,基于前馈神经网络设计了一种用于故障诊断的递归结构识别 (RESID) 算法,数字仿真验证成功后,对该算法进行自动代码生成并与手写的驱动代码相结合;最后,搭建了硬件在环仿真平台,对 RESID 算法进行了验证。

1 故障诊断器设计

1.1 故障诊断器架构设计

液体火箭发动机故障诊断器主要完成传感器信号采集和外部输入指令的录取功能,运行故障诊断算法,显示故障诊断结果并实现报警功能。本文采用 FPGA + DSP 双系统来对故障诊断器进

行设计。
FPGA + DSP 双系统的最大优点是结构灵活,有较强的通用性,适合模块化设计;同时开发周期较短,易于维护和扩展^[6]。系统的核心由芯片和可以重构的器件组成,另外还包括一些外围的辅助电路。系统中各有优势,利用 FPGA 强大的并行处理能力进行数据采集并对系统外围电路进行控制,利用 DSP 的运算能力对信号进行实时处理并运行故障诊断算法,同时与 FPGA 之间也有数据信息的交换。根据需求和标准化、模块化、通用化的设计思想,根据故障诊断器的功能划分确定系统总体设计框图,如图 1 所示,其主要包括电源模块、A/D 采集模块、开关量信号输入、数字处理模块和输出接口组成。

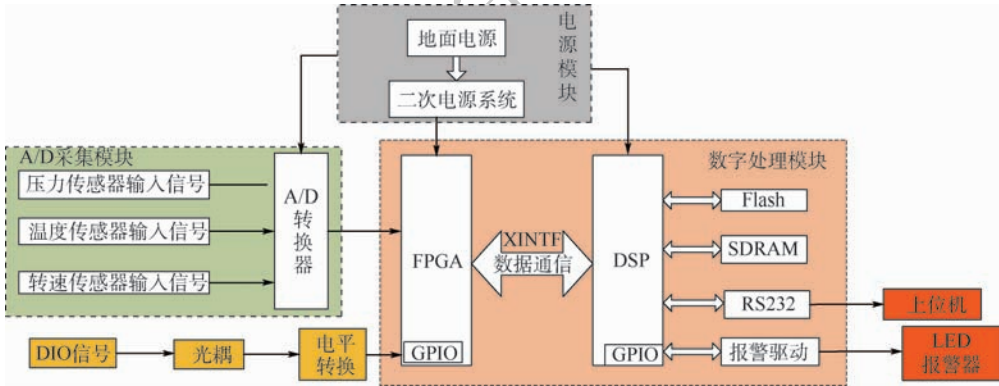


图 1 故障诊断器总体架构
Fig.1 Overall architecture of fault diagnosis device

1.2 故障诊断器硬件设计

在故障诊断器设计中,FPGA 的功能主要包括 24 通道数据采集、内存扩展、DSP 数据通信及在内部开辟数据缓冲区等。综上考虑,选用 Xilinx 公司的 Spartan-6 系列芯片,型号为 XC6SLX16-2FT256I,采用 256 脚的 BGA 封装,其内部包含 14 579 个逻辑单元,32 个硬件乘法器,136 Kbit 可配置逻辑块,576 Kbit 的 Block RAM,可用 IO 数量为 232 个^[7]。

在故障诊断器系统中,DSP 主要的功能是运算故障诊断算法,并且与 FPGA 通信。由于系统本身需要进行故障诊断算法,其包含大量浮点运算,所以选择 C2000 系列的浮点 DSP。综上,最终选择的型号为 TI 公司生产的 TMS320F28335,其是一款哈佛结构的高性能 32 位浮点 DSP,主频为 150 MHz,广泛用于工业控制领域。

本文对外部传感器信号进行采集,核心芯片采用 16 位分辨率的 AD7606。AD7606 是一款集成式 8 通道同步采样数据采集系统,选用 3 片 AD7606,通过并行通信接口与 FPGA 进行数据交

互,核心芯片采用 16 位分辨率的 AD7606。3 片 AD7606 共用 16 位数据线,分别在一个指令周期内进行轮训采样。图 2 为 A/D 采集实现框图。利用 FPGA 的可编程逻辑控制特点,采用状态机的思想,通过编程设计 AD7606 的读写控制时序,将采样的结果实时保存在双口 RAM 里面,进而实现高精度的 A/D 采样。

故障诊断器的显示模块主要包括状态指示灯和故障指示灯 2 部分。图 3 为显示模块及报警模块实现框图。DSP 的 GPIO 口与 LED 相连,显示故障诊断结果功能,绿色为无故障状态,红色为故障状态;状态指示灯主要显示工作状态和各模块自检的状态等;DSP 的 GPIO 口与蜂鸣器相连,

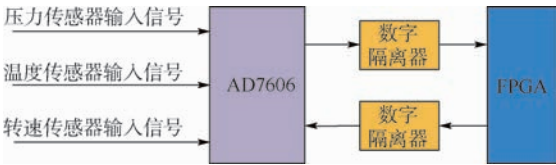


图 2 A/D 采集实现框图
Fig.2 Block diagram of A/D acquisition implementation

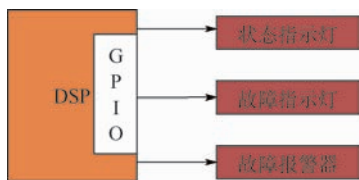


图 3 显示模块及报警模块实现框图

Fig. 3 Block diagram of display module and alarm module implementation

实现报警功能,当故障发生时,蜂鸣器响 30 s 加以提示。

1.3 故障诊断器软件设计

1.3.1 FPGA 控制数据采集

通过高精度、高速率的 A/D 转换芯片将发动机传感器传来的模拟量转化为数字量,实时传给诊断处理芯片,中央控制器通过实时的传感器信息,采用故障诊断算法在线判断发动机状态,给出估计,判断故障信息。FPGA 通过并口对多个传感器数据进行读取,相比于串行数据传输,并行传输速率更快,在实时性要求较高的故障诊断系统中,需要传输多组传感器数据进行处理,采用并行传输成为了最好的选择^[8]。

AD7606 可以对所有 8 路的模拟输入通道进行同步采样。当 2 个 CONVST 引脚 (CONVSTA 和 CONVSTB) 连在一起时,所有通道同步采样。

在此共用 CONVST 信号的上升沿启动对所有模拟输入通道的同步采样 (V1 ~ V8)。在 FPGA 里通过有限状态机来实现对 AD7606 的控制,CS 和 RD 均处于逻辑低电平时,则会使能输出总线 DB[15:0],使转换结果输出在并行总线上,AD_BUSY 输出保持高电平,直到所有通道的转换过程完成为止,通过 RANGE 引脚来选择模拟输入的范围。AD7606 有限状态机的结构框图如图 4 所示。根据 AD7606 的工作原理和时序图,先复位信号,再转换,完成后等待一段时间进行 A/D 数据的读取^[9]。具体的状态机轮转图如图 5 所示。

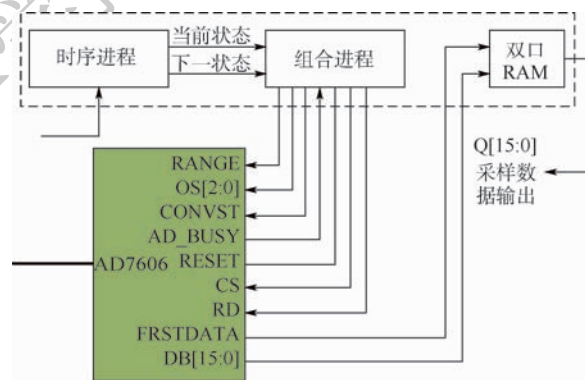


图 4 AD7606 有限状态机结构框图

Fig. 4 Structural block diagram of AD7606 finite state machine

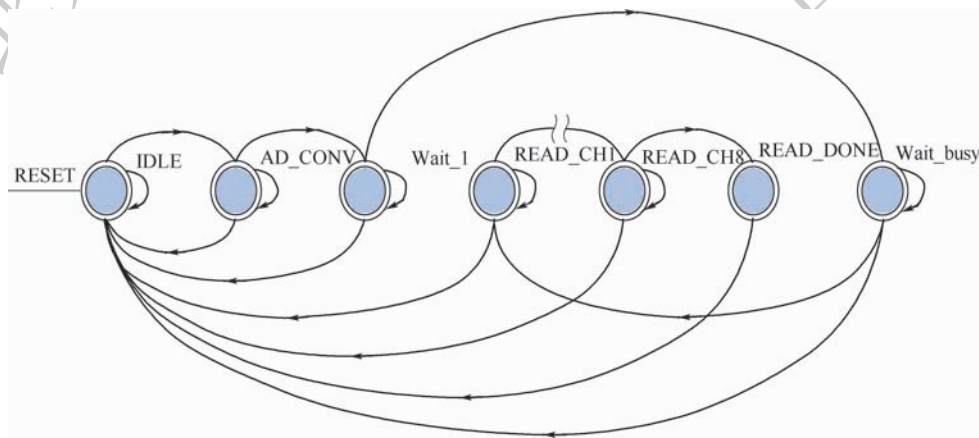


图 5 FPGA 控制 AD7606 有限状态机

Fig. 5 FPGA controlling AD7606 finite state machine

1.3.2 FPGA 与 DSP 通信

通过 FPGA 进行数据采集之后,需要把采样数据发送给 DSP 处理器进行故障诊断。在 FPGA 里通过 IP 核构建一块双口 RAM,用于存储 AD7606 的数据输出结果。DSP 作为故障诊断的主控处理器,可以任意时刻读取双口 RAM 里的采样数据,与通过中断方式或者查询方式读取采样数据相比,减小了对 DSP 资源的占用,提高了

DSP 效率。在 FPGA 里采用 IP 核构造双口 RAM 和 PLL 模块,通过 PLL 分频和倍频,分别构造 10、50、75、100 MHz 的时钟作为双口 RAM 的时钟。

通过选择 XINTF 接口,DSP 可以很方便地读取或者写入 FPGA 的寄存器,保证了系统的实时性^[10]。通过对这些数据进行处理并将其作为故障诊断算法所需的输入,进行故障诊断,再分别通过故障指示灯和故障报警器来显示故障诊断结果

并报警,报警和指示灯的软件采用基于模型的设计,将其加入算法模型,利用 RTW 工具箱生成可执行的嵌入式代码并下载至 DSP 中。

2 故障诊断算法设计

本文根据已有的液体火箭发动机模型来进行故障诊断的研究^[11],在建模时不考虑冗余,传感器测点参数如表 1 所示,主要包含甲烷涡轮泵转速、液氧涡轮泵转速、甲烷泵后压力、液氧泵后压力、副系统甲烷喷嘴前压力、副系统液氧喷嘴前压力、主系统甲烷喷嘴前压力、主系统液氧喷嘴前压力、燃气发生器压力、燃气发生器温度和燃烧室压力。根据目前国内外的研究,液体火箭发动机系统的故障主要发生在推力室和涡轮泵系统。发动机主级阶段主要发生堵塞、泄漏和涡轮泵系统故障等,本文模拟液体火箭发动机主要的故障类型,包括阀门开度故障、泵气蚀、涡轮效率损失和管路泄漏。

RESID 算法基于自适应学习网络概念,将复杂的特征阵列的非线性关系用一个模型进行表达,该特征阵列即为传感器参数,其通过递归的方法建立了不同传感器之间的相互关系。RESID 算法的特征是在递归过程中一步步确立出最优的模型。RESID 算法结构如图 6 所示。图中: E 为 RESID 算法中每一步采用的训练算法, x_i 为输入

表 1 液体火箭发动机传感器测点

Table 1 Sensor measurement points for liquid rocket engines

参数	变量符号	单位
甲烷涡轮泵转速	n_f	r/min
液氧涡轮泵转速	n_o	r/min
甲烷泵后压力	P_{ef}	MPa
液氧泵后压力	P_{eo}	MPa
副系统甲烷喷嘴前压力	P_{gf}	MPa
副系统液氧喷嘴前压力	P_{go}	MPa
主系统甲烷喷嘴前压力	P_{cf}	MPa
主系统液氧喷嘴前压力	P_{co}	MPa
燃气发生器压力	P_g	MPa
燃气发生器温度	T_g	K
燃烧室压力	P_c	MPa

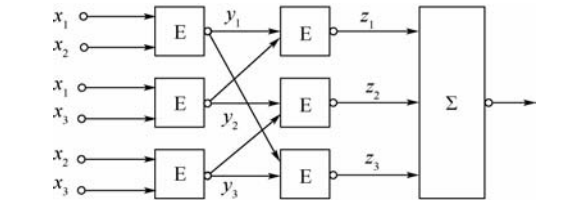


图 6 两级 RESID 网络示意图

Fig. 6 Schematic of two-level RESID network

传感器参数。对输入传感器参数的所有成对组合都确定函数 E 。选择过程中,在最小平方误差条件下表现差的元素被淘汰,其余元素成为下一阶段训练和选择步骤的输入。这些训练和选择步骤重复,直到性能度量没有显示任何进一步的改进。在 RESID 算法中,需要对模型中每一组参数进行训练。相较于传统的非线性回归方法,本文采用了性能更优的前馈神经网络算法^[12]。

RESID 算法通过选择 n 个传感器数据作为建模的预测变量 X , 1 个传感器数据作为建模的响应变量 y , 建立方程 $y = \text{func}(X(1), X(2), \dots, X(n))$, $X(1) \sim X(n)$ 为预测变量中的参数组, func 为建模方程。通过标称数据的响应变量和预测变量,求得建模方程。当系统工作正常时,新的输入变量代入方程中计算得到的预测值,应与实际传感器测试值具有较高的一致性;否则,认为发生故障。

图 7 为液体火箭发动机 RESID 算法框图。第 1 步,对数据进行预处理。第 2 步,根据训练数据集,对所有可以配对的 RESID 模型进行建模,输入为 n 个传感器数据,输出为推力室室压。第 3 步,选择所有模型中预测均方估计误差最小的模型,确定下来该模型。第 4 步,根据确定下来的模型计算出的推力室室压,作为下一级迭代建模用的推力室室压数据(第 3 步、第 4 步迭代进行,直到 m 个模型建立成功)。第 5 步,将被测数据代入建好的 RESID 模型中,该模型利用选定的 n 个传感器数据,预测推力室室压,将预测值与输入的推力室室压数据相比较,得到残差数据。第 6 步,对所有模型的残差计算平均绝对误差,若该值超过诊断阈值,则认为有故障发生。

故障判定时,计算训练数据建立的模型得到的预测值与实际值之间的平均绝对误差,平均绝对误差为全部模型误差的均值。该误差作为故

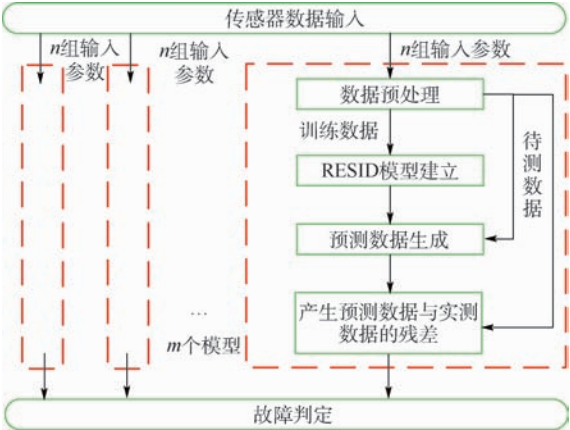


图 7 RESID 算法框图

Fig. 7 Block diagram of RESID algorithm

障诊断判定的基准。在测试时,将新输入的数据代入模型,得到推力室室压的预测值,并将预测值与实测值比较得到该测试数据的平均绝对误差。将测试平均绝对误差与训练平均绝对误差相比较。若超过训练时平均绝对误差值的 n 倍,则认为有故障发生。通过查阅文献,通常 n 取 3 或 5,可根据实际发动机状态进行调整。

针对某型液体火箭发动机模型,在第 4 s 注入故障,使流量系数由 1 衰减为 0.9。在数字仿真下观察结果。使用 GUI 界面可以进行通道选择和参数设置,图 8 为 GUI 界面显示的训练残差。取训练残差最大值 0.020 6 为阈值。图 9 为 GUI 界

面显示的故障诊断参数设置和诊断结果界面,上面为参数设置,下面诊断结果界面中红线和蓝线代表残差和阈值。可以看出,在第 4 s 时阈值很明显高于残差,因此可以判定在第 4 s 发生了故障。发生故障前后分别进行计时,诊断出故障的时间为 4.006 s。在发生故障之后,可以在 6 ms 诊断出来,验证了在数字仿真下 RESID 算法可以进行液体火箭发动机故障诊断。

3 HIL 试验验证

3.1 仿真平台总体架构设计

发动机故障诊断器需要对液体火箭发动机各部件的工作状态进行故障诊断。针对发动机机载故障诊断需求,研究机载故障诊断器研制过程中的关键技术,并采用通用硬件、软件技术构建故障诊断器,验证其平台功能(硬件构架、软件功能)及诊断算法性能(实时性等),为故障诊断工程样机研制奠定技术基础^[13]。本文构建了对应的 HIL 仿真平台,主要由实际控制器和虚拟对象组成,其可以模拟发动机实际运行过程和故障特征,实现故障诊断器状态监控、故障诊断和健康管理功能及性能的在线、实时试验验证。

搭建的 HIL 仿真平台主要包括故障诊断端、被监控端和上位机端,其分别采用设计的故障诊断器、工控机及 PC 机模拟实现^[14]。其中,故障诊断端与被监控端之间采用 D/A 和 A/D 通信方式实现数据实时交互,故障诊断与主工作站之间采用 RS232 串口通信实现数据实时交互,被监控端与主工作站之间采用 TCP/IP 通信实现数据实时交互;通过 TCP/IP 通信,将诊断结果传输给 TCP/IP 的上位机端。具体的 HIL 仿真平台的总体硬件结构框图如图 10 所示。

3.2 RESID 算法 HIL 验证

图 11 为本文设计的液体火箭发动机故障诊断器与搭建的 HIL 平台实物图,主要包含液体火

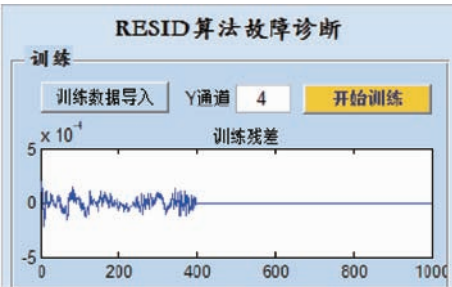


图 8 RESID 算法训练残差
Fig. 8 Training residuals of RESID algorithm

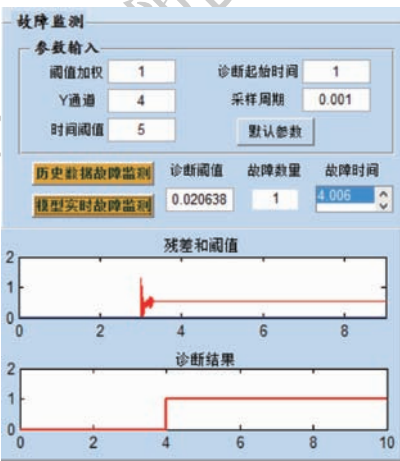


图 9 RESID 算法故障诊断结果
Fig. 9 Fault diagnosis results of RESID algorithm

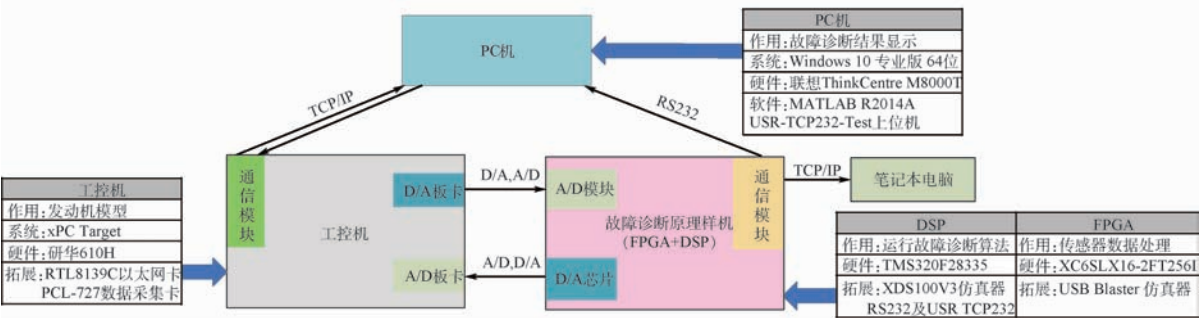


图 10 HIL 仿真平台总体结构设计框图
Fig. 10 Overall structure design block diagram of HIL simulation platform

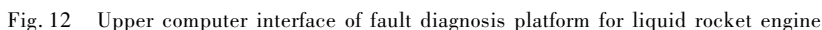
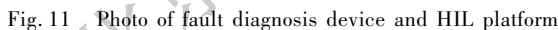
箭发动机实时状态参数显示、PCL-727 模块产生的模拟量输出、上位机的显示和故障诊断器。

为了实现在线故障诊断多种功能需求,构建的故障诊断需要尽可能满足复杂算法实时性要求。本文设计了一种双线程机制来完成程序结构设计,缩短程序运行时间^[15]。这2种机制分别为HWI硬件中断线程和TSK任务线程。HWI硬件中断线程具有最高的优先级和严格的实时性,中断服务程序采用HWI硬件中断线程,通过定时器中断完成,主要实现实时数据采集、在线故障诊断算法及控制算法。TSK任务线程优先级较低,在运行过程中可以等待,直到所需要的资源可用。由于主程序的特点就是运行的非连续性及中断

位置的随机性,因此采用了 TSK 任务线程这部分程序实现信息的传输功能,即将故障诊断结果及故障诊断端接收到的传感器数据传输到主工作站。

由上所述,将搭建的故障诊断软件系统分为 2 个部分:主程序和中断服务程序。在主程序中,主要进行各个模块和算法的初始化,采用 CPU 定时器中断,20 ms 自动触发一次,在中断服务函数中运行故障诊断算法并将其结果发送至上位机界面,在算法开始前和开始后分别在某一定时器寄存器中写值,这样就可以观察算法的运行时间。故障诊断算法基于模型设计的自动代码生成方式进行生成,先对其进行 DO-178C 标准检查,再替换不符合标准的模块,使其生成嵌入式 C 代码,最后将生成的代码嵌入至 DSP 代码中。

通过 RS422 串口将 A/D 采集到的值和故障诊断结果发送至液体火箭发动机故障诊断平台的上位机上进行观察,如图 12 所示。将 A/D 采集到的数据与 xPC 端的数据进行对比,精度可以达到 1 mV,在某一时间植入故障,上位机中各参数发生变化,并显示发生故障,算法在 20 ms 的中断中只运行了 3.9 ms。因此,采用本文设计的 DSP + FPGA 双系统架构可以使数据采集精度更高并且减少 CPU 占用率,采用手写代码与嵌入式代码相结合的方式方便 HIL 平台的开发。



4 结 论

本文基于 FPGA 和 DSP 双处理器架构设计了具有通用性的故障诊断器,可用于液体火箭发动机机载在线故障诊断,对故障诊断器软硬件和故障诊断算法进行了研究和试验验证。主要结论如下:

1) 对故障诊断器各模块软硬件进行了设计,利用 FPGA 灵活的配置能力,基于有限状态机实现了对 AD7606 进行传感器数据采集的控制。在 FPGA 端构造双口 RAM,DSP 通过 XINTF 总线成功地从双口 RAM 中读取 A/D 转换器的数据。

2) 提出了一种 RESID 算法用于液体火箭发动机故障诊断,将算法自动生成代码并与手写嵌入式代码相结合,然后下载至故障诊断器中进行验证。

3) 利用故障诊断器、工控机和上位机搭建了用于算法验证的 HIL 仿真平台,证实了设计算法在故障诊断器中实现的可靠性,与数字仿真结果一致。

本文为液体火箭发动机故障诊断工程样机的设计提供了有效途径,促进了液体火箭发动机故障诊断领域的研究工作。本文是在液体火箭发动机的数学模型下进行的,后续可在地面试车台对故障诊断器进行测试并改进,将其应用于机载装置中。

参考文献 (References)

- [1] LI Y J, PENG X H, CHENG Y Q, et al. Research of pipeline fault diagnosis for liquid rocket propulsion system[J]. Applied Mechanics and Materials, 2012, 232: 305-309.
- [2] LI Y J, PENG X H, CHENG Y Q, et al. Research of faulty sensors data reconstructed for liquid-propellant rocket engines[J]. Applied Mechanics and Materials, 2012, 229-231: 1449-1453.
- [3] ZHU F, WANG Q, SHEN Z. APSO-RVM for fault detection of liquid rocket engines test-bed[J]. Information Technology Journal, 2012, 11(10): 1496-1501.
- [4] DOU W, SUN L. Analysis of on-site fault diagnosis technology of a liquid rocket engine[J]. Strength and Environment, 2010, 37(5): 46-51.
- [5] JI W, TAN D L, NDAGIJIMANA E, et al. An embedded hardware-in-the-loop testbed for aeroengine simulation education: AIAA-2018-4621[R]. Reston: AIAA, 2018.
- [6] CHI Z, JIA Q, WANG X, et al. An intelligent interface design method based on DSP and FPGA[C] // International Conference on Electrical, Mechanical and Industrial Engineering,

2016: 318-321.

- [7] 徐加彦, 张之万, 陈兴林, 等. 基于 FPGA 的高速多通道 AD 采样系统的设计与实现[J]. 自动化与仪表, 2014, 29(9): 49-52.
- XU J Y, ZHANG Z W, CHEN X L, et al. Design and implementation of high speed multichannel AD sampling system based on FPGA[J]. Automation and Instrumentation, 2014, 29(9): 49-52 (in Chinese).
- [8] 姜楠, 马迎建, 冯翔. DSP 和 FPGA 并行通信方法研究[J]. 电子测量技术, 2008, 31(10): 146-148.
- JIANG N, MA Y J, FENG X. Research on parallel communication method between DSP and FPGA[J]. Electronic Measurement Technology, 2008, 31(10): 146-148 (in Chinese).
- [9] 郭迎清, 冯健朋, 张书刚. 涡扇发动机故障诊断硬件在回路实时仿真平台[J]. 航空工程进展, 2014, 5(2): 165-170.
- GUO Y Q, FENG J P, ZHANG S G. Hardware-in-the-loop simulation platform for turbofan engine fault diagnosis[J]. Journal of Aeronautical Engineering, 2014, 5(2): 165-170 (in Chinese).
- [10] WU J. Liquid-propellant rocket engines health-monitoring—A survey[J]. Acta Astronautica, 2005, 56(3): 347-356.
- [11] LEE K, CHA J, KO S, et al. Fault detection and diagnosis algorithms for an open-cycle liquid propellant rocket engine using the Kalman filter and fault factor methods[J]. Acta Astronautica, 2018, 150: 15-27.
- [12] ZHUO P C, ZHU Y, WU W X, et al. Real-time fault diagnosis for gas turbine blade based on output-hidden feedback Elman neural network[J]. Journal of Shanghai Jiaotong University (Science), 2018, 23(Suppl): 95-102.
- [13] THOMAS G L, CULLEY D E, BRAND A. The application of hardware in the loop testing for distributed engine control[C] // AIAA/SAE/ASEE Joint Propulsion Conference, 2016: 1-11.
- [14] KULIKOV G G, AR'KOV V Y, ABDULNAGIMOV A I. Hardware-in-the-loop testing technology for integrated control and condition monitoring systems of aircraft gas turbine engines[J]. Russian Aeronautics, 2008, 51(1): 47-52.
- [15] 吕升, 郭迎清, 孙浩. 航空发动机故障诊断装置硬件在环实时仿真平台[J]. 航空发动机, 2017, 43(3): 43-49.
- LÜ S, GUO Y Q, SUN H. Hardware-in-the-loop real-time simulation platform for aero-engine fault diagnosis device[J]. Aero Engine, 2017, 43(3): 43-49 (in Chinese).

作者简介:

赵万里 男, 博士研究生。主要研究方向: 液体火箭发动机故障诊断算法及其硬件在环仿真平台设计。

郭迎清 男, 教授, 博士生导师。主要研究方向: 航空发动机先进控制技术、故障诊断与健康维护。

杨菁 女, 博士研究生。主要研究方向: 复杂系统的故障诊断与预测。

Design of liquid rocket engine fault diagnosis device and its HIL verification

ZHAO Wanli¹, GUO Yingqing^{1,*}, YANG Jing¹, XUE Wei², WU Xiaoping²

(1. School of Power and Energy, Northwestern Polytechnical University, Xi'an 710129, China;

2. Beijing Aerospace Propulsion Institute, Beijing 100076, China)

Abstract: In order to realize the real-time fault diagnosis of a liquid rocket engine onboard, a fault diagnosis device is designed by combining FPGA and DSP as the hardware architecture. The FPGA controls the high-precision A/D for sensor data acquisition, the DSP runs the fault diagnosis algorithm and outputs the result. The hardware and software of the fault diagnosis device were designed separately. A recursive structure identification (RESID) algorithm is proposed for liquid rocket engine fault diagnosis. The algorithm can diagnose traffic attenuation faults in 6 ms. Based on the hardware-in-the-loop (HIL) test platform of fault diagnosis and industrial computer, the algorithm was tested and verified by the combination of automatic code generation technology and handwritten code, and observed through the upper computer interface. The results show that the RESID algorithm can accurately diagnose the common faults of the engine and realize it on the fault diagnosis device. The running time of the algorithm is 3.9 ms. The fault diagnosis device can realize real-time data monitoring and fault diagnosis, which is more compact and economical than the traditional platform. It can be used both as an onboard device and as a general platform to develop new algorithms.

Keywords: liquid rocket engine; fault diagnosis; fault diagnosis device; FPGA + DSP; code generation; HIL platform

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0050

基于 Gram-Schmidt 变换的有监督变量聚类



刘瑞平¹, 王惠文^{1,2}, 王珊珊^{1,3,*}

1. 北京航空航天大学 经济管理学院, 北京 100083; 2. 北京航空航天大学 大数据科学与脑机智能
高精尖创新中心, 北京 100083; 3. 城市运行应急保障模拟技术北京市重点实验室, 北京 100083)

摘 要: 为进一步研究回归模型中高维数据的降维方法, 提出基于 Gram-Schmidt 变换的新的有监督变量聚类(SCV-GS)方法。该方法未采用以潜变量为聚类中心的层次聚类, 而是借用变量扫描思想, 依次挑出对响应变量有重要贡献的关键变量, 并将其作为聚类中心。SCV-GS 方法基于 Gram-Schmidt 变换, 对变量之间的高度相关性进行批量处理, 并得到聚类结果; 同时, 结合偏最小二乘思想, 提出新的同一性度量, 并以此来选取最佳聚合参数。SCV-GS 不仅可以快速得到变量聚类结果, 而且可识别出对响应变量的解释及预测起关键作用的变量类。仿真表明该聚类方法运算速度显著提升, 而且所得潜变量对应的回归系数的估计结果与对照方法表现一致; 实例分析表明该方法具有更好的解释性和预测能力。

关 键 词: 降维; 变量聚类; 回归; 高度相关; Gram-Schmidt 变换

中图分类号: O212.4

文献标识码: A **文章编号:** 1001-5965(2019)10-2003-08

随着信息技术的快速发展和大数据时代的来临, 数据信息的收集与存储变得极其便捷, 各行各业已经形成规模巨大、亟待开发的数据。如何灵活运用这些数据宝藏, 快捷有效地探究数据的本质特征与运行规律, 这已经成为大数据分析中的重要课题, 催生新一轮统计数据建模理论与方法的创新高潮。其中建立回归模型或分类模型是最常用的统计分析方法之一, 而此类模型通常面临两大挑战: 第一, 在成千上万维的变量中可能仅有少数变量起关键作用, 即变量维数 p 远远大于关键变量的维数 d ; 第二, 变量之间往往存在高度相关性。对此, 需要对数据进行降维处理。文献中基于回归模型的高维数据降维技术主要包含 2 种方法: 第 1 种是变量选择(或称特征筛选), 即从原始变量集合中筛选出维数较小的变量子集。这类方法包括经典的惩罚类方法(如 LASSO (Leaset Absolute Shrinkage and Selection Operator)^[1]、elastic-

net^[2]等)和扫描类方法^[3-4]。这类方法可以达到很好的预测效果, 但在进行变量筛选的同时可能会忽略掉与关键变量强相关的重要变量, 导致模型解释性的缺失。第 2 种是进行变量综合, 即将原始 p 维变量投影到低维空间, 得到原始变量的线性组合, 再将这些组合作为潜变量(latent variable)进行回归建模。比如经典的主成分回归、偏最小二乘回归等。这类方法虽考虑了变量之间的高度相关性, 同时实现了降维, 但因其将所有变量都考虑进来, 为模型的解释性增加了困难。

为了兼顾两类数据降维方法的优点, 实现既能挑选出关键变量, 又能将变量之间的高度相关性考虑进来, 一些稀疏综合方法被相继提出, 比如稀疏主成分分析(Sparse Principal Component Analysis, SPCA)方法^[5]、稀疏偏最小二乘(Sparse Partial Least Squares regression, SPLS)^[6]等。此类方法通过将原始变量进行稀疏组合, 得到稀疏潜变

收稿日期: 2019-02-16; 录用日期: 2019-03-15; 网络出版时间: 2019-06-10 09:15

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190606.1053.002.html

基金项目: 国家自然科学基金(71420107025, 11701023)

* 通信作者。E-mail: sswang@buaa.edu.cn

引用格式: 刘瑞平, 王惠文, 王珊珊. 基于 Gram-Schmidt 变换的有监督变量聚类[J]. 北京航空航天大学学报, 2019, 45(10): 2003-2010. LIU R P, WANG H W, WANG S S. Supervised clustering of variables based on Gram-Schmidt transformation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2003-2010 (in Chinese).

量,再进行建模。另一种思路是对所有变量进行聚类。2016年,Chen和Vigneau^[7]首次提出“有监督变量聚类”,以求同时提升模型的预测性与解释性,该方法与SPCA、SPLS有所区别,其目标是实现对所有变量的聚类,同时从变量类中筛选出对模型预测性能起关键作用的变量组。本文第2节将详细介绍有监督变量聚类方法。

事实上,由于变量聚类有助于探索多元数据的内部结构,便于后续的解释和分析。因此,在Chen和Vigneau^[7]之前,已有相关学者对变量聚类进行探索研究。只是在此之前的变量聚类大多属于“无监督变量聚类”。在20世纪70年代,Jolliffe^[8]总结了对变量的层次聚类的步骤。首先定义2个类之间的相似度;然后从每个变量自成一类开始,对所有类计算两两之间的相似性;最后将相似性最大的2个类合并为一类,循环上述步骤,直到所有变量聚为一类。Hastie等^[9]提出了Tree harvesting方法,并将其应用于基因数据的分析研究。Tree harvesting方法对 p 维变量施加层次聚类法,由此可得到 $2p-1$ 个变量类;然后采用向前法,在每一步中,挑选当前候选类中最小化残差平方和的变量类,将 $2p-1$ 个变量类依次以变量平均的形式加入模型。Tree harvesting方法思想朴素,步骤简单。之后Vigneau和Qannari^[10]对变量聚类进行了深入的研究,提出围绕潜变量的聚类,并给出 K 个类的聚类算法,目标为最大化所有类潜变量与对应的各类中自变量的相关程度 T 。求解方法是迭代算法,在聚类的不同阶段允许变量 x_j 进来或出去,在每一个阶段会增大准则 T 的值。当前有关变量聚类研究的文献相对较少,主要是Vigneau和Qannari团队在研究,且多数为无监督变量聚类。Vigneau等^[11-12]给出了围绕潜变量聚类的R程序,并将变量聚类扩展至含缺失数据的情形,Cariou等^[13]将围绕潜变量的聚类方法应用于结构方程模型中。

Vigneau等^[7,10]提出的聚类方法均围绕潜变量展开。具体来说,是希望同时找到 K 个变量类和 K 个潜变量,使得每个变量类与对应的潜变量具有强相关。但由于该方法步骤复杂,且包含多次重复循环,因此计算速度较低。为此,本文提出了一种新的快速有监督变量聚类方法。具体地,该方法借用变量扫描思想,依次挑出关键变量,并将其作为聚类中心;基于Gram-Schmidt变换,对变量之间的高度相关性进行批量处理;同时,结合偏最小二乘思想,提出新的同一性度量,并以此来选取最佳聚合参数。因此,它不仅可以快速得到

变量聚类结果,以及变量是以什么样的结构对响应变量起作用的。

1 有监督变量聚类

Chen和Vigneau^[7]首次提出基于回归模型的有监督围绕潜变量的变量聚类(SCV-LV)方法。本文采用层次聚类法,通过最大化提取出的变量组的组内相关性以及最大化变量组对响应变量 y 的贡献,依次迭代提取出一系列变量组。聚类准则为:最大化变量组的全局同一性,同时根据响应变量来控制变量组局部同一性的损失。聚类目标是:得到变量之间的聚类信息,提升模型的预测能力,同时提升模型的可解释性。做法是:首先将响应变量 y 和所有解释变量 $\{x_1, x_2, \dots, x_p\}$ 放在一起,然后进行层次聚类。得到多层聚类结果后,接着确定最佳聚合水平,即选出满足要求的最优一层聚类结果。具体做法如下:

步骤1 初始化。

将 y 和所有解释变量 $\{x_1, x_2, \dots, x_p\}$ 放在一起,即 $U = \{x_1, x_2, \dots, x_p, y\}$ 。

步骤2 对 U 进行层次聚类。

需将初始的 $p+1$ 个类,一层一层聚类,直到所有 $p+1$ 个变量成为一类。从上一层到下一层的过程中(例如,从 $p+1$ 个类到 p 个类的过程),有 C_{p+1}^2 种可能性,此时需选取全局同一性下降最小的那一种。假设当前层有 K 个变量类,记作 $G_1, G_2, \dots, G_K$,则全局同一性定义为该层各组所得潜变量 c_k 与对应各组中各个变量之间的相关性之和,其中潜变量是对应变量组中所含变量的线性组合 $c_k = G_k \mathbf{v}_k$, $\mathbf{v}_k$ 为组合系数向量。

全局同一性准则 T 的表达式为

$$T = \sum_{k=1}^K \sum_{j=1}^{|G_k|} \text{Cov}^2(x_j, c_k) \quad \|\mathbf{v}_k\| = 1 \quad (1)$$

式中: $|G_k|$ 为变量组 G_k 中所含变量的个数。

容易证明最大化准则 T 所得潜变量 c_k 恰好为 G_k 中变量所得的第一主成分,因此该准则等价于最大化各变量组协方差矩阵的第一特征值,即 $\max \sum_{k=1}^K \lambda_1^{G_k}$, λ_1 为第一特征值。考虑当前聚合层中的任意2个变量类 G_A 和 G_B ,则将 G_A 与 G_B 合并为同一类的过程中,同一性的下降量为

$$\Delta = \lambda_1^{G_A} + \lambda_1^{G_B} - \lambda_1^{G_A \cup G_B} \quad (2)$$

因此,从当前聚类层到下一聚类层的划分依据为:将 Δ 取最小值时对应的2个类合并为一类。

步骤 3 确定最佳聚合水平。

得到各个层的聚类结果后,为了确定最佳一层聚合水平,需要选取局部同一性变化最小的一层。考虑第 $l-1$ 层及第 l 层中 y 所在的变量组,并记对应的潜变量为 c_{l-1} 和 c_l 。则局部同一性定义为

$$\begin{cases} T_{G_{l-1}} = n \left[\text{Cov}^2(y, c_{l-1}) + \sum_{j | a_{l-1,j}=1} \text{Cov}^2(x_j, c_{l-1}) + \sum_{j | a_{l-1,j}=0; a_{lj}=1} \text{Cov}^2(x_j, c_{*j}) \right] \\ T_{G_l} = n \left[\text{Cov}^2(y, c_l) + \sum_{j | a_{lj}=1} \text{Cov}^2(x_j, c_l) \right] \end{cases} \quad (3)$$

式中: n 为样本容量; c_{*j} 为变量 x_j 在与 y 合并(第 l 层)之前(第 $l-1$ 层)所在的变量组对应的潜变量。于是相邻两层之间的局部同一性变化率为

$$\Gamma_l = \frac{T_{G_{l-1}} - T_{G_l}}{T_{G_l}} \quad (4)$$

可以看出局部同一性准则反映了 2 个相邻聚合水平同一性的损失比率。给定损失阈值,最佳聚合水平定义为 $l^* = \max l(\Gamma_l \leq \gamma)$, γ 通常设为 5% 或 10%。

SCV-LV 方法虽然可以达到很好的聚类效果,但其步骤复杂,计算量庞大,导致计算成本过高。本文将结合变量扫描的思想,同时借助 Gram-Schmidt 变换,提出一种新的快速有监督变量聚类算法。

2 基于 Gram-Schmidt 变换的回归模型的快速变量聚类

本节首先介绍 Gram-Schmidt 变换及其信息分解准则,然后提出新的快速有监督变量聚类算法。该算法借用变量扫描思想,依次挑出关键变量;同时基于 Gram-Schmidt 变换,对变量之间的高度相关性进行批量处理。最后给出基于变量聚类的回归模型。

2.1 Gram-Schmidt 变换

Gram-Schmidt 变换作为线性代数中的经典算法^[14-15],已被应用于多种领域。比如,Chen 等^[16]将之应用于非线性系统的识别;Stoppiglia 等^[17]首次提出基于 Gram-Schmidt 变换的特征选择;王惠文等^[18-19]提出了基于 Gram-Schmidt 过程的回归模型及判别模型的变量筛选方法;Liu 等^[20]研究了基于 Gram-Schmidt 过程的函数型数据回归模型的变量扫描。

给定 $n \times p$ 矩阵 $M = [w_1, w_2, \dots, w_p]$, M 的

秩为 $s(s \leq \min(n, p))$, 则 Gram-Schmidt 正交化过程将 w_j 转换为 u_j :

$$\begin{cases} u_1 = w_1 \\ u_j = w_j - \sum_{k=1}^{j-1} \frac{w_j^T u_k}{u_k^T u_k} I(u_k^T u_k > 0) \quad j = 2, 3, \dots, p \end{cases} \quad (5)$$

式中: $I(\circ)$ 为示性函数。

易知,在所得到的 $u_j(j = 1, 2, \dots, p)$ 中,有 s 个非零向量,其余均为零向量。设 u, w 分别为变量 u, w 的样本实现,可证 w_k 所携带的信息可以分解为两部分:一部分由 u_k 承载,另一部分由 w_k 和 $u_l(l = 1, 2, \dots, k-1)$ 共同承担。 w_k 携带的信息有如下分解准则:

$$\text{Var}(w_k) = \text{Var}(u_k) + \sum_{l=1}^{k-1} \text{Cor}^2(w_k, u_l) I(u_l^T u_l > 0) \quad (6)$$

u_k 的方差满足:

$$\text{Var}(u_k) = \text{Cor}^2(w_k, u_k) I(u_k^T u_k > 0) \quad (7)$$

式中: Var 表示变量对应的样本方差; Cor 表示样本相关系数。其中 $\text{Var}(u_k) = u_k^T u_k / n$, 且 $\text{Cor}(w_k, u_l) = w_k^T u_l / \sqrt{w_k^T w_k u_l^T u_l}$ 。

从信息分解式(5)可知,经过 Gram-Schmidt 变换后, u_k 承载的信息量不会多于原始变量 w_k 承载的信息量。而且, w_k 携带的信息量恰好等于原始变量 w_k 和变换后的变量 u_k 之间的样本相关系数的平方和。因此由 u_k 承载的信息可以由 $\text{Cor}(w_k, u_l)$ 间接表示,式(5)和式(6)称为 Gram-Schmidt 变换的信息分解准则。

2.2 基于 Gram-Schmidt 变换的变量聚类

本文将结合扫描思想,提出基于 Gram-Schmidt 变换的新的有监督变量聚类(SCV-GS)方法。假设收集的样本数据为 $y = [y_1, y_2, \dots, y_n]^T$, $X_{n \times p} = [x_1, x_2, \dots, x_n]^T$, 其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^T$ 。不失一般性,假设响应变量和解释变量分别经过中心化和标准化处理:

$$\begin{cases} \sum_{i=1}^n y_i = 0 \\ \sum_{i=1}^n x_{ij} = 0 \\ \sum_{i=1}^n x_{ij}^2 = n \end{cases} \quad j = 1, 2, \dots, p \quad (8)$$

SCV-GS 方法步骤如下:

步骤 1 变量扫描。

对于初始变量集 $\Lambda^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \dots, x_p^{(0)}\} = \{x_1, x_2, \dots, x_p\}$, 设 $k = 1$, 利用回归系数的显著性

检验,将最解释响应变量 y 的自变量挑出来。

1) 对任意 $x_j^{(k-1)} \in \Lambda^{(k-1)}$, 关于 y 做一元回归,可得回归系数 $\hat{\beta}_j^{(k-1)}$ 。

2) 计算 $T_j^{(k-1)}$, 其中 $T_j^{(k-1)}$ 为

$$T_j^{(k-1)} = \frac{\hat{\beta}_j^{(k-1)}}{S_n(\hat{\beta}_j^{(k-1)})} \sim N(0, 1) \quad n \rightarrow \infty \quad (9)$$

式中: $S_n(\hat{\beta}_j^{(k-1)})$ 为 $\hat{\beta}_j^{(k-1)}$ 的标准误。

3) 计算 $\eta = \arg \max_{j \in J} |T_j^{(k-1)}|$, 其中 $J = \{j | x_j^{(k-1)} \in \Lambda^{(k-1)}\}$ 。令 $z_k \leftarrow x_\eta^{(k-1)}$, 则 z_k 为第 k 个 GS 变量, 更新变量集 $\tilde{\Lambda} \leftarrow \Lambda^{(k-1)} \setminus x_\eta^{(k-1)}$ 。

步骤 2 变量聚类。

1) 将所有其余变量 $x_j^{(k-1)} \in \tilde{\Lambda}$ 关于步骤 1 挑出的变量 z_k 作 Gram-Schmidt 正交化:

$$x_j^{(k)} = x_j^{(k-1)} - \frac{z_k^T x_j^{(k-1)}}{z_k^T z_k} z_k \quad (10)$$

2) 根据信息分解准则式(6), 衡量变换后的变量 $x_j^{(k)}$ 与原始变量的相似性。不失一般性, 假设 $x_1, x_2, \dots, x_k, x_{k+1}$ 是对应于 $z_1, z_2, \dots, z_k, x_j^{(k)}$ 的原始变量。接下来判断 $x_j^{(k)}$ 所对应的原始变量能否被看作冗余变量。

需注意的是, $x_j^{(k)}$ 所对应的样本可能为零向量, 即不再承担任何信息, 因此与之对应的原始变量 x_{k+1} 可被看作冗余变量。另外, 若 $x_j^{(k)}$ 的方差近似于零, x_{k+1} 也可被看作冗余, 即认为该变量的信息几乎可由其他之前的变量代表, 可将之与上一步挑出的 GS 变量 z_k 所对应的原始变量聚成一组, 得到第 k 组变量。为了判断某变量能否被当作冗余, 即 $|\text{Cor}(x_{k+1}, x_j^{(k)})|$ 是否为零, 这里引入相关性检验, 采用 Fisher Z 变换^[21]:

$$S(x_{k+1}, x_j^{(k)}) = \frac{1}{2} \ln \left(\frac{1 + \text{Cor}(x_{k+1}, x_j^{(k)})}{1 - \text{Cor}(x_{k+1}, x_j^{(k)})} \right) \quad (11)$$

Gaussian 情形下可得如下准则:

准则 1 对于原假设 $H_0: \rho(x_{k+1}, x_j^{(k)}) = 0$ 和双边备择假设 $H_A: \rho(x_{k+1}, x_j^{(k)}) \neq 0$, 若 $\tilde{T} = (n-3)^{1/2} |S(x_{k+1}, x_j^{(k)})| > \Phi^{-1}(1 - \alpha/2)$, 则拒绝原假设。这里 ρ 为变量间的相关系数, α 为显著性水平, $\Phi(\cdot)$ 为标准正态累积分布函数。

3) 根据方差分解式(5)、式(6)和准则 1, 判断 $x_j^{(k)}$ 能否被看作冗余变量, 同时更新 $\tilde{\Lambda} \leftarrow \{x_j^{(k)} | \tilde{T}_j > \Phi^{-1}(1 - \alpha/2)\}$, 这里 $\tilde{T}_j$ 的定义参见准则 1。若 $\tilde{\Lambda} = \emptyset$, 转至步骤 3; 否则, 令 $\Lambda^{(k)} = \tilde{\Lambda}$, $k \leftarrow k+1$, 返回步骤 1。

步骤 3 在步骤 1 和步骤 2 迭代结束后, 提

取潜变量 LV。

这里潜变量的提取可考虑两种方法: 主成分分析和偏最小二乘回归。

注: 相关性检验显著性水平的确定。在变量聚类过程中, 将经过 Gram-Schmidt 变换后, 通过相关性检验 (即被看作无信息的变量, 满足 $\tilde{T}_j < \Phi^{-1}(1 - \alpha/2)$) 的那些变量 x_j , 与 GS 变量 z 所对应的原始变量聚为一类。但需注意, 数据不同, 相应的最佳显著性水平不同。为了求得每组数据的最佳阈值, 将 α 设定在范围 $\{10^{-h}, h = 1, 2, \dots, 10\}$, 然后从该范围中选取一个最佳值。易知 α 越小, 聚类所得变量类数目越小; α 越大, 聚类所得类数目越少。我们希望聚类所得的结果, 能够尽可能表述原始变量的组结构, 同时聚类所得的潜变量要尽可能解释响应变量 y , 因此考虑偏最小二乘准则^[22]。原始的偏最小二乘是为了找出已知变量集合 X 的投影方向, 使得该方向最大化所求成分的方差以及成分与响应变量 y 的相关性, 即求解问题为

$$\max \text{Var}(Xv) \cdot \text{Cor}^2(Xv, y) \quad \|v\| = 1 \quad (12)$$

式中: X 为 p 维变量集合; v 为 p 维投影向量。

受偏最小二乘准则启发, 这里提出一种新的同一性准则, 称之为“综合偏最小二乘准则”:

$$\sum_{k=1}^K \frac{\text{Var}(LV_k) \cdot \text{Cor}^2(LV_k, y)}{K} \quad (13)$$

式中: LV_k 为第 k 个变量组对应的第一偏最小二乘成分。于是将综合偏最小二乘准则(12)作为选取 α 的最优准则, 即在给定的 α 范围内, 选取使得综合偏最小二乘准则(12)最大化的那个 α 值。

2.3 基于 Gram-Schmidt 变换变量聚类的预测

在变量聚类算法结束后, 可以得到一系列的变量组。每个变量组 G_k 可由对应的潜变量 LV_k 来表示, 其中 LV_k 为第 k 组变量的线性组合, 即 $LV_k = X_k v_k$ 。这里 X_k 为变量组 G_k 中的 p_k 个变量对应的 $n \times p_k$ 维数据矩阵, 组合系数向量 $v_k = [v_{k1}, v_{k2}, \dots, v_{kp_k}]^T$ 为 X_k 的第一特征向量。由算法可知, LV_1 代表了与 y 最相关的一组变量, 最后一组 LV_K 则代表了最不解释 y 的一组变量。

因此, 得到有监督变量聚类的结果后, 接下来对各个潜变量采用向前扫描法^[4]。设最终扫描得到 $m (m \leq K)$ 个潜变量, 则最终回归模型为

$$\hat{y} = \sum_{k=1}^M LV_k \hat{\beta}_{LV_k} \quad (14)$$

由此还可同时得到回归模型关于原始变量的回归系数。

3 仿真研究

3.1 仿真设置

本文参考 Chen 和 Vigneau^[7]的仿真设置。设样本容量 $n = 50$, 变量维数 $p = 80$, 变量分成 5 个组: G_1, G_2, G_3, G_4, G_5 。每组变量由隐变量 Z_k ($k = 1, 2, \dots, 5$) 生成。各组中变量个数为 $|G_1| = 20, |G_2| = 20, |G_3| = 10, |G_4| = 10, |G_5| = 20$ 。隐变量 Z_1, Z_2, Z_3, Z_4 和 Z_5 之间的相关系数设置为

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.5 & 0.3 & 0 \\ 0 & 0.5 & 1 & 0.2 & 0 \\ 0 & 0.3 & 0.2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

变量 x_j 由 Z_k 生成, 即

$$x_j = \begin{cases} s_j Z_1 + \varepsilon_j & j = 1, 2, \dots, 20 (G_1) \\ s_j Z_2 + \varepsilon_j & j = 21, 22, \dots, 40 (G_2) \\ \vdots & \\ s_j Z_5 + \varepsilon_j & j = 61, 62, \dots, 80 (G_5) \end{cases}$$

其中: $s_j \in \{+1, -1\}; \varepsilon_j \sim N(0, 0.4^2)$ 。

考虑如下模型:

$$y = 6Z_1 + 1.5Z_2 + 2Z_3 + \varepsilon \quad \varepsilon \sim N(0, 1) \quad (15)$$

即响应变量 y 只与变量组 G_1, G_2 和 G_3 有关, 而且变量的重要性为 $G_1 > G_3 > G_2$ 。下面将仿真生成 100 组数据, 对每组数据施行 SCV-LV 方法和 SCV-GS 方法。

3.2 仿真结果

仿真考虑如下指标: 变量聚类结果的准确性, 聚类算法运行时间, 变量类对应的回归系数 β 的估计。

1) 变量聚类结果的准确性

在 100 次仿真结果中, SCV-LV 方法可以得到 100 次正确聚类结果; 而 SCV-GS 方法所得结果稍有差错。表 1 展示了施行 SCV-GS 方法后各变量在所得 5 个变量组中出现的频次。可以看出, 对于原仿真设置中的第 1 组变量 ($x_1 \sim x_{20}$), 100 次仿真实验中, 除了有一次未将变量 (x_{17}, x_{19}) 包含在内, 第 1 组聚类结果与真实设置一致。第 2 组聚类结果有 97 次聚类成功, 100 次聚类结

表 1 SCV-GS 方法变量聚类结果
Table 1 Variable clustering results by SCV-GS

变量	G_1	G_2	G_3	G_4	G_5	变量	G_1	G_2	G_3	G_4	G_5	变量	G_1	G_2	G_3	G_4	G_5
x_1	100	0	0	0	0	x_{28}	0	98	1	0	0	x_{55}	0	0	0	100	0
x_2	100	0	0	0	0	x_{29}	0	99	0	0	0	x_{56}	0	0	0	100	0
x_3	100	0	0	0	0	x_{30}	0	99	1	0	0	x_{57}	0	0	0	100	0
x_4	100	0	0	0	0	x_{31}	0	99	1	0	0	x_{58}	0	0	0	100	0
x_5	100	0	0	0	0	x_{32}	0	98	2	0	0	x_{59}	0	0	0	100	0
x_6	100	0	0	0	0	x_{33}	0	99	1	0	0	x_{60}	0	0	0	100	0
x_7	100	0	0	0	0	x_{34}	0	100	0	0	0	x_{61}	0	0	0	0	100
x_8	100	0	0	0	0	x_{35}	0	100	0	0	0	x_{62}	0	0	0	0	100
x_9	100	0	0	0	0	x_{36}	0	100	0	0	0	x_{63}	0	0	0	0	100
x_{10}	100	0	0	0	0	x_{37}	0	100	0	0	0	x_{64}	0	0	0	0	100
x_{11}	100	0	0	0	0	x_{38}	0	99	0	0	1	x_{65}	0	0	0	0	100
x_{12}	100	0	0	0	0	x_{39}	0	100	0	0	0	x_{66}	0	0	0	0	100
x_{13}	100	0	0	0	0	x_{40}	0	100	0	0	0	x_{67}	0	0	0	0	100
x_{14}	100	0	0	0	0	x_{41}	0	0	100	0	0	x_{68}	0	0	0	0	100
x_{15}	100	0	0	0	0	x_{42}	0	0	100	0	0	x_{69}	0	0	0	0	100
x_{16}	100	0	0	0	0	x_{43}	0	0	100	0	0	x_{70}	0	0	0	0	100
x_{17}	99	0	0	0	0	x_{44}	0	0	100	0	0	x_{71}	0	0	0	0	100
x_{18}	100	0	0	0	0	x_{45}	0	0	100	0	0	x_{72}	0	0	0	0	100
x_{19}	99	0	0	0	0	x_{46}	0	1	99	0	0	x_{73}	0	0	0	0	100
x_{20}	100	0	0	0	0	x_{47}	0	0	99	0	1	x_{74}	0	0	0	0	100
x_{21}	0	97	2	0	0	x_{48}	0	0	100	0	0	x_{75}	0	0	0	0	100
x_{22}	0	100	0	0	0	x_{49}	0	0	100	0	0	x_{76}	0	0	0	0	100
x_{23}	0	99	1	0	0	x_{50}	0	0	100	0	0	x_{77}	0	0	0	0	100
x_{24}	0	100	0	0	0	x_{51}	0	0	0	100	0	x_{78}	0	0	0	0	100
x_{25}	0	99	0	0	0	x_{52}	0	0	0	100	0	x_{79}	0	0	0	0	100
x_{26}	0	100	0	0	0	x_{53}	0	0	0	100	0	x_{80}	0	0	0	0	100
x_{27}	0	99	1	0	0	x_{54}	0	0	0	100	0						

果中出现原本来自第2组的变量落入第3组的情形,主要是由于仿真设置中第2组变量与第3组变量之间存在较强的相关性(相关系数为0.5)。类似的,第3组变量也出现两次误入第2类的情形。第4组变量及第5组变量在100次实验中,聚类结果完全正确。

2) 聚类算法运行时间

在100次仿真实验中,SCV-LV方法运行时间均在15 s以上;本文提出的SCV-GS算法运行时间不超过0.1 s,而且非常稳定。运行时间结果详细对比图参见图1。

3) 变量类对应的回归系数的估计

表2分别给出经SCV-LV方法和SCV-GS方法进行变量聚类后,由向前扫描法所得潜变量对应的回归系数相对于真值(6,1.5,2)的估计结果(包含估计值的偏差和标准差)。这里 $\hat{\beta}$ 为经变量聚类后,对挑选出的变量组求第一偏最小二乘成分后所得潜变量对应的回归系数。可以看出,在 β 估计值的大小和稳定性方面,SCV-LV方法和SCV-GS方法所得回归结果基本一致。

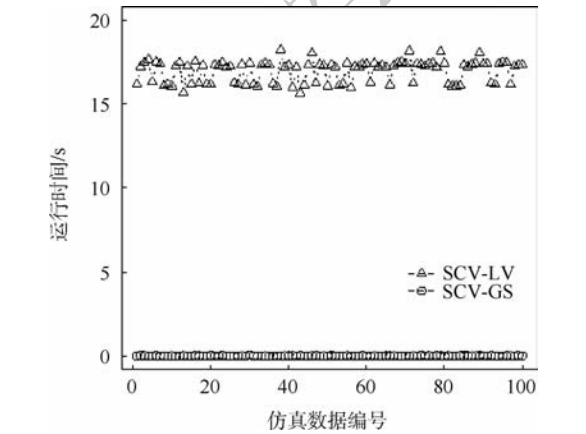


图1 SCV-LV与SCV-GS方法运行时间对比

Fig.1 Comparison of computation time (in seconds) between SCV-LV and SCV-GS

表2 SCV-LV与SCV-GS方法所得潜变量回归系数的估计结果

Table 2 Estimated regression coefficients by SCV-LV and SCV-GS as a function of latent variables			
方法	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
SCV-LV	-0.07(0.44)	-0.08(0.43)	0.03(0.28)
SCV-GS	-0.10(0.44)	-0.10(0.43)	0.02(0.28)

4 实例分析

为了进一步验证方法的有效性,下面分别考察SCV-LV及SCV-GS方法在实例数据上的表现。该数据集选自文献[23]中的教育数据集。

数据来自1996—1999年美国新闻“美国最佳大学”和美国教育部综合高等教育数据系统(IPEDS),共涉及94所高校。因变量Y为1996—1998年全国大学体育协会(NCAA)高等大学中的平均6年毕业率,19个解释变量包含学校情况、学生的个人情况和学生在校际育项目中的表现。分析目的是找出影响毕业率的重要决定因素。

对该数据集分别施行SCV-LV和SCV-GS方法,结果表明SCV-GS优于SCV-LV方法。表3给出了具体聚类结果。表3中指标K表示聚类后所得变量类的数目,“ $\text{Cor}(y, \hat{y})$ ”为施行算法后所得预测值与真值y之间的相关系数,第3列“变量”表示模型中涉及的变量。可以看出,SCV-LV方法可得到一个变量组,即将所有变量聚为一类;而SCV-GS方法可得到16个变量类,因此SCV-GS方法所得聚类结果可以给出更为详细的变量之间的联系信息。另外,尽管SCV-GS方法所得模型涉及的变量数目少(仅包含变量 x_2, x_3),但SCV-GS方法的预测效果要优于SCV-LV方法(预测值更接近真值)。2种方法均选出变量 x_2, x_3 ,表明“综合ACT(American College Test)分数”和“住校生比例”对毕业率有较强的积极影响。

表3 SCV-LV与SCV-GS方法作用于实例数据集所得结果

Table 3 Results on real dataset by SCV-LV and SCV-GS			
方法	K	$\text{Cor}(y, \hat{y})$	变量
SCV-LV	1	0.83	$x_1 \sim x_{19}$
SCV-GS	16	0.87	x_2, x_3

5 结论

在预测模型中,随着自变量维数的急剧式增大,模型的解释性和稳定性都将受到影响,有监督变量聚类有助于解决此问题。本文提出了一种基于Gram-Schmidt变换的新的有监督变量聚类方法,具有以下优点:

- 1) 避开层次聚类,且未采用围绕潜变量的思想。
- 2) 可实现较为准确的变量聚类。
- 3) 借助Gram-Schmidt变换,可以显著提高变量聚类方法的运行时间。
- 4) 结合偏最小二乘回归,提出新的聚合同一性准则。
- 5) 可扩展至其他模型,比如基于判别模型的有监督变量聚类,基于复杂数据的变量聚类等。

参考文献 (References)

- [1] TIBSHIRANI R. Regression shrinkage and selection via the lasso; A retrospective[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2011, 73(3): 273-282.
- [2] ZOU H, HASTIE T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67(2): 301-320.
- [3] FAN J Q, LV J C. Sure independence screening for ultrahigh dimensional feature space[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2008, 70(5): 849-911.
- [4] WANG H S. Forward regression for ultra-high dimensional variable screening[J]. Journal of the American Statistical Association, 2009, 104(488): 1512-1524.
- [5] ZOU H, HASTIE T, TIBSHIRANI R. Sparse principal component analysis[J]. Journal of Computational and Graphical Statistics, 2006, 15(2): 265-286.
- [6] CHUN H, KELES S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2010, 72(1): 3-25.
- [7] CHEN M K, VIGNEAU E. Supervised clustering of variables[J]. Advances in Data Analysis and Classification, 2016, 10(1): 85-101.
- [8] JOLLIFFE I T. Discarding variables in a principal component analysis. I: Artificial data[J]. Applied Statistics, 1972, 21(2): 160-173.
- [9] HASTIE T, TIBSHIRANI R, BOTSTEIN D, et al. Supervised harvesting of expression trees[J]. Genome Biology, 2001, 2(1): research0003-1.
- [10] VIGNEAU E, QANNARI E. Clustering of variables around latent components[J]. Communications in Statistics-Simulation and Computation, 2003, 32(4): 1131-1150.
- [11] VIGNEAU E, CHEN M, QANNARI E M. ClustVarLV: An R package for the clustering of variables around latent variables[J]. The R Journal, 2015, 7(2): 134-148.
- [12] VIGNEAU E. Segmentation of a panel of consumers with missing data[J]. Food Quality and Preference, 2018, 67: 10-17.
- [13] CARIOU V, QANNARI E M, RUTLEDGE D N, et al. ComDim: From multiblock data analysis to path modeling[J]. Food Quality and Preference, 2018, 67: 27-34.
- [14] BJÖRCK Å. Numerics of gram-schmidt orthogonalization[J]. Linear Algebra and Its Applications, 1994, 197-198: 297-316.
- [15] LEON S J, BJÖRCK A, GANDER W. Gram-Schmidt orthogonalization; 100 years and more[J]. Numerical Linear Algebra with Applications, 2013, 20: 492-532.
- [16] CHEN S, BILLINGS S A, LUO W. Orthogonal least squares methods and their application to non-linear system identification[J]. International Journal of Control, 1989, 50(5): 1873-1896.
- [17] STOPPIGLIA H, DREYFUS G, DUBOIS R, et al. Ranking a random feature for variable and feature selection[J]. Journal of Machine Learning Research, 2003, 3: 1399-1414.
- [18] 王惠文, 仪彬, 叶明. 基于主基底分析的变量筛选[J]. 北京航空航天大学学报, 2008, 34(11): 1288-1291.
- WANG H W, YI B, YE M. Variable selection based on principal basis analysis[J]. Journal of Beijing University of Aeronautics and Astronautics, 2008, 34(11): 1288-1291 (in Chinese).
- [19] 王惠文, 陈梅玲, SAPORTA G. 基于 Gram-Schmidt 过程的判别变量筛选方法[J]. 北京航空航天大学学报, 2011, 37(8): 958-961.
- WANG H W, CHEN M L, SAPORTA G. Variable selection in discriminant analysis based on Gram-Schmidt process[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(8): 958-961 (in Chinese).
- [20] LIU R P, WANG H W, WANG S S. Functional variable selection via Gram-Schmidt orthogonalization for multiple functional linear regression[J]. Journal of Statistical Computation and Simulation, 2018, 88(18): 3664-3680.
- [21] FISHER R. On the probable error of a coefficient of correlation deduced from a small sample[J]. Metron, 1921, 1(4): 3-32.
- [22] FRANK L E, FRIEDMAN J H. A statistical view of some chemometrics regression tools[J]. Technometrics, 1993, 35(2): 109-135.
- [23] MANGOLD W D, BEAN L, ADAMS D. The impact of intercollegiate athletics on graduation rates among major ncaa division I universities: Implications for college persistence theory and practice[J]. Journal of Higher Education, 2003, 74(5): 540-562.

作者简介:

刘瑞平 女, 博士研究生。主要研究方向: 高维数据的降维方法及应用。

王惠文 女, 博士, 教授, 博士生导师。主要研究方向: 经济管理中复杂数据统计分析的理论、方法与应用。

王珊珊 女, 博士, 助理教授, 硕士生导师。主要研究方向: 高维复杂数据分析、半参数统计、机器学习、统计算法及应用。

Supervised clustering of variables based on Gram-Schmidt transformation

LIU Ruiping¹, WANG Huiwen^{1,2}, WANG Shanshan^{1,3,*}

(1. School of Economics and Management, Beihang University, Beijing 100083, China;

2. Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100083, China;

3. Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing 100083, China)

Abstract: In order to study the dimension reduction method of high-dimensional data based on regression model further, and the supervised clustering of variables algorithm based on Gram-Schmidt transformation (SCV-GS) is proposed. SCV-GS uses the key variables selected in turn by the variable screening idea as the clustering center, which is different from the hierarchical variable clustering around latent variables. High correlation among variables is processed based on Gram-Schmidt transformation and the clustering results are obtained. At the same time, combined with the concept of partial least squares, a new criterion for “homogeneity” is proposed to select the optimal clustering parameters. SCV-GS can not only get the variable clustering results quickly, but also identify the most relevant variable groups and in what kind of structure the variables work to influence the response variable. Simulation results show that the calculation speed is significantly improved by SCV-GS, and the estimated regression coefficients corresponding to the latent variables are consistent with the comparison method. Real data analysis shows that SCV-GS performs better in interpretation and prediction.

Keywords: dimension reduction; variable clustering; regression; high correlation; Gram-Schmidt transformation

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0019

PMMWI 与 VI 优势互补的人体隐蔽违禁物检测与定位



赵国, 秦世引*

(北京航空航天大学 自动化科学与电气工程学院, 北京 100083)

摘 要: 根据公共场所人体安检的性能要求和技术需求, 将被动毫米波成像(PMMWI)的可透视成像性能优势与可见光成像(VI)的细节高分辨性能优势相结合, 提出一种基于 PMMWI 与 VI 优势互补的人体隐蔽违禁物检测与定位算法。首先, 提出一种基于低层特征融合的改进 U-Net 以增强深度神经网络(DNN)对 PMMWI 中弱小目标轮廓的敏感度, 提高 PMMWI 中人体轮廓和隐蔽违禁物的分割精度, 并同时实现 VI 中人体轮廓的像素级分割; 然后, 在 PMMWI 和 VI 中的人体轮廓分割基础上, 通过基于人体轮廓的尺度变换与滑动适配实现 PMMWI 人体轮廓和 VI 人体轮廓的良好配准, 根据配准结果实现单帧图像中人体隐蔽违禁物的高效检测; 最后, 通过序列图像检测结果的对比融合与优化决策给出隐蔽违禁物的定位结果。一系列综合实验与对比分析结果, 验证了提出的人体隐蔽违禁物检测与定位算法的性能优势。

关键词: 毫米波安检; 被动毫米波成像(PMMWI); 人体轮廓分割; 深度学习; 深度神经网络(DNN); 隐蔽违禁物检测与定位

中图分类号: TP391

文献标识码: A

文章编号: 1001-5965(2019)10-2011-15

在人流密集的公共场所及交通站点实施人体安全检查对保障中国社会稳定和人民生命财产安全具有重大意义。目前, 针对地铁、火车站和大型场馆等人流密集区域的安检技术, 主要包括立式人证核验机、金属安检门、手持金属探测器、液体检测仪和通道式 X 射线安检机^[1]。此外, 人工安检对受检人员的触摸不可避免的会引发个人隐私问题, 因此更加高效的毫米波成像技术得到了深入研究和广泛应用^[2]。由于毫米波不会对人体造成辐射损害, 且兼具成像速度快等优势, 现已得到大力推广和使用。2018 年 6 月, 中国民用航空局颁布相关安检设备标准文件大力支持和推进毫

米波人体安检设备在民航安检中的应用。2018 年 11 月 5 日, 首届中国国际进口博览会在上海开幕, 多套基于被动太赫兹成像技术的人体安检系统成功部署并应用, 实现了整个安检过程中的受检人员快速无停留通过以及无接触安检。

针对人体安检中隐蔽违禁物的检测和识别, 根据毫米波成像体制的不同, 可供分析的成像数据分为主动毫米波成像(Active Millimeter Wave Imaging, AMMWI)和被动毫米波成像(Passive Millimeter Wave Imaging, PMMWI)。由于 AMMWI 中人体隐蔽物图像细节清晰, AMMWI 安检系统在国内机场和重要会议等场合得到广泛应

收稿日期: 2019-01-17; 录用日期: 2019-05-28; 网络出版时间: 2019-06-18 13:38

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190617.0922.001.html

基金项目: 国家自然科学基金(61731001)

* 通信作者: E-mail: qsy@buaa.edu.cn

引用格式: 赵国, 秦世引. PMMWI 与 VI 优势互补的人体隐蔽违禁物检测与定位[J]. 北京航空航天大学学报, 2019, 45(10): 2011-2025. ZHAO G, QIN S Y. Detection and localization of concealed forbidden objects on human body based on complementary advantages of PMMWI and VI [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2011-2025 (in Chinese).

用^[3-4]。针对 AMMWI 的人体隐蔽违禁物自动检测, Du 等^[5]提出一种基于快速小波变换的隐蔽违禁物检测方法,其利用快速小波变换实现隐蔽违禁物的边缘检测,从而达到对隐蔽违禁物的检测。与利用传统图像处理算法的隐蔽物检测算法相比,随着深度神经网络(Deep Neural Networks, DNN)在目标检测上的深入研究和应用,与 Faster RCNN (Faster Region-based Convolutional Neural Networks)^[6]检测算法的思路类似,文献[7]提出一种基于候选框和 DNN 的人体隐蔽违禁物的快速检测方法。

虽然 AMMWI 细节信息丰富,但是由于成像速度较慢,一般需要受检人员站立数秒等待系统完成 AMMWI,而 PMMWI 速度较快,能够满足大流量安检场合的应用需求。因此,基于 PMMWI 的枪支等危险隐蔽物检测也逐渐得到了广泛研究和应用^[8-10]。Yeom 等^[9]设计了一种面向室外实时隐蔽违禁物检测的 PMMWI 系统,并且提出一种基于两阶段图像分割的人体隐蔽违禁物检测算法。文献[10]则是在人体隐蔽违禁物分割基础上,通过计算不同隐蔽违禁物形状特征间的欧氏距离实现了枪支和金属板两类违禁物的识别。针对毫米波透视成像引起的人体隐私问题,周健等^[11]提出一种基于 AMMWI 的男女性别识别算法,以针对不同性别人群采取不同隐私部位保护措施,但是毫米波图像依然是公开显示的,仍然无法减少公众对人体隐私的担忧。

传统的基于毫米波图像的人体隐蔽违禁物检测一般是单独利用 AMMWI 或 PMMWI 进行违禁物检测,人体隐蔽违禁物的检测与识别将在一套主被动毫米波复合的安检系统上实现,该系统主要利用 PMMWI 和可见光成像(Visible Imaging, VI)进行人体隐蔽违禁物的检测和定位。在准确定位基础上,利用 AMMWI 系统对该定位区域进行小范围快速扫描成像,最终通过分析定位区域的 AMMWI 完成人体隐蔽违禁物的属性识别。

为了解决人体隐私和基于单一 PMMWI 的违禁物检测结果不准确问题,本文提出一种基于 PMMWI 和 VI 优势互补的隐蔽违禁物检测与定位算法。与文献[11]中采用对人体隐私部位进行遮挡处理的方法不同,本文将在后台完成人体隐蔽违禁物的检测与定位,并在 PMMWI 与 VI 配准基础上直接将人体隐蔽违禁物检测结果显示在 VI 中,从根本上解决毫米波图像的人体隐私泄露问题。PMMWI 与 VI 技术原理上的差异性,导致其视场角、视场范围都存在明显不同。因此,为了

实现 PMMWI 和 VI 的配准,本文首先提出一种改进 U-Net 网络模型以实现 PMMWI 人体轮廓的像素级分割,并通过与经典 DNN 图像分割模型,比如 FCN (Fully Convolutional Networks)^[12]、Seg-Net^[13]和 U-Net^[14]等进行人体轮廓分割对比实验,验证了本文提出的 DNN 模型的性能优势。其次,在人体轮廓分割基础上,通过尺度变换与滑动配准方法实现 VI 和 PMMWI 人体轮廓的良好配准。然后,利用 VI 和 PMMWI 优势互补特性,实现单帧图像的人体隐蔽违禁物检测与定位。最后,通过多帧检测结果的对比融合决策与复合定位,完成对人体隐蔽违禁物的精确定位。

1 人体隐蔽违禁物检测技术挑战与关键性问题

1.1 常规检测方式的技术挑战

毫米波成像技术可分为 PMMWI 和 AMMWI 技术,目前常规的毫米波检测方法是采用单一的技术方案,图 1(a)所示为 AMMWI 安检设备,虽然其成像质量好,对隐蔽违禁物的透视成像清晰度高,但其成像速度较慢,一般需要受检人员静止站立数秒时间;而 PMMWI 速度较快,可以采用如图 1(b)所示的通道式设计,使受检人员无停留通过安检通道^[15],但其成像效果欠佳,对违禁物形状特征的分辨率较低,甚至模糊不清。

图 2 所示为本文所采用的主/被动毫米波成像相结合互补的安检方案,其使 PMMWI 速度快和 AMMWI 清晰的优势得以综合利用。由于 PMMWI 成像速度快,可实现人体目标的实时透视成像,为人体衣物中的隐蔽违禁物的高效检测提供可靠的图像信息。在 PMMWI 初检的基础上,为了能够增强检测和定位的精度并有效排除虚警,本文将 PMMWI 透视成像和 VI 人体轮廓清晰成像的优势相结合,提出一种 PMMWI 与 VI 优势互补的人体隐蔽违禁物检测与定位算法。这将为 AMMWI 的进一步定向扫描并进行违禁物检测

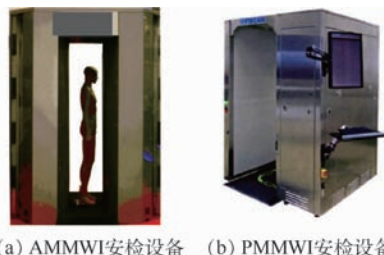


图 1 主/被动毫米波成像安检设备示例

Fig. 1 Illustration for AMMWI and PMMWI security check equipment

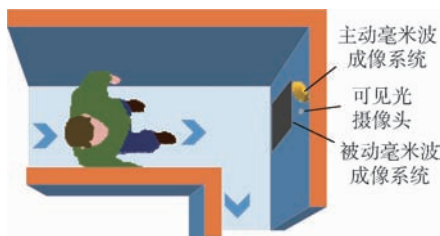


图 2 主/被动毫米波成像结合互补的安检方案

Fig. 2 Security check scheme based on combination and complementation of AMMWI and PMMWI

确认与分类识别奠定基础。

1.2 需要解决的关键性问题

由于采用主/被动毫米波成像结合互补的安检方案,相比常规采用单一成像技术的检测方式,将面临更多复杂的问题。本文主要解决的是基于 PMMWI 和 VI 的隐蔽违禁物检测和定位问题,其亟需解决的关键性问题如下: PMMWI 与 VI 人体轮廓的像素级分割; PMMWI 与 VI 中人体轮廓的配准;面向 PMMWI 与 VI 优势互补的人体隐蔽违禁物检测;基于序列图像检测结果对比融合优化决策和复合定位。

其中,首先需要解决的关键性问题是通过分析 PMMWI 与 VI 人体目标成像特性,实现人体轮廓的像素级分割,其分割结果将直接影响人体隐蔽违禁物的检测与 PMMWI 和 VI 人体轮廓的良好配准,因此 PMMWI 与 VI 人体轮廓的分割是本文需要解决的核心问题。

2 基于优势互补的检测/定位方案与系统架构

为了解决隐蔽违禁物检测与定位中的关键性问题,本节将在 PMMWI 与 VI 特性分析基础上,给出基于优势互补的单帧图像检测与定位方案,以及整个算法的系统架构。

2.1 PMMWI 和 VI 的特性分析

图 3(a) 所示为室内安检环境实拍得到的 PMMWI,室内背景温度的变化导致背景噪声也较大, PMMWI 中的人体轮廓明显不如图 3(b) 所示同一时刻的 VI 清晰,甚至腿部区域已经完全不可辨。

图 3 中矩形框标记区域分别为 PMMWI 与 VI 中人体手臂与躯干形成的虚警区域,而箭头所指向的黑色区域则是隐蔽违禁物区域。由图 3(a) 可以看出,虚警区域和违禁物区域存在相似的图像像素值,并且与背景区域的像素值也相似,仅从 PMMWI 中很难分辨出人体目标区域内的黑色区

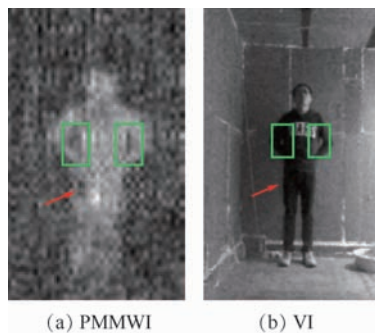


图 3 PMMWI 和 VI 示例

Fig. 3 Samples of PMMWI and VI

域是否为隐蔽违禁物。而在 VI 中可以明显看出是背景区域,因此可以利用 PMMWI 与 VI 优势互补特性实现隐蔽违禁物的检测与定位。

2.2 基于单帧图像优势互补的检测/定位方案

在 PMMWI 和 VI 的特性分析基础上,本文提出一种基于 PMMWI 和 VI 优势互补的隐蔽违禁物检测与定位算法,如图 4 所示,其包括 PMMWI 和 VI 人体轮廓提取、PMMWI 和 VI 的人体轮廓配准和基于人体轮廓配准的隐蔽违禁物的检测与定位等 3 个主要模块。

PMMWI 和 VI 人体轮廓提取的目的是将人体目标从背景中分割出来,并同时实现 PMMWI 中隐蔽违禁物的分割。需要特别说明的是,在 PMMWI 中违禁物区域和人体肢体动作造成的虚警区域均设为非人体区域。此外,由于 PMMWI 与 VI 成像原理的不同,人体轮廓存在较大差异,需要实现 PMMWI 和 VI 中人体轮廓的配准,以保证虚警目标得到有效排除,最终实现人体隐蔽违禁物的检测。

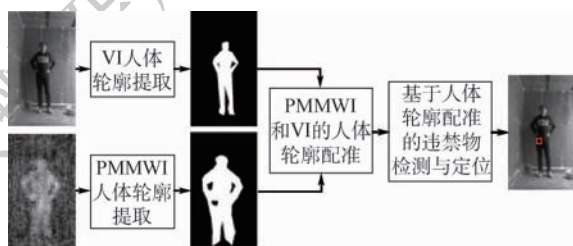


图 4 基于 PMMWI 和 VI 优势互补的检测与定位方案

Fig. 4 Detection and localization scheme based on complementary advantage of PMMWI and VI

2.3 核心模块及整体系统的组织架构

针对实际应用中单帧图像的隐蔽违禁物检测与定位存在较高的虚警率和误检率,本文在单帧违禁物检测结果上,提出一种多帧违禁物检测结果对比融合的检测方法,其整体组织结构如图 5 所示,包括基于单帧图像优势互补的检测和基于序列图像检测结果对比融合优化决策两大主体

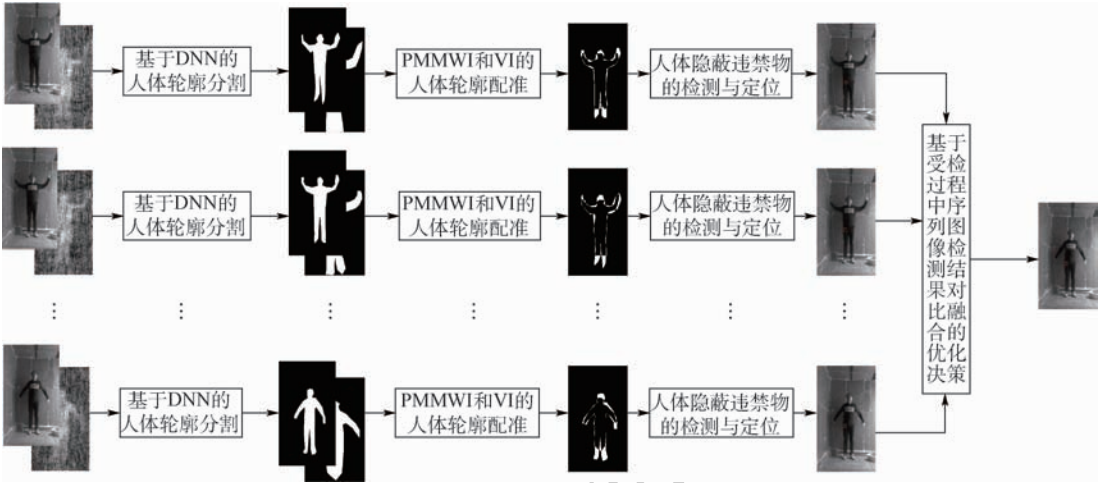


图5 基于 PMMWI 和 VI 优势互补的隐蔽违禁物检测与定位系统的组织结构与核心模块

Fig.5 Organization structure and core modules of concealed forbidden object detection and localization system based on complementary advantage of PMMWI and VI

结构。其中基于单帧图像的违禁物检测方案如图4所示,其详细分析见 2.2 节。在单帧图像的隐蔽违禁物检测基础上,通过对连续多帧的检测结果进行对比融合分析,以提高人体隐蔽违禁物定位精度,减少单帧图像检测结果不稳定性。

3 基于改进 U-Net 的人体轮廓高精度分割

3.1 U-Net 网络及其性能优势

U-Net 是由 Ronneberger 等^[14]为解决医学图像分割而提出的一种图像分割模型,并获得 2015 年 ISBI(International Symposium on Biomedical Imaging)细胞跟踪挑战赛冠军。如图 6 所示^[14],U-Net 网络结构呈现出明显的“U”型特征,这也是称其为 U-Net 的原因,同时可以看出 U-Net 具有结构对称性。U-Net 分为编码部分和解码部分,左侧为编码部分,主要是针对输入图像进行特征提取,右侧解码部分是由提取出的特征进行上采

样逐渐解码成与输入图像相同大小的二值掩码图像,并且,在每次上采样之后,就和编码部分对应的特征图进行拼接。

由于 U-Net 引入的跨越连接操作,使得在反向传播过程中能够较快地对整个网络权重进行更新,因此具有非常优秀的易训练特性。同时,跨越连接使得高分辨图像得到的特征能够将更多的细节信息传输到上采样层,实现更加精确的图像分割。此外,U-Net 能够在较小的数据集上实现高精度的图像分割性能。

由于 U-Net 具有以上性能优势,在 Kaggle 图像分割竞赛中得到了广泛应用,同时越来越多的研究学者对 U-Net 进行了改进,比如 TernaNet 网络使用预训练 VGG11 结构代替 U-Net 中特征提取部分^[16],获得了更加优异的分割性能,并且获得 2017 年 Carvana Image Masking Challenge 挑战赛冠军。

3.2 U-Net 模型的低层特征卷积融合与整体性能改进

3.2.1 低层特征的 1×1 卷积融合

Lin 等^[17]在提出的 NIN(Network in Network)图像分类模型中首次使用 1×1 卷积,1×1 卷积得到了大量研究人员的关注,当 1×1 卷积与 3×3 卷积或 5×5 卷积叠加一起使用时,就会产生更加复杂的特征抽象过程,并且能够将模型参数降到原始模型的十分之一甚至更少,为深度学习应用扩展到移动终端提供了可能性。目前,1×1 卷积已经成为 DNN 中常见的结构组件之一,包括 GoogLeNet^[18-19]、ResNet^[20]和 ResXNet^[21]等优秀 DNN 结构中大量使用 1×1 卷积实现了整体网络

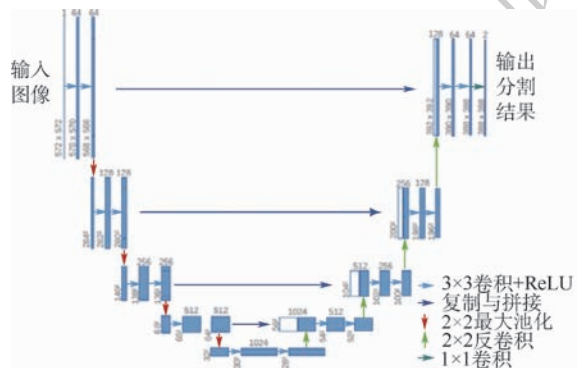


图6 U-Net 结构^[14]

Fig.6 U-Net architecture^[14]

性能的提升。

根据 1×1 卷积原理,其具有如下 2 个显著特点:①实现特征图跨通道间的特征交互和融合,增强孤立特征通道之间的通信;②实现特征图的降维/升维。利用 1×1 卷积可以对多通道的特征进行卷积融合,同时实现特征图维度的优化,比如降维可以减少网络计算量,提高运算速度;升维可以增加提取特征维度,提高网络预测能力。

由于 PMMWI 中人体目标内隐蔽违禁物区域较小或成像效果较差,导致 U-Net 网络分割性能表现不佳,本文提出一种基于 1×1 卷积的低层特征卷积

融合结构以改进整体网络分割性能,与原始 U-Net 网络直接将编码与解码部分特征图进行拼接不同,本文利用 1×1 卷积对编码部分的高维特征图进行跨通道卷积融合,再与解码部分进行拼接。

如图 7 所示,假设编码部分的特征图为 $64 \times 64 \times 128$,使用 $1 \times 1 \times 128$ 的卷积核与其卷积运算,可得大小为 $64 \times 64 \times 1$ 的特征图,图 7 使用 256 个 $1 \times 1 \times 128$ 卷积核对编码部分特征图进行卷积融合,可得大小为 $64 \times 64 \times 256$ 的特征图,其与解码部分对应特征图大小一致,因此可与解码部分特征图直接进行通道维度上的拼接。

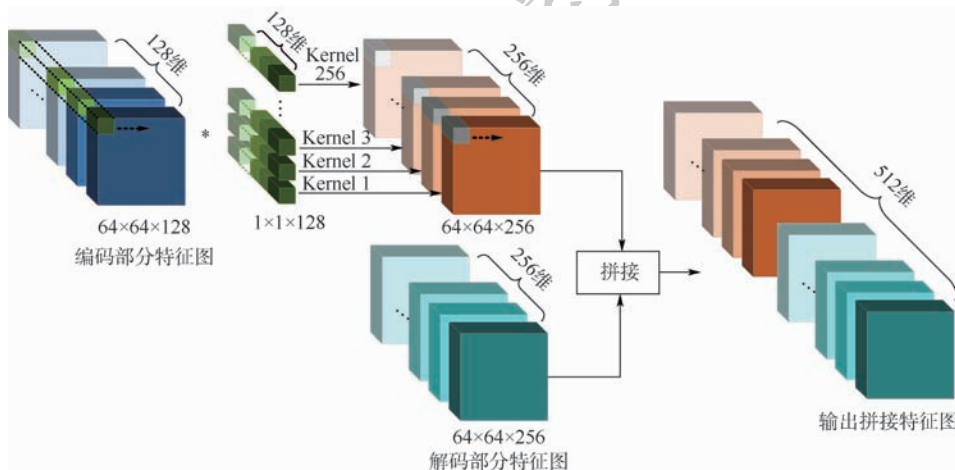


图 7 1×1 卷积特征的拼接

Fig. 7 Concatenation of features by 1×1 convolution

3.2.2 改进的 U-Net 网络结构与性能优势

本节将在 U-Net 图像分割模型基础上,提出一种基于 1×1 低层特征卷积融合图像分割模型,以实现 PMMWI 中人体轮廓的像素级分割。需要说明的是,本文提出的 PMMWI 人体轮廓分割模型结构并不是在原始 U-Net 结构上直接增加低层特征融合结构,只是整体网络结构与 U-Net 具有类似的编解码框架。

如图 8 所示,针对 PMMWI 人体轮廓的分割,本文提出一种基于低层特征融合的 U-Net 模型,称之为 LFF-UNet (Low-layer Feature Fusion based U-Net) 模型。本文提出的 LFF-UNet 模型共有 20 层卷积 (Conv) 层、5 层 BN (Batch Normalization) 层、4 层 Pooling 层和 4 层转置卷积 (Deconv) 层,不同的层以不同的颜色标记加以区分,并且每一层中的数字表示当前层输出的特征图尺寸大小。

此外,低层特征将经过 4 个低层特征融合块 (Low-layer Feature Fusion Block, LFFB) 进行卷积融合后,再与上采样层特征图拼接。由于低层

特征图具有更好的图像细节信息,通过对编码部分的低层特征图进行跨通道卷积融合,可提高网络对图像细节的特征提取能力。尤其在室内毫米波安检中,受毫米波成像器件和现场环境等多种因素影响,PMMWI 中人体隐蔽违禁物会出现成像较弱或轮廓模糊情况。如果编码部分无法有效提取弱小目标的轮廓特征,则会出现漏分割情况,而通过对低层特征图的卷积融合可增加网络对 PMMWI 中弱小目标轮廓的敏感度,减少背景噪声对弱小目标的干扰,提高毫米波安检中人体隐蔽违禁物的分割精度。

如图 9 所示,针对 PMMWI 人体轮廓的分割,本文提出 3 种 LFFB 结构,主要区别是 1×1 卷积层数的差异。以图 9(a) 所示的一层 1×1 卷积 LFFB 结构为例, W 、 H 和 C 分别表示当前层输出特征图的宽、高和通道维度,图 8 中的所有 LFFB 全部为一层 1×1 卷积,本文以 LFF1-UNet 表示当前网络结构。类似地,LFF2-UNet 表示 LFFB 模块为两层 1×1 卷积,LFF3-UNet 表示 LFFB 模块为三层 1×1 卷积。

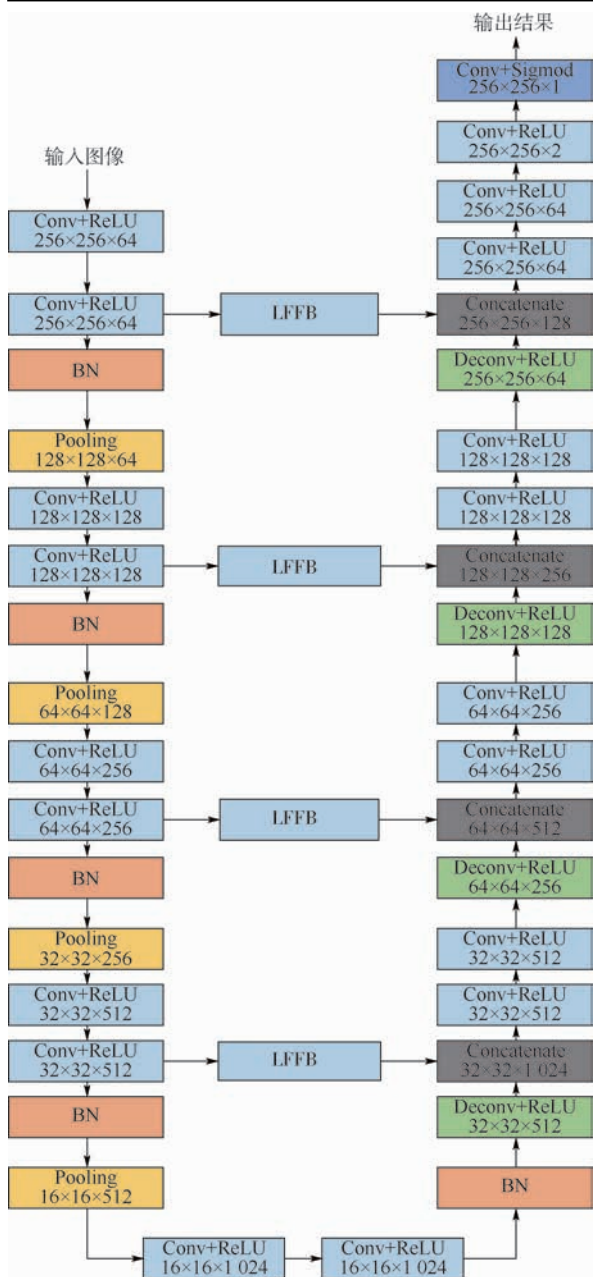


图8 基于低层特征融合的 LFF-UNet 网络结构

Fig. 8 LFF-UNet architecture based on low-layer

feature fusion

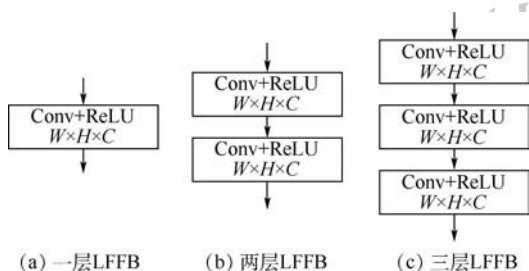


图9 本文提出的 3 种 LFFB 结构

Fig. 9 Three types of LFFB structure proposed in our work

3.3 基于 LFF-UNet 的人体轮廓高精度分割

3.3.1 样本数据的采集与标注

为了验证本文提出的 LFF-UNet 结构的性能

优势,利用现有毫米波安检系统进行真实数据集采集,全部样本数据集共 386 张图像,其中 PMMWI 和 VI 各 193 张,并且为了后续隐蔽违禁物的检测与定位,同时进行 PMMWI 和 VI 的样本数据的采集与标注。需要特别说明的是,LFF-UNet 主要针对 PMMWI 的人体轮廓分割,而 VI 人体轮廓分割则直接使用 U-Net 网络^[14]。

图 10 为采集的 PMMWI 和 VI 样本数据集和人工标注结果图像,需要特别说明的是,针对 PMMWI 人体轮廓的标注,由于肢体动作导致的人体轮廓内的空洞区域与隐蔽违禁物呈现出相同的成像特性,因此 PMMWI 的人工标注结果排除了肢体动作造成的空洞区域和隐蔽违禁物。

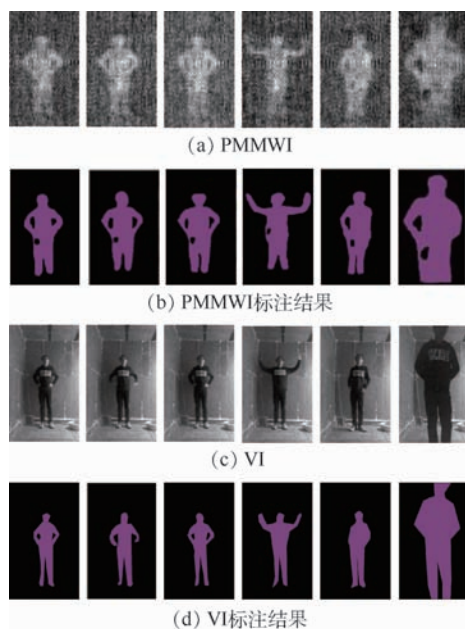


图 10 PMMWI 和 VI 样本数据集与人工标注图像

Fig. 10 Dataset and manually annotated image of PMMWI and VI samples

3.3.2 改进 U-Net 网络的离线监督训练

为了能够有效地利用 PMMWI 数据集对 LFF-UNet 进行训练和模型评估,PMMWI 数据集集中的 160 张图像归为训练集,剩余 33 张图像为测试集。本文将利用训练集对 LFF-UNet 进行离线监督训练,测试集作为对离线监督训练后的 DNN 模型实施测试与评估。在离线监督训练过程中,采用交叉熵函数为损失函数,借助于人工标记图像,利用自适应矩估计 (Adaptive moment estimation, Adam)^[22] 优化方法对网络模型进行参数更新,交叉熵的解析表达式为

$$L = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1)$$

式中: n 为图像像素点的个数; $\hat{y}_i$ 为第 i 个像素的模型预测值; y_i 为第 i 个像素的类别标签。同时,为了与经典 DNN 图像分割模型进行对比分析,进一步验证本文提出的 LFF-UNet 分割性能优势,在 PMMWI 数据集上分别进行了 FCN-8^[12]、SegNet^[13]、U-Net^[14]、LFF1-UNet、LFF2-UNet 和 LFF3-UNet 等 6 种模型的训练,6 种模型的性能进化曲线如图 11 所示。与经典 U-Net 模型相比,本文提出的 3 种 LFF-UNet 模型在训练过程中,交叉熵损失下降更快,而 FCN-8 模型训练时间比其他模型更长,训练迭代次数超过 5 000 余次才开始收敛。

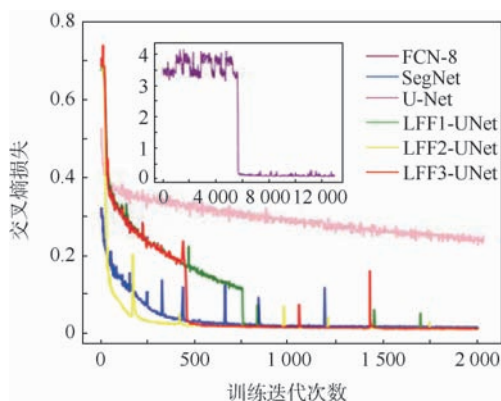


图 11 PMMWI 中人体轮廓分割 DNN 模型性能进化曲线
Fig. 11 Performance evolution curves of human contour segmentation DNN models in PMMWI

3.3.3 基于 LFF-UNet 的人体轮廓分割算法

在毫米波安检中,PMMWI 中人体轮廓的分割精度不仅影响 PMMWI 与 VI 人体轮廓配准精度,更直接影响隐蔽违禁物的检测与定位。借助于深度学习理论,利用 1×1 卷积的跨通道特征融合特性,设计出分割性能更优的 LFF-UNet 分割网络,可以实现 PMMWI 中人体轮廓的高精度分割。

图 12 所示为基于 LFF-UNet 的 PMMWI 人体轮廓分割算法。首先将毫米波安检系统采集的 PMMWI 训练数据集输入 LFF-UNet,利用交叉熵测度对网络输出结果与示教信息进行监控与评判;然后由参数调理算法对 DNN 进行优化,当交叉熵损失下降到最小且不再变化时,即完成 PMMWI 的人体轮廓分割 LFF-UNet 模型的训练。在实际应用过程中,将毫米波安检设备获取的 PMMWI 图像输入训练好的 LFF-UNet 模型,即可生成对应大小的人体轮廓分割结果。

3.3.4 分割性能的测评与比较分析

为了对 DNN 的图像分割性能进行评估,首先将训练好的 DNN 模型在 PMMWI 测试集上进行实验,然后将分割结果与人工标注结果对比分析,以综合评判 DNN 图像分割网络的性能优劣。不

同 DNN 模型的 PMMWI 人体轮廓分割对比结果如图 13 所示,基于 FCN-8 和 SegNet 的人体轮廓分割精度较差,隐蔽违禁物的分割也不够准确,U-Net 则在人体轮廓上分割精度较好,但是隐蔽违禁物区域分割效果较差。通过对低层特征的卷积融合优化,能更好地处理细节特征,在人体轮廓分割精度不变的前提下,显著提升了隐蔽违禁物的分割,尤其 LFF3-UNet 在人体轮廓和隐蔽违禁物分割效果上达到最优。

为了能够从量化指标的角度对比分析各 DNN 网络模型的分割性能,本文将对分割结果进

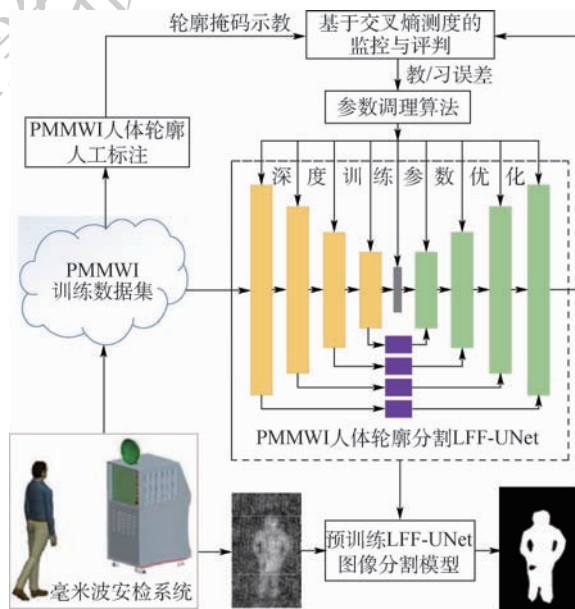


图 12 基于 LFF-UNet 的 PMMWI 人体轮廓分割算法
Fig. 12 Segmentation algorithm of human contour in PMMWI based on LFF-UNet architecture



图 13 不同 DNN 模型 PMMWI 人体轮廓分割对比
Fig. 13 Comparison of human contour segmentation in PMMWI among different DNN models

行交并比(Intersection-over-Union,IoU)计算:

$$\text{IoU} = \frac{\text{area}(\text{RoI}_p \cap \text{RoI}_c)}{\text{area}(\text{RoI}_p \cup \text{RoI}_c)}$$

(2)

式中:RoI_p表示DNN生成的PMMWI人体轮廓预测分割结果;RoI_c表示人工标注的人体轮廓区域;area表示图像面积大小。IoU值域范围为[0,1],当IoU值越接近于1时,表示DNN模型的分割性能越好。图14所示为不同DNN模型的PMMWI人体轮廓分割IoU对比结果,相比于经典的分割模型,本文提出的3种LFF-UNet网络在测试集上的IoU得分都较高。

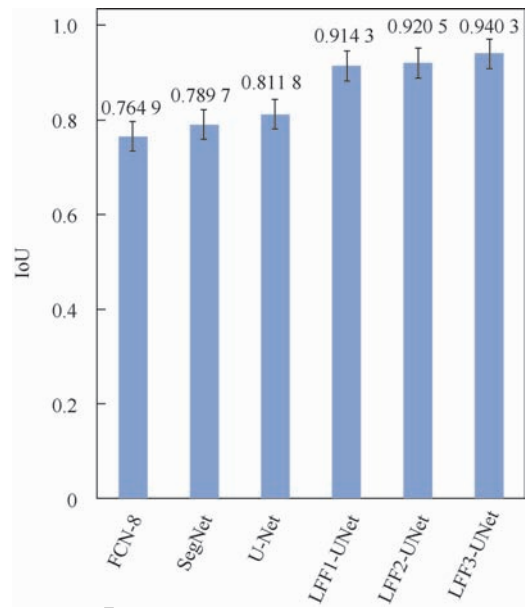


图14 不同DNN模型分割结果IoU对比
Fig. 14 Comparison of IoU of segmentation result among different DNN models

4 基于单帧受检图像的隐蔽违禁物检测和定位

4.1 PMMWI和VI差异性分析

在毫米波安检过程中,受检人员会步行通过安检通道,受检人员与PMMWI和VI设备的距离也是随之变化的。图15所示为不同距离下的VI和PMMWI人体轮廓分割对比结果,由于毫米波成像与VI技术的差异性,当受检人员与成像设备处于不同距离时,PMMWI和VI中人体轮廓尺寸大小和图像视场角均有较大不同。因此,PMMWI和VI的人体轮廓配准是必不可少的。

4.2 面向优势特征互补关系的尺度变换人体轮廓配准算法

为了有效利用PMMWI和VI优势互补特性,本文提出一种基于人体轮廓的尺度变换与滑动适

配算法,以实现人体轮廓的良好配准。

图16为PMMWI与VI人体轮廓的尺度变换与滑动适算法。首先利用预训练好的U-Net模型实现VI人体轮廓提取,并对人体轮廓内的空洞填充处理。同时,利用训练好的LFF3-UNet模型对PMMWI进行人体轮廓提取,由于人体轮廓在PMMWI中的尺度和位置与VI存在较大的差异,因此首先需要对PMMWI进行尺度变换,以保证PMMWI和VI中人体轮廓尺度相同,然后再进行空洞填充。最后,以VI为基准图像,将PMMWI中人体轮廓进行滑动适配和萃取。需要说明的是,上述步骤中空洞填充的目的是避免在滑动适配过程中,过大的内部空洞很容易导致适配的不准确。

PMMWI与VI人体轮廓的尺度变换与滑动适配方法主要目的是在滑动适配过程中寻找到与VI人体轮廓最佳配准的PMMWI图像,算法1为尺度变换与滑动适配算法,最终输出为与VI实现最佳配准的PMMWI分割掩码图像,为下一步人体隐蔽违禁物的检测做好准备。

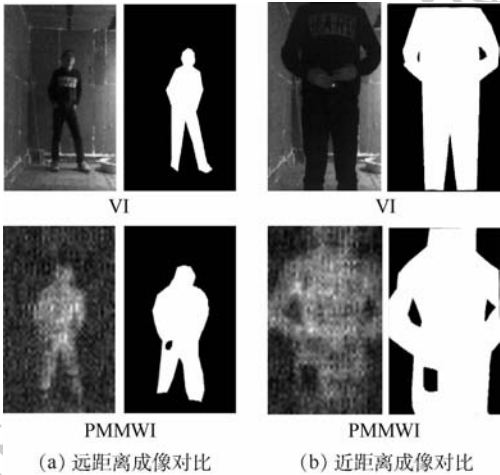


图15 不同距离下的VI和PMMWI人体轮廓分割对比
Fig. 15 Comparison of segmentation of human contours between VI and PMMWI in different distances

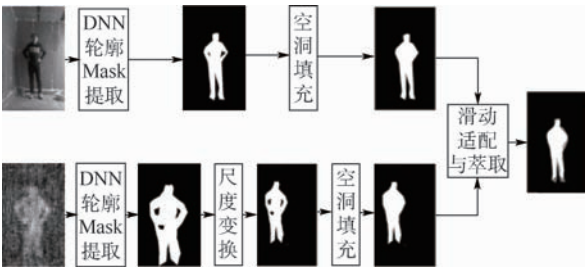


图16 PMMWI与VI人体轮廓尺度变换与滑动适配
Fig. 16 Scale transform and sliding fit of human contour in PMMWI and VI

算法 1 PMMWI 与 VI 人体轮廓的尺度变换与滑动适配。

输入:算法原始被动毫米波图像 P , 原始可见光图像 V 。

输出:与 VI 实现最佳配准的 PMMWI 分割掩码图像 $\tilde{P}_{seg}$ 。

初始化:图像大小为 $W \times H$, 尺度变换因子 $\varepsilon \in [0.5, 1.0]$, 滑动步数 $k \in [1, W \times H]$ 。

1. 利用训练好的 DNN 模型对 PMMWI 和 VI 进行人体轮廓分割和空洞填充, 其空洞填充后的结果图像分别为 P_{seg} 和 V_{seg}

2. 选取最佳尺度变换经验值 ε^* , 对 P_{seg} 中人体轮廓进行尺度变化得 $P_{seg}^{\varepsilon^*}$

3. For $k \leftarrow 1$ to $W \times H$:

4. 将 $P_{seg}^{\varepsilon^*}$ 中人体轮廓进行滑动 k 步长得 $P_{seg}^{\varepsilon^*, k}$

5. 将 $P_{seg}^{\varepsilon^*, k}$ 与 V_{seg} 进行异或运算, 得异或图像 $X^{\varepsilon^*, k}$

6. 将异或图像中所有像素值求和, 得 $S^{\varepsilon^*, k}$

7. if $k = 1$: $S^{\varepsilon^*, k^*} = S^{\varepsilon^*, k}$

Else: $S^{\varepsilon^*, k^*} = \min(S^{\varepsilon^*, k}, S^{\varepsilon^*, k^*})$

End If

8. End For

9. 最终获得最佳尺度因子 ε^* 与滑动步数 k^* , 即可得到与 VI 最佳配准的 PMMWI 分割掩码图像 $\tilde{P}_{seg}$ 。

为了能够有效评估滑动适配过程中配准的

准确率, 对 PMMWI 和 VI 进行图像异或处理, 如果两幅图像对应位置均为白色或黑色, 则异或结果为黑色; 如果对应位置颜色值不同, 则异或结果为白色。图 17 展示了 PMMWI 和 VI 滑动适配过程中的部分异或图像, 其白色区域越大, 异或图像中所有像素值和值越大。如图 17 第 4 张图所示, 当异或图像白色区域最小时, 即异或图像的和值取值最小, PMMWI 中的人体轮廓与 VI 中的人体轮廓实现最佳配准。



图 17 PMMWI 和 VI 匹配过程展示

Fig. 17 Exhibition of image registration process of PMMWI and VI

4.3 基于人体轮廓配准的单帧受检图像中隐蔽违禁物检测与定位

由于毫米波成像系统可以穿透人体衣物进行成像, 因此其透视成像特性也必然会引起公众对人体隐私安全的担忧, 本节将利用 VI 与 PMMWI 的优势互补特性, 排除 PMMWI 中的虚警目标, 并将隐蔽违禁物检测结果标记在 VI 中, 防止被动毫米波透视成像造成人体隐私的泄露。

图 18 所示为基于人体轮廓配准的单帧受检

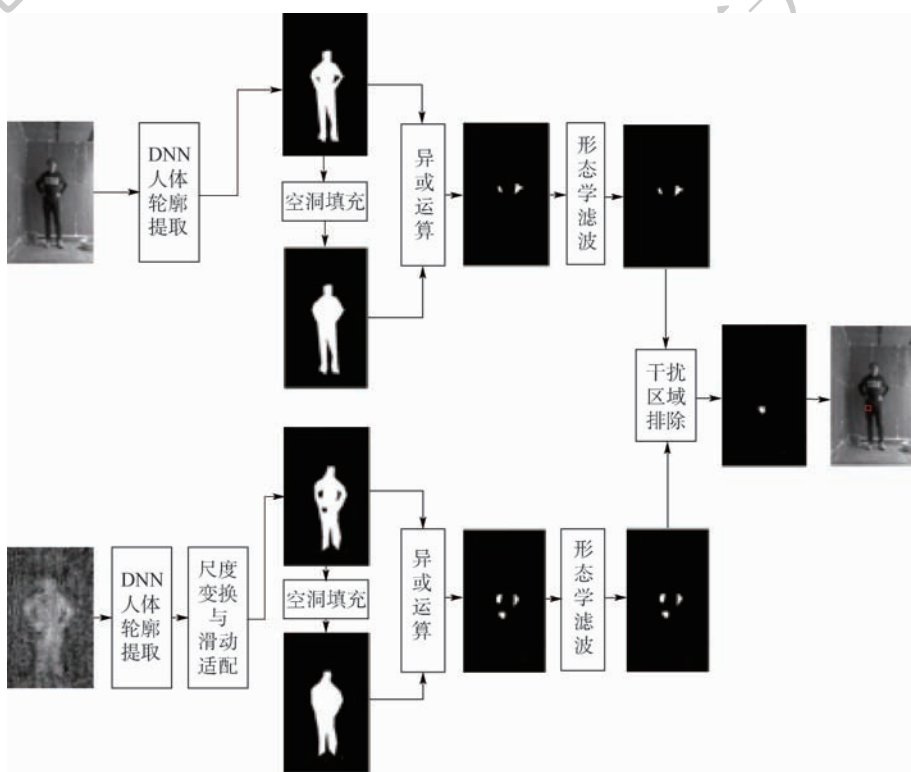


图 18 基于人体轮廓配准的隐蔽违禁物检测与定位

Fig. 18 Detection and localization of concealed forbidden object based on human contour registration

图像中隐蔽违禁物的检测与定位算法,其步骤为:
①对 VI 人体轮廓分割图像与其空洞填充图像进行异或运算可得到人体轮廓内的所有空洞区域,其目的是利用 VI 排除 PMMWI 中受检人员肢体动作形成的虚警区域;②利用同样的方法可以得到 PMMWI 中所有空洞区域,其中包含了虚警区域和隐蔽违禁物区域;③由于毫米波安检针对较大的违禁物,因此分别对前两步得到的图像进行腐蚀和膨胀的数学形态学滤波处理,以减少噪声点对检测结果的干扰;④通过对 PMMWI 和 VI 中可疑区域进行虚警排除,最终将 PMMWI 中真实隐蔽违禁物检测结果标记在 VI 中。

4.4 检测与定位的性能测评

为了验证基于单帧受检图像的人体隐蔽违禁物检测与定位算法的高效性,受检人员右侧裤子口袋携带违禁物进行测试,隐蔽违禁物的检测与定位结果均显示在 VI 中,避免隐私泄露。

通过对比图 19(c)和(d)可以明显看出,与 VI 相比,PMMWI 人体目标内出现了明显的隐蔽违禁物。图 19(e)为 4.2 节提出的基于多尺度变换人体轮廓配准算法得到的与 VI 最佳配准的 PMMWI,通过与图 19(c)对比可知,人体轮廓在尺度大小和位置上都实现了良好配准。由图 19(f)可以看出,受检人员隐藏违禁物已被

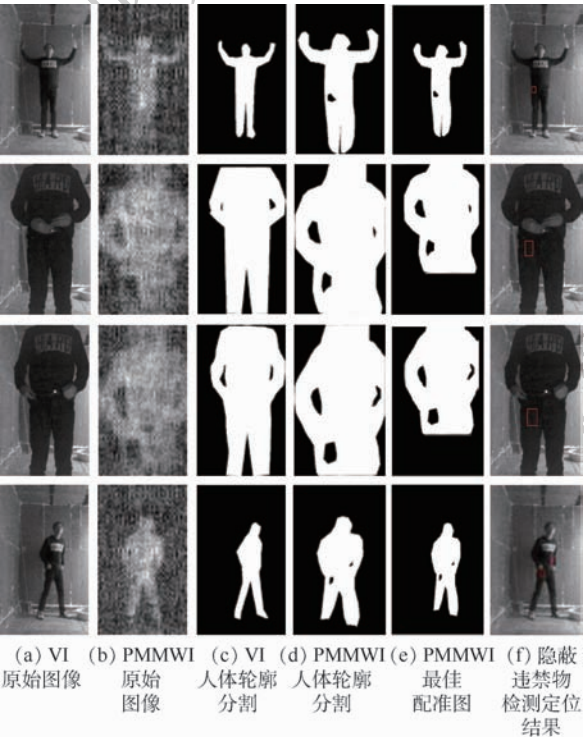


图 19 人体隐蔽违禁物检测与定位结果
Fig. 19 Results of detection and localization of concealed forbidden objects on human body

检测与标记出来。此外,仅从图 19(b)PMMWI 中很难分辨出虚警目标与违禁物,本文提出的隐蔽违禁物检测算法有效利用了 VI 和 PMMWI 的优势特征互补特性,实现了对虚警区域的有效排除。

5 基于受检过程序列图像检测结果对比融合的优化决策

5.1 受检过程序列图像的检测结果对比分析与融合

图 20 为多帧人体隐蔽违禁物检测结果,由图中可以看出,PMMWI 中隐蔽违禁物的检测框大小和位置呈现不稳定性,而且时间跨度越大,检测框位置差异性越大,因此必须对多帧图像的检测结果进行对比融合和综合分析,对隐蔽违禁物进行复合定位,以提高隐蔽违禁物的定位精确度。

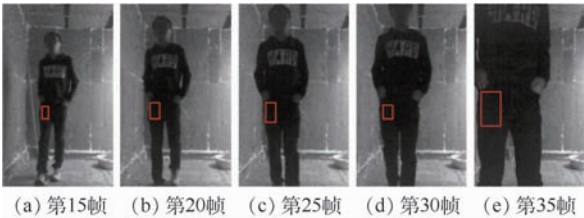


图 20 多帧人体隐蔽违禁物检测结果对比
Fig. 20 Comparison of multi-frame detection results of concealed forbidden objects on human body

5.2 基于序列图像检测结果对比融合的优化决策与复合定位

由于受检人员通过毫米波安检通道的时间较短,复合定位算法参考的帧数越多,违禁物定位误差将越大,从而导致违禁物定位坐标失效,因此本文提出的违禁物复合定位算法如算法 2 所示。

算法 2 基于序列图像检测结果对比融合的优化决策与复合定位算法。

输入:原始被动毫米波图像 P ,原始可见光图像 V 。

输出:人体隐蔽违禁物复合定位结果 C 。

初始化:对比融合的参考帧数 N 设为 5,对比融合图像 R 初始化为与 V 同样大小的 0 值矩阵。

1. For $t = 1$ to N
2. 获取当前时刻 PMMWI $P^{[t]}$ 和 VI $V^{[t]}$
3. 将当前帧 $P^{[t]}$ 和 $V^{[t]}$ 输入训练好的 DNN 图像分割模型得到人体轮廓掩码图像 $P_{seg}^{[t]}$ 和 $V_{seg}^{[t]}$
4. 依据 4.2 节提出的 PMMWI 与 VI 人体轮廓配准算法,获得与 VI 实现最佳配准的 PMMWI $\tilde{P}_{seg}^{[t]}$
5. 依据 4.3 节提出的单帧图像隐蔽违禁物检测算法,获得人体隐蔽违禁物检测结果图像 $D^{[t]}$

6. 对单帧检测框内部和外部区域分别标记为 1 和 0,得到单帧检测结果二值图像 $M^{[t]}$
7. 将前 t 帧的检测二值图像累加,即 $R = R + M^{[t]}$
8. End For
9. 寻找融合图像 R 中像素值为 5 的区域 R_5
10. 人体隐蔽违禁物复合定位结果为 R_5 的中心点,即人体隐蔽违禁物定位结果 $C = \text{centroid}(R_5)$ 。

图 21 左图为 5 帧对比融合后的细节放大图,右图为其 3D mesh 图,本文将每一帧的检测掩码部分标记为 1,这样 5 帧融合结果就分为 5 个等级,等级越高表示该区域被检测到的次数越多,等级为 5 表示该区域存在隐蔽违禁物的可能性最大。因此,本文将等级为 5 的区域中心点设为复合定位中心点,该中心坐标将传输给 AMMWI 系统获得主动毫米波图像,进而实现违禁物的属性识别。

5.3 算法与模型的综合集成与优化

整体集成优化算法如图 22 所示,包括基于

DNN 的单帧 PMMWI 和 VI 人体轮廓分割、基于人体轮廓的 PMMWI 和 VI 配准、单帧图像的隐蔽违禁物检测与定位、基于序列图像检测结果对比融合的优化决策和复合定位。最终,复合定位算法给出人体轮廓区域内隐蔽违禁物的中心坐标,AMMWI 系统将对此坐标进行局部范围扫描成像,以实现隐蔽违禁物的属性识别。

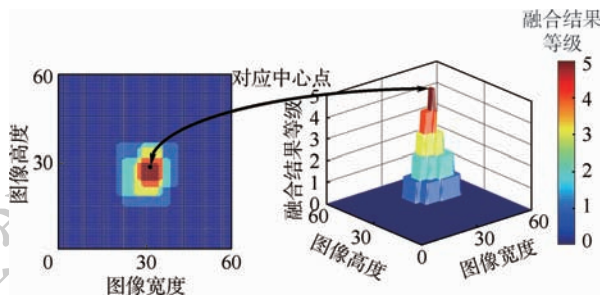


图 21 基于检测结果对比融合的复合定位

Fig. 21 Composite localization based on contrasting fusion of detection results

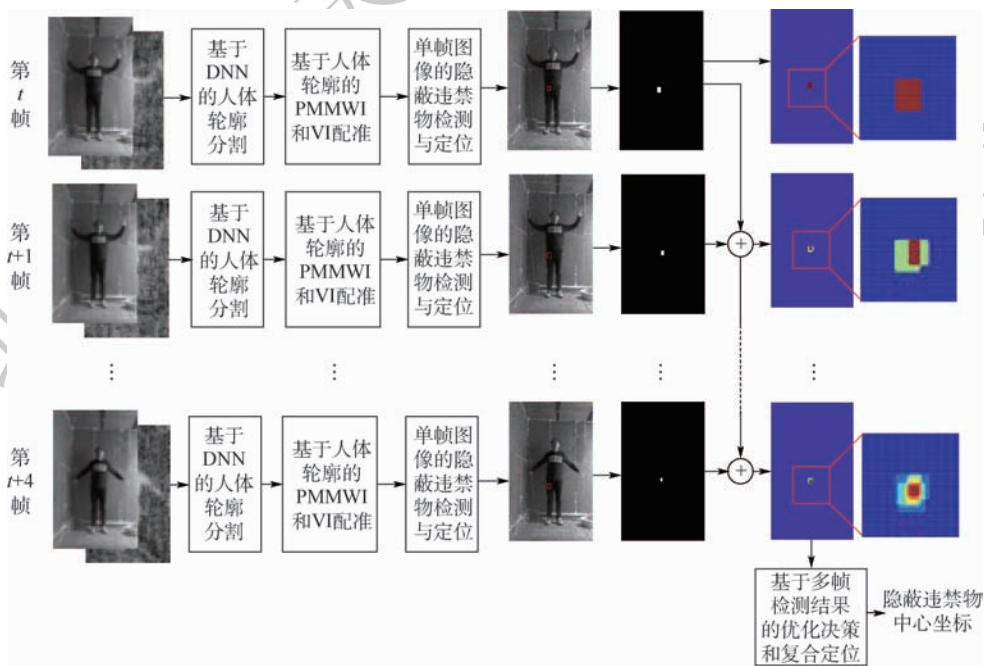


图 22 算法与模型的综合集成与优化

Fig. 22 Comprehensive integration and optimization of algorithms and models

6 整体检测算法的性能测试与评价

6.1 数据集与实验平台

本文数据集采集自实际毫米波安检系统,共包含 386 张图像,其中包含 193 张 PMMWI 和 193 张 VI,并且利用图像分割数据集标记工具对其全部进行了人工数据集制作,详细信息可参考 3.3.1 节。本文所有的实验均以 Python 作为开发语言,基于 TensorFlow 1.9 进行 DNN 模型开发和应用,在 Ubuntu 16.04 系统上进行训练和测试,

并利用 Nvidia 1080ti 显卡加速 DNN 模型的运算。

6.2 实验框架设计与参数设置

针对基于 LFF-UNet 网络的 PMMWI 人体轮廓分割,训练策略参数设置见 3.3.2 节,而在基于尺度变换和滑动适配的人体轮廓配准算法中,尺度变换参数范围为 $[0.5, 1]$,尺度变化步长为 0.1,通过对 PMMWI 中人体轮廓进行滑动适配,以实现与 VI 中人体轮廓的最佳配准,其中滑动步长设为 2。最后,基于多帧检测结果的对比融合优化策略和复合定位算法主要利用 5 帧图像进行

实验。

6.3 实验结果与性能评估

由于 PMMWI 缺乏显著性的辨识度,一般缺乏先验知识的人很难从 PMMWI 中定位隐蔽违禁物的位置,因此实验结果展示了远距离和近距离下的人体隐蔽违禁物检测结果,其中对 PMMWI 和 VI 均进行了隐蔽违禁物的标识,以方便与检测结果进行对比。

图 23 和图 24 分别为远距离和近距离下的人体隐蔽违禁物检测结果,其前两行分别为 PMMWI 和 VI 原始图像,并且同时对隐蔽违禁物进行了人工标记,第3行为综合集成算法的隐蔽违禁物检

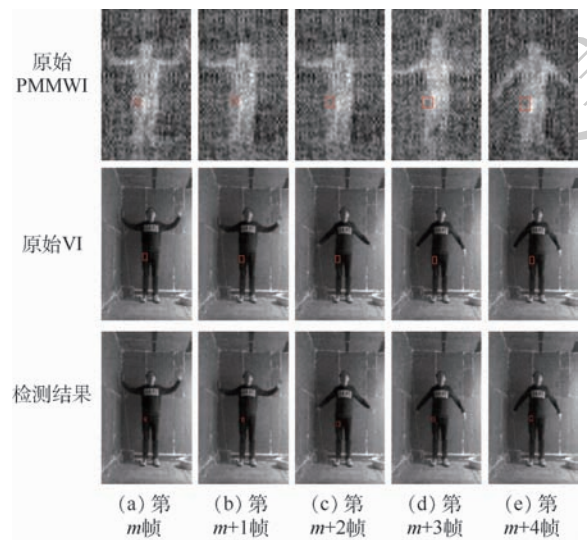


图 23 远距离人体隐蔽违禁物检测

Fig. 23 Detection of concealed forbidden objects on human body from far distance

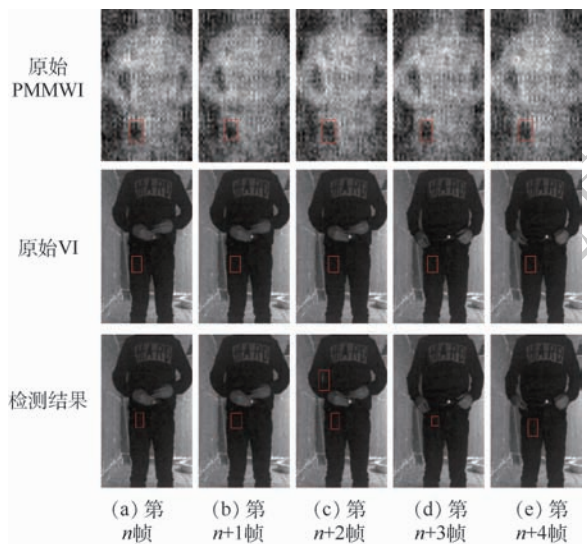


图 24 近距离人体隐蔽违禁物检测

Fig. 24 Detection of concealed forbidden objects on human body from close distance

测结果。在图 24 所示近距离检测 results 中,受到 PMMWI 的影响,每一帧检测框大小变化更加明显,而且第 $n+2$ 帧图像还有误检情况的发生,主要是 VI 中右侧肢体围成的区域由于面积过小,没有实现精确分割,但是经过序列图像检测结果对比融合的优化决策与复合定位算法可以排除误检导致的隐蔽违禁物定位不准确问题。

表 1 给出了不同距离下的人体隐蔽违禁物检测 IoU 对比结果,表中 a~e 对应图 23 和图 24 中 (a)~(e)。通过分析连续帧的检测 results 可知,单帧检测结果不够稳定,尤其近距离下图 24(c) 检测结果远低于其他帧的 IoU 值,而通过多帧检测 results 的优化决策,可以避免单帧检测不稳定对最终隐蔽违禁物定位坐标的影响。

为了清晰地对比人体隐蔽违禁物定位 results,通过计算真实定位坐标与算法得到的定位坐标之间的距离来衡量定位算法的优劣,即真实世界中两点之间的距离,单位为 cm。图 25 中 a~e 分别对应于图 23 和图 24 中的不同帧图像,图中虚线表示复合定位坐标与真实定位坐标之间的误差值。

图 25 的单帧图像隐蔽违禁物定位误差是根据实际违禁物的中心坐标与本文定位算法的定位 results 之间的距离进行测算的。由图中可以看出,近距离定位误差较小,而远距离定位误差偏大,其中,远距离第 $m+2$ 帧的定位误差达 9 cm。在实际应用中,采取复合定位的原则,通过多帧图像的复合定位来确定隐蔽违禁物的中心坐标位置。从而通过复合过程的叠加过滤而排除单帧图像定位误差偏大的情况,确保定位 results 的可靠性,为进一步启动主动毫米波扫描成像确定精准的区域指向。

为了更加清晰地观察复合定位 results,图 26(a) 为原始 VI,箭头所指为隐蔽违禁物中心位置,图 26(b) 为 5 帧检测结果复合定位 results 图,为了方便观察定位 results,图 26(c) 为图 26(a) 和图 26(b) 相加图像,其中红点标记为最终隐蔽违禁物的复合定位 results,经过多帧的复合定位,可以有效避免单帧检测不精确和不稳定问题。

表 1 人体隐蔽违禁物检测 IoU 对比

检测方式	a	b	c	d	e
远距离检测	0.642	0.859	0.782	0.490	0.648
近距离检测	0.786	0.810	0.362	0.584	0.853

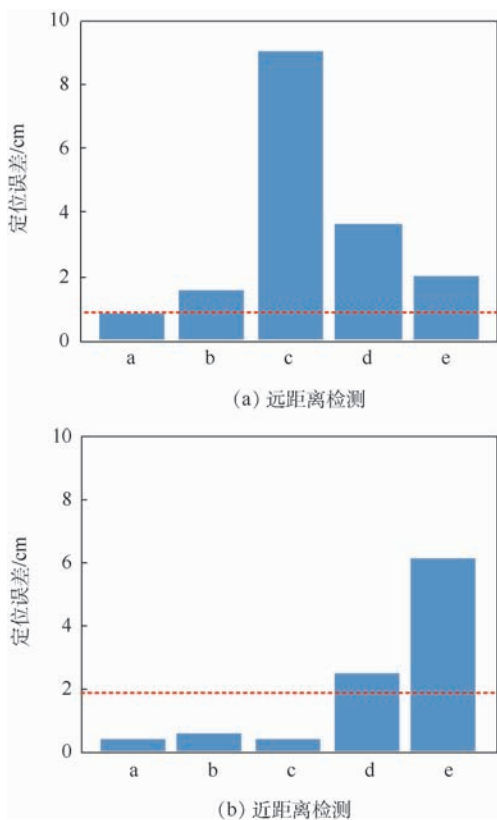


图 25 人体隐蔽违禁物定位误差对比

Fig. 25 Comparison of localization error of concealed forbidden objects on human body

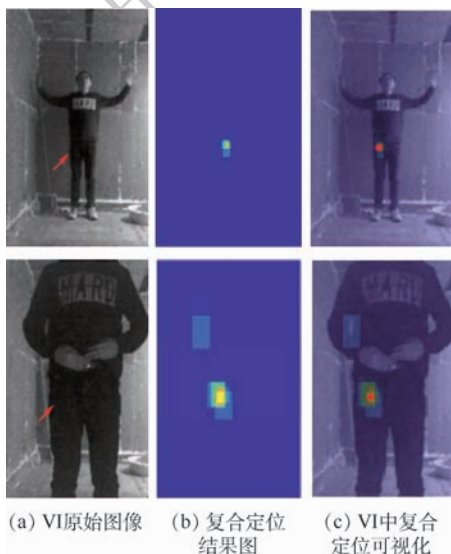


图 26 人体隐蔽违禁物定位可视化结果

Fig. 26 Visualization of localization result for concealed forbidden objects on human body

6.4 述评与注解

本节主要对 LFF3-UNet 与 U-Net 进行对比分析,通过利用特征图可视化技术^[23]对跨越连接层进行特征图可视化,以分析 LFFB 模块的作用。如图 27 所示,与文献[23]不同的是,使用 3D 网

格化方法对特征图进行可视化。需要特别说明的是,由于隐藏层特征图维度一般都很高,因此可视化结果是所有特征图在通道维度上相加并归一化后的结果。

图 27 展示了不同网络模型在第一次跨越连接时的特征图可视化结果。通过对比图 27(a) 和

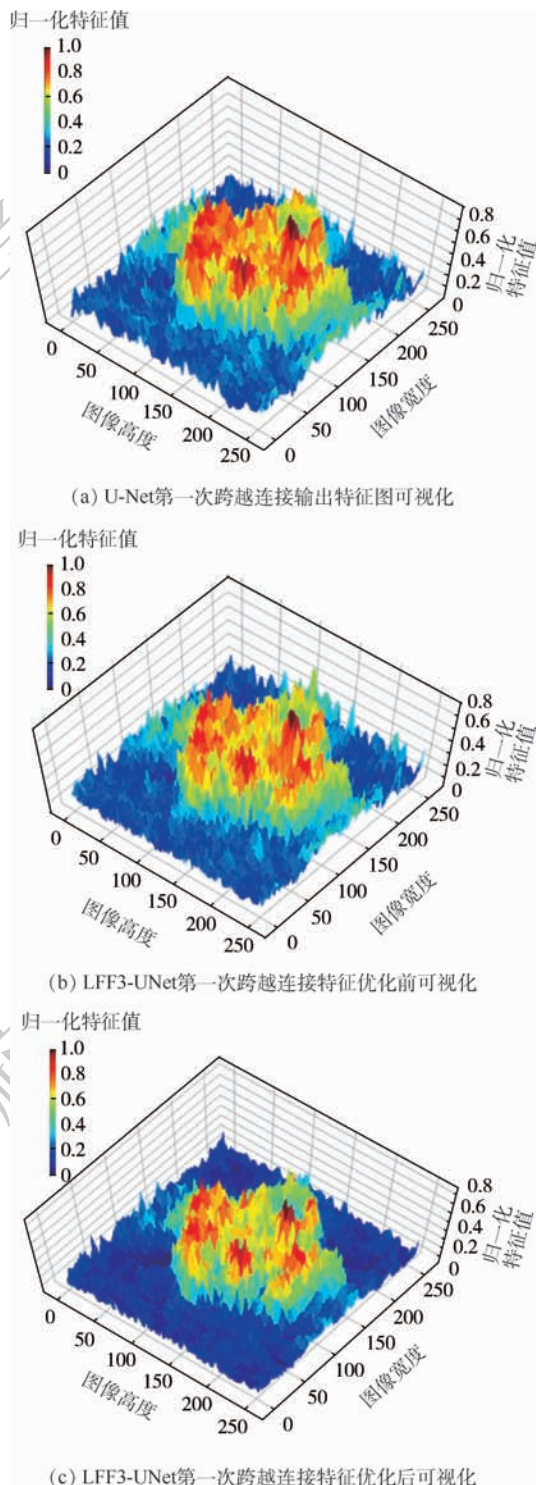


图 27 U-Net 与 LFF3-UNet 网络跨越连接层特征图可视化对比

Fig. 27 Comparison of feature map visualization at skip-connection layer between U-Net and LFF3-UNet

图 27(b) 可知, U-Net 与 LFF3-UNet 在跨越连接前传输的特征图是基本一致的, 而经过本文提出的特征优化之后, 可以明显的看出, 卷积融合后的背景区域噪声更少, 人体轮廓更加清晰。低层特征图中干扰因素越少, 对最终分割结果影响也就越小, 这也与最终 PMMWI 分割实验结果一致。

7 结 论

本文针对面向大流量安检场景的毫米波安检系统, 提出一种基于 PMMWI 和 VI 优势互补的人体隐蔽违禁物检测与定位算法。

1) 利用 PMMWI 和 VI 优势互补特性, 提出一种基于人体轮廓的 PMMWI 和 VI 配准方法, 以排除可疑区域, 同时将检测结果显示在 VI 中, 避免由于毫米波的透视成像特性引起公众对人体隐私安全的担忧。

2) 为了解决传统 U-Net 对 PMMWI 中弱小目标分割性能较差问题, 提出一种基于低层特征融合的 LFF-UNet 以增强 U-Net 网络对 PMMWI 中弱小目标轮廓的敏感度, 显著提高了 PMMWI 中人体轮廓和隐蔽违禁物的分割精度。

3) 提出一种基于序列图像检测结果对比融合的方法, 避免 PMMWI 不稳定导致的单帧检测结果不准确问题, 以实现人体隐蔽违禁物的高精度定位。实验结果也验证了本文算法具有良好的鲁棒性。

在本文算法对人体隐蔽违禁物检测与定位的基础上, 如何继续利用 AMMWI 系统对定位区域进行二次扫描, 最终完成基于 AMMWI 的人体隐蔽违禁物识别, 将是下一步研究的重点。

参考文献 (References)

[1] 王任肩, 王宇光, 陈志福, 等. 大型活动中的安检排爆工作及保障研究[J]. 警察技术, 2018(5): 19-22.
WANG R J, WANG Y G, CHEN Z F, et al. Research on exploder clearing work and technical support in security inspection of large activities[J]. Police Technology, 2018(5): 19-22 (in Chinese).

[2] 费鹏, 方维海, 温鑫, 等. 用于人员安检的主动毫米波成像技术现状与展望[J]. 微波学报, 2015, 31(2): 91-96.
FEI P, FANG W H, WEN X, et al. State of the art and future prospect of the active millimeter wave imaging technique for personnel screening[J]. Journal of Microwaves, 2015, 31(2): 91-96 (in Chinese).

[3] GROSSMAN E N, MILLER A J. Active millimeter-wave imaging for concealed weapons detection[C] // Proceedings of Conference on Passive Millimeter-Wave Imaging Technology VI and Radar International Society for Optics and Photonics. Belling-

ham: SPIE, 2003, 5077: 62-70.

[4] ANDREWS D A, HARMER S W, BOWRING N J, et al. Active millimeter wave sensor for standoff concealed threat detection[J]. IEEE Sensors Journal, 2013, 13(12): 4948-4954.

[5] DU K, ZHANG L, CHEN W, et al. Concealed objects detection based on FWT in active millimeter-wave images[C] // Proceedings of 7th International Conference on Electronics and Information Engineering. Bellingham: SPIE, 2017, 10322: 103221O.

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

[7] 骆尚, 吴晓峰, 杨明辉, 等. 基于卷积神经网络的毫米波图像人体隐蔽物检测[J]. 复旦学报(自然科学版), 2018, 57(4): 442-452.
LUO S, WU X F, YANG M H, et al. Convolutional-neural-network-based human concealed object detection for millimeter wave images[J]. Journal of Fudan University (Natural Science), 2018, 57(4): 442-452 (in Chinese).

[8] SINCLAIR G N, ANDERTON R N, APPLEBY R. Passive millimeter-wave concealed weapon detection[C] // Proceedings of Enabling Technologies for Law Enforcement and Security. Bellingham: SPIE, 2001, 4232: 142-152.

[9] YEOM S, LEE D, SON J, et al. Real-time outdoor concealed-object detection with passive millimeter wave imaging[J]. Optics Express, 2011, 19(3): 2530-2536.

[10] YEOM S, LEE D, JANG Y, et al. Real-time concealed-object detection and recognition with passive millimeter wave imaging[J]. Optics Express, 2012, 20(9): 9371-9381.

[11] 周健, 叶金晶, 孙谦晨, 等. 主动毫米波成像性别识别算法研究[J]. 红外, 2018, 39(9): 34-40.
ZHOU J, YE J J, SUN Q C, et al. Study of a gender identification algorithm for active millimeter-wave imaging[J]. Infrared, 2018, 39(9): 34-40 (in Chinese).

[12] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[13] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[14] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C] // Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.

[15] 桑伟, 岳胜利. 毫米波成像技术在人体安全检查领域的应用[J]. 中国安防, 2013(4): 83-87.
SANG W, YUE S L. Application of millimeter wave imaging technology on the area of human body security inspection[J]. China Security & Protection, 2013(4): 83-87 (in Chinese).

[16] IGLOVNIKOV V, SHVETS A. TeraNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation[EB/OL]. (2018-01-17) [2018-12-20]. <https://arxiv.org/abs/>

- 1801.05746.
- [17] LIN M, CHEN Q, YAN S C. Network in network [EB/OL]. (2014-03-04) [2018-12-20]. <https://arxiv.org/abs/1312.4400>.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015:1-9.
- [19] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [EB/OL]. (2015-12-11) [2018-12-20]. <https://arxiv.org/abs/1512.00567v2>.
- [20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [EB/OL]. (2015-12-10) [2018-12-20]. <https://arxiv.org/abs/1512.03385>.
- [21] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [EB/OL]. (2017-04-11) [2018-12-20]. <https://arxiv.org/abs/1611.05431>.
- [22] KINGMA D P, BA J L. Adam: A method for stochastic optimization [EB/OL]. (2017-01-30) [2018-12-20]. <https://arxiv.org/abs/1412.6980>.
- [23] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C] // Proceedings of European Conference on Computer Vision. Berlin: Springer, 2014: 818-833.
- 作者简介:
- 赵国 男, 博士研究生。主要研究方向: 深度学习、图像语义分割、目标的检测与识别。
- 秦世引 男, 博士, 教授, 博士生导师。主要研究方向: 图像处理、模式识别、智能优化控制等。

Detection and localization of concealed forbidden objects on human body based on complementary advantages of PMMWI and VI

ZHAO Guo, QIN Shiyin*

(School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China)

Abstract: According to the performance requirements and technology demands of human security check in public place, combining the performance advantages of perspective imaging of passive millimeter wave imaging (PMMWI) and high discriminability in image details of visible imaging (VI), an approach for detection and localization of concealed forbidden objects on human body is presented in this paper based on the complementary advantages of PMMWI and VI. Firstly, an improved U-Net based on feature fusion in low layers is presented to enhance the sensitivity of deep neural networks (DNN) to the contour of dim small targets, and improve the accuracy of segmentation of human contours and concealed forbidden objects. Meanwhile, the pixel-level segmentation of human contours in VI is also implemented. Then, the contours of human body in PMMWI are registered with some corresponding ones in VI by scale transform and sliding fit, so the concealed forbidden objects on human body can be detected from a single frame with a high precision. Finally, the concealed forbidden objects are localized by contrasting fusion and optimizing decision according to detection results with sequence images. A series of comprehensive experiments and comparative analysis results validate the good performance of the proposed detection and localization algorithm of concealed forbidden objects on human body towards security check of public places.

Keywords: millimeter wave security check; passive millimeter wave imaging (PMMWI); contour segmentation of human body; deep learning; deep neural network (DNN); detection and localization of concealed forbidden objects

Received: 2019-01-17; **Accepted:** 2019-05-28; **Published online:** 2019-06-18 13:38

URL: kns.cnki.net/kcms/detail/11.2625.V.20190617.0922.001.html

Foundation item: National Natural Science Foundation of China (61731001)

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

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0020

非结构重叠网格显式装配算法



宣传伟, 韩景龙*

(南京航空航天大学 航空学院 机械结构力学及控制国家重点实验室, 南京 210016)

摘 要: 针对重叠网格中洞映射法占用过多物理内存的问题, 发展了一种改进型洞映射法; 基于相邻单元搜索法, 发展了一种基于相邻阵面的贡献单元搜索法; 通过将割补法与隐式切割技术相结合提出了一种非结构重叠网格显式装配算法。该算法首先生成一套包围物面的笛卡儿网格, 其次存储所有与物面边界相交的笛卡儿网格信息, 最后根据所存储笛卡儿网格与所需判断的网格单元的相对位置来判断其是否为洞内单元。在成功判断出所有洞内单元后, 以当前洞边界为初始阵面推进, 同时以各个网格单元的物面距离为判别标准对重叠区域进行优化, 生成最终插值边界。所提算法优化了传统非结构重叠网格装配过程, 具有物理内存占用低, 贡献单元搜索次数少以及计算效率高等特点。通过 2 个典型复杂流动算例验证了所提算法的准确性与适用性。

关 键 词: 重叠网格; 割补法; 洞映射; 非结构网格; 物面距离

中图分类号: V211.3

文献标识码: A **文章编号:** 1001-5965(2019)10-2026-09

在计算流体力学中, 重叠网格技术是用于模拟复杂流动的重要途径之一。相较于 Steger 等^[1]提出的结构网格, 非结构网格没有网格节点的结构限制, 具有更好的灵活性, 对复杂构型的适应能力更强。与结构重叠网格不同, Nakahashi 和 Gumiya^[2]提出的非结构重叠网格技术没有显式的挖洞过程, 得到广泛研究^[3-6]。该技术通过在宿主单元搜索的过程中对网格节点进行分类(洞内点、洞边界点和洞外点), 并将网格节点到自身物面的最近物面距离定义为节点分类的准则参数。该过程实际上是一种隐式切割技术^[7], 即通过贡献单元的品质来进行选择, 这些品质可以是网格的物面距离、单元体积或者二者的乘积等。

理论上, 结构重叠网格技术同样适用于非结构重叠网格。但非结构网格分布的随意性以及隐式装配方法过程的相对简单性, 使得非结构重叠网格技术的研究大都基于 Nakahashi 所提出的隐

式装配算法。近年来, 该方法已被成功应用于多体相对运动^[8]、高超声速边界层转捩^[9]、旋翼翼型^[10]、桨-涡干扰^[11]等研究。Lohner 等^[12]和 Luo^[13]在 Nakahashi 方法的基础上将网格的尺度与物面距的乘积量作为网格类别的判断标准, 提高了插值稳定性。Landmann 和 Montagnac^[14]、Xu 等^[15]分别通过几种不同方法提高了隐式装配算法的鲁棒性, 其中 Xu 等^[15]更是成功地将该方法应用于飞行器的抖振研究。田书玲^[16]发展了一种基于阵面的相邻单元搜索法, 提高了隐式装配效率。

贡献单元的搜索占据了非结构重叠网格过程的大部分时间, 且非结构重叠网格隐式算法需要针对所有网格点搜索贡献单元。然而, 在结构重叠网格技术中, 首先通过显式的挖洞过程识别出所有位于物面内部的单元, 然后可通过割补法^[17]等阵面推进方法^[18]对网格重叠区域进行优化, 并

收稿日期: 2019-01-17; 录用日期: 2019-04-05; 网络出版时间: 2019-04-30 16:02

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190429.1444.002.html

基金项目: 国家自然科学基金(11472133); 江苏高校优势学科建设工程基金

* 通信作者: E-mail: hjlae@nuaa.edu.cn

引用格式: 宣传伟, 韩景龙. 非结构重叠网格显式装配算法[J]. 北京航空航天大学学报, 2019, 45(10): 2026-2034.

XUAN C W, HAN J L. Explicit assembly algorithm of unstructured overset grid[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2026-2034 (in Chinese).

且只需在网格割补的同时搜索重叠区域网格的贡献单元,而非所有网格点的贡献单元。显然,相较于结构重叠网格的显式装配算法,传统非结构重叠网格隐式装配算法针对所有网格点进行贡献单元搜索的过程是低效的。

作为显式重叠网格装配算法的首要步骤,挖洞过程可以由多种方法实现,如洞映射法^[19]、射线求交法^[20]、矢量判别法^[21]、直接法^[22]和目标 X 射线法^[23]等。其中,洞映射法采用辅助笛卡儿网格近似挖洞曲面,进而将点与曲面的关系转化为点与洞映射单元的简单关系。由于效率高、算法实现简单等,洞映射法已被广泛应用于旋翼流场计算^[24]、叶轮机流场分析^[25]、飞行器外挂物分离^[26]和网格生成^[27]等领域。但是,洞映射法要求物面必须封闭,且当笛卡儿网格具有一定精度时所消耗的物理内存也急剧升高。淮洋等^[28]发展了一种改进型洞映射法,解决了物面必须封闭的问题。袁武^[29]、王文^[30]等提出了一种“最小洞映射”的概念,节约了内存的开销。

本文首先针对洞映射法内存占用过大的问题进行了研究,发展了一种改进型洞映射法,降低了内存的占用率;其次,通过将割补法与非结构重叠网格中隐式切割技术相结合提出了一种非结构重叠网格显式装配算法;然后,为进一步优化贡献单元搜索效率,对相邻单元搜索法进行了改进,发展了一种基于相邻阵面的贡献单元搜索法;最后,将该显式网格装配算法应用于 2 个典型流动算例,验证了本文算法的有效性。

1 洞映射法及其改进

1.1 传统洞映射法

传统洞映射法一般首先根据洞曲面的坐标范围建立包围盒,以洞曲面单元的平均尺寸对包围盒进行划分从而形成笛卡儿网格。然后,查找与洞曲面接触的网格单元并标记为边界单元。接着,由笛卡儿网格各个角点开始向内推进并标记洞外单元,最后剩余单元则为洞内单元。映射完成后,对于任一点 $P(X, Y, Z)$,其对应笛卡儿网格编号的计算式为

$$\begin{cases} I = \text{INT}\left(\frac{X - X_0}{\Delta X}\right) + 1 \\ J = \text{INT}\left(\frac{Y - Y_0}{\Delta Y}\right) + 1 \\ K = \text{INT}\left(\frac{Z - Z_0}{\Delta Z}\right) + 1 \end{cases} \quad (1)$$

式中: I, J 和 K 分别为点 P 在 X, Y 和 Z 3 个方向上的编号; $\text{INT}()$ 表示取整运算; X_0, Y_0 和 Z_0 为笛

卡儿网格起点坐标; $\Delta X, \Delta Y$ 和 ΔZ 分别为笛卡儿网格在 X, Y 和 Z 3 个方向上的分辨率。

1.2 改进型洞映射法

传统洞映射法需要针对所有笛卡儿网格进行遍历并计算其属性(洞内单元、边界单元或洞外单元);另外,对于物面特征复杂或存在狭小缝隙的模型,笛卡儿网格分辨率必须提高很多倍才能满足挖洞需求。因此,为存储所有笛卡儿网格单元是否洞内单元的属性而所消耗的内存空间通常是庞大的。

实际上,针对一个封闭洞面,在查找出所有与洞曲面接触的边界单元之后,若点 $Q(X_q, Y_q, Z_q)$ 位于 2 个或多个边界单元之间,则其必是洞内单元;反之,则为洞外单元。即是说,仅存储笛卡儿网格中边界单元的属性即可,而针对所有笛卡儿网格单元属性的存储不是必须的。由此,便可大大地节约洞映射法对物理内存的消耗。本文对传统洞映射法做出如下改进。

首先,查找出所有与洞面相交的边界单元,并根据其编号 $N(I, J, K)$ 顺序存储边界单元。对于任一点 $Q(X_q, Y_q, Z_q)$,首先根据其编号 $N_q(I_q, J_q, K_q)$ 判断是否位于某个边界单元内;若不是,则从所存储边界单元信息中查找 I_q 行所包含的不同边界单元编号,若同时存在编号大于 N_q 和小于 N_q 的不同边界单元,则点 Q 为洞内单元,反之则为洞外单元。图 1 所示为改进型洞映射法产生的笛卡儿网格示意图,图中黑色单元表示边界单元。对于洞内任一点 O ,编号记为 $N_o(I_o, J_o, K_o)$,在 I_o 行必然存在编号小于 N_o 的边界单元 BO1 和大于 N_o 的边界单元 BO2。而对于点 P ,由于其编号小于其所在行编号最小的边界单元 BP,故点 P 为洞外点;对于点 Q ,由于其编号大于其所在行编号最大的边界单元 BQ,故点 Q 也为洞外点。

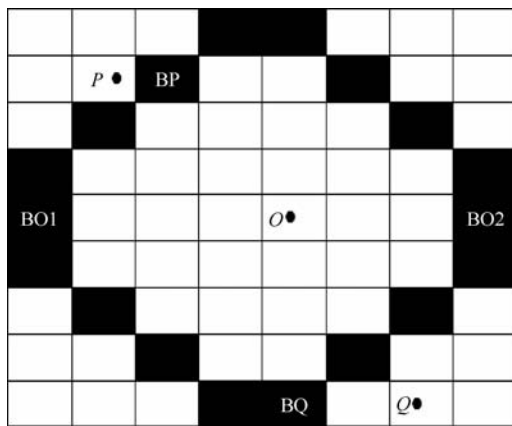


图 1 改进型洞映射法产生的笛卡儿网格示意图

Fig. 1 Cartesian grids of improved hole mapping method

可见,相较于传统洞映射过程中需要存储所有笛卡儿网格是否位于物面内的信息,本文改进型洞映射法仅需按一定顺序存储边界单元的信息即可,因此降低了内存空间的占用率。

2 显式网格装配算法

2.1 割补法

割补法是在初始挖洞的基础上经“切割”和“填补”2个阶段使得各子网格间的洞边界最优化的过程,贡献单元的搜索也在该过程中完成。

在切割阶段,所有洞边界点依次寻找贡献单元,若贡献单元的所有节点都为正常点,则该洞边界点转变为洞内点,与其相邻的正常点转变为洞边界点,由此将洞边界推离物面;若贡献单元的节点不全为正常点,则停止该过程。

经切割之后,没有网格单元重叠,而填补阶段本质上是切割阶段的逆过程,其目的是将洞边界逐步收缩从而实现网格单元的重叠。在填补阶段,所有洞边界点依次寻找贡献单元,若贡献单元的节点不全为正常点,则该洞边界点转变为正常点,与其相邻的洞内点转变为洞边界点,由此收缩洞边界,直至再无洞内点转变为洞边界。

2.2 基于相邻阵面的贡献单元搜索技术

在割补过程中,ADT(Alternating Data Tree)寻点技术^[31]是常用的贡献单元搜索方法。但是,考虑到割补法本质上是一种阵面推进技术;在阵面推进过程中,对于当前阵面中的网格点,若将上一阵面中与其相邻的点的贡献单元作为其贡献单元搜索的起始单元,便可采用相邻单元搜索法^[32]通过很少的几步迭代快速找到当前阵面中点的贡献单元。

相较于ADT寻点技术,田书玲等^[33]提出的基于阵面推进的相邻单元搜索法是一种更高效的贡献单元搜索方法,其结合了阵面推进技术与相邻单元搜索法的特点,核心步骤与上述思想类似:首先确定任一网格点的贡献单元,同时以该节点为首节点建立一个链表;然后循环搜索该点周围邻居节点的贡献单元,并将所有邻居节点推入链表,在搜索完所有邻居节点的贡献单元后删除首节点,接着搜索新的首节点的邻居节点;如此循环维护该链表中节点的增减,直至链表为空,贡献单元搜索结束。

图2所示为由红蓝两套网格组成的重叠网格系统。可看出蓝色网格点 P_0 的贡献单元为红色网格单元 C 、 D 、 I 。在搜索 P_0 点的相邻节点 $q_1 \sim q_5$ 的贡献单元时,将单元 C 、 D 、 I 作为搜索的起始

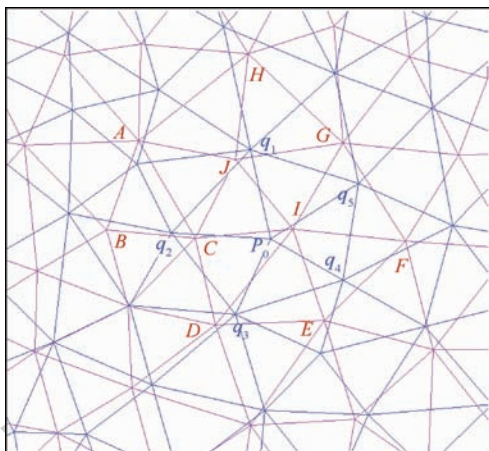


图2 基于阵面推进的相邻单元搜索法
Fig. 2 Neighbor-to-neighbor search algorithm
based on advancing front

单元,采用相邻单元搜索法通过很少的几步迭代便可确定出点 $q_1 \sim q_5$ 的贡献单元。如:在单元 C 、 D 、 I 的相邻单元中, D 、 E 、 I 是距离点 q_4 最近的单元,再由 D 、 E 、 I 查找到其相邻单元 I 、 E 、 F 即是点 q_4 的贡献单元;再如:从 C 、 D 、 I 出发通过一步迭代便可查找到其相邻单元 D 、 E 、 I 即是点 q_3 的贡献单元。

基于阵面推进的相邻单元搜索法会搜索所有网格点的贡献单元。但在割补法过程中仅须在切割过程中搜索当前洞边界点的贡献单元,因此本文在该方法基础上发展了一种基于相邻阵面的贡献单元搜索法以适配割补法。具体步骤如下:

1) 在初始挖洞结束之后,建立2个空链表LI和LJ,将所有当前洞边界点($I_0, I_1, \dots, I_n$)推入链表LI。

2) 查找链表LI中首节点 I_0 的贡献单元 D_0 。

3) 以 D_0 为起始单元通过相邻单元搜索法查找 I_1 的贡献单元 D_1 ,再以 D_1 为起始单元搜索 I_2 的贡献单元 D_2 ,如此推进直至遍历链表LI中所有节点并对所有已搜索过贡献单元的节点进行标记。

4) 清空链表LJ,遍历链表LI,查询每个节点 I_i 的相邻节点 J_i ,若 J_i 未被标记则以 I_i 的贡献单元 D_i 为起始单元通过相邻单元搜索法搜索节点 J_i 的贡献单元,然后将其推入链表LJ。

5) 清空LI,将LJ中所有节点推入LI,重复步骤4),如此通过2个链表互相迭代推进洞边界位置。

2.3 网格装配算法

贡献单元的搜索是传统非结构重叠网格隐式挖洞技术的核心,同时也是耗时最多的环节。隐式挖洞方法仅仅通过比较网格质量(如体积、物面距离等)区分插值单元与计算单元,不会挖去

落在物体内部的非计算单元。虽然不影响计算结果的准确性,但这些洞内单元却参与到了贡献单元的搜索过程中,降低了装配效率。本文基于结构重叠网格理论,首先通过改进型洞映射法排除洞内单元,其次通过将割补法与隐式切割技术相结合进一步优化非结构重叠网格中贡献单元的搜索过程,提高网格装配效率。具体过程如下:

- 1) 循环所有网格,计算各网格单元到物面的距离并将所有单元标记为正常单元。
- 2) 采用本文改进型洞映射法挖去所有洞内单元,将位于物面内部的网格单元标记为洞内单元,与洞内单元相邻的正常单元标记为洞边界单元。

3) 以上述洞边界单元为初始阵面推进,并以本文所发展的基于相邻阵面的贡献单元搜索法搜索并保存其贡献单元信息。在阵面推进过程中,比较该洞边界单元与其贡献单元的物面距大小。若前者大于后者,则将该洞边界单元标记为洞内单元,而与其相邻的正常单元标记为洞边界单元。当洞边界单元的物面距小于或等于其贡献单元的物面距时结束该过程。

- 4) 为保证二阶计算精度,将与洞边界单元相邻的 2 个洞内单元标记为插值单元,同时将该洞边界单元标记为正常单元。

图3所示为本文算法过程,该套非结构重叠

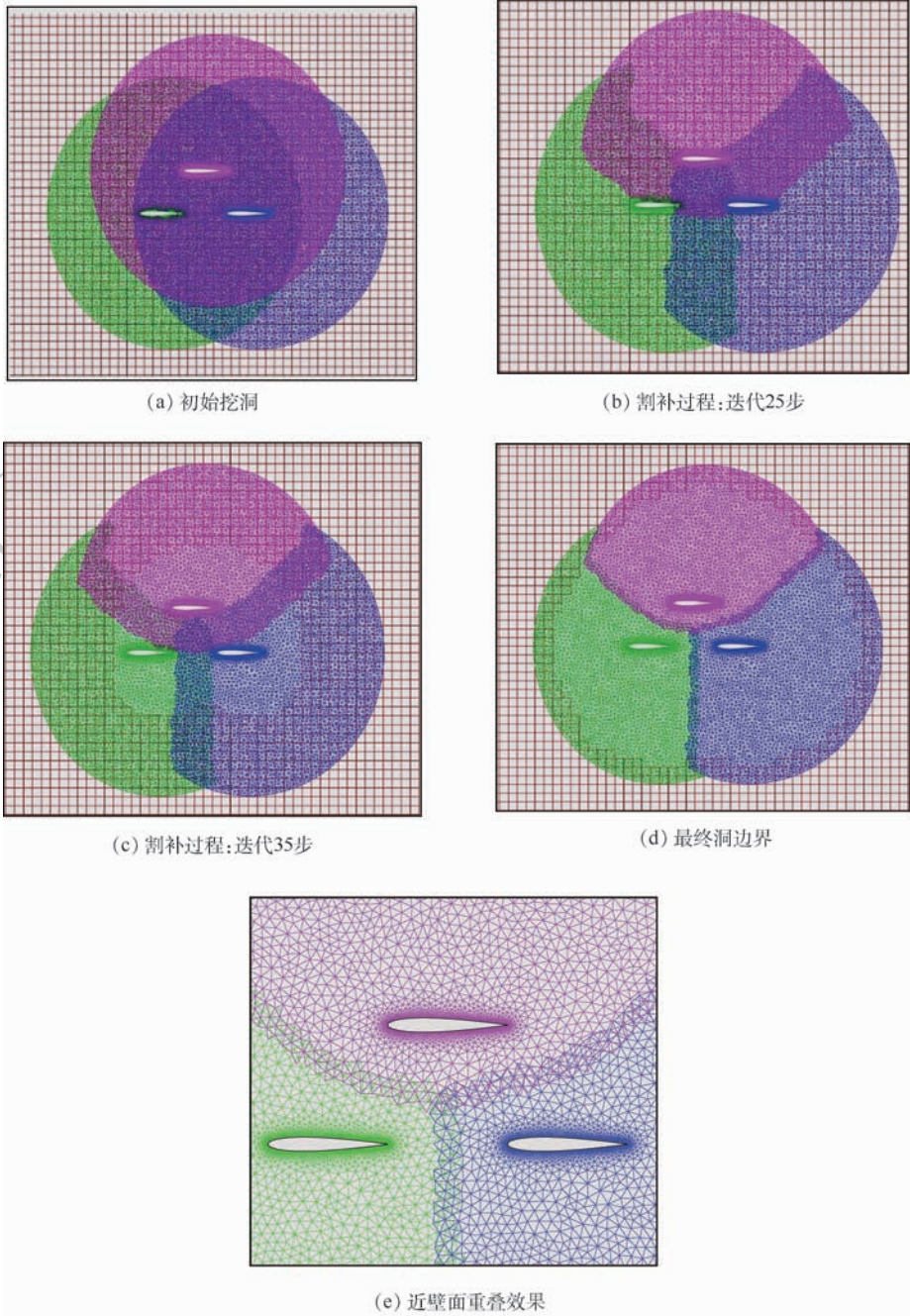


图 3 非结构重叠网格显式装配技术

Fig. 3 Explicit assembly technique of unstructured overset grid

网格系统由 3 块二维 NACA0012 翼型子网格和一块矩形背景网格组成。每块子网格包含 11 930 个三角形单元,背景网格包含 3 025 个四边形单元,共计 38 815 个单元。图 3(a)所示为以物面边界为初始挖洞边界的结果示意图,从图中可以看出,位于物面内部的网格均已被挖去,将已挖去网格单元标记为洞内单元。接着,将与洞内单元相邻的正常单元标记为洞边界单元。以所有洞边界单元为初始阵面开始推进,并使用基于相邻阵面的贡献单元搜索法搜索所有洞边界单元的贡献单元,比较洞边界点与其贡献单元的物面距大小,迭代更新洞边界,图 3(b)和图 3(c)分别表示迭代推进 25 步和 35 步时的效果图。可以看出,重叠区域逐渐减小。当所有洞边界点的物面距小于或等于其贡献单元的物面距时,停止迭代。最终的洞边界位置如图 3(d)所示。图 3(e)所示为 3 个翼型之间近壁面处的局部放大图。可以看出,洞边界位于不同物面的中间位置,重叠效果理想。为保证二阶计算精度,将与洞边界点相邻的 2 个洞内点标记为插值点。

使用传统隐式算法对该套网格系统进行了装配,其与本文算法的效率对比情况如表 1 所示。可以看出,本文算法所需进行的贡献单元搜索次数明显少于传统隐式算法,因而也具有更高的效率。本文所有计算基于 Intel Xeon CPU E5-2650 平台,主频为 2.00 GHz。

表 1 不同算法的效率比较

Table 1 Comparison of efficiency between different algorithms		
算法	贡献单元搜索次数	时间/s
本文算法	4 134	1.25
传统隐式算法	38 815	5.12

3 算 例

3.1 计算方法

本文采用有限体积法对 Navier-Stokes 控制方程进行离散求解,空间离散采用 Roe-FDS (Flux Difference Splitting) 格式、Piecewise Linear 插值方法,提供二阶空间离散精度;湍流模型采用 Menter SST (Shear Stress Transport) 模型;时间离散采用 LUSGS (Lower-Upper Symmetric Gauss-Seidel) 隐式计算方法。

3.2 30P30N 三段式机翼

对经典的二维三段式机翼 30P30N^[34]进行了数值模拟。来流马赫数 $Ma = 0.2$,雷诺数 $Re = 8.7 \times 10^6$,迎角 $\alpha = 8^\circ$ 。流场网格由 4 部分组成,

前缘缝翼网格包含 9 996 个三角形单元、主翼面网格包含 48 361 个三角形单元,后缘襟翼网格包含 25 203 个三角形单元,背景网格为 348 401 个四边形单元。采用本文算法后的重叠效果如图 4(a)所示。图 4(b)和图 4(c)分别为前缘和后缘的局部放大图。可以看出,重叠网格的洞边界清晰整洁,位于不同翼面的中间位置。图 5 所示为该三段翼型马赫数云图,从图中可看出,流场分布合理,重叠边界处等马赫线过渡光滑。图 6 为翼面压力系数 C_p 的数值与试验结果对比,可以看出计算结果与试验值吻合良好,表明本文算法准确可靠。

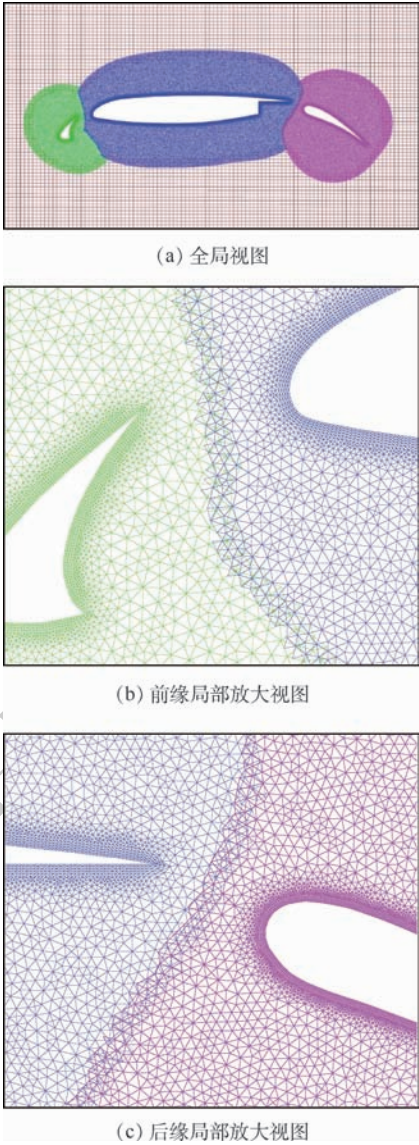


图 4 30P30N 机翼重叠网格图
Fig. 4 Overset mesh of 30P30N wing

3.3 Titan IV 运载火箭

为进一步验证本文算法在三维复杂流场中的适用性,对美国大力神 Titan IV 运载火箭^[35]超声速外流场进行了数值仿真。计算条件为:马赫数

$Ma = 1.6$, 迎角 $\alpha = 0^\circ$, 雷诺数 $Re = 1.15 \times 10^7$ 。图 7 所示为 Titan IV 火箭非结构重叠网格系统,由 3 部分组成:一个芯级网格以及 2 个助推器网格,共 4 532 680 个四面体单元。采用本文算法进行挖洞后的对称面网格重叠效果如图 8 (a) 所示,图 8(b)为助推器局部放大视图。对称面流场压力分布如图 9 所示。可以看出,网格重叠区域流场分布合理、光滑。图 10 给出了芯级中心线压力数值结果与试验值的比较,其中 p 为芯级中心线压强, p_{ref} 为来流远场压强, x 为沿芯级中心线

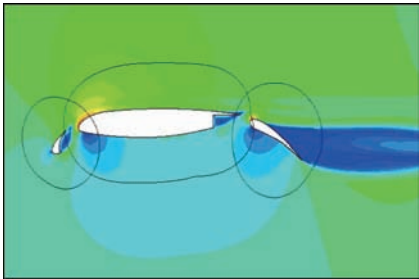


图 5 30P30N 翼面马赫数云图

Fig. 5 Mach number contour of 30P30N wing surface

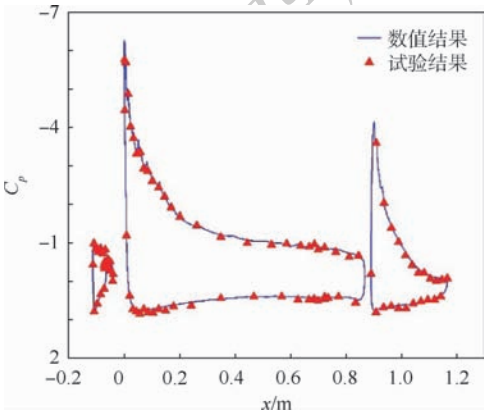


图 6 30P30N 翼面压力系数的数值与试验结果对比

Fig. 6 Comparison of pressure coefficient of 30P30N wing surface between numerical and experimental results

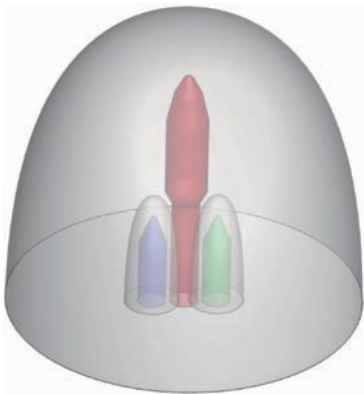
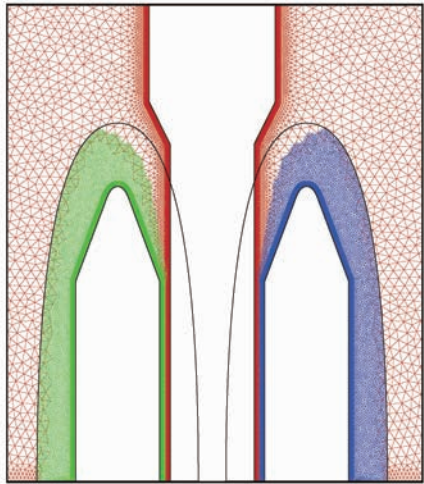
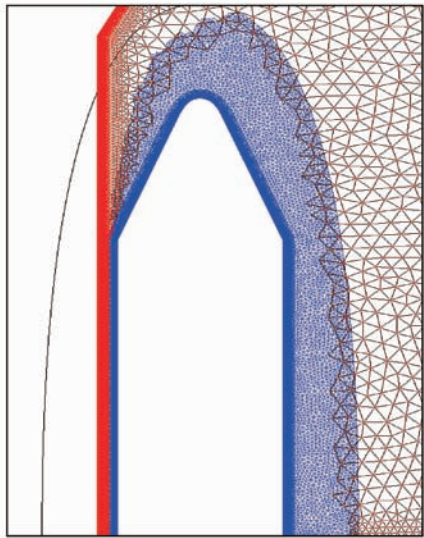


图 7 Titan IV 非结构重叠网格系统

Fig. 7 Unstructured overset mesh system for Titan IV



(a) 全局视图



(b) 局部放大视图

图 8 Titan IV 对称面重叠网格

Fig. 8 Overset mesh of Titan IV's symmetry plane

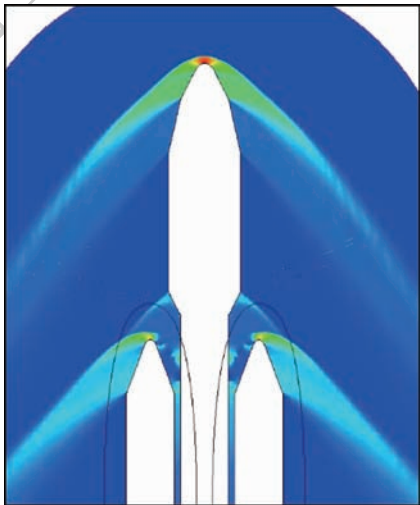


图 9 Titan IV 对称面压力云图

Fig. 9 Pressure contour of Titan IV symmetry plane

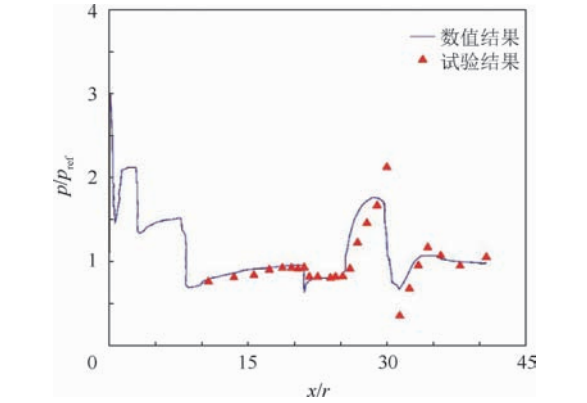


图 10 芯级中心线压力分布

Fig.10 Pressure distribution along rocket center line

位置, r 为芯级半径。计算结果与试验值吻合良好。芯级与助推级之间存在连接机构, 导致试验的峰值点略高, 且在连接机构附近数值结果与实验值之间的偏差稍大(图中 x/r 在 25 ~ 35 之间的区域)。

针对上述 2 种模型, 分别采用本文算法与传统隐式算法进行了重叠网格装配, 效率对比情况如表 2 所示。从表中可以看出, 本文所发展的改进型洞映射法的内存占用率更低; 同时相较于传统隐式装配算法, 本文算法具有更好的效率。

表 2 不同算法的性能比较

Table 2 Comparison of performance among different algorithms			
性能指标	算法	30P30N	Titan IV
内存占用/MB	本文算法	77.66	1023.77
	传统洞映射	151.23	1903.58
贡献单元搜索时间/s	本文算法	2.12	29.50
	传统隐式算法	10.35	126.43

4 结 论

本文通过对传统非结构重叠网格装配过程进行优化, 提出了一种显式非结构重叠网格装配算法, 并得出以下几点结论:

- 1) 通过对传统洞映射过程进行优化, 发展了一种改进型洞映射法, 并大大减少了洞映射过程中对计算机内存空间的占用。
- 2) 结合割补法与相邻单元搜索法的特点, 发展了一种基于相邻阵面的贡献单元搜索法, 优化了贡献单元搜索过程, 提高了搜索效率。
- 3) 与传统非结构重叠网格隐式装配过程相比, 本文所提显式装配过程具有更高的效率。
- 4) 通过 30P30N 三段式机翼和 Titan IV 运载火箭 2 个经典算例验证了本文算法的准确性与可靠性。

参考文献 (References)

[1] BENEK J A, STEGER J L, DOUGHERTY F C. A flexible grid embedding technique with applications to the Euler equations; AIAA-1983-1944 [R]. Reston, VA: AIAA, 1983.

[2] NAKAHASHI K, GUMIYA T. An intergrid boundary definition method for overset unstructured grid approach; AIAA-1999-3304 [R]. Reston, VA: AIAA, 1999.

[3] TOGASHI F, NAKAHASHI K, ITO Y, et al. Flow simulation of NAL experimental supersonic airplane/booster separation using overset unstructured grids; AIAA-2000-1007 [R]. Reston, VA: AIAA, 2000.

[4] TOGASHI F, ITO Y, MURAYAMA M, et al. Flow simulation of flapping wings of an insect using overset unstructured grid; AIAA-2001-2619 [R]. Reston, VA: AIAA, 2001.

[5] TOGASHI F, ITO Y. Extensions of overset unstructured grids to multiple bodies in contact; AIAA-2002-2809 [R]. Reston, VA: AIAA, 2002.

[6] TOGASHI F, ITO Y. Overset unstructured grids method for viscous flow computations; AIAA-2003-3405 [R]. Reston, VA: AIAA, 2003.

[7] LEE Y L, BAEDER J D. Implicit hole cutting-a new approach for overset grid connectivity; AIAA-2003-4128 [R]. Reston, VA: AIAA, 2003.

[8] XIAO T, QIN N, LUO D, et al. Deformable overset grid for multibody unsteady flow simulation [J]. AIAA Journal, 2016, 54 (8) : 2392-2406.

[9] XU X, WANG X D, ZHANG M, et al. A parallelized hybrid N-S/DSMC-IP approach based on adaptive structured/unstructured overlapping grids for hypersonic transitional flows [J]. Journal of Computational Physics, 2018, 371 (15) : 409-433.

[10] MISHRA A, BAEDER J D. Coupled aeroelastic prediction of the effects of leading-edge slat on rotor performance [J]. Journal of Aircraft, 2012, 53 (1) : 141-157.

[11] LEE Y, BAEDER J. High-order overset method for blade vortex interaction [C] // AIAA Aerospace Sciences Meeting & Exhibit. Reston, VA: AIAA, 2013: 1-10.

[12] LOHNER R, SHAROV D, LUO H. Overlapping unstructured grids; AIAA-2001-0439 [R]. Reston, VA: AIAA, 2001.

[13] LUO H. An overlapping unstructured grid method for viscous flows; AIAA-2001-2603 [R]. Reston, VA: AIAA, 2001.

[14] LANDMANN B, MONTAGNAC M. A highly automated parallel Chimera method for overset grids based on the implicit hole cutting technique [J]. International Journal for Numerical Methods in Fluids, 2011, 66 (6) : 778-804.

[15] XU J, CAI J, LIU Q, et al. Flow simulations by enhanced implicit-hole-cutting method on overset grids [J]. Journal of Aircraft, 2014, 51 (5) : 1401-1409.

[16] 田书玲. 基于非结构网格方法的重叠网格算法研究 [D]. 南京: 南京航空航天大学, 2008.

[17] TIAN S L. Investigation of overset unstructured grids algorithm [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2008 (in Chinese).

[17] CHO K W, KWON J H, LEE S. Development of a fully system-

- ized Chimera methodology for steady/unsteady problems [J]. Journal of Aircraft, 1999, 36(6): 973-980.
- [18] WEY T C. Development of a mesh interface generator for overlapped structured grids: AIAA-1994-1924 [R]. Reston, VA: AIAA, 1994.
- [19] BENEK J A, STEGER J L, DOUGHERTY F A. A flexible grid embedding technique with application to the Euler equations: AIAA-1983-1944 [R]. Reston, VA: AIAA, 1983.
- [20] LABOZZETTA W F, GATZKE T D. MACGS-towards the complete grid generation system: AIAA-1994-1923 [R]. Reston, VA: AIAA, 1994.
- [21] CHIU I T, MEAKIN R. On automating domain connectivity for overset grids: AIAA-1995-0854 [R]. Reston, VA: AIAA, 1995.
- [22] CRABILL J A, WITHERDEN F D, JAMESON A. A parallel direct cut algorithm for high-order overset methods with application to a spinning golf ball [J]. Journal of Computational Physics, 2018, 374: 692-723.
- [23] MEAKIN R L. Object X-rays for cutting holes in composite overset structured grid: AIAA-2001-2537 [R]. Reston, VA: AIAA, 2001.
- [24] YANG X, YANG A, SI J. Efficient numerical techniques for simulating a rotorcraft flow field with overlapping grids [J]. AIAA Journal, 2015, 53(5): 1372-1383.
- [25] NINI M, MOTTA V, BINDOLINO G, et al. Three-dimensional simulation of a complete vertical axis wind turbine using overlapping grids [J]. Journal of Computational & Applied Mathematics, 2014, 270: 78-87.
- [26] DI Y C, GAO G, XU J L, et al. Time-accurate simulation of the aircraft external store separation [J]. Applied Mechanics and Materials, 2013, 444-445: 854-859.
- [27] ZHAO Y, YU B, YU G. A new two-dimensional hybrid grid generation method based on improved hole cutting [J]. International Communications in Heat and Mass Transfer, 2014, 54: 103-114.
- [28] 淮洋, 郝海兵, 姚冰, 等. 一种改进型洞映射法 [J]. 航空计算技术, 2015, 45(2): 31-34.
- HUAI Y, HAO H B, YAO B, et al. An improved method of hole-map [J]. Aeronautical Computing Technique, 2015, 45(2): 31-34 (in Chinese).
- [29] 袁武, 阎超, 席柯. 洞映射方法的研究和改进 [J]. 北京航空航天大学学报, 2012, 38(4): 563-568.
- YUAN W, YAN C, XI K. Investigation and enhancement of hole mapping method [J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(4): 563-568 (in Chinese).
- [30] 王文, 阎超, 袁武, 等. 鲁棒的结构网格自动化重叠方法 [J]. 航空学报, 2016, 37(10): 2980-2991.
- WANG W, YAN C, YUAN W, et al. A robust and automatic structured overlapping grid approach [J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(10): 2980-2991 (in Chinese).
- [31] BONET J, PERAIRE J. An alternating digital tree (ADT) algorithm for 3D geometric searching and intersection problems [J]. International Journal for Numerical Methods in Engineering, 1991, 31: 1-17.
- [32] 董国国, 王立强. 基于动态结构重叠网格的三维外挂物分离的数值模拟 [J]. 航空兵器, 2011(2): 21-25.
- DONG G G, WANG L Q. Three-dimensional store separation simulation using moving structured overlapping grid technique [J]. Aero Weaponry, 2011(2): 21-25 (in Chinese).
- [33] 田书玲, 伍贻兆, 夏健. 用动态非结构重叠网格法模拟三维多体相对运动绕流 [J]. 航空学报, 2007, 28(1): 46-51.
- TIAN S L, WU Y Z, XIA J. Simulation of flows past multi-body in relative motion with dynamic unstructured overset grid method [J]. Acta Aeronautica et Astronautica Sinica, 2007, 28(1): 46-51 (in Chinese).
- [34] CHIN V D, PETER D W, SPAID F W. Flow field measurements about a multi-element airfoil at high Reynolds numbers: AIAA-1993-3137 [R]. Reston, VA: AIAA, 1993.
- [35] BIEDRON S, JOHNSON C T, WANG L V. Simulations using a concurrent implicit Navier-Stokes solver: AIAA-1995-223 [R]. Reston, VA: AIAA, 1995.

作者简介:

宣传伟 男, 博士研究生。主要研究方向: 气动弹性力学。

韩景龙 男, 博士, 教授, 博士生导师。主要研究方向: 气动弹性力学。

Explicit assembly algorithm of unstructured overset grid

XUAN Chuanwei, HAN Jinglong*

(State Key Laboratory of Mechanics and Control of Mechanical Structures, College of Aerospace Engineering,
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To solve the problem that the hole mapping method occupies too much physical memory, an improved hole mapping method was developed. Based on the neighbor-to-neighbor search algorithm, a donor search method based on adjacent front was developed. An explicit assembly algorithm of unstructured overset grid was presented by combining the cut-paste method with the implicit cutting technique. First, the algorithm generated a set of Cartesian grids surrounding the wall surface. Second, those Cartesian cells intersecting the wall surface were stored. Finally, relative positions of the stored Cartesian cells were used to determine whether a grid point was inside the wall. After successfully determining all the grid points inside the wall, the current fringe grid points were used as the initial front, and the overlapping area was optimized by the wall distance of each grid point to generate the final interpolation boundary. The proposed explicit algorithm optimizes the traditional implicit assembly process of unstructured overset grid. It features in low physical memory occupation, low cost of donor searching and high computational efficiency. The accuracy and applicability of the proposed explicit method were verified by two typical complex flow examples.

Keywords: overset grid; cut-paste method; hole mapping; unstructured grid; wall distance

Received: 2019-01-17; **Accepted:** 2019-04-05; **Published online:** 2019-04-30 16:02

URL: kns.cnki.net/kcms/detail/11.2625.V.20190429.1444.002.html

Foundation items: National Natural Science Foundation of China (11472133); the Priority Academic Program Development of Jiangsu Higher Education Institutions

* **Corresponding author.** E-mail: hjlae@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0028

阵列射流冲击复合不同肋化表面的沸腾特性



张添^{1,2}, 张畅^{1,2}, 谢荣建^{1,2}, 董德平^{1,2,*}

(1. 中国科学院上海技术物理研究所, 上海 200083; 2. 中国科学院大学, 北京 100049)

摘

要: 阵列射流冲击冷却技术可以有效地解决高热流密度器件的散热问题, 为了验证受冲击表面强化传热结构对优化两相射流冷却性能的有效性, 结合高速显微摄像手段, 研究了不同肋化表面结构形态对受限式阵列射流冷却的流动、传热特性的影响。设计了2种含不同肋化表面形态: 光滑切割肋(0.6 mm × 0.6 mm × 1.0 mm)、外覆多孔烧结层的粗糙肋(粒径为73 ~ 53 μm)。实验使用无水乙醇为工质, 以光滑表面的射流冷却热沉为对照组, 入口温度均为20℃, 在固定工质流量7.5 mL/s下, 随着加热热流密度由5 W/cm²增加至100 W/cm²时, 热沉的换热系数均持续上升但增幅逐渐减小, 未明显观察到沸腾相变的发生。对固定热流密度82.6 W/cm²、80.5 W/cm²改变工质流量(射流雷诺数)的实验工况, 当工质流量由7.5 mL/s逐渐降低至1.0 mL/s时, 可以非常明显地观测到射流腔内部工质由分层湍流逐步进入泡状流、弹状流及环状流, 其分别对应起始沸腾区、核态沸腾区及膜态沸腾区。

关键词: 高热流密度; 受限式阵列射流; 多孔介质; 沸腾强化; 可视化

中图分类号: TK124

文献标识码: A

文章编号: 1001-5965(2019)10-2035-09

伴随着机载设备、光电子元器件高集成化、小型化发展趋势, 各类元件功耗与发热量的同步增加, 由此而产生的百瓦/平方厘米级高热流密度散热需求愈加迫切^[1-3], 行之有效的散热措施成为了保证电子设备正常、可靠运行的核心技术之一^[4-7]。针对不同散热需求差异, 常见散热技术其散热能力由小到大分别为: 外加翅片自然对流散热、强迫风冷散热、微通道热沉水冷散热^[3]、喷雾冷却散热^[6]和射流冷却散热^[2-3]等, 以上技术均利用了流体的自然/强迫对流换热的换热系数较高的特点。

根据文献[8]中提出的“场协同”原理可知, 通过改变速度方向与热流方向的夹角, 可以控制对流换热的强度, 且两者夹角越小换热强度越大。对于射流冲击冷却, 在孔口下方滞止点附近, 速度

梯度与温度梯度夹角较小, 此处的换热系数高于贴壁射流处^[9]。基于以上特点, 射流冷却散热技术可用于高热流密度热源散热, 并达到较高换热系数。现阶段, 针对射流冷却技术的研究主要集中于不同冷却工质种类^[10]、射流孔板结构^[11-12]、射流距离、射流表面形貌^[13-14]及系统参数控制等对热沉性能的影响情况^[15-18]。

Chang等^[10]使用3M公司的PF5060介电液体作为工质, 对不同雷诺数($Re = 2\,000, 3\,000, 5\,000$)下, 采用不同射流高度、孔径比的单孔受限射流冷却装置的性能进行了可视化实验测试, 结果表明沸腾的起始点(Onset Nuclear Boiling, ONB)与 Re 成正比, 与射流高度、孔径比成反比。Sung和Mudawar^[11-12]对采用矩形狭缝、等截面/渐变截面圆形阵列射流孔对阵列射流复合微通道

收稿日期: 2019-01-22; 录用日期: 2019-02-16; 网络出版时间: 2019-06-13 14:40

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190613.0718.001.html

* 通信作者: E-mail: dongdeping@mail.sip.ac.cn

引用格式: 张添, 张畅, 谢荣建, 等. 阵列射流冲击复合不同肋化表面的沸腾特性[J]. 北京航空航天大学学报, 2019, 45(10): 2035-2043. ZHANG T, ZHANG C, XIE R J, et al. Boiling characteristics of array jet impingement with various pin-finned surfaces[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2035-2043 (in Chinese).

冷却模组在单相、两相工况下的性能进行了相关数值模拟及实验研究。以 HFE 7100 电子氟化液为工质,当入口温度 $T_{in} = 20^{\circ}\text{C}$ (过冷度约 80°C) 时可实现 CHF (Critical Heat Flux) 最高可达 $1\,127\text{ W/cm}^2$ (热源尺寸为 $1\text{ cm} \times 2\text{ cm}$)。在射流速度较低时,CHF 主要由微通道流特性决定,当速度较高时,其主要由射流特性决定。Ruander 和 Viond^[17] 使用 FC-72 为工质,对单孔冲击光滑表面的浸没射流冷却性能进行了研究,其可视化实验结果表明,整个表面均实现充分沸腾所需的时间与初始壁面过热度大小成反比。

上述研究表明,合理的射流孔板结构、强化射流表面形貌设计可以在一定程度上提升射流热沉的冷却能力。同时,笔者考虑到由于工质汽化时相变潜热的存在,能够在较小的壁面温升下增加散热能力。因此,在前人研究基础上,设计了 2 种复合不同肋化表面的可视化阵列射流热沉:光滑切割针肋、外覆多孔烧结层的粗糙针肋。同时结合高速摄像机,以无水乙醇为工质,对其在固定入口温度、固定背压下的换热能力进行了性能测试及可视化实验,对 2 种肋化表面结构相较光滑射流表面对于强化沸腾换热的有效性进行验证。

1 复合阵列射流热沉设计

本文设计了 2 种不同可视化复合阵列射流热沉,热沉整体结构示意图如图 1 所示。热沉整体呈左右轴对称结构,工质由上部入口(蓝色箭头处)进入热沉后首先流经整流段(填充不锈钢丝网),实现均匀分配通过各射流孔的工质流量;然后经射流孔板入射至无氧铜基板表面,此时冷流体与热壁面完成能量交换;最后再由两侧出口(红色箭头处)分别流出。其中,盖板和支撑板层间使用硅橡胶垫片,结合螺栓连接实现机械密封。

1.1 整流腔设计

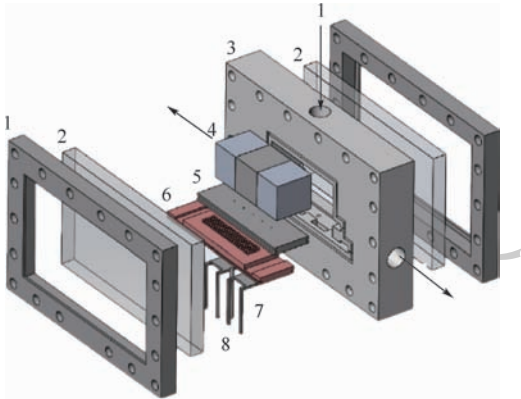
整流腔用于实现通过各射流孔的工质流量的均匀分配,采用了填充不同目数丝网的方案,将整流腔内分成阻力损失不同的三等份。中间部分填充 150 目不锈钢丝网,增大流动阻力;两侧选择孔隙率更大、流动阻力更小的 100 目不锈钢丝网,引导中心位置工质向两侧阻力较小区流动。使用 FLUENT 对整流腔流体流动情况进行了仿真分析,结果显示上述整流腔设计可以有效避免入口正下方区域工质流速过大。

1.2 受冲击表面肋化结构设计

可视化复合阵列射流热沉采用 1×5 的阵列

射流孔板,底部有效肋化区域的投影面尺寸为 $6\text{ mm} \times 30\text{ mm}$,形成 5 个射流单元,如图 2 所示。每个射流单元对应基板表面 4×4 的阵列针肋。在 2 种肋化结构方案中,针肋尺寸均为 $0.6\text{ mm} \times 0.6\text{ mm} \times 1.0\text{ mm}$ (长 $\times$ 宽 $\times$ 高),粗糙肋化表面采用在机械加工针肋表面再烧结一层多孔铜粉颗粒层。其中,铜粉粒径为 $75 \sim 53\text{ }\mu\text{m}$,多孔层总厚度约为 2 倍的粒径。射流结构参数见表 1。

针肋结构可以通过破坏层流底层,在较低雷诺数下增加扰动促进湍流的发生;烧结铜粉多孔介质层表面可以形成多个汽化核心,进而促进相变的发生。因此,2 种肋化表面可以同时实现对单相、两相工况的性能提升。粗糙肋化表面实物图及扫描电子显微镜 (SEM) 照片见图 3。



1—不锈钢盖板;2—石英玻璃视窗;3—不锈钢支撑板;
4—不锈钢丝网;5—不锈钢孔板;6—无氧铜基板;
7—K型铠装测温热电偶;8—陶瓷加热片。

图 1 可视化复合阵列射流热沉结构示意图
Fig.1 Schematic diagram of visualized hybrid array jet heat sink structure

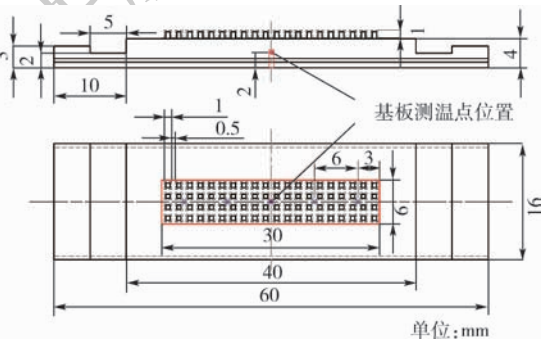


图 2 肋化射流表面的基板尺寸示意图
Fig.2 Schematic diagram of baseplate dimension of pin-fined jet surface

表 1 阵列射流结构参数

Table 1 Array jet structure parameters mm				
参数	射流孔径	孔间距	孔板厚度	射流距离
数值	0.5	6	3	3

注:射流距离为射流孔板下表面至基板上表面(肋底部)。

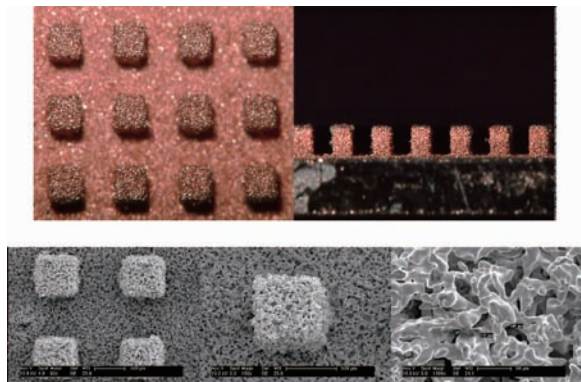


图 3 粗糙肋化表面实物图及 SEM 照片

Fig. 3 Photographs and SEM images of porous coated pin-finned surface

2 实验系统

实验系统可分为性能测试及可视化拍摄两部分。其中,性能测试系统示意图如图 4 所示,采用闭式泵流体回路作为测试系统。

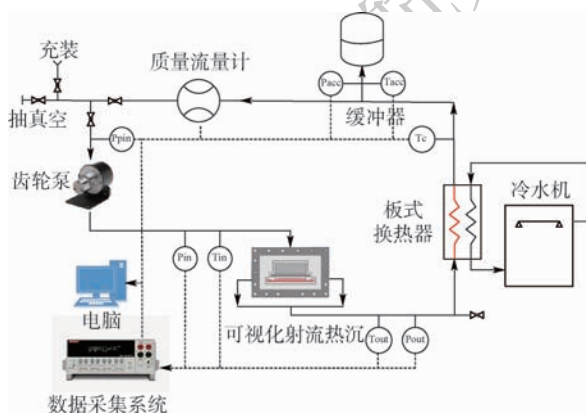


图 4 性能测试系统示意图

Fig. 4 Schematic diagram of performance test system

2.1 性能测试系统

性能测试系统中主要部件包括:可视化射流热沉、板式换热器、质量流量计、齿轮泵、缓冲器和数据采集系统。系统工质使用机械泵驱动循环,经过滤器进入测试段,工质受热后进入板式换热器进行冷凝冷却,最终流回泵输入端完成整个循环。热沉使用陶瓷加热片作为模拟热源,实验时贴于基板下表面。系统使用二次冷却回路,板式换热器热侧为循环工质,冷侧连接恒温冷水机,通过控制循环水温度对测试系统温度进行控制。

实验时,保持射流结构和系统输入条件不变,即热沉孔径、孔数、孔间距、射流距离、射流速度(系统工质流量)、射流温度指标保持不变。观测在使用不同表面强化方式时,射流区域内液相工质随着加热热流密度的变化的流动、传热特性。

2.2 可视化拍摄系统

可视化实验的核心观测是与常规光滑表面基板相比较,2 种复合不同肋化表面的阵列射流冷却热沉内工质的流动状态及流型转化特点。

拍摄系统采用 KYENCE 公司 VW-600C 彩色高速显微摄像机主机、VW-Z2 长距离高清显微变焦镜头(60x)、主体部分内置专用金卤灯光源。

3 数据处理与误差分析

3.1 数据处理

施加于基板的有效加热热流密度 Φ 通过陶瓷加热片供电功率 UI (电压 U 与电流 I 的乘积)经校核后折算得到。在相同的环境条件下,热沉向环境的漏热量仅与装置外表面的温度有关,与热沉内部工质状态无关。使用单相稳态热平衡关系式,计算工质实际吸收的加热功率 Q 与陶瓷加热片供电功率 UI 的比值作为折算系数 ξ 。

$$\Phi = \frac{Q}{A} = \xi \frac{UI}{A} \quad (1)$$

式中: A 为热源面积。

将基板上表面与射流工质间的换热温差记为 ΔT_m ,壁面温度 T_w 由傅里叶导热定律^[9]计算得到:

$$\Delta T_m = T_w - T_{in} \quad (2)$$

$$T_w = T - \frac{\Phi \Delta x}{\lambda} \quad (3)$$

式中: T_{in} 为工质入口温度; T 为正对中部射流下方,基板内布置的热电偶探头实测得到的固体温度; Δx 为测温点所在位置距基板表面的距离, $\Delta x = 2 \text{ mm}$; λ 为基板无氧铜材料的热导率,取 $\lambda = 381 \text{ W/(m} \cdot \text{K)}$ 。

使用壁面温度 T_w 与工质入口温度 T_{in} 的差值 ΔT_m 计算对流换热系数 h ,作为衡量热沉综合传热性能的指标:

$$h = \frac{\Phi}{\Delta T_m} \quad (4)$$

此外,为了考核不同射流速度 U_{jet} 或工质流量 q_v 对热沉内部流动及传热特性的影响,采用无量纲雷诺数 Re 综合考虑二者的影响。

$$Re = \frac{U_{jet} D_{jet}}{\nu} \quad (5)$$

$$U_{jet} = \frac{q_v}{A_{jet}} = \frac{q_v}{N \cdot \frac{1}{4} \pi D_{jet}^2} \quad (6)$$

$$q_v = \frac{q_m}{\rho} \quad (7)$$

式中: D_{jet} 为射流孔直径; A_{jet} 为所有射流孔面积和;射流孔数 $N = 5$; ρ 为工质密度; q_m 为工质质量

流量; ν 为运动黏度。以工质入口温度 T_{in} 为上述物性参数的定性温度。

使用进、出口工质压力损失 ΔP 作为衡量热沉流动阻力特性的指标。

$$\Delta P = P_{in} - P_{out}$$

(8)

式中: P_{in} 和 P_{out} 分别为测量得到的热沉进、出口工质压力。

3.2 误差分析

假设不存在粗大误差与系统误差的前提下,根据系统中各测量仪表的精度得到直接误差 σ_{xi} ,基于误差分析理论采用误差传递函数公式(9),对各间接误差 σ_y 进行分析计算:

$$\sigma_y = \sqrt{\sum_{i=1}^n \left(\frac{\partial f(x)}{\partial x_i} \right)^2 \sigma_{xi}^2}$$

(9)

式中: x_i 为直接测量参数; $f(x)$ 为间接测量参数。

系统中温度测量分别使用了:2种K型热电偶(底面温度、工质温度),精度分别为 $\pm 2\%$ 、 $\pm 0.75\%$;压阻式压力传感器,精度 $\pm 5\%$;科里奥利质量流量计,液体流量测量精度 $\pm 0.2\%$,密度测量误差 $< \pm 5\text{ kg/m}^3$ 。根据各参数可能出现的最大值,最终计算得各参数误差如表2所示。

表 2 实验不确定度

Table 2 Uncertainties of experiment		%
误差	参数	数值
直接相对误差	T	± 2
	$T_{in}、T_{out}$	± 0.75
	$x、D_{jet}$	± 2
	$q_m、q_v$	± 0.2
	ρ	± 0.6
	P	± 0.25
	Φ	± 0.13
	T_w	± 1
	ΔT_m	± 0.96
	h	± 2.26
间接相对误差	U_{jet}	± 0.2
	Re	± 2.2
	ΔP	± 0.25

4 结果与讨论

实验使用无水乙醇为工质,各工况下均控制工质入口温度 $T_{in} = (20 \pm 1)^\circ\text{C}$ 不变。根据性能测试系统采集到的入口温度/压力、出口温度/压力、底板温度等系统测量数据,可以得到相关热沉换热特性(如:换热系数、壁面过热度、压力损失等)。包含对照组在内,光滑表面、光滑肋化表面和粗糙肋化表面3个不同可视化热沉分别命名为KS0、KS1和KS2。

各实验工况分为两大类:第1类基于常规沸腾曲线测试方法,测试段入口工质流量保持不变,逐渐增加加热热流密度;第2类选定加热热流密度不变,逐步减小入口工质流量。

实验的过程中,当系统各温度、压力指标趋于稳定,则认为达到热平衡状态,此时进行针对热沉射流腔的高速拍摄,用以观测各工况下工质的流型转化特性。

4.1 固定工质流量

首先,对3种可视化热沉在保持相同的入口条件(入口温度 $T_{in} = (20 \pm 1)^\circ\text{C}$ 、体积流量 $q_v = (7.5 \pm 0.5)\text{ mL/s}$)下,热沉性能随加热热流密度 Φ 增加($\Phi = 5 \sim 100\text{ W/cm}^2$)时的变化情况进行实验研究。

图5为壁面过热度随热流密度的变化。实验结果显示:在 q_v 保持固定,只增加 Φ 时, ΔT_m 随 Φ 增加而逐渐增大。大部分工况下KS1、KS2性能均优于光滑射流表面结构KS0,在相同热流密度下换热温差更小(见图5)、换热系数更大(见图6)。当 $\Phi < 70\text{ W/cm}^2$ 时, ΔT_m 值 $\text{KS0} > \text{KS1} > \text{KS2}$;当 Φ 进一步增加时,三者性能趋于一致,壁面温度趋近于对应压力下的饱和温度。

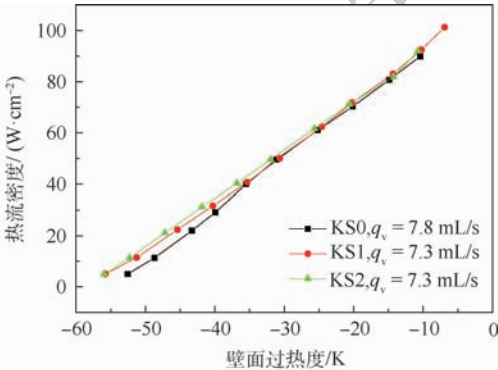


图 5 壁面过热度随热流密度的变化

Fig. 5 Influence of heat flux on wall surface degree of superheat

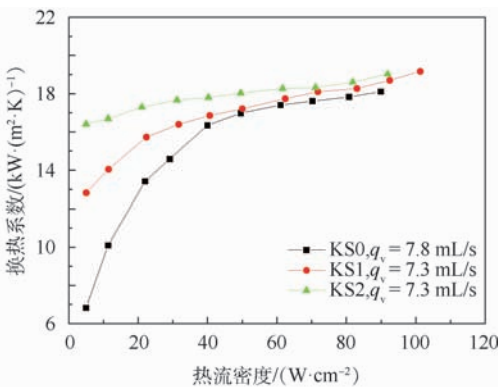


图 6 换热系数随热流密度的变化

Fig. 6 Influence of heat flux on heat transfer coefficient

图 6 为换热系数随热流密度的变化。由图中可知:① $h-\Phi$ 曲线在初始工况下,2 种肋化表面 KS1、KS2 均大于对照组 KS0 的 h ,且 h 值 $KS2 > KS1 > KS0$;②各热沉 h 均随着热流密度 Φ 增加而增大,其中 KS0 换热系数近似抛物线增大,增长率逐渐放缓,而肋化表面 KS1、KS2 的换热系数 h 从起始小热流密度工况始终能保持较高值,且随 Φ 增大后续小幅度波动增加,增长率小幅度增加。

可视化实验结果显示,在上述各工况下,对照组 KS0 内部均为液相,未出现明显相变。而对于 2 种肋化表面的热沉 KS1、KS2,随着加热热流密度 Φ 的增加,可以观察到射流腔体内部由于工质温度梯度的存在,导致了冷、热流体分层流动及相互掺混形成湍流。但由于此时工质流量相对于总发热功率 Q 较大,在测试工况内,射流腔体内部未见明显的沸腾现象发生。与常规沸腾曲线单相段传热系数 (Heat Transfer Coefficient, HTC) 随热流密度几无变化现象相悖,结果表明在热流密度小于 60 W/cm^2 时,换热系数随着热流密度的增加而增大。分析认为,上述现象产生的原因主要是本文采用的射流速度 (射流雷诺数) 较小,当加热热流密度较小时,腔体内工质主要保持层流状态,此时换热强度和换热系数均较小,对于无表面强化结构的 KS0,其换热系数最小;随着加热热流密度的增加,距离加热壁面距离不同位置处工质温度不同,对应的黏度、密度等均发生变化,此时流体掺混增加,工质与壁面间由单相层流换热转换为单相湍流换热,因此出现了换热系数上升。但单相湍流换热能力不会无限制增长,因此传热系数值最终趋于常数。

4.2 固定加热热流密度

实验测试了 2 种不同肋化表面的热沉 KS1、KS2 在固定热流密度 Φ 时的传热性能及压力损失随入口工质流量 q_v 变化的情况。其中工质入口温度 $T_{in} = (20 \pm 1)^\circ\text{C}$ 保持不变, $\Phi = 82.6 \text{ W/cm}^2$ (KS1), $\Phi = 80.5 \text{ W/cm}^2$ (KS2), $q_v = 1.0 \sim 7.5 \text{ mL/s}$,其他系统参数保持不变。换热系数 h 随射流雷诺数 Re 的变化情况见图 7。

使用射流雷诺数对应不同的工质流量,对比不同肋化表面结构受工质流量影响的情况。根据定热流密度变工质流量的实验结果可知,热沉整体压力损失 ΔP 与射流雷诺数 Re 成正比,减小系统流量可以显著降低热沉压力损失,即可降低对泵驱动能力的要求,如图 8 所示。综合图 7 及图 8 实验结果可以看出,虽然降低工质流量,热沉

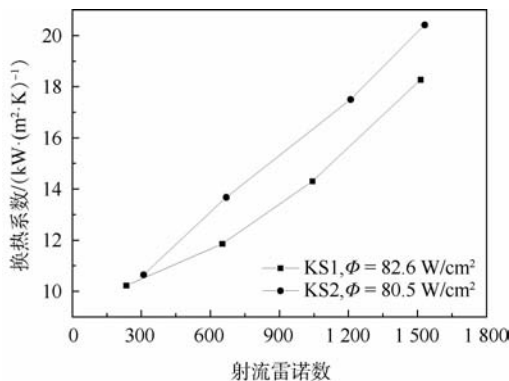


图 7 换热系数随射流雷诺数的变化

Fig. 7 Influence of jet Reynolds number on heat transfer coefficient

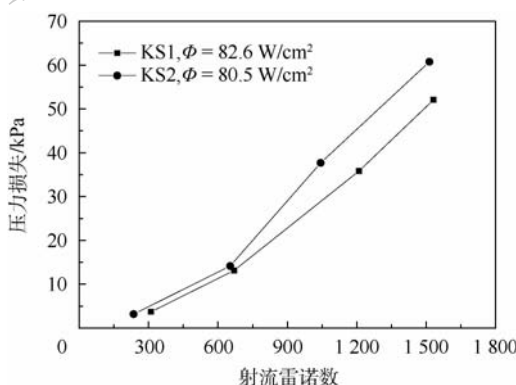


图 8 压力损失随射流雷诺数的变化

Fig. 8 Influence of jet Reynolds number on pressure loss

换热系数会出现一定程度下降,但在一定范围内,可以在满足需求的前提下降低热沉压力损失。

可视化实验观测结果分为稳态分析和瞬态分析两部分。稳态分析侧重于比较不同工况间的流动状态差异,瞬态分析侧重于展示某一工况时腔体内气泡生成随时间的变化情况。

4.2.1 KS1 可视化结果

对于 KS1,固定热流密度 $\Phi = 82.6 \text{ W/cm}^2$ 不变,观察在 4 种不同流量: $q_v = 7.3, 5.0, 3.1, 1.1 \text{ mL/s}$ 时左半边射流腔体内部工质流动情况,结果如图 9 所示。由图中可知:①随着 q_v 的降低,射流腔体上部及出口处均处于单相状态,仅在肋间通道底部出现一定的气体薄层,如图 9(b) 中 A 处所示;②在发生相变之前,由于工质在垂直基板表面方向出现明显的温度梯度,射流腔体内可观察到明显的分层流动产生,从图 9(c) 可以观测到,在针肋上方,工质流动的混乱程度远大于下游靠近出口处。上述现象表明在底部针肋结构的作用下,一方面可以强化液态工质与固体基板间的热交换,另一方面也强化了各层流体工质的掺混。③进一步降低工质流量至 $q_v = 1.1 \text{ mL/s}$,结果如

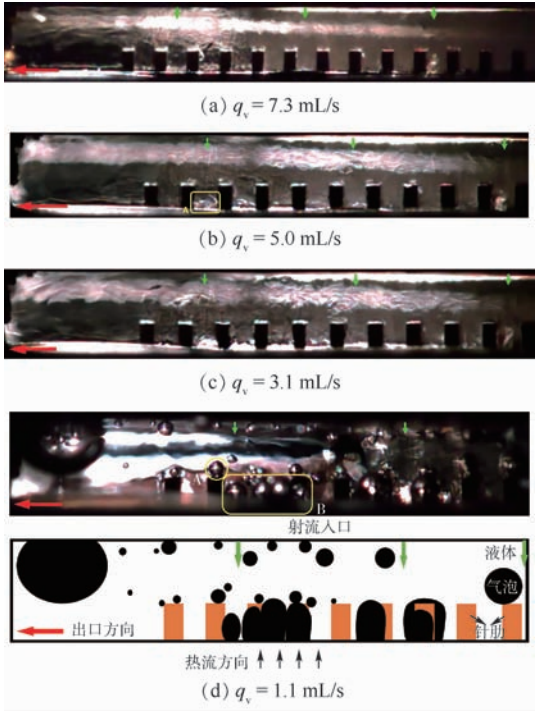


图 9 $\Phi = 82.6 \text{ W/cm}^2$ 、流量 q_v 减小的稳态工况气泡分布
Fig. 9 Images of bubble distribution with decreasing flux q_v and $\Phi = 82.6 \text{ W/cm}^2$ at steady-state conditions

图 9(d) 所示,当系统达到热平衡后,观察可见在针肋的顶部、侧面(A 处)及肋间底部(B 处)均会周期性的生成气泡,即为泡状流。其中,肋顶及肋侧气泡尺寸较小且发生速率较大,肋间底部气泡尺寸较大且生成速率较小。

详细分析了 KS1 在 $q_v = 1.1 \text{ mL/s}$ 工况下针肋顶部、侧部、肋间底部不同位置工质流型处随时间 t 变化的情况如图 10 所示,并综合分析了气泡的形态、生成周期等差异。A 处为针肋顶部、侧部气泡生长随时间变化的规律,可以观察到:①当小气泡在肋顶部生成后会持续长大,如图 10(a) 和 (b) 所示;②由于液态工质的扰动,气泡的位置会发生一定移动但始终未脱离固体壁面如图 10(c) 所示;③随着相邻位置气泡持续生长,当两气泡相遇,会“融合”成为一个更大的气泡,如图 10(d) 和 (e) 所示;④“融合”后的大气泡会受到更大的浮力作用,最终脱离固体壁面溢出。B 处为针肋底部,气泡呈竖直柱状生长,单个气泡生成至溢出周期更长。

4.2.2 KS2 可视化结果

对于 KS2,固定热流密度 $\Phi = 80.5 \text{ W/cm}^2$,在 4 种不同流量: $q_v = 7.4, 5.8, 3.2, 1.5 \text{ mL/s}$ 时,射流腔体内部工质流动情况如图 11 所示。可视化观测结果显示:①与 KS1 相同,在 q_v 较小时,如图 11(a) 所示,射流腔内工质保持液相,仅出现了

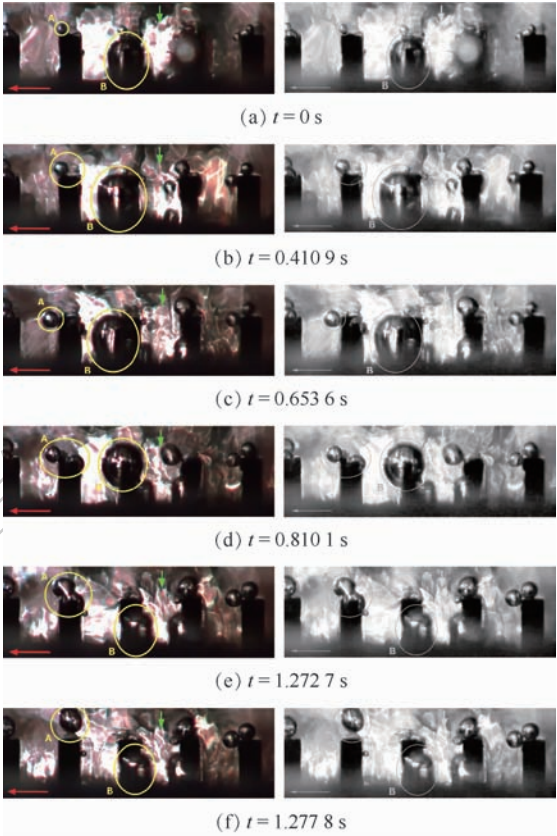


图 10 $\Phi = 82.6 \text{ W/cm}^2$ 、 $q_v = 1.1 \text{ mL/s}$ 瞬态工况气泡分布
Fig. 10 Images of bubble distributions with $\Phi = 82.6 \text{ W/cm}^2$ and $q_v = 1.1 \text{ mL/s}$ at transient conditions

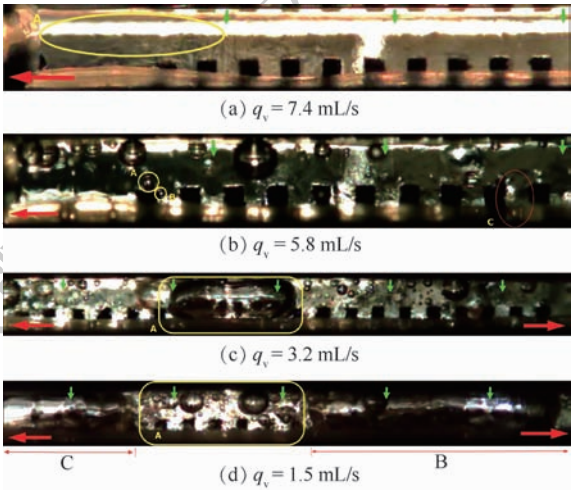


图 11 $\Phi = 80.5 \text{ W/cm}^2$ 、流量 q_v 减小的稳态工况气泡分布
Fig. 11 Images of bubble distribution with decreasing flux q_v and $\Phi = 80.5 \text{ W/cm}^2$ at steady-state conditions

由于密度不同而导致的分层流动现象;②随着 q_v 的减小,当在 $q_v = 5.8 \text{ mL/s}$ 时,如图 11(b) 所示,可观察到明显的相变发生,远离射流孔的针肋顶部(A 处)和侧部(B 处)分别出现孤立小气泡,肋间通道(C 处)有更大尺寸的气泡生成,此时已达

到沸腾起始点,流动状态以泡状流为主;③当 $q_v = 3.2 \text{ mL/s}$ 时,如图 11(c)所示,由于 q_v 的减小,射流速度及滞止压力均下降,相变产生的气泡无法及时脱离壁面,相邻小气泡“融合”成为更大的气柱,由于此时热沉入口压力有一定波动,气柱会周期性的被排出并重新产生,此时整个射流腔体内部形成与通道内弹状流流动状态类似的气柱;④当 q_v 进一步减小至 $q_v = 1.5 \text{ mL/s}$ 时,如图 11(d)所示,此时冷流体射流已无法直接接触底部壁面,冷流体无法及时补充,靠近左右对称轴线 A 处的工质继续保持核态沸腾状态,随着饱和工质在下游持续吸热,上游孤立气泡逐渐合并形成连续“气塞”,在水平方向生成更大的气柱(B、C 区),下游处工质的流动以环状流为主。

对比 KS1 和 KS2 定热流密度 Φ 变工质流量 q_v 工况的可视化观测结果可知,在降低工质流量的过程中,KS2 可以更早地观测到持续的气泡生成,即相较 KS1 其更快进入两相工况,由此证明,多孔层的存在可以更有效的强化换热,促进两相工况的发生。同时,对比 2 种不同强化热沉可知,沸腾产生气泡的尺寸只与相变发生的位置有关,与有无多孔层无关。

图 12 展示了 KS2 在 $\Phi = 80.5 \text{ W/cm}^2$ 、 $q_v = 5.8 \text{ mL/s}$ 工况下的气泡生长规律。A 处为针肋顶部,可以观察到:①气泡会周期性的生成并逐步增大(图 12(a)~(c));②同一时刻,不同位置会有多个处于不同生长阶段的小气泡(图 12(b)、(c));③各个气泡持续生长的过程中,相邻两小气泡接触会相互“融合”变成一个大气泡(图 12(d)、(e)),随着时间推进大气泡继续与旁边小气泡“融合”成为一个更大的气泡。B 处为针肋间柱状气泡,由于拍摄时间较短,其形态尺寸并未出现明显变化。

图 13 展示了 KS2 在相同热流密度,流量更小时的流动特性,此时 $\Phi = 80.5 \text{ W/cm}^2$ 、 $q_v = 1.5 \text{ mL/s}$ 。由于该工况下流量更小,射流腔内“气塞”的生长和移动规律与图 12 差异较大。可以观察到:①热沉中部约四分之一区域始终处于核态沸腾状态为泡状流;②热沉靠近下游两侧“气塞”形成后向下游逐渐生长,生长至一定尺寸后内部气态工质从出口排出,热沉中部重新生成新的“气塞”,继续重复上述过程,此时流动状态主要为环状流。

结合前述性能测试结果综合分析可视化观测结果,对于定流量 q_v 工况:①当热流密度较小时,3种热沉内均处于强迫对流换热,无工质相变发

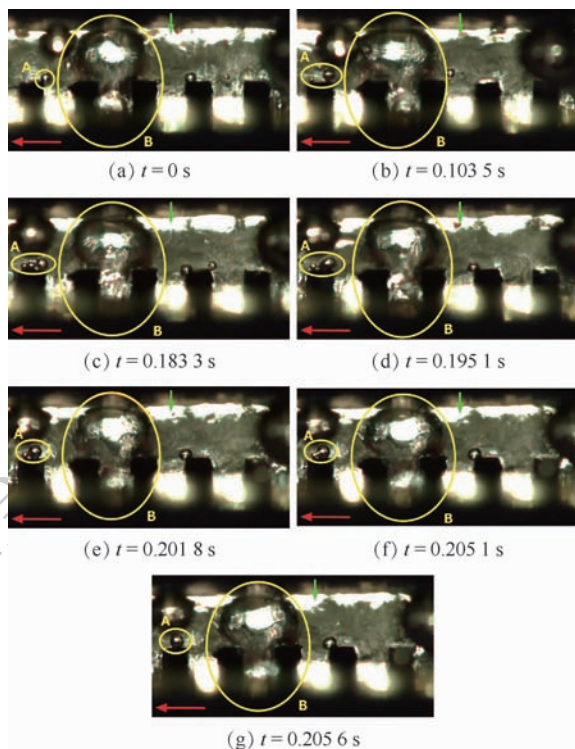


图 12 $\Phi = 80.5 \text{ W/cm}^2$ 、 $q_v = 5.8 \text{ mL/s}$ 瞬态工况气泡分布

Fig. 12 Images of transient bubble distribution when $\Phi = 80.5 \text{ W/cm}^2$ and $q_v = 5.8 \text{ mL/s}$ at transient conditions

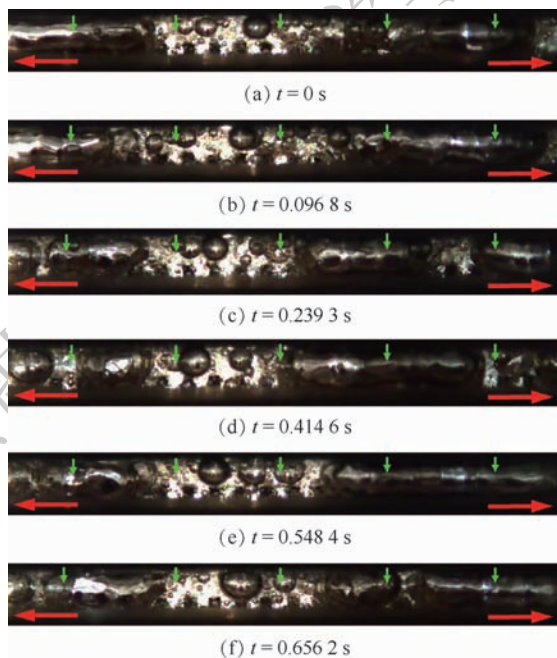


图 13 $\Phi = 80.5 \text{ W/cm}^2$ 、 $q_v = 1.5 \text{ mL/s}$ 瞬态工况气泡分布

Fig. 13 Images of transient bubble distribution with $\Phi = 80.5 \text{ W/cm}^2$ and $q_v = 1.5 \text{ mL/s}$ at transient conditions

生;②KS0 由于基板上表面光滑无促进扰动发生的结构,工质基本保持层流状态,而 KS1、KS2 针肋结构可以有效地促进湍流的发生,强化单相对

流换热系数;③当热流密度足够大时,由于KS2表面形貌更加不规则,随着热流密度 Φ 的进一步增大,针肋上表面更易形成汽化核心促进局部相变的发生。对于固定热流密度 Φ 工况:随着流量的降低,KS1、KS2两种强化热沉均能促进相变的发生,但KS2进入两相工况的时间更早,气泡的生成速率更快。

5 结 论

本文结合高速显微摄像手段,研究了不同肋化表面(光滑肋化表面/粗糙肋化表面)结构形态对受限式阵列射流冷却的流动、传热特性的影响。为了对其展开研究,设计了一种可用于高热流密度散热的可视化复合阵列射流冷却热沉,其中针肋尺寸均为 $0.6\text{ mm} \times 0.6\text{ mm} \times 1.0\text{ mm}$ 的矩形针肋;粗糙肋化表面的针肋外覆多孔介质层颗粒尺寸 $73 \sim 53\text{ }\mu\text{m}$,厚度约为2倍的颗粒直径。

实验测试的结果表明,2种肋化射流表面相较于对照组光滑射流表面,均可以同时提升射流冷却热沉的单相强迫对流换热性能,并促进两相工况沸腾相变的发生。

在使用无水乙醇为工质,保持入口温度恒定 $T_{\text{in}} = (20 \pm 1)\text{ }^\circ\text{C}$ 时,分别对加热热流密度逐渐增加、固定入口工质流量 $q_v = (7.5 \pm 0.5)\text{ mL/s}$;及固定加热热流密度、入口工质流量持续变化两种工况下对射流腔内流动与传热情况展开了性能测试及可视化观测研究。结果表明:

1) 在固定工质流量不变的工况中,由于工质流量相对较大,各热沉内均未观察到射流腔内有显著的相变发生。

2) 在固定加热热流密度的工况中,随着工质流量的减小,可以在两种肋化表面射流热沉的射流腔内观察到泡状流、弹状流、环状流的发生。

3) 粗糙肋化表面相较于光滑肋化表面传热性能更优,且其射流腔内可以更早地观察到沸腾现象的出现,即烧结多孔层可以有效地促进相变的发生,并优化整体性能。

参考文献 (References)

[1] RIOFRIO M C, CANEY N, GRUSS J A. State the art of efficient pumped two-phase flow cooling technologies[J]. Applied Thermal Engineering, 2016, 104: 333-343.

[2] 刘亮堂, 王安良. 星载电子器件用空气射流散热特性[J]. 北京航空航天大学学报, 2015, 41(8): 1553-1559.

LIU L T, WANG A L. Characteristic of air jet impingement cooling performance for electronic equipment of satellite[J]. Journal of Beijing University of Aeronautics and astronautics,

2015, 41(8): 1553-1559 (in Chinese).

- [3] MIRA-HERNANDEZ C, CLARK M D, WEIBEL J A, et al. Development and validation of a semi-empirical model for two-phase heat transfer from arrays of impinging jets[J]. International Journal of Heat and Mass Transfer, 2018, 124: 782-793.
- [4] 张添, 王仕越, 芮骥才, 等. 不同工质下带蒸汽腔的 Ω 形微通道热沉特性[J]. 化工进展, 2018, 37(8): 2954-2961.
- ZHANG T, WANG S Y, RUI J C, et al. Characteristic of an Ω -shape microchannel heatsink with different working fluid[J]. Chemical Industry and Engineering Progress, 2018, 37(8): 2954-2961 (in Chinese).
- [5] 孟恒辉, 谭沧海, 耿利寅, 等. 激光通信终端主体热设计与热分析[J]. 北京航空航天大学学报, 2013, 39(9): 1222-1227.
- MENG H H, TAN C H, GENG L Y, et al. Thermal control design and analysis for laser communication terminal[J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(9): 1222-1227 (in Chinese).
- [6] 孙少鹏. 高热流密度电子元件喷雾相变冷却系统的研究[D]. 重庆: 重庆大学, 2010.
- SUN S P. Research on spray cooling system for electronics with high heat flux[D]. Chongqing: Chongqing University, 2010 (in Chinese).
- [7] 万忠民, 刘靖, 陈敏, 等. 高热流密度散热的多孔微热沉流动与传热实验研究[J]. 中国电机工程学报, 2011, 31(29): 74-78.
- WAN Z M, LIU J, CHEN M, et al. Experimental investigation of flow and heat transfer in a porous micro heat sink for dissipating high heat flux[J]. Proceedings of the CSEE, 2011, 31(29): 74-78 (in Chinese).
- [8] GUO Z Y. Mechanism and control of convective heat transfer—Coordination of velocity and heat flow fields[J]. Chinese Science Bulletin, 2001, 46(7): 597-600.
- [9] 杨世铭, 陶文铨. 传热学[M]. 4版. 北京: 高等教育出版社, 2006.
- YANG S M, TAO W Q. Heat transfer[M]. 4th ed. Beijing: Higher Education Press, 2006 (in Chinese).
- [10] CHANG H S, KYUNG M K, SUNG H L, et al. Influences of nozzle-plate spacing on boiling heat transfer of confined planar dielectric liquid impinging jet[J]. International Journal of Heat and Mass Transfer, 2009, 52(23-24): 5293-5301.
- [11] SUNG M K, MUDAWAR I. Effects of jet pattern on single-phase cooling performance of hybrid micro-channel micro-circular-jet-impingement thermal management scheme[J]. International Journal of Heat and Mass Transfer, 2009, 52(13-14): 3364-3372.
- [12] SUNG M K, MUDAWAR I. CHF determination for high-heat flux phase change cooling system incorporating both micro-channel flow and jet impingement[J]. International Journal of Heat and Mass Transfer, 2009, 52(3-4): 610-619.
- [13] JOSHI S N, DEDE E M. Two-phase jet impingement cooling for high heat flux wide band-gap devices using multi-scale porous surfaces[J]. Applied Thermal Engineering, 2017, 110: 10-17.
- [14] 崔付龙, 詹可敬, 洪芳军. 射流-针肋微通道混合型蒸发器换热特性的实验研究[J]. 制冷技术, 2017, 37(4): 1-6.
- CUI F L, ZHAN K J, HONG F J. Experimental study on heat

- transfer performance of jet impingement-pin fin microchannel hybrid evaporator[J]. Chinese Journal of Refrigeration Technology, 2017, 37(4): 1-6 (in Chinese).
- [15] 刘明艳. 微小通道与射流相结合的高热流密度热沉结构的数值模拟[D]. 北京:清华大学, 2010.
- LIU M Y. Numerical simulation of heat sink with combined micro channels and jet arrays for high heat flux density[D]. Beijing: Tsinghua University, 2010 (in Chinese).
- [16] 马朝, 严超, 曹学伟, 等. 阵列空气射流传热均匀性问题的数值研究[J]. 工程热物理学报, 2016, 37(11): 2378-2385.
- MA C, YAN C, CAO X W, et al. Numerical study on array air jet heat transfer uniformity[J]. Journal of Engineering Thermophysics, 2016, 37(11): 2378-2385 (in Chinese).
- [17] RUANDER C, VIOND N. Heat transfer characteristics of submerged jet impingement boiling of saturated FC-72[J]. International Journal of Heat and Mass Transfer, 2012, 55(15-16): 4217-4231.
- [18] SHAILESH N J, DEDE E M. Two-phase jet impingement cooling for high heat flux wide band-gap devices using multi-scale porous surfaces[J]. Applied Thermal Engineering, 2017, 110: 10-17.

作者简介:

张添 女, 博士研究生。主要研究方向: 用于高热流密度散热的先进热管理技术。

董德平 男, 博士, 研究员, 博士生导师。主要研究方向: 空间载荷先进热管理技术。

Boiling characteristics of array jet impingement with various pin-finned surfaces

ZHANG Tian^{1,2}, ZHANG Chang^{1,2}, XIE Rongjian^{1,2}, DONG Deping^{1,2,*}

(1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Array jet impingement cooling technology can effectively solve the heat dissipation problem of high heat flux devices. In order to verify the effectiveness of heat transfer enhancement on the impacted surface for optimizing cooling performance of two-phase jet cooling, this article studied the effects of different pin-finned surface structures on the flow and heat transfer characteristics of confined array jet cooling combined with high-speed microscopic imaging methods. Two kinds of pin-finned surface morphology were designed: smooth cutting needle rib (0.6 mm × 0.6 mm × 1.0 mm) and rough needle rib with porous sintered layer (particle size 73 ~ 53 μm). In the experiment, jet cooling heat sink with smooth surface was used as the control group, anhydrous ethanol was used as the working medium, and all the inlet temperatures were the same (20°C). When the flow rate is 7.5 mL/s and the heating heat flux increases from 5 W/cm² to 100 W/cm², the heat transfer coefficient of the heat sink continues to increase but the increase rate gradually decreases, and no phase change is observed. Under the experimental conditions of changing the fluid flow rate (fluid Reynold number) with fixed heat flux 82.6 W/cm², 80.5 W/cm², when the flow rate decreases from 7.5 mL/s to 1.0 mL/s, it can be clearly observed that the working fluid in the jet cavity gradually enters bubble flow, slug flow and annular flow from stratified turbulence flow, which correspond to the initial boiling zone, nuclear boiling zone and membrane boiling zone, respectively.

Keywords: high heat flux; confined array jet; porous media; boiling enhancement; visualization

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0039



一种基于 HXDSP 的移位器查找表技术

叶鸿¹, 顾乃杰^{1,*}, 林传文², 张孝慈¹, 陈瑞¹

(1. 中国科学技术大学 计算机科学与技术学院, 合肥 230027; 2. 合肥学院 计算机科学与技术系, 合肥 230601)

摘 要: 高性能信号处理应用的快速发展, 对相应处理器的运算速度及吞吐效率提出了巨大挑战。移位器是数字信号处理器(DSP)上的重要部件, 通过为移位器设计额外专用随机存取存储器(RAM)和查找表(LUT), 并对其指令集及架构进行优化调整, 从而达到提高处理器使用效率和传输速率的目的。此外, 基于移位器与相应查找表指令, 可在数据暂存的同时进行移位、提取、算术与逻辑运算处理, 将部分数据运算的过程直接合并在对移位器 RAM 的数据存取过程中, 显著地提高了运算部件的使用效率。结果表明: 基于移位器查找表的暂存技术可以达到与传输总线接近的吞吐率, 对信号处理算法快速傅里叶变换(FFT)可以达到加速比约为1.15~1.20的性能提升效果。

关 键 词: 数字信号处理器(DSP); 移位器; 查找表(LUT); 单指令多数据流(SIMD); 超长指令字(VLIW)

中图分类号: TP402

文献标识码: A **文章编号:** 1001-5965(2019)10-2044-07

数字信号处理器(DSP)芯片是对信号和图像实时处理的一类芯片, 其具有高性能、低功耗和低成本等特点, 在通信、雷达、电子对抗、仪器仪表等领域有着广泛的应用^[1]。随着现代通信技术的发展, 数字电视广播(DVB)、数字媒体播放(DMB)、数字音频广播(DAB)等技术的全面应用与拓展, 处理器对数字信号处理领域算法的性能需求也愈发强烈, 如快速傅里叶变换(FFT)、有限冲激响应(FIR)、无限冲激响应(IIR)等。高性能 DSP 技术已成为目前集成电路领域发展速度最快、竞争最激烈的研究方向。

数字信号处理应用通常涉及大量的数据传输与运算。对于 FFT、FIR 等数据密集型算法, 数据吞吐率很大程度上决定了算法的直接效率。输入数据需经过多次读取、计算后, 再重新写回内存中。对于部分计算较为简单的运算环

节, 同一数据的高频次的取数、写回过程会导致数据传输拥堵、运算部件闲置、总线带宽负载过重等问题, 进而引发流水线停滞, 带来大量时间、空间上的开销和浪费。因此, 采用高效便捷的传输策略, 在数字信号处理领域尤为重要。

在处理 DSP 芯片的传输瓶颈问题时, 通常会采用调整访存策略、提高传输带宽、扩充寄存器、增加 Cache 等方式来提高吞吐量^[2]。Lee 和 Sunwoo^[3] 基于 DSP 芯片提出数据处理单元(DPU), 可以让现有 DSP 芯片的访存速度提高一倍。Nishitsuji 等^[4] 则提出通过增加额外 bank, 利用 5-bank 的错位内存结构来缓解访存冲突问题。Chang 和 Sung^[5] 采用设计额外的访存 Cache, 用以解决非连续和非对齐数据访问的问题。这些方式大多需要对专用芯片进行大量的硬件调整, 更多地侧重于通过扩充硬件或调

收稿日期: 2019-01-29; 录用日期: 2019-03-22; 网络出版时间: 2019-06-20 14:01

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190619.1542.001.html

基金项目: 安徽省科技重大专项(18030901011); 合肥学院科研发展基金(19ZR03ZDA)

* 通信作者。E-mail: gunj@ustc.edu.cn

引用格式: 叶鸿, 顾乃杰, 林传文, 等. 一种基于 HXDSP 的移位器查找表技术[J]. 北京航空航天大学学报, 2019, 45(10): 2044-2050. YE H, GU N J, LIN C W, et al. A shifter look-up table technique based on HXDSP[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2044-2050 (in Chinese).

整策略的方式提高传输总量。但在同一块处理器上,硬件的过分扩充会影响其通用性,且额外增加的通道相互之间可能产生相干或冲突^[6]。Le 等^[7]提出一种高级访存管理方案,该方案可以动态计算并有效预测地址,从而实现访存的规范管理。Luo^[8]和 Hsiao^[9]等则通过针对数字信号处理领域中特定应用的算法特点,进行访存策略设计,从而达到充分利用现有访存资源的目的。然而,这些方式大多数基于通用访存设备和额外硬件。在充分利用现有资源的同时,还可以深挖其他部件的使用价值,使资源利用效率达到更高的程度。

本文首先提出了一种基于移位器的查找暂存技术,通过为移位器额外添加专用暂存空间随机存取存储器(RAM),从而形成传输缓存区。其次,通过移位器查找表(LUT)的方式对相应区域进行灵活存读取,并对运算中间过程数据进行暂存,以达到缓解系统访存压力的目的。此外还充分利用移位器数据暂存过程中对数据进行简单移位、提取、算术与逻辑运算等能力,将部分数据运算处理的过程直接融合在数据存读取过程中,取数、计算、写回的过程完全在移位器中实现,显著地提高了运算部件的使用效率。

1 移位器

在普通微处理器中,一般没有单独的移位器单元,相应的移位运算通过逻辑运算单元实现。但在数字信号处理领域,移位操作被广泛使用在计数器、分频器及译码器上。信号处理算法普遍涉及大量的移位操作,单纯通过算术逻辑单元(ALU)来实现移位功能不仅无法满足运算需求,同时会造成运算资源的浪费。所以一般 DSP 芯片中都设计有专用的移位器,其性能直接影响到所在 DSP 的处理能力和执行速度^[10]。

移位器是 DSP 中的重要部件,其可以实现对源操作数进行任意裁减、分解、移位和拼接等功能。在传统 CPU 中,移位器的设计一般采用矩阵结构和树状结构。而在 DSP 芯片上,为满足移位器由一个时钟周期内完成一次简单的移位发展到一个时钟周期内可进行大量位数的移位操作的需求,桶形移位器^[11-12]与对数移位器^[13]先后被提出和实现,并发展成混合译码移位器。如今的移位器,在逻辑/算数运算、压缩及拓展、移位裁减、位域操作、数据传输及校验等方向有着广泛的

应用。

在 DSP 领域的移位器的研究中,大量工作的研究方向都是提高其运算速度,以及降低移位器功耗。除此之外,为移位器设计专用暂存 RAM,可在部分算法应用,运算部件负载压力较小但传输带宽总线压力较大时,充当数据中转区,以实现数据暂存。使用专用 RAM 和 LUT 辅助移位器的设计思路,在现场可编程门阵列(FPGA)领域有一定的应用。Xilinx 公司的 Block ram^[14](基于 RAM 模拟的移位器)与 Altera 公司的 Altshift_taps^[15]通过在 FPGA 中模拟实现类似功能,但它们为一个 IP 核结构,并非移位器专用部件,在专用 DSP 芯片中也没有类似结构。因为移位器暂存部件与数据访存总线之间并不冲突,在同一指令发射上,同时采用移位器查找存储和数据访存的方式,可以极大效率地提高数据吞吐效率,而不产生相干。

通过移位器查找表进行的数据暂存,利用移位器取数的灵活特性,不仅可以实现数据的暂存,还可以在寄存器使用压力较大的算法中,分担部分寄存器存储压力。同时,在分簇 DSP 的单指令多数据流(SIMD)运算的过程中,常常会因为数据宏间传输限制、从属于不同 bank 的相邻数据之间运算等问题,造成流水线停滞,浪费大量时间和空间。利用移位器查找表作为宏间数据传输中转站,还可以达到提高数据访问效率,减少运算 bank 冲突的作用。

2 平台

HXDSP 是国内某研究所研制的第二代高性能超长指令字(VLIW)DSP 芯片^[16],其结构如图 1 所示。它是一款 32 位静态超标量处理器,内部集成 2 个新一代处理器核 eC104+,工作主频为 800 MHz~1 GHz。采用 16 纳米工艺,16 发射、SIMD 架构,通信带宽为 252 Gbit/s,存储容量 105 Mbit,运算性能可达 134 GFLOPS(GFLOPS 表示每秒 10 亿次的浮点运算数)。

HXDSP 的每个核心处理器如图 2 所示,其采用哈佛体系结构,有独立的程序总线 and 数据总线。HXDSP 处理器有宽度为 512 bit 的指令总线和非对称全双工的内部数据总线。其包括 4 个基本执行宏(element operation macro),每个执行宏由 8 个 ALU、8 个乘法器(MUL)、4 个移位器(SHF)、1 个超算器(SPU)以及 128 个通用寄存器组成。

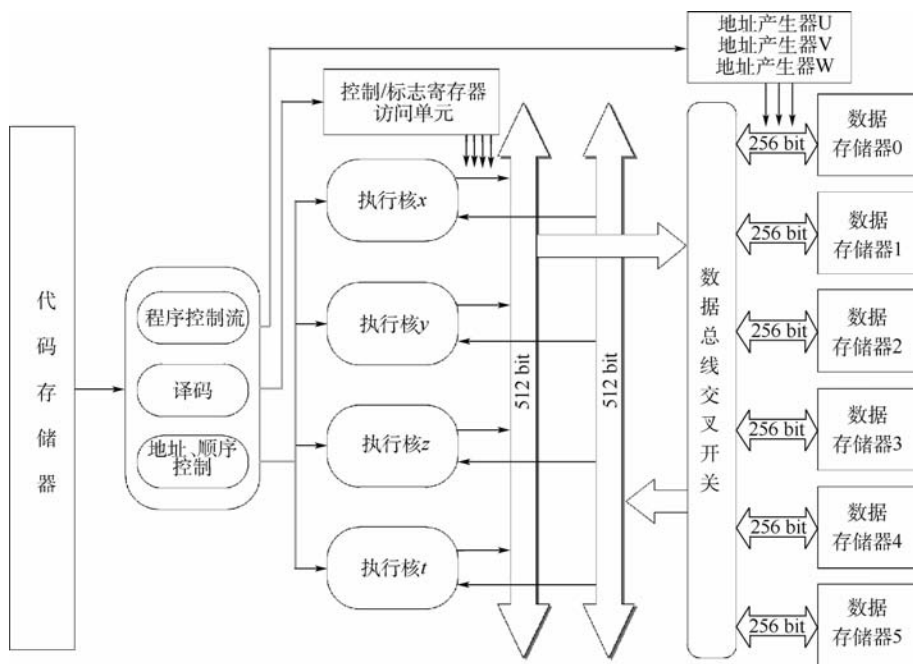


图 1 HXDSP104x 系统架构

Fig. 1 Architecture of HXDSP104x system

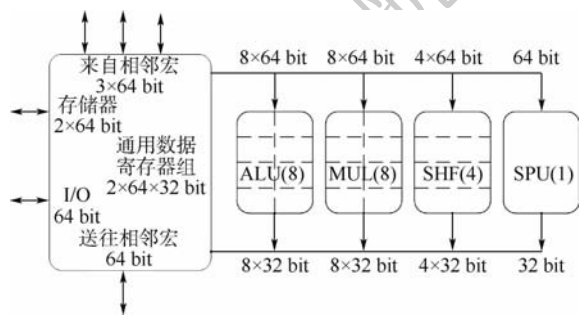


图 2 HXDSP 处理器向量处理单元

Fig. 2 Vector processing unit for HXDSP Processor

3 移位器 RAM 及指令

3.1 移位器 RAM 结构

在HXDSP的每个分簇中(HXDSP的4个执行宏,即4个分簇,分别称为Cluster x, y, z, t),包含4个移位器(Shifter 0 ~ Shifter 3)^[17]。为移位器增加的额外暂存空间结构如图3所示,移位器查找表TAB是移位器内部对应的RAM,每个移

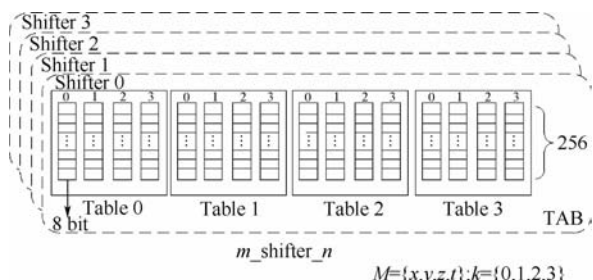


图 3 移位器查找表结构

Fig. 3 Structure of shifter look-up table

位器对应一个 TAB,用参数 k 来区分各移位器。每个移位器包含 4 块 8 bit RAM (Table 0 ~ Table 3),每块 8 bit RAM 又分为 4 个块(用参数 n 表示),共计 16 块,每块 256×8 bit。

当用户从移位器取一个标准 32 bit 数数据时,从 4 块 RAM 的每个第 n 位置,分别取一个 8 bit 数据,组成一个 32 bit。当用户从移位器取一个标准 16 bit 数时,有 2 种方式:①从第 0、1 块 RAM 的第 n 个位置,每块取一个 8 bit 数据,组成一个 16 bit;②从第 2、3 块 RAM 的第 n 个位置,每块取一个 8 bit 数,组成一个 16 bit。第 0、1 块 RAM 取出的 16 bit 可与 32 bit 通用寄存器的低 16 bit 做运算,第 2、3 块 RAM 取出的 16 bit 数可与 32 bit 通用寄存器的高 16 bit 做运算。

3.2 LUT 指令

基于这样的硬件结构, HXDSP 的移位器查找表通用访问指令为

$$\begin{aligned} R_s &= \text{STAB}_n(\text{LLOR}_m)(k) \setminus \text{STAB}_n(\text{LLOR}_s) = R_m(k) \\ R_s &= \text{STAB}_n(\text{HLOR}_m)(k) \setminus \text{STAB}_n(\text{HLOR}_s) = R_m(k) \\ R_s &= \text{STAB}_n(\text{LHOR}_m)(k) \setminus \text{STAB}_n(\text{LHOR}_s) = R_m(k) \\ R &= \text{STAB}_n(\text{HHOR}_m)(k) \setminus \text{STAB}_n(\text{HHOR}_s) = R_m(k) \end{aligned}$$

通用寄存器 R_m 划分为 4 个 8 位区域: $[7:0]$ 、 $[15:8]$ 、 $[23:16]$ 、 $[31:24]$, 分别称为低位、次低位、次高位、高位, 并通过读取指定区域上的 8 位数值, 映射查找表中的具体地址。同一个 8 bit RAM 中的 4 块分别读出一个 8 bit 数据, 即可拼接成一个需要的 32 bit 数。

4 移位器查找表的扩展使用

4.1 移位器查找表的暂存作用

DSP 芯片作为数字信号处理领域的通用处理器, 经常会面对各种算法的需求与性能考量^[18]。有些算法特点为数据密集型, 即在函数运算的过程中, 运算部件的运算占用率不高, 此时性能瓶颈主要在带宽总线的吞吐上; 有些算法为运算密集型, 此时至少有一种运算部件在核心周期指令发射上接近满负荷, 运算占用率接近 1, 此时性能瓶颈主要在算法与运算部件上。还有一些算法, 在计算的不同阶段, 对传输和运算有不同的需求, 可能在算法的不同时期, 分别造成不同瓶颈。如基-2 点的 FFT 算法, 会在不同的运算层, 产生传输瓶颈与运算瓶颈。

FFT 算法在计算和实现的过程中, 需要对大量旋转因子 (twiddle) 及中间层数据进行暂存。对于这类数据密集型运算, 单纯依靠总线带宽, 往往会造成运算部件的闲置和资源浪费。HXDSP 平台每个执行宏中, 含有 4 个移位器, 所以在同一条指令行可以执行 4 条移位器查找表指令。4 个执行宏同时执行时, 每个时钟周期可以读取 4×4 个 32 bit 数据。同时, HXDSP 拥有 3 条 64 bit 非对称全双工的内部数据总线, 可同时进行两读一写或两写一读操作, 每次 2 个 32 bit 数据。

当 4 个执行宏同时执行运算时, 可以对存储器读取或写入 $2 \times 4 \times 2 \times 32$ bit 数据。即采用移位器查找表暂存方式, 有着与访存指令相同的吞吐能力。基于 HXDSP 的 16 发射、SIMD 架构, 移位器查找表的暂存指令可以和数据存储指令同时执行, 使得访存速度提升至原来的 2 倍。这显著地提升了平台数据吞吐能力, 有助于解决内存读写速度与运算部件计算速度不匹配的问题。对于 FFT、FIR 等访存密集型算法, 有着较大的性能提升作用。

4.2 移位器查找表在存读取过程中的额外运算支持

在基-2 或者基-2ⁿ 的 FFT 运算中, 数据地址寻址会随着层数的变化而变化。图 4 为 16 点

DIT-FFT 算法实现图, DIT 模式算法, 在数据读入时, 需要对输入地址进行位反序寻址。寻址过程通常会选择将基地址依层次变化加上相应偏移地址形成实际访存地址。这就需要针对相应层次, 为基地址加上 32、16、8 等偏移量。但如表 1 所示, 考虑到偏移地址依层次变化, 十进制中变址跃迁的 8、16、32 等相对距离, 转换成二进制机器数据, 就是针对寄存器中二进制表示的数据的其中某一位去置 0 或置 1。这些被读取数据地址的变化, 可利用移位器作为暂存寄存器处理。移位器可在作为替代内存进行访存指令数据暂存的同时, 额外进行一定的逻辑或算数运算。从而在不产生额外时钟周期消耗的同时, 完成一些需要的运算。

为增强对移位器查找表访存过程中额外运算的支持, 文中提出一种改良微指令 $R_s = \text{STAB}_n(R_m)(k)$ 。

该指令以 R_m 寄存器的高位与低位分别存储映射查找表中具体地址的 8 bit 数值及调整自增数值的偏移位, 使寄存器在一条指令中能够在存/读取移位器查找表的同时, 完成偏移量修正自增工作。在 FFT 运算中, 常常会出现 4 个移位器全部用于移位器查找表的读写; 8 个 ALU 全部用于蝶形计算。所以如果中间再加入移位器查找表中的标志寄存器自增指令, 能节省不必要的开销。

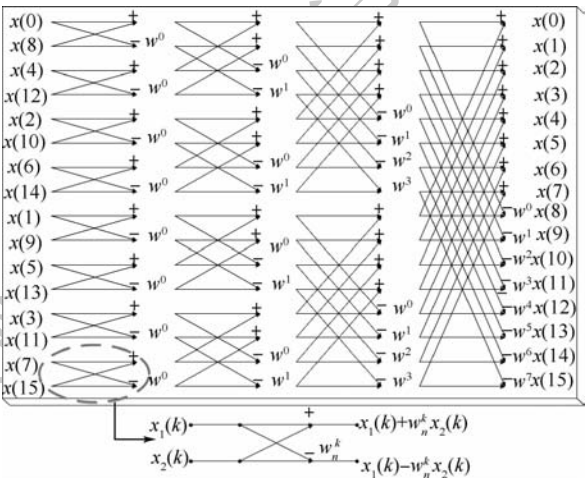


图 4 16 点 DIT-FFT

Fig. 4 16-DIT-FFT

表 1 64 点 FFT 偏移地址

Table 1 64-FFT offset address

层数	偏移 (D)	偏移 (O)
1	32	100000
2	16	010000
3	8	001000
4	4	000100
5	2	000010
6	1	000001

微指令 $R_s = \text{STAB}_n(R_m)(k)$ 完成过程:

1) 以源寄存器 R_m 的低位 $[7:0]$ 为地址, 读出需访存移位器查找表中的目的位置 s 。

2) 从 4 块 RAM 的每个第 s 位置, 每块取一个 8 bit 数据, 组成一个 32 bit 数并将这个 32 bit 数赋给目的寄存器堆 R_s 。

3) 同时以寄存器 R_m 的次高位 $[23:16]$ 为地址, 读出下一次源寄存器低位 $[7:0]$ 读取变更数据位。当算法需求为自增时, 最低位置 1, 当算法需求为 FFT 等呈 2^n 次幂形式递增时, 修改相应数据位。

4) 读取 STAB 查找表中相应数据并赋给目的寄存器堆 R_s 的同时, 在同一个时钟周期内, 更新源寄存器 R_m , 进入下一次运算。

在这样的指令形式下, FFT 算法可对中间层临时数据进行按需读取或存储操作。并且在暂存器访存指令执行的同时对相应偏移地址进行调整, 从而达到在同一条指令中实现取数、取值、运算、更新的全部过程。这一过程显著地提升了运算效率, 为访存部件和运算器都减轻了工作压力。而当变更寄存器置位为最低位为 1 时, 可利用移位器部件, 实现循环自增寻址工作。这一功能可以广泛运用在数字信号处理, 数字图像处理等领域。

本文还提出另一种改良微指令

$$\text{HSTAB}_n(\text{LOR}_s) + = \text{HR}_m(U, k)$$

该指令利用 R_s 的次低位 $[15:8]$ 和低位 $[7:0]$ 位寻址高 16 bit 和低 16 bit, 分别取出第 n (n 可以取 0、1、2、3) 块 STAB 中的 2 个 16 位数, 并分别与 $R_m[19:16]$ 和 $[3:0]$ 累加, 累加结果存入原位置。高效直接地对移位器查找表中的数据进行调整修改, 从而提高了工作效率。

4.3 移位器查找表在存读取过程中的额外寄存器支持

移位器查找表的 RAM 的存在, 类似于寄存器与内存之间的 Cache, 能够完成部分数据的暂存作用。但是因为移位器自身的特点, 它可以灵活地对暂存的数据进行选择输出或处理, 甚至在寄存器使用瓶颈的时候, 起到部分替代寄存器的作用。

HXDSP 原有微指令 $\text{STAB}_d(R_n, R_s, R_m)(k)$ 的作用是把寄存器 R_m 的值写入到 STAB 中, 同时把 STAB 中某个位置的值读出到寄存器 R_s 。 k 指定本指令在哪个移位器上执行, 在第 2 个字的 $[1:0]$ 位。 d 指定本指令占用的 STAB 块, 在第 2 个字 $[3:2]$ 位。 $R_n[7:0]$ 位指定读出 STAB 中某一块的位置。寄存器 $R_n[23:16]$ 指定写入 STAB 中某一块的位置。

本文同时提出对 $\text{STAB}(R_n, R_s, R_m)$ 这一微指

令的修改方案。

在目前的这条指令中, R_n 的低位 $[7:0]$ 用于移位器查找表中读位置的标识, 次高位 $[23:16]$ 用于移位器查找表中写位置的标识, 如:

	写位置标识		读位置标识
--	-------	--	-------

此时次低位 $[15:8]$, 高位 $[32:24]$ 空闲。对于寄存器资源紧张的算法应用, 如 FFT 算法的中间层代码编写, 是极大的浪费。

因此可调整成:

$$\text{STAB}(\text{HR}_n, R_s, R_m)$$

写位置标识		读位置标识	
-------	--	-------	--

$$\text{STAB}(\text{LR}_n, R_s, R_m)$$

	写位置标识		读位置标识
--	-------	--	-------

这样可以充分利用 R_n 寄存器, 将一个寄存器划分成 2 个, 使寄存器使用紧张的情况得以缓解。同时由于移位器查找表的灵活寻址取数方式, 它实际上构成了寄存器与内存区域中间一层灵活的缓冲区。在高效完成大量移位操作的同时, 为 DSP 芯片的数据传输、运算部件使用, 寄存器分配充分地分担了压力。

5 实验结果与分析

5.1 移位器查找表吞吐率分析

HXDSP 的微指令 $R_s = \text{STAB}_n(\text{LLOR}_m)(k)$ 每次执行, 只占用一个移位器, 指令周期为 1 个 cycle。因为 HXDSP 是一个支持 4 个执行宏 4 路 SIMD 指令的处理器, 每个执行宏又包括 4 个移位器, 所以在每一个时钟周期中, 最多可以 4 路并行执行 4 条上述指令, 即一个周期可存取 $4 \times 4 \times 32$ bit。因此, 通过移位器查找表指令, 可实现一个时钟周期对 16 个寄存器的暂存访问。

同时, HXDSP 拥有 3 条非对称全双工总线, 可以在同一指令行同时进行两读一写或两写一读操作。HXDSP 的通用读写指令 $\{x, y, z, t\} R_s + 1: s = [U_n + = U_m, C]$ 支持双字读写, 即每次执行可以存取连续的 2 个 32 bit 数据, 指令周期也为 1 个 cycle。其中 U_n 表示寻址的起始地址, U_m 表示每次寻址之后的基地址跃迁地址, 立即数 C 表示寻址过程中的偏移地址。因此每个时钟周期最多可以进行 $2 \times 2 \times 4 \times 32$ bit 的读操作或者写操作, 即在一个时钟周期内, HXDSP 可对 16 个寄存器进行访存读写操作。

上述对比表明, 移位器查找表有着和 HXDSP

访存总线相同的传输效率和吞吐能力。当面对数据密集型运算时,可利用传输指令和移位器的暂存 RAM 进行复合读写操作,存读取速度可扩展到之前的 2 倍,极大地缓解数据访存压力。

5.2 移位器查找表应用分析

基于移位器查找表,可以对数字信号处理应用中许多算法进行加速优化。对于数字信号处理算法如 FFT,移位器查找表可以作为 DIF-FFT 寻址中相应偏移地址形成的加法器;也可以作为旋转因子及中间层运算结果的暂存器,以提高运算过程中的访存效率。

如图 5 所示,对于基-2 点的 FFT 算法,移位器查找表的暂存功能对 FFT 的访存性能有显著地提升。图中,横坐标表示 FFT 算法的具体运算规模,左侧纵坐标表示算法运行的实际运算周期数,右侧纵坐标表示使用移位器查找表前后加速比。当点数为 64 时,由于 FFT 点数较少,部分中间运算可以直接用寄存器存储,所以加速比不高。当 FFT 点数小于等于 1024 点时,FFT 效率显著提升,加速比约为 1.20。这是因为所有数据的中间计算暂存可以在移位器查找表中完成,可充分提高访存效率。而当 FFT 点数大于 1024 点时,加速比略有下降,但是依然稳定在 1.15 左右。这是因为对于大点数的 FFT,数据访问无法完全在移位器查找表中实现,其访存过程需要和访存总线同时进行。

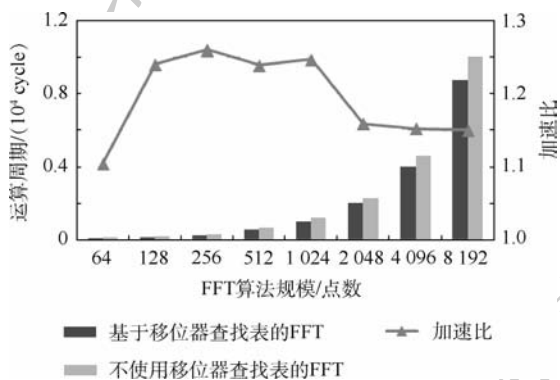


图 5 使用移位器查找表前后对比

Fig. 5 Comparison with and without shifter look-up table

6 结 论

HXDSP 是一款高性能的 VLIW DSP 芯片,基于其开发设计的移位器暂存 RAM 与移位器查找表的架构,是一种新颖而灵活的移位器拓展结构。其在高效完成数字信号处理应用中大量移位操作的同时,还可以在在一定程度上分担芯片的传输与运算压力。HXDSP 的移位器查找表在同一周期

内,有着和系统总线同样的传输带宽,不仅可以分担传输压力,还可以将部分简单运算过程融入其中,极大地提高了芯片的使用效率。

目前这一技术尚在试用及继续开发中,关于移位器查找表的适应指令集依然在调整和扩充。本文提出一种对移位器查找表的高效使用,并对现有指令集进行了两处调整和升级。相信在不久的将来,会有更多更高效的创新,让这一技术和结构,能够在 DSP 领域,发挥更出色的作用。

参考文献 (References)

- [1] EYRE J, BIER J. The evolution of DSP processors[J]. IEEE Signal Processing Magazine, 2000, 17(2): 43-51.
- [2] YE H, GU N, ZHANG X, et al. Design and implementation of a conflict-free memory accessing technique for FFT on multicluster VLIW DSP[J]. IEICE Electronics Express, 2018, 15(18): 20180674.
- [3] LEE J S, SUNWOO M H. Design of new DSP instructions and their hardware architecture for high-speed FFT[J]. Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology, 2003, 33(3): 247-254.
- [4] NISHITSUJI T, KAKUE T, SHIMOBABA T, et al. Conflict-free FFT circuit using loop architecture by 5-bank memory system[C]// IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). Piscataway, NJ: IEEE Press, 2014: 523-526.
- [5] CHANG H, SUNG W. Efficient vectorization of SIMD programs with non-aligned and irregular data access hardware[C]// Proceedings of the 2008 International Conference on Compilers, Architectures and Synthesis for Embedded Systems. New York: ACM, 2008: 167-176.
- [6] YE H, GU N, ZHANG X, et al. An efficient conflict-free memory-addressing unit for SIMD VLIW DSP[C]// 2017 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS). Piscataway, NJ: IEEE Press, 2017: 1-7.
- [7] LE G B, CASSEAU E, HUET S. Dynamic memory access management for high-performance DSP applications using high-level synthesis[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2008, 16(11): 1454-1464.
- [8] LUO H F, LIU Y J, SHIEH M D. Efficient memory-addressing algorithms for FFT processor design[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2015, 23(10): 2162-2172.
- [9] HSIAO C F, CHEN Y, LEE C Y. A generalized mixed-radix algorithm for memory-based FFT processors[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2010, 57(1): 26-30.
- [10] PENG Y X, ZOU J J. Design and implementation of ALU and shifter in X-DSP[J]. Journal of Computer Applications, 2010, 30(7): 1978-1982.
- [11] PEREIRA R, MICHELL J A, SOLANA J M. Fully pipelined TSPC barrel shifter for high-speed applications[J]. IEEE Jour-

nal of Solid-State Circuits,1995,30(6):686-690.

[12] ACKEN K P,IRWIN M J,OWENS R M. Power comparisons for barrel shifters[C] //Proceedings of 1996 International Symposium on Low Power Electronics and Design. Piscataway, NJ: IEEE Press,1996:209-212.

[13] WESTE N,ESHRAGHIAN K. Principles of CMOS VLST design [M]. Boston:Addison Wesley,1993.

[14] Xilinx. PG058-block memory generator v8. 4 product guide (v8.4) [EB/OL]. (2017-10-04) [2019-01-28]. https://china.xilinx.com/support/documentation/ip_documentation/blk_mem_gen/v8_4/pg058-blk-mem-gen.pdf.

[15] Altera. RAM-based shift register (ALTSHIFT_TAPS) IP core user guide [EB/OL]. (2014-08-18) [2019-01-28]. https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/ug/ug_shift_register_ram_based.pdf?wapkw=altshift_taps&_ga=2.155121400.1796521913.1564220766-1713056724.1564220743.

[16] 中国电子科技集团公司第三十八研究所. BWDSP104x 软件用户手册[Z].2015.

CETC-38. Software user's manual BWDSP104x[Z]. 2015 (in Chinese).

[17] 中国电子科技集团公司第三十八研究所. BWDSP104x 硬件用户手册[Z].2015.

CETC-38. Hardware user's manual BWDSP104x[Z]. 2015 (in Chinese).

[18] 刘余福,郎文辉,贾光帅. HXDSP 平台上矩阵乘法的实现与性能分析[J]. 计算机工程,2019,45(4):25-29.

LIU Y F,LANG W H,JIA G S. Implementation and performance analysis of matrix multiplication on the platform HXDSP [J]. Computer Engineering, 2019, 45 (4): 25-29 (in Chinese).

作者简介:

叶鸿 男,博士研究生。主要研究方向:并行计算、体系架构优化。

顾乃杰 男,博士,教授,博士生导师。主要研究方向:网络计算、大数据处理和分析、云计算与应用,在线算法、软件和代码优化、大型软件代码检测,深度学习软硬件系统、并行和分布式计算。

林传文 男,博士。主要研究方向:编译优化、深度学习训练优化及应用。

陈瑞 男,硕士。主要研究方向:体系架构优化。

A shifter look-up table technique based on HXDSP

YE Hong¹, GU Naijie^{1,*}, LIN Chuanwen², ZHANG Xiaoci¹, CHEN Rui¹

(1. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China;
2. Department of Computer Science and Technology, Hefei University, Hefei 230601, China)

Abstract: With the development of digital signal processing technology, the application of high-performance signal processing has attracted more and more attention, which also poses great challenges to the computing speed and throughput efficiency of the corresponding processors. The shifter unit is an important component on the digital signal processor (DSP). By designing additional dedicated random access memory (RAM) and look-up table (LUT) for the shifter unit, this paper optimizes and adjusts its instruction set and architecture, so as to improve the use efficiency and transmission rate of the processor. In addition, based on the shifter and the corresponding look-up table instruction, it can carry out shift, extraction, arithmetic and logical operation processing at the same time of data temporary storage. And the process of the partial data operation is directly merged into the data read/write process of the shifter RAM, which greatly improves the efficiency of arithmetic unit. Experiments show that the temporary storage technology based on the shifter look-up table can achieve the throughput rate close to the transmission bus, and the signal processing algorithm fast Fourier transformation (FFT) can achieve the performance improvement of the acceleration ratio of 1.15 to 1.20.

Keywords: digital signal processor (DSP); shifter; look-up table (LUT); single instruction multiple data (SIMD); very long instruction word (VLIW)

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0042

基于地面基站的定位系统构建和方案



耿珂, 黄智刚*, 苏雨, 石培辰, 高强, 熊华钢

(北京航空航天大学 电子信息工程学院, 北京 100083)

摘

要: 针对大型、复杂、多功能建筑,其内部信号环境恶劣,卫星导航的信号衰减较大,较难稳定捕获跟踪,建筑内多径效应严重,短多径对定位精度的影响较大,直接使用卫星导航信号进行定位难度大的问题,提出了一种基于地面基站的大区域(建筑群)室内定位方案。根据频率与信号穿透性能和空间衰减之间的关系,选择甚高频频段作为信号载波,采用码伪距和载波伪距联合定位的方式,可同时兼顾覆盖性和定位性能。利用所提出的新型定位方式,搭建了一个基本测试系统,通过信号的产生、发射、无线传播,进行了信号的捕获、跟踪和伪距求差解算,初步验证了本文方法的可行性及覆盖能力。

关键词: 大区域; 高精度定位; 码伪距; 载波伪距; 地面基站

中图分类号: V221+.3; TB553

文献标识码: A

文章编号: 1001-5965(2019)10-2051-07

随着社会的发展进步及信息化时代的到来,人们的工作、生活等社会活动越来越依赖于位置信息,位置服务逐渐成为大家关注的焦点。目前室外定位导航技术大多依赖于卫星导航系统,经过几十年的发展,到现在已经比较成熟,如今在运载体导航、人员跟踪定位及紧急救助服务等方面发挥着巨大的作用^[1],并且实现了广域精准定位,广泛应用于工农业生产和军事等领域。

相比较于室外广域无线定位,室内定位有自己的特点,其范围一般较小,范围通常在几百米以内;另一方面由于室内环境的复杂,干扰因素较多,容易对定位信号产生影响。而卫星导航定位在室内由于受到建筑物、家具等的遮挡,卫星导航信号衰减很快,而且即使有一些信号,由于环境因素复杂,导致多径现象严重,使其无法在室内定位应用。

自20世纪90年代末起,室内无线定位技术开始受到人们的广泛关注。经过近20年的研究,

室内定位技术已经取得了一些成果^[2],有一些比较成熟的技术得到了初步应用。室内定位技术分为局部室内定位和大范围室内定位,目前最常见的局域定位技术是WiFi室内定位,其基于信号能量的指纹匹配定位技术,定位精度能够达到3 m,但是信号受环境影响较大,而且需要前期采集大量的数据,定位时需要与数据库的数据进行比对。另一种室内定位技术UWB(Ultra Wideband)能够达到很高的定位精度,典型值为0.3 m,但缺点也很明显,即覆盖范围小。除此之外,还有一些新的局域室内定位技术,比如Zig Bee、蓝牙、射频标签等^[3,4],这些局域室内定位技术几乎都不是专门用作定位应用的,主要作用还是通信应用。

目前能够实现大范围室内定位的技术主要是架设地面伪卫星系统和利用通信基站的室内定位系统。北京邮电大学的时分 & 码分正交频分复用(TC-OFDM)^[5]广域室内定位技术利用现有的通信基站,通过将定位信号承载到手机通信的信

收稿日期: 2019-01-30; 录用日期: 2019-03-08; 网络出版时间: 2019-06-18 10:45

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190617.1040.004.html

* 通信作者: E-mail: baahzg@163.com

引用格式: 耿珂, 黄智刚, 苏雨, 等. 基于地面基站的定位系统构建和方案[J]. 北京航空航天大学学报, 2019, 45(10): 2051-2057.
GENG K, HUANG Z G, SU Y, et al. Positioning system construction and scheme based on ground base station[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2051-2057 (in Chinese).

号上,能够实现大范围室内定位,系统精度虽然较高,但是大范围室内定位的技术复杂,而且需要较长的布设周期和较大的成本。地面伪卫星系统实现室内定位的代表是澳大利亚的 Locata 系统^[6],通过增大地面伪卫星发射信号的功率,在室内也可达到较高的定位精度,但其覆盖范围有限或需架设很多伪卫星基站。

目前,针对大型、复杂、多功能建筑,其内部结构复杂、信号环境恶劣、导航信号衰减幅度大,而且多径影响严重,无法直接利用卫星导航信号进行定位;并且在大型火灾、暴力恐怖袭击等灾害现场或事件发生地,电力供应常常被切断或无法使用,停电范围可能波及多个街区,导致目前已有的室内定位手段无法展开应用。针对突发事件现场,对定位系统的基本要求是独立、快速、易用及高精度,目前还缺乏成熟的技术或系统可以应用^[7]。

当前室内定位方案难以成熟应用的问题,主要是覆盖范围与定位精度不可兼得^[8]。如 UWB 的精度较高,但穿透能力有限;Locata 和 TC-OFDM 系统为了保证覆盖,需要布设较多基站^[9];惯导模块或设备没有覆盖范围的限制,但存在积累误差^[10]。

针对以上问题,本文提出了一种基于地面基站的大区域室内定位方案,不同于发射卫星信号的伪卫星,地面基站发送的是专门设计的导航信号,在频率、信号体制、定位方式等方面进行改进。考虑到频率与信号穿越性能的关系,选择了较低频段作为信号载波,以及采用码伪距和载波伪距联合定位的方式。本文利用所提出的新型定位方式,搭建了一个基本测试系统,通过信号的产生、发射,以及空间无线传播,进行信号的捕获、跟踪和伪距求差解算,初步验证了本文方法的可行性及有效性。

1 系统模型

系统的设计采用覆盖与精度统筹考虑的策略,选用了 76 MHz 的载波频率,具有良好的信号穿透能力以及障碍翻越能力。采用伪随机码进行初步定位,获取大致的用户位置。采用载波进行精细测距,最终实现高精度的定位。

根据以上设计思路,专门研制了实际系统。系统包括多个发射基站及接收终端,下面分别进行介绍。发射端的组成框图如图 1 所示。接收端的组成框图如图 2 所示。

图中各部分的功能介绍如下:

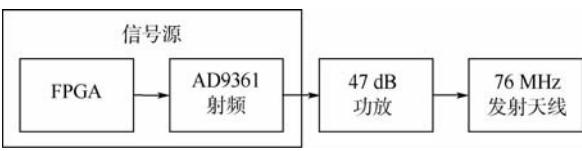


图 1 发射端组成
Fig. 1 Transmitter structure

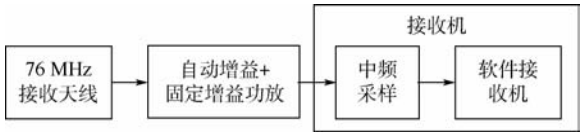


图 2 接收端组成
Fig. 2 Receiver structure

现场可编程门阵列 (FPGA): 完成数字基带信号的生成与调制,生成的数字信号包括导航电文、伪随机码和信号载波,采用低数字中频方式,将导航电文、伪随机码对数字载波进行调制。

AD9361 射频: 完成数字信号向射频信号的转换,数字基带信号输入给 AD9361 模块后,将进行滤波、数模转换、上变频等一系列处理,产生所需射频信号。对 AD9361 射频本振频率进行设置,产生中心频率为 76 MHz 的射频信号。

47 dB 功放: 将 AD9361 输出的模拟信号功率放大到 1 W。

自动增益 + 固定增益功放: 将天线接收后的信号进行功率放大,使信号的适应性增加,能够正常进行采样。

中频采样: 将天线接收到的模拟信号使用 62 MHz 进行采样。采样后数据使用软件接收机进行捕获、跟踪,输出码相位和载波相位观测值。

2 定位方案

系统研制完成后,通过在地面合适位置布设 4~6 个地面基站,使得在所服务的区域内都可以收到至少 4 个基站的定位信号,接收机对接收到的信号进行捕获跟踪,得到到达时间 (TDOA) 即距离的测量值,并进行定位解算,就可以获得用户的实际位置。定位过程如下:

- 1) 按布局设计搭建多个基站,使得服务区域内的 GDOP 值尽量小。
- 2) 通过地理测绘或全球导航卫星系统 (GNSS) 差分技术等,对基站的真实位置进行标定。
- 3) 通过光纤技术、时间双向比对技术等,实现对基站之间的高精度时间同步。
- 4) 各基站同步发射导航信号,利用低频载波增加覆盖范围。

5) 接收机接收 4 个质量较好、GDOP 值较小的基站信号进行距离测量并实时定位,并且通过载波相位测量提高定位精度。

本系统的定位算法原理如下:利用载波波长较小,载波跟踪精度高和受多径影响小的特点,使用载波来进行高精度定位,缺点是载波整周是位置的。本定位方案使用码伪距来确定未知的载波整周,将载波伪距与码伪距确定的载波整周相结合,得到较精确的目标到信号源距离差,使用下面的 TDOA 方法完成定位结算。

基于 TDOA 的无线电定位是一种重要的目标三维定位方法,被广泛应用于导航、雷达等定位系统,定位方法主要分为 2 种:一是迭代算法^[11],二是解析算法^[12-13]。迭代算法的优越性体现在能充分利用系统的冗余信息,计算精度高,但计算复杂,实时性差。解析算法精度较低,但实时性好。考虑到室内运动速度不高,本文采用迭代算法。

当已知信号源的位置坐标为 $S_i(x_i, y_i, z_i)^T$, 目标位置坐标为 $T_r(x_r, y_r, z_r)^T$, 得到目标到信号源的距离差为

$$d_{ij} = r_i - r_j = |T_r - S_i| - |T_r - S_j| \quad i \neq j \quad (1)$$

对目标到信号源的距离取平方:

$$r_i^2 = (x_r - x_i)^2 + (y_r - y_i)^2 + (z_r - z_i)^2 \quad (2)$$

通过对目标接收机码相位及载波相位观测值分别两两做差,可以得到距离差 d_{ij} ,在实际测试中,距离差含有各种因素引起的误差 ε_{ij} ,假设距离差的理论值为 d_{ij}^t ,所以距离差测量值可以表示为

$$d_{ij} = d_{ij}^t + \varepsilon_{ij} \quad (3)$$

将式(1)移项平方,得到 $r_i^2 = (d_{ij} + r_j)^2$,将式(2)代入,可得

$$0 = p_{ij}^2 - d_{ij}^2 - 2d_{ij}r_j - 2S_{ij}^T q_{ij} \quad (4)$$

式中: $p_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$, $S_{ij} = S_i - S_j$, $q_{ij} = T_r - S_j$, 目标的位置与径向距离呈线性关系。

由于测试误差的存在,式(4)的等号左边并不等于 0,可用微小量 δ_{ij} 代替。实际情况下,可以将式(4)表示成矩阵形式,通过最小二乘法就可以求解目标位置。

这种实现方案,优点是避免卫星导航中的主要误差因素,如星历、星钟、对流层、电离层等误差,导航信号的完好性也大大增强,并且利用低频载波可以增加覆盖范围和穿透能力,缺点是在

地球表面的多径信号较多,需要通过抗多径技术减弱多径的影响。

与其他室内定位技术相比^[14],比如 WiFi 定位,本文方法利用伪随机码的相关性进行测距和定位,精度明显高于利用 WiFi 信号的功率测距方法,所以 WiFi 定位目前大多通过指纹匹配技术来提高精度;而 UWB 定位,其测距精度最高,但 UWB 工作频率也很高,受周围环境的影响很大,覆盖范围十分受限;Locata 系统是一种地基伪卫星定位系统,目前有一定应用,但其在每个室内都需要布设多个伪卫星基站并实现同步,成本及布局问题限制了其进一步应用^[15]。

3 测试结果处理与分析

3.1 在室外空旷条件下测试

首先测试信号源间码相位和载波相位同步情况,接着测试 76 MHz 发射频率时中垂线处和偏移处接收码相位和载波相位能否体现距离差的变化。

以操场为测试环境(见图 3),在主席台右侧看台最高处台阶放置两台信号发射源,发射功率为 30 dBm,信号源 1、2 号星间距为 45 m。在对侧跑道上取 3 个等间隔的测试点(接收机位置),分别为 1 号点、2 号点、3 号点。信号源与测试点构成长方形。2 号点(取 1 号星、2 号星中垂线与长方形的交点)距两信号源中点 87 m。1 号点、3 号点与两信号源的距离差约为 11 m。信号源与接收机的位置如图 3 所示。

1) 首先进行 2 个信号源的同步测试,启动 1 号星与 2 号星,经过同步模块进行时钟校准后,使用电脑同时配置信号源,使两台信号源分别产生给定的定位信号。

将两台信号源通过导线同时连接接收机,



图 3 操场测试环境

Fig. 3 Playground test environment

通过接收机显示界面观察码相位,并采集码相位及载波相位观测量,其码相位差和载波相位差的测试曲线分别如图 4 和图 5 所示(测试 2 次)。其对应的测试数据如表 1 所示。

测试结果表明,系统工作正常,但不同组测试时可能会有半周差,可在后期处理时校正。可以看出,在导线连接条件下测试的精度较高,原因是由于没有引入空间传播方面的误差。

2) 下面进行运动状态下的测试,分别连接好发射、同步、接收设备,并启动其正常工作,在操场携带接收端设备首先从 1 号点沿直线走到 2 号点,再走到 3 号点,在该过程中在 3 个测试点各停留约 200 s 左右,实时采集并处理 2 个信号源发出

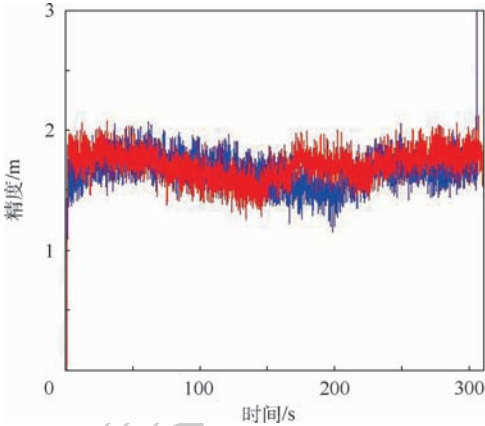


图 4 码相位差值(同步测试)
Fig. 4 Code phase difference value (synchronous test)

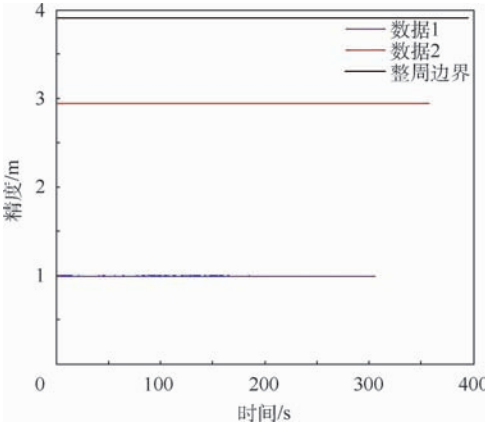


图 5 载波相位差值(同步测试)
Fig. 5 Carrier phase difference value (synchronous test)

表 1 测试数据(同步测试)

Table 1 Test data (synchronous test)			
观测量	测试分组	均值/m	标准差/cm
码相位差	第 1 次	1.63	57.7
	第 2 次	1.68	53.7
载波相位差	第 1 次	0.991 5	0.18
	第 2 次	2.943 1	0.11

的定位信号,其码相位差和载波相位差的测试曲线分别如图 6 和图 7 所示。其对应的测试数据如表 2 所示。

测试结果表明:码相位差值在 1 号点有较大的波动,在 2 号点、3 号点比较平稳,标准差符合预期。同样,载波相位差在 1 号点时出现了几次跳跃,与静止状态不符,但 2 号点、3 号点基本平稳,标准差也小,符合预期,并且比较理想地反映了从 2 号点到 3 号点的运动过程。

3) 最后再从 3 号点沿直线返回 2 号点,再返回 1 号点,同样在 3 个测试点各停留约 200 s 左右。实时采集并处理 2 个信号源发出的定位

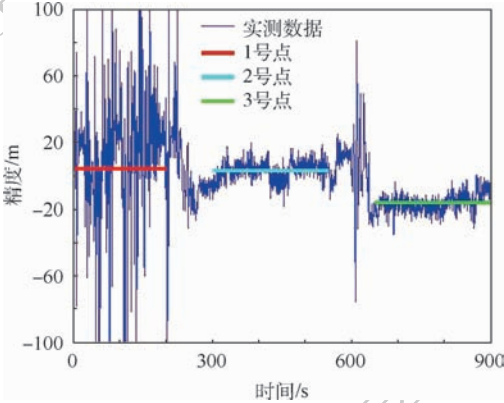


图 6 码相位差值(运动状态下测试,1 号点~3 号点)
Fig. 6 Code phase difference value (test under motion, point 1 to 3)

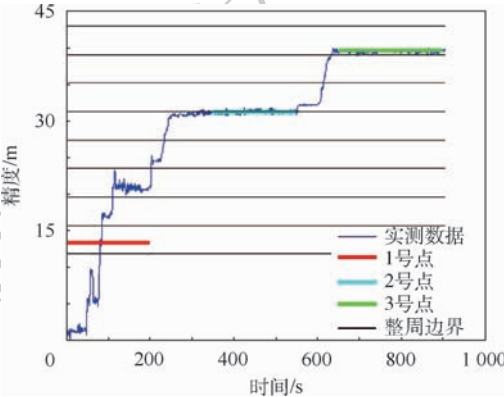


图 7 载波相位差值(运动状态下测试,1 号点~3 号点)
Fig. 7 Carrier phase difference value (test under motion, point 1 to 3)

表 2 测试数据(运动状态下测试,1 号点~3 号点)

Table 2 Test data (test under motion, point 1 to 3)				
测试点	码相位差		载波相位差	
	均值	标准差	均值	标准差
1	3.66	66.8	13.2	8.397
2	2.83	5.6	31.1	0.19
3	-16.6	5.4	39.5	0.13

信号,其码相位差和载波相位差的测试曲线分别如图 8 和图 9 所示。其对应的测试数据如表 3 所示。

测试结果表明:在 2 号点、3 号点及其连接路径上的定位精度比较好,同样在 1 号点出现了大的误差和波动,从 3 号点走到 2 号点的路径刻画十分理想,说明了测试数据的可信性。虽然行走的方向与 1 号点~3 号点相反,但 2 个方向的测试结果是一致的,2 号点和 3 号点载波相位差对应的两次位置差值分别为 8.4、8 m,比较接近,说

明了测试的可重复性及有效性。

3.2 在复杂室内条件下测试

在前期室外测试验证了系统工作原理及测试结果正确性的基础上,开展了实际复杂室内条件下的测试,测试场景为北京航空航天大学新主楼,如图 10 所示。

图中发射信号源放在位于大楼中间露天的地面绿点处,发射功率 4 W,测试点为 2 楼室内的红点处,可以看到环境比较复杂,尤其外圈的一些红点,信号要穿过多重建筑实体才能够到达。

通过在红点处逐个测试,采集信号源穿过楼体的到达信号,通过事后的相应处理,得到以下结果:

1) 在所有测试点(红点处)都可以接收、捕获到信号源发出的导航信号,说明本文的低频策略是有效的,该方法的覆盖能力较强。如在测试点 2 对 3 个基站信号的捕获情况如图 11 所示。

2) 由于环境复杂、多径严重等原因,测试数据的跟踪结果表明码相位和载波相位的误差都比较大,尤其信号路径长的测试点还出现了大的偏差,严重的已无法用于定位解算,需要进一步进行抗多径等处理。同样在测试点 2,对 3 个基站的码跟踪曲线及基站间的码伪距差曲线如图 12 所示。图 12(d)为 PRN1-2 码相位示意图,最大值和最小值分别为 83.071 m 和 72.570 m,标准差为 1.8479 m,均值为 78.605 4 m;图 12(e)为 PRN1-3 码相位示意图,最大值和最小值分别为 -21.904 7 m 和 -121.623 9 m,标准差为 34.529 4 m,均值为 -63.764 4 m;图 12(f)为 PRN1-3 码相位示意图,最大值和最小值分别为 57.223 8 m 和 -44.515 9 m,标准差为 34.770 9 m,均值为 14.840 9 m。从图中可以看出,各基站间的码伪距差波动较小,但其

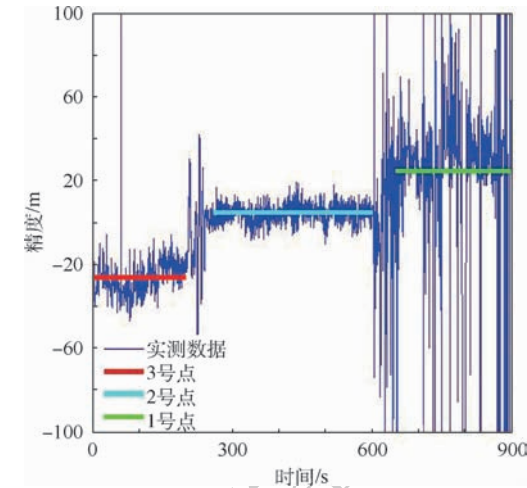


图 8 码相位差值(运动状态下测试,3 号点~1 号点)
Fig. 8 Code phase difference value (test under motion, point 3 to 1)

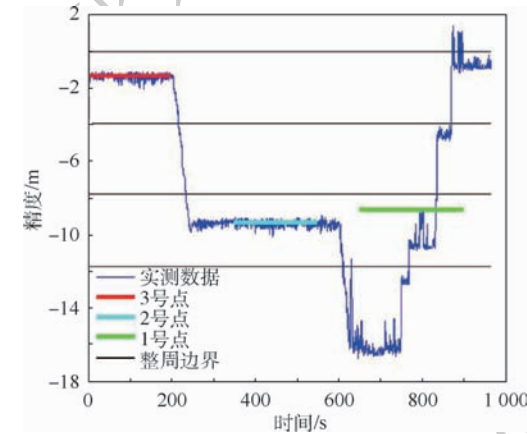


图 9 载波相位差值(运动状态下测试,3 号点~1 号点)
Fig. 9 Carrier phase difference value (test under motion, point 3 to 1)

表 3 测试数据(运动状态下测试,3 号点~1 号点)
Table 3 Test data (test under motion, point 3 to 1)

测试点	码相位差		载波相位差	
	均值	标准差	均值	标准差
3	-26.5	7.1	-1.38	0.138
2	4.4	4	-9.38	0.125
1	24.5	53.8	-8.67	6.4

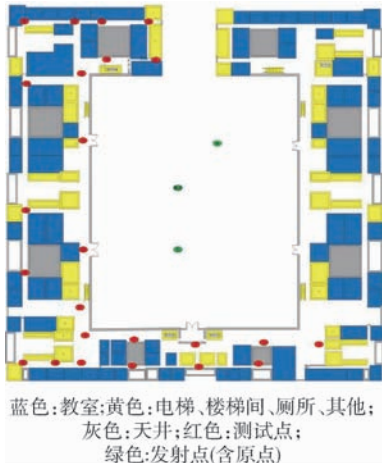


图 10 新主楼室内测试环境
Fig. 10 New main building indoor test environment

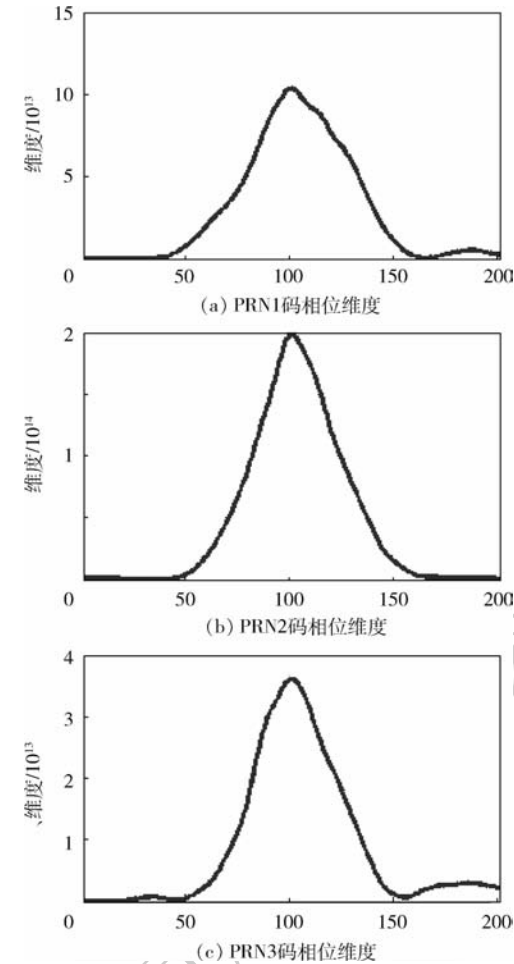


图 11 测试点 2 的捕获相关峰

Fig. 11 Capture correlation peak at test Point 2

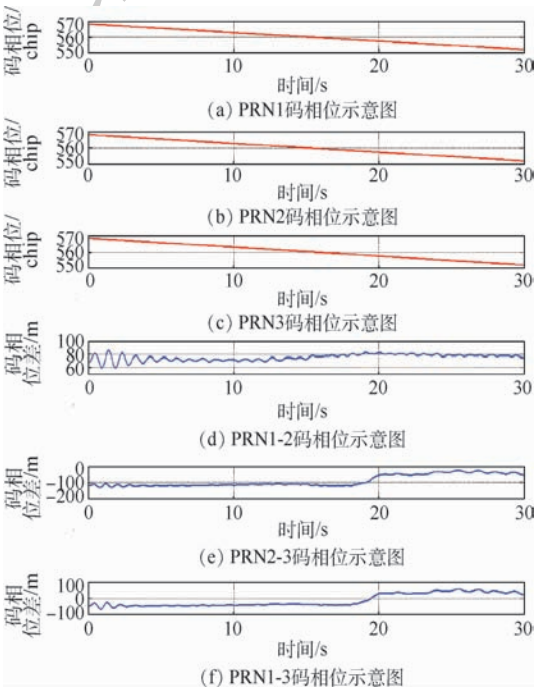


图 12 测试点 2 跟踪曲线及基站间距离差曲线

Fig. 12 Tracking curve and distance difference curves between base stations at test point 2

均值与事先标定的真值偏差较大,分析是由多径因素造成的固定偏差。

4 结 论

本文经过不同环境的测试并对结果进行分析,初步验证了所提方法的可行性,证明了系统定位的可行性和良好的覆盖能力,但也发现了系统进一步应用需解决的问题。综合分析本文测试结果,可以得出如下结论:

- 1) 在室外开阔环境(操场 2 号点、3 号点)的测试,位置差的精度较高,动态测试符合预期,说明了方法的可行性和有效性。
- 2) 如果接收天线被行人遮挡(操场 1 号点,有跑步人员),会出现码相位误差变大和载波相位跳变的情况,在使用中应该尽量避免。
- 3) 本文信号的穿透能力较强,对复杂室内的覆盖范围较大,但需要进一步解决多径严重,以及信号衰减大、信号质量下降等问题,才可以实现大区域的室内定位,这也是下一步的工作重点。

参考文献 (References)

[1] 刘蔚,康永. 室内定位技术应用分析研究[J]. 现代导航, 2016,7(2):86-93.
LIU W, KANG Y. Application analysis of indoor positioning technology[J]. Modern Navigation, 2016,7(2):86-93 (in Chinese).
[2] 阮陵,张翎,许越,等. 室内定位:分类、方法与应用综述[J]. 地理信息世界, 2015,22(2):8-14.
RUAN L, ZHANG L, XU Y, et al. Indoor positioning: A review of classification, methods and applications [J]. Geomatics World, 2015, 22(2):8-14 (in Chinese).
[3] LIU H, DARABI H, BANERJEE P, et al. Survey of wireless indoor positioning techniques and systems[J]. IEEE Transactions on Systems Man & Cybernetics Part C, 2007, 37(6):1067-1080.
[4] MAINETTI L, PATRONO L, SERGI I. A survey on indoor positioning systems [C] // International Conference on Software, Telecommunications and Computer Networks. Piscataway, NJ: IEEE Press, 2015:111-120.
[5] 邓中亮,尹露. 基于 TC-OFDM 体制的室内定位系统[J]. 电信网技术, 2015(3):32-35.
DENG Z L, YIN L. Indoor positioning system based on TC-OFDM system[J]. Information and Communications Technology and Policy, 2015(3):32-35 (in Chinese).
[6] BARNES J, RIZOS C, WANG J, et al. High precision indoor and outdoor positioning using LocataNet[J]. Positioning, 2003, 2(2):73-82.
[7] 赵锐,钟榜,朱祖礼. 室内定位技术及应用综述[J]. 电子科技, 2014,27(3):154-157.
ZHAO R, ZHONG B, ZHU Z L. Overview of indoor positioning

- technology and application[J]. Electronic Science and Technology, 2014, 27(3): 154-157 (in Chinese).
- [8] 许华. 连续相位调制技术研究及其实现[D]. 西安: 西安电子科技大学, 2009.
- XU H. Research and implementation of continuous phase modulation technology[D]. Xi'an: Xidian University of Electronic Technology, 2009 (in Chinese).
- [9] 冯思泉. 连续相位调制(CPM)技术研究[D]. 昆明: 昆明理工大学, 2007.
- FENG S Q. Research on continuous phase modulation (CPM) technology[D]. Kunming: Kunming University of Science and Technology, 2007 (in Chinese).
- [10] 黄嵩. CPFSK 调制解调的研究与实现[D]. 绵阳: 西南科技大学, 2011.
- HUANG S. Research and implementation of CPFSK modulation and demodulation[D]. Mianyang: Southwest University of Science and Technology, 2011 (in Chinese).
- [11] FOY W H. Position-location solutions by Taylor-series estimation[J]. IEEE Transactions on Aerospace and Electronic Systems, 1976, 12(2): 187-194.
- [12] 王成, 李少洪, 王鑫全, 等. 测时差被动定位算法的研究[J]. 系统工程与电子技术, 2001, 23(11): 9-12.
- WANG C, LI S H, WANG X Q, et al. Research on time difference differential positioning algorithm[J]. Systems Engineering and Electronics, 2001, 23(11): 9-12 (in Chinese).
- [13] HUANG Y T, BENESTY J, ELKO G W, et al. Real-time passive source localization: A practical linear-correction least-squares approach[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(8): 943-956.
- [14] 齐恒, 彭林宁, 姜禹, 等. 基于临近听域超声波 TDOA 室内定位的实现[J]. 数据采集与处理, 2018, 33(6): 1094-1100.
- QI H, PENG L N, JIANG Y, et al. Realization of indoor positioning based on near-listening ultrasonic TDOA[J]. Journal of Data Acquisition and Processing, 2018, 33(6): 1094-1100 (in Chinese).
- [15] 张令文, 杨刚. 超宽带室内定位关键技术[J]. 数据采集与处理, 2013, 28(6): 706-713.
- ZHANG L W, YANG G. Ultra-wideband indoor positioning key technology[J]. Journal of Data Acquisition and Processing, 2013, 28(6): 706-713 (in Chinese).

作者简介:

耿珂 男, 博士研究生。主要研究方向: 协作通信、无线电导航。

黄智刚 男, 博士, 教授, 博士生导师。主要研究方向: 通信与信息系统学科的无线电导航、卫星导航、自动测试与控制、航空电子。

Positioning system construction and scheme based on ground base station

GENG Ke, HUANG Zhigang*, SU Yu, SHI Peichen, GAO Qiang, XIONG Huagang

(School of Electronic and Information Engineering, Beihang University, Beijing 100083, China)

Abstract: For the harsh internal signal environment and the large signal attenuation of satellite navigation with the difficulties of capturing and tracking signals stably inside large-scale, complex and multifunctional buildings, the multipath effect is serious, and the short multipath has a great influence on the positioning accuracy. Taken in this sense, it is difficult to locate directly with the navigation signals. This paper proposes a method of indoor positioning method based on a large area of ground base station (building group). According to the relationship between frequency and signal penetration performance and spatial attenuation, the frequency band of very high frequency is selected as the signal carrier, with the combination of the code pseudorange and the carrier pseudorange, to achieve the coverage and positioning performance. In this paper, a basic test system is built by using the proposed new positioning method. The signal acquisition, tracking and pseudorange differential are carried out through signal generation, transmission and wireless propagation, which preliminarily verifies the feasibility and the coverage ability of the proposed method.

Keywords: large area; high-precision positioning; code pseudorange; carrier pseudorange; ground base station

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0030

边条/鸭翼对前掠翼和后掠翼气动特性的影响



张冬¹, 陈勇^{2,*}, 胡孟权², 付向恒³

(1. 空军工程大学 研究生院, 西安 710051; 2. 空军工程大学 航空工程学院, 西安 710038;

3. 北京航空航天大学 自动化科学与电气工程学院, 北京 100083)

摘 要: 为分析前掠翼气动布局设计在航空工业中无法得到推广运用的原因, 将前掠翼和后掠翼通过加装边条和鸭翼形成简化的边条翼布局、鸭式布局和边条/鸭式布局, 从而深入认识前掠翼和后掠翼两种不同布局之间的流动特点以及涡系干扰机理。首先进行算例数值计算, 通过对比分析计算结果与试验数据, 验证了数值计算方法的可靠性和准确性; 然后对不同布局进行数值计算, 得到各布局的升力系数曲线; 最后通过压力分布云图和流线图对各布局中复杂涡系的干扰机理进行分析。结果表明: 基于后掠机翼形成的边条翼布局、鸭式布局和边条/鸭式布局中的涡系之间通过诱导和卷绕作用, 涡系相互增强, 大幅提高了布局的升力系数并推迟失速迎角, 同时加装边条和鸭翼效果更加明显; 基于前掠机翼形成的边条翼布局、鸭式布局和边条/鸭式布局中的涡系之间不存在卷绕作用, 涡系之间存在碰撞挤压的不利干扰, 这使得前掠翼布局在大迎角时的升力系数远远低于相应的后掠翼布局。前掠翼气动布局中的机翼前缘涡在大迎角时无法同鸭翼涡和边条涡相互耦合增强, 不能充分地利用非线性升力, 这是前掠翼气动布局设计中的一些不足。

关 键 词: 边条; 鸭翼; 前掠翼; 后掠翼; 干扰机理

中图分类号: V221⁺.3; TB553

文献标识码: A

文章编号: 1001-5965(2019)10-2058-11

前掠翼气动布局具有优越的气动性能, 可大大提高飞机的低速操纵性能, 显著减小跨声速飞行时的阻力并增强机动性, 在大迎角下具有良好的失速特性和横航向可控性^[1]。但是, 在主流的飞机气动布局设计中, 几乎都没有采用前掠翼气动布局设计, 美国的“X-29”前掠翼验证机和俄罗斯的“金雕”前掠翼验证机都未能进入最终的定型量产阶段, 已有文献给出最多的解释是前掠翼布局存在的结构气动弹性发散问题。但随着复合材料技术的迅速发展, 前掠翼气动弹性发散问题得到了较好的解决, 且带来了结构重量轻、隐身性能高等一系列优点^[2-3]。尽管如此, 前掠翼气动

布局设计依然没有得到广泛运用, 气动弹性发散问题的解决并没有重新引起航空界对前掠翼气动布局的重视和研究, 前掠翼布局无法得到推广运用的原因尚未得到有力解释。为此, 本文将从前掠翼在非线性升力利用方面的特点来分析该种布局的一些不足, 以揭示前掠翼气动布局设计未能得到广泛运用的另一重要原因。

非线性升力是由旋涡空气动力带来的气动收益。为了充分利用非线性升力, 进一步提高飞机的大迎角性能, 后来出现了边条翼布局和鸭式布局, 即通过机翼边条和近距耦合鸭翼产生边条涡和鸭翼涡, 使其与机翼前缘涡产生有利干扰, 增强

收稿日期: 2019-01-22; 录用日期: 2019-05-28; 网络出版时间: 2019-06-11 18:03

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190611.1040.002.html

基金项目: 国家自然科学基金 (61473307)

* 通信作者: E-mail: chen_y043@163.com

引用格式: 张冬, 陈勇, 胡孟权, 等. 边条/鸭翼对前掠翼和后掠翼气动特性的影响[J]. 北京航空航天大学学报, 2019, 45(10): 2058-2068. ZHANG D, CHEN Y, HU M Q, et al. Effect of strake and canard on aerodynamic characteristics of forward-swept wing and back-swept wing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2058-2068 (in Chinese).

并稳定机翼前缘涡,从而提高飞机的升力和失速迎角^[4-5]。

由于前后掠机翼上气流的流动形态不同,采用相同的边条和鸭翼会对后面机翼的流场造成不同的影响,两者之间有着不同的干扰机理。关于后掠翼气动布局已有大量研究,Gloss^[6]、Hummel^[7-8]和刘沛清^[9-12]等通过实验的手段研究了边条和鸭翼的位置、大小和后掠角度等参数对鸭式布局 and 边条翼布局的气动影响。研究表明,在大迎角时,鸭翼涡和边条涡能够同后掠翼的机翼前缘涡相互耦合增强,从而提高后掠翼布局的升力系数和失速迎角。关于前掠翼气动布局也有丰富的研究,Spacht^[13]和 Mann^[14]研究了当前技术的发展对前掠翼气动布局设计的影响,结果表明复合材料、电传操纵和鸭翼布局等技术能够使前掠翼布局的优势得到进一步拓展。通过对“X-29”前掠翼验证机的研究,Moore 和 Frei^[15]、Kehoe 等^[16]指出前掠翼的展向流动是由翼尖流向翼根,使机翼展向升力分布更接近椭圆,因而具有较小的诱导阻力,且前掠翼翼根后置,有利于采用鸭翼耦合设计;Saltzman 和 Hicks^[17]研究发现前掠翼的中、外翼展向流动具有较好的分离特性,机翼失速性能良好。张彬乾等^[18-19]采用边条和鸭翼对前掠翼根部流动分离进行控制,研究结果表明边条和鸭翼能够有效控制前掠翼翼根的流动分离。王晋军等^[20-21]对鸭翼-前掠翼布局进行了实验研究,结果表明,加装鸭翼能够从整体上提高前掠翼的升力系数。李岸一等^[22]研究了鸭翼涡与边条涡对前掠翼布局的增升作用,结果表明鸭翼和边条都能提升前掠翼布局的升力系数。以上研究对分析边条和鸭翼对前掠翼布局的气动影响都有一定的参考价值,但都未提及鸭翼涡和边条涡在大迎角时对机翼前缘涡的不利干扰,这是前掠翼布局中存在的一个明显缺陷。早期的研究由于方法的限制,大多只给出了测力结果,而没有具体的涡系干扰分析,且在研究时把边条和鸭翼的升力计入后再与单独前掠翼作对比,这只能看出边条和鸭翼对整机的气动影响,而无法看到边条和鸭翼对单独前掠翼的气动影响。另外,在研究过程中没能和后掠翼布局对比分析,也不易发现边条和鸭翼对前掠翼布局的气动影响。

随着计算机性能的提升以及计算流体力学的发展,人们越来越多地采用数值模拟方法进行气动特性的研究以及复杂涡系的分析^[23-24]。因此,本文采用数值模拟方法,分别分析边条和鸭翼对后掠翼布局和前掠翼布局的气动影响,通过对各

布局计算结果的比较和涡系干扰机理的分析,揭示前掠翼气动布局设计中的一些不足。

1 计算模型和方法

1.1 计算模型

计算采用平板模型,基于前后掠机翼,再搭配边条和鸭翼形成不同简化布局形式,为方便叙述,采取如下简记方法:B 代表单独后掠机翼,基于后掠机翼的边条翼布局用 BS 表示,鸭式布局用 BC 表示,边条翼/鸭式布局用 BSC 表示;F 代表单独前掠机翼,基于前掠机翼的边条翼布局用 FS 表示,鸭式布局用 FC 表示,边条翼/鸭式布局用 FSC 表示。机翼根弦长为 124 cm,鸭翼根弦长为 73 cm,模型厚度为 1 cm,边条、鸭翼和机翼前缘皆倒角 45°,保持前后掠机翼翼面积、根梢比、展弦比、1/4 弦线斜掠角相同,其中翼面积为 1.26 m²,展弦比为 3.2,根梢比为 2.38,1/4 弦线斜掠角为 40°。鸭翼、边条与前后掠机翼在同一平面,前后掠机翼采用相同的边条和鸭翼,鸭翼前缘后掠 56°,边条后掠 70°,图 1 为 BCS 和 FCS 布局平面示意图。计算区域取机翼根弦长的 50 倍,来流速度为 68 m/s。网格剖分时采用结构网格,考虑到模型的几何对称性,在计算时均采用半模,边界层内第 1 层网格高度为机翼平均气动弦长的 10⁻⁶,边界层内第一层高度 y^+ 值控制在 0~1,以满足飞机表面黏性边界层的计算要求^[25]。图 2 为计算模型的网格分布示意图,图 2(a)为 FSC 表面网格分布,图 2(b)为近壁面网格分布。

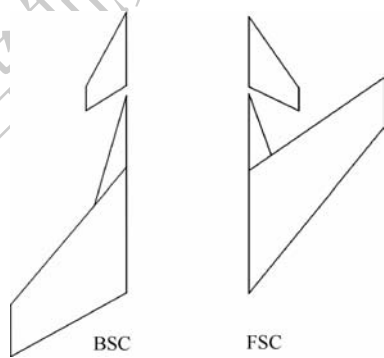


图 1 BCS 和 FCS 布局平面示意图

Fig. 1 Plan sketches of BCS and FCS configurations

1.2 计算方法

数值模拟时采用无热源的三维 Navier-Stokes 方程。在笛卡儿坐标系 (x_1, x_2, x_3) 中,定义速度分量 (u_1, u_2, u_3) 无热源的三维 Navier-Stokes 方程守恒形式为^[26]

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{f}_i}{\partial x_i} = \frac{\partial \mathbf{f}_{vi}}{\partial x_i} \quad (1)$$

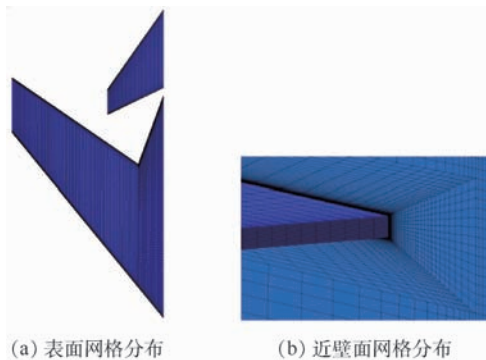


图2 计算模型网格分布

Fig. 2 Grid distribution of computational model

式中: w 为状态矢量; f 为无黏(对流)通矢量项; f_v 为黏性(扩散)通矢量项。各项具体表达式分别为

$$w = [\rho \quad \rho u_1 \quad \rho u_2 \quad \rho u_3 \quad \rho E]^T$$

$$f_i = \begin{bmatrix} \rho u_i \\ \rho u_i u_1 + p \delta_{i1} \\ \rho u_i u_2 + p \delta_{i2} \\ \rho u_i u_3 + p \delta_{i3} \\ \rho u_i H \end{bmatrix}$$

$$f_{vi} = \begin{bmatrix} 0 \\ \tau_{ij} \delta_{j1} \\ \tau_{ij} \delta_{j2} \\ \tau_{ij} \delta_{j3} \\ u_j \tau_{ij} + k \frac{\partial T}{\partial x_i} \end{bmatrix}$$

其中: ρ 、 E 、 H 、 p 和 T 分别为密度、总能、总焓、压强和温度; k 为热传导系数; δ_{ij} 为克罗尼柯尔符号; τ_{ij} 为黏性应力。采用有限体积法将控制方程离散为差分方程,其中对流项和扩散项分别采用二阶迎风格式和中心差分格式。湍流模型选用 SST (Shear Stress Transport) $k-\omega$ 模型,即剪切应力输运模型。SST $k-\omega$ 模型在标准 $k-\omega$ 模型上进行了改进,合并了来源于 ω 方程中的交叉扩散,湍流黏度考虑到了湍流剪应力的传播。这些改进,使得 SST 模型比标准 $k-\omega$ 模型具有更广泛的应用范围和更高的精度^[27]。控制方程采用有限体积法离散,对流项选用二阶迎风差分格式离散^[28]。物面为无滑移壁面条件,远场条件为压力远场,对称面为对称边界条件,计算残差收敛精度为 10^{-5} 。

1.3 方法验证与确认

为验证数值计算方法的可行性和精确性,对文献[29-30]的前掠翼模型进行数值计算验证,如

图 3(a) 所示,机翼前缘前掠 40° 、后缘后掠 52° ,展弦比为 3.81,梢根比为 0.4。后掠鸭翼的前、后缘掠角分别为 49° 和 24° ,展弦比为 3.08,梢根比为 0.3。机翼和鸭翼在顺气流方向均采用 NACA 64A010 翼型。计算采用结构网格,单元个数约 500 万,附面层最底层网格厚度控制在 1.68×10^{-7} 以下(L_F 为机身长度),见图 3(b)。计算采用 SST $k-\omega$ 湍流模型,结果如图 4 所示,并与文献[30]的风洞试验数据进行对比,图中, α 为迎角, C_L 为升力系数。从图中可以看出,数值计算结果与试验数据吻合较好,说明所采用的数值计算方法具有较高的精度,适用于本文研究模型的气动特性和流动机理的计算和分析。

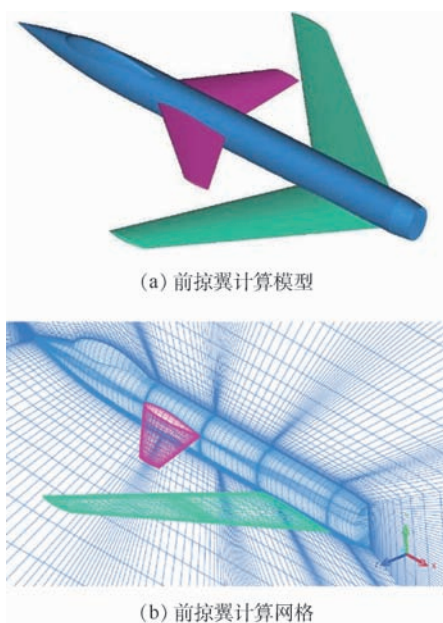


图3 前掠翼计算模型和计算网格

Fig. 3 Computational model and grid of forward-swept wing

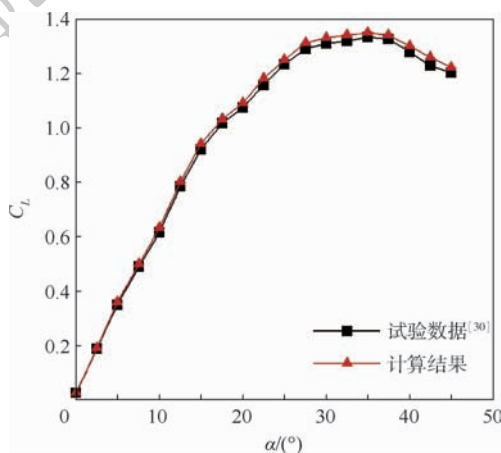


图4 计算结果与试验数据的升力系数曲线

Fig. 4 Lift coefficient curves of calculation results and experimental data

2 结果分析与讨论

2.1 边条/鸭翼对后掠翼气动特性的影响

本文数值模拟的模型采用图 1 所示的平板模型,针对不同的布局单独进行数值模拟研究,先单独计算后掠机翼的升力系数,然后分别对 BC、BS 和 BSC 布局进行数值模拟研究。

图 5 给出了后掠翼各布局升力系数的计算结果,选择主机翼面积作为参考面积。图 5(a)计入了边条和鸭翼的升力贡献,反映整个布局的升力系数随迎角的变化趋势。为更加直观地反映边条和鸭翼对后面机翼的干扰作用,图 5(b)仅考虑机翼上的升力,不计入边条和鸭翼的升力。从图 5(b)可以看出,BC、BS、BSC 布局都能明显提高后掠机翼的最大升力系数,并推迟失速迎角。单独后掠机翼的最大升力系数为 0.788,失速迎角为 16° ;加装边条后(BS 布局),机翼的最大升力系数提高到 0.98,失速迎角推迟到 18° ;加装

鸭翼后(BC 布局),机翼的最大升力系数提高到 1.05,失速迎角推迟到 23° ;同时加装边条和鸭翼(BSC 布局),机翼的最大升力系数提高到 1.11,失速迎角推迟到 28° 。3 种方案中,BSC 方案对后掠机翼气动特性的改善效果最好,不仅大幅提高机翼的最大升力系数,推迟失速迎角,还使得失速时升力系数没有出现剧烈下降。

2.1.1 单独后掠机翼的流动分析

由于采用的计算模型属于中等掠角机翼,机翼的升力系数受旋涡外侧气流的卷洗作用影响,在分析过程中发现,旋涡涡核破裂后,旋涡外侧气流依然保持有规则的旋转运动,还有较大的能量,对翼面的控制作用较强,只有当旋涡外侧气流能量下降,对翼面的控制能力减弱时,机翼升力系数才开始下降,所以本文主要通过旋涡外侧气流的形态来分析旋涡的发展。

图 6 给出了不同迎角下单独后掠机翼表面压力云图和空间流线图。从图中可以看到,当 $\alpha = 5^\circ$ 时,机翼上表面开始形成机翼前缘涡,此时旋涡能量很低,靠近机翼前缘;当 $\alpha = 10^\circ$ 时,机翼前缘涡能量增强,控制区域扩大;当 $\alpha = 16^\circ$ 时,旋涡对翼面的控制能力达到最强,升力系数达到最高点;继续增大迎角,机翼前缘涡能量下降,出现倒流,机翼进入失速状态。为便于分析,全文采用与图 4 相同的压力云图进行渲染,后文压力云图参考此处图例。

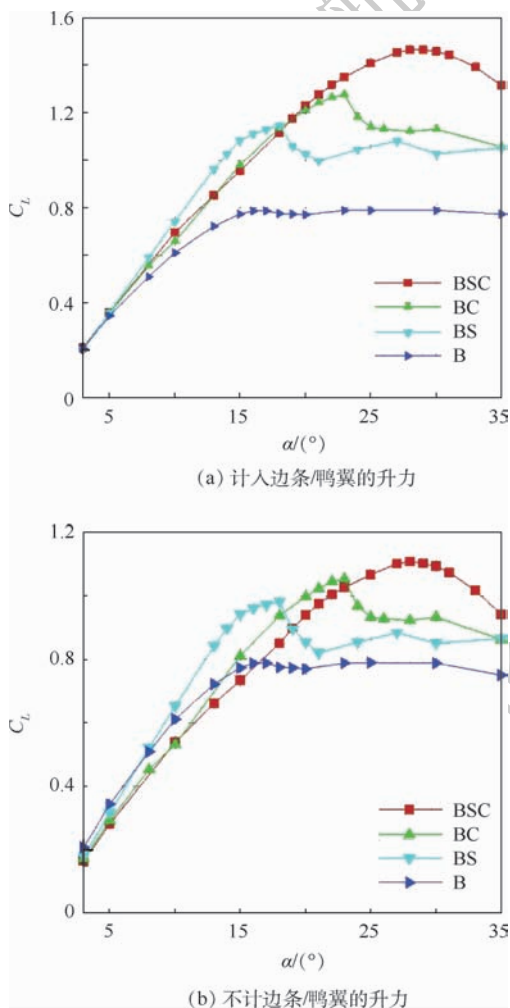


图 5 后掠机翼各布局的升力系数曲线
Fig.5 Lift coefficient curves of various configurations based on back-swept wing

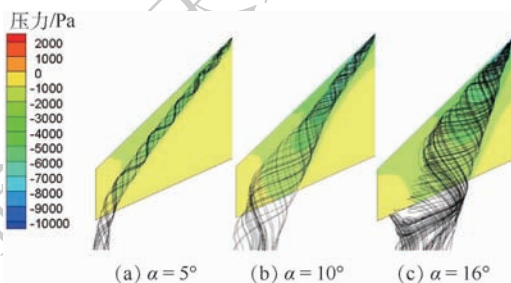


图 6 单独后掠机翼表面压力云图和空间流线图
Fig.6 Surface pressure contours and streamlines in flow field of separate back-swept wing

2.1.2 BS 布局的流动分析

图 7 给出了不同迎角下 BS 布局的空间流线图。从图中可以看出,在 $\alpha = 5^\circ$ 时,BS 布局的边条涡与机翼前缘涡还未耦合,二者相对独立,能量较为分散,所以在小迎角下 BS 方案中机翼的升力系数比单独机翼的升力系数略低。当 $\alpha = 10^\circ$ 时,BS 布局中的边条涡与机翼前缘涡开始卷并在一起,旋涡能量明显增强,控制区域扩大,升力系数提高,并超过单独机翼的升力系数。继续增大

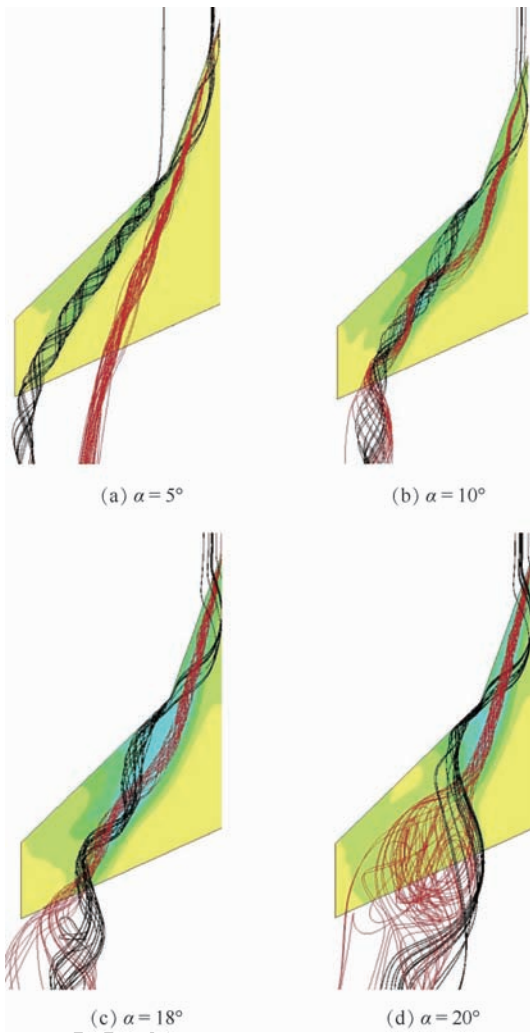


图 7 BS 布局空间流线图

Fig. 7 Streamlines in flow field of BS configuration

迎角,边条涡与机翼前缘涡的能量继续增强,并且二者的耦合作用进一步加强,两涡卷并得更紧,对翼面的控制范围更大,使得机翼的升力系数进一步提高。当 $\alpha = 18^\circ$ 时,边条涡与机翼前缘涡耦合作用达到最强,对翼面的控制作用达到最大,升力系数达到最高点。迎角进一步增大,边条涡与机翼前缘涡耦合后的旋涡开始从机翼后缘向前缘破裂,破裂后旋涡的能量急剧降低,如图 7(d) 所示,在 $\alpha = 20^\circ$ 时,旋涡破裂点已出现在机翼前部,旋涡破裂的区域占机翼面积很大部分,因而此时升力系数急剧下降,出现明显失速。

2.1.3 BC 布局的流动分析

图 8 给出了 $\alpha = 10^\circ$ 时 BC 布局中机翼 20% 和 60% 翼根弦长处的截面流线图以及空间流线图。从图中可以看出,在机翼前部,鸭翼涡处于机翼前缘涡外侧,对机翼前缘涡起到下洗作用,使得该区域有效迎角降低,抑制了机翼前缘涡的发展,造成升力损失;在机翼后部,鸭翼涡处于机翼前缘涡内侧,

鸭翼涡对机翼前缘涡产生上洗作用,输入能量的同时将机翼前缘涡外推,一方面有利于机翼前缘涡的稳定,但另一方面却使得机翼前缘涡控制的区域变小,使其对机翼升力贡献减少。上述原因使得中小迎角下 BC 布局中机翼的升力系数小于单独机翼的升力系数。随着迎角增大,鸭翼和机翼前缘涡能量增强,鸭翼涡对机翼前缘涡的作用以内侧上洗为主,稳定机翼前缘涡并输入能量,机翼的升力系数显著提高,并逐渐超过单独机翼的升力系数。当迎角达到 23° 时,BC 布局中旋涡发展如图 9(a) 所示,此时机翼的升力系数达到最大,旋涡对翼面的控制作用最强;当迎角继续增大,鸭翼涡和机翼前缘涡开始破裂,两涡的耦合作用降低,机翼出现明显失速现象,图 9(b) 为 $\alpha = 25^\circ$ 时翼面旋涡破裂时的流线图。

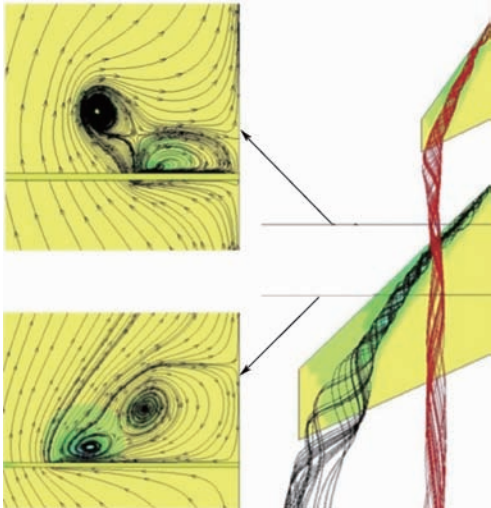


图 8 BC 布局截面及空间流线图($\alpha = 10^\circ$)

Fig. 8 Streamlines over clip plane and flow field of BC configuration ($\alpha = 10^\circ$)

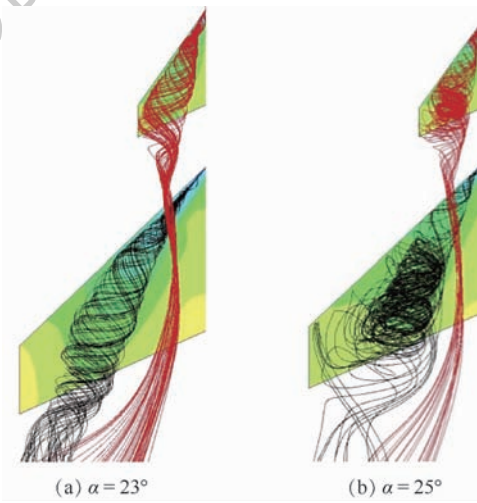


图 9 BC 布局空间流线图

Fig. 9 Streamlines in flow field of BC configuration

2.1.4 BSC 布局的流动分析

图 10 给出迎角 $\alpha = 10^\circ$ 时 B、BS 和 BSC 布局表面压力云图,BSC 布局同时给出鸭翼涡流线图。从 BSC 布局鸭翼涡形态和位置可以看出,在机翼前面部分,BSC 布局中的鸭翼涡位于边条外侧,边条几乎全部位于鸭翼涡的下洗区域,因而边条涡的发展受到很强的抑制作用。当鸭翼涡发展到机翼时,鸭翼涡位于机翼前缘涡内侧,对机翼前缘涡产生上洗,增加旋涡的能量,但由于鸭翼涡对机翼前缘涡有一个外推作用,使得机翼前缘涡控制的区域较小,因而 BSC 布局中机翼的升力系数比单独机翼的升力系数小。从图 10 可以看到,在 BSC 布局中边条的压力明显高于 BS 布局中边条压力,这是由鸭翼涡下洗作用造成。从图中还可以看到,BSC 布局中机翼上表面的低压区比单独机翼要少一些,这是由于鸭翼涡外推机翼前缘涡,使得机翼前缘涡在机翼上表面控制的区域减少。随着迎角增大,边条涡与机翼前缘涡逐渐增强,并相互卷并,卷并后旋涡的能量增强,鸭翼涡对卷并后旋涡的上洗作用逐渐占主导地位,一方面增强了旋涡的能量,提高了机翼的升力系数,另一方面使得旋涡变得更加稳定,延迟了旋涡破裂。

图 11 为 BSC 布局表面压力云图和空间流线图。从图 11 中可以看到,旋涡向机翼前缘破裂速度变得缓慢,迎角从 25° 增加到 30° ,旋涡破裂点还保持在机翼中部,这就使得机翼可以在失速迎角前后很大的范围保持较大的升力系数,并且变化平缓。

2.2 边条/鸭翼对前掠翼气动特性的影响

本节和 2.1 节一样,都采用平板模型进行数值

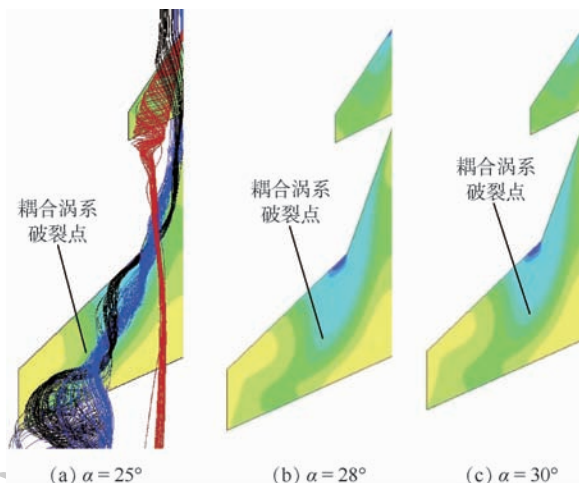


图 11 BSC 布局表面压力云图和空间流线图

Fig. 11 Surface pressure contours and streamlines in flow field of BSC configuration

模拟。针对不同的布局单独进行数值模拟研究,先单独计算前掠机翼的升力系数,然后分别对 FC、FS 和 FSC 布局进行数值模拟研究。

图 12 给出了前掠机翼各布局升力系数的计算结果,选择单独前掠翼面积作为参考面积。和分析后掠机翼一样,图 12(a) 考虑了边条和鸭翼的升力贡献,而图 12(b) 仅考虑前掠机翼上的升力,不计入边条和鸭翼的升力。从图 12(a) 可以看出,相对于单独前掠机翼,FS、FC、FSC 布局的最大升力系数都得到提高,失速迎角有所推迟,且在失速迎角附近,升力系数变化平缓,这和大多数研究结果一致。但去除边条和鸭翼的升力贡献,如图 12(b) 所示,FS、FC 和 FSC 布局中前掠机翼的升力系数低于单独前掠机翼的升力系数,FSC 布局只是在更大迎角时略高于单独前掠机翼的升力系数。从这里可以看出,去除边条和鸭翼的升力贡献,边条和鸭翼对前掠机翼起到不利的干扰作用。

对比图 5(b) 和图 12(b) 可以看出,不计边条和鸭翼的升力贡献,加装边条和鸭翼后,在大中迎角范围,后掠机翼的升力系数都高于单独后掠机翼,而前掠机翼在加装边条和鸭翼后,其升力系数基本低于单独前掠机翼。从这可以看出,后掠机翼可以很好利用边条涡和鸭翼涡的有利干扰,从而提高最大升力系数和失速迎角,而前掠机翼在边条涡和鸭翼涡的利用方面明显不如后掠翼。但对比前掠翼和后掠翼各种布局在失速时的变化曲线可以看出,前掠翼呈现出缓失速的特点,在失速时没有出现后掠翼那样大幅的下降,这有利于飞行的稳定与控制。

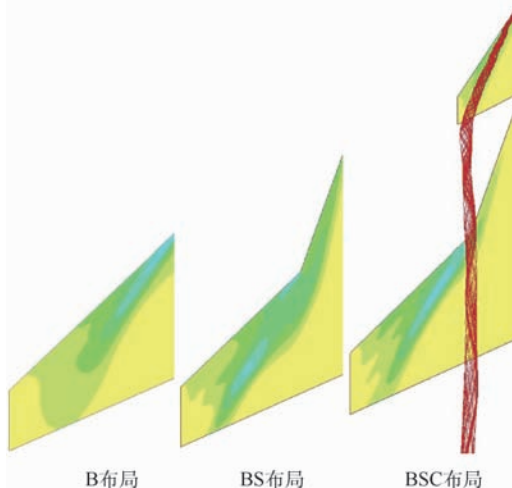
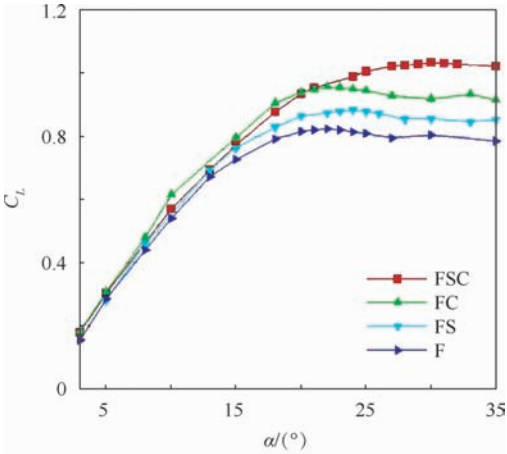
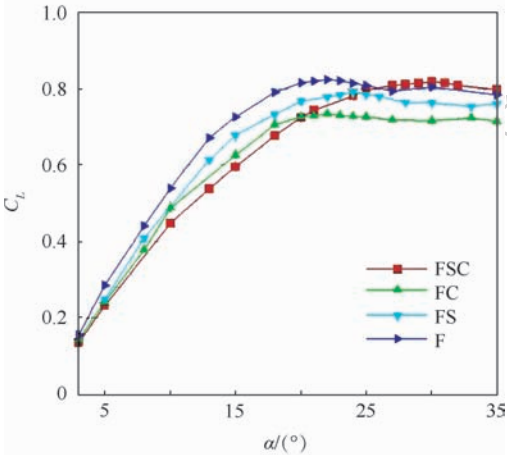


图 10 不同布局表面压力云图 ($\alpha = 10^\circ$)

Fig. 10 Surface pressure contours of different configurations ($\alpha = 10^\circ$)



(a) 计入边条/鸭翼的升力



(b) 不计边条/鸭翼的升力

图 12 前掠机翼各布局的升力系数曲线
Fig. 12 Lift coefficient curves of various configurations based on forward-swept wing

2.2.1 单独前掠机翼的流动分析

图 13 给出了单独前掠机翼在不同迎角下机翼前缘涡的发展变化。当 $\alpha = 5^\circ$ 时，机翼上表面开始形成机翼前缘涡，此时旋涡能量较低；当 $\alpha = 10^\circ$ 时，机翼前缘涡能量增强，控制区域扩大；当 $\alpha = 22^\circ$ 时，旋涡对翼面的控制作用达到最大，升力系数达到最高点；继续增大迎角，机翼前缘涡能量下降，对翼面的控制能力减弱并出现倒流，机翼进入失速状态。

从计算结果可以看到，单独前掠机翼比单独后掠机翼的失速迎角更大，从图 6 和图 13 中机翼空间流线图可以看到，单独前掠机翼在 $\alpha = 22^\circ$ 时机翼前缘涡依然稳定，但单独后掠机翼在 $\alpha = 20^\circ$ 就已经出现大规模倒流。首先可以从几何参数分析，在保持前后掠机翼 1/4 弦线斜掠角相等时，虽然后掠机翼比前掠机翼的前缘斜掠角大，但前掠机翼的后缘斜掠角却比后掠机翼的大，前掠机翼有效斜掠角更大，较大的斜掠角更有利于旋涡的稳定；另外，从旋涡本身特点来看，后掠机翼的

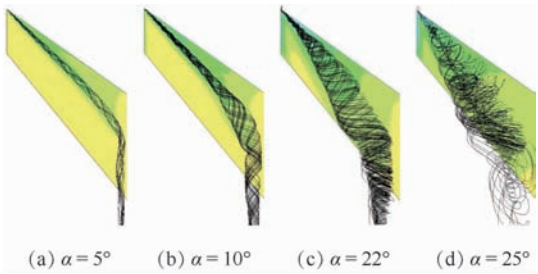


图 13 单独前掠机翼空间流线图
Fig. 13 Streamlines in flow field of separate forward swept wing

旋涡从翼根流向翼尖，在旋涡破裂过后，呈自由流向后发展，而前掠机翼的旋涡是由翼尖流向翼根，在对称面处左右机翼的前缘涡相互切洗，产生有利干扰，使旋涡变得更加稳定。

2.2.2 FS 布局的流动分析

图 14 给出了 $\alpha = 10^\circ$ 时 FS 布局中机翼 20% 翼根弦长处的截面流线图以及空间流线图。从图中可以看出，在 FS 布局中的机翼受机翼前缘涡和边条涡控制，两涡相对独立，不像 BS 布局的机翼前缘涡和边条涡之间存在相互卷绕和卷并作用，两涡不能通过耦合作用增强自身能量。相反，在 FS 布局中，由于边条涡和机翼前缘涡产生的位置较远，且两涡轴向成一定角度向后发展，在机翼上两涡相互碰撞挤压，造成能量损失，这对两涡的发展起到不利影响。但从图 14 的截面流线图可以看出，机翼前缘涡和鸭翼涡旋转方向相反，且两涡外切，相互上洗，能量得到增强，使得旋涡变得稳定，这是两涡之间的有利干扰。

从计算结果来看，FS 布局中机翼的失速迎角为 25° ，且升力系数在失速迎角附近变化平缓。图 15 给出了 FS 布局在失速迎角附近旋涡的发展变化。从图中可以看到，在迎角为 23° 时，机翼前缘涡中心已出现倒流，但由于受到边条涡的

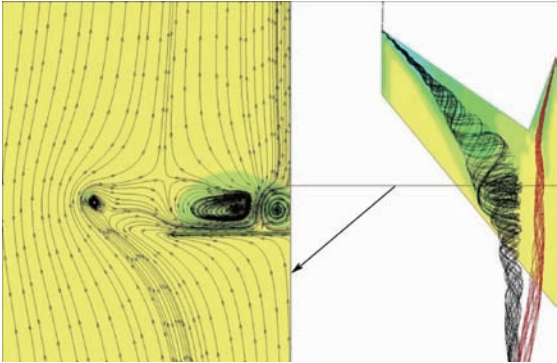


图 14 FS 布局截面及空间流线图 ($\alpha = 10^\circ$)
Fig. 14 Streamlines over clip plane and flow field of FS configuration ($\alpha = 10^\circ$)

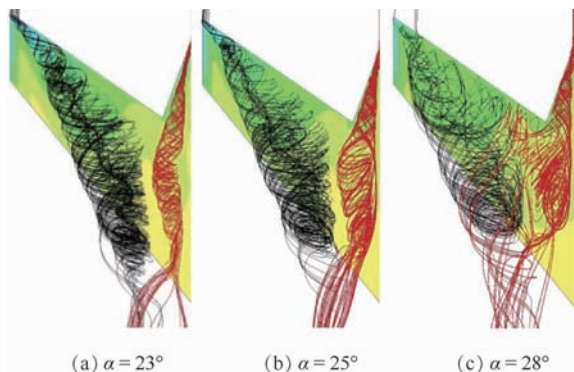


图 15 FS 布局空间流线图

Fig. 15 Streamlines in flow field of FS configuration

上洗作用,机翼前缘涡外侧气流能量得到增强,所以旋涡并没有迅速破裂,没有造成升力系数下降。另外,边条涡在此时依然保持稳定,其控制着机翼翼根上翼面流动,也使得机翼升力系数保持稳定。当迎角达到 25° 时,旋涡对翼面的控制作用达到最强,升力系数达到最高点。继续增大迎角,机翼前缘涡和边条涡开始破裂,机翼升力系数逐渐下降。

对比 BS 和 FS 两种布局可以看到,尽管 FS 布局在大迎角附近具有缓失速的特点,但其机翼前缘涡与边条涡不能相互耦合增强,无法像 BS 布局中的那样显著提高最大升力系数。对于 BS 布局,去除边条升力贡献,最大升力系数提高了 0.2,计入边条升力贡献,最大升力系数提高了 0.35。对于 FS 布局,去除边条升力贡献,最大升力系数下降了 0.04,计入边条升力贡献,最大升力系数也仅仅提高了 0.06。BS 布局在大迎角附近的升力系数得到显著提升,但在耦合涡系破裂后升力系数会出现急剧下降。

2.2.3 FC 布局的流动分析

图 16 给出了 $\alpha = 15^\circ$ 时 FC 布局机翼 20% 根弦长处的截面流线图和空间流线图。从图中可以看出,鸭翼涡对机翼前缘涡有一个明显的外推作用,这使得机翼前缘涡的发展受到不利影响,对翼根的控制能力减弱,从而引起较大的升力损失,从图 12 可以看出,去除鸭翼自身的升力贡献,FC 布局中机翼的升力系数明显低于单独机翼的升力系数。从截面流线图可以看到,鸭翼涡与机翼前缘涡旋转方向相反,两涡相互外切,能量相互增强,这是两涡之间的有利干扰,使得旋涡外侧的气流变得稳定,旋涡破裂变得缓慢。

图 17 给出了 FC 布局在失速迎角附近旋涡的发展变化图。从图中可以看到,当迎角为 20° 时,机翼前缘涡的涡核就出现倒流,而单独前掠机

翼的机翼前缘涡发展良好,涡核还没出现倒流。当迎角达到 22° 时,鸭翼涡和机翼前缘涡涡核都出现倒流,但由于两涡外侧气流相互切洗,能量有所增强,机翼升力系数在此时达到最大。继续增大迎角,鸭翼涡和机翼前缘涡外侧气流能量降低,对翼面控制能力减弱并出现大规模倒流,升力系数下降。

对比 BC 和 FC 两种布局可以看到,BC 布局中的机翼前缘涡与鸭翼涡可以通过相互耦合作用,明显提升最大升力系数,去除鸭翼升力贡献,最大升力系数提高了 0.205,计入鸭翼升力贡献,最大升力系数提高了 0.48。而 FC 布局中的机翼前缘涡与鸭翼涡则不能相互耦合增强,去除鸭翼升力贡献,最大升力系数下降了 0.1,计入鸭翼升力贡献,最大升力系数提高了 0.13。BS 布局在大迎角附近的升力系数得到大幅提升,但在耦合效果破裂后升力系数同样会出现急剧下降。

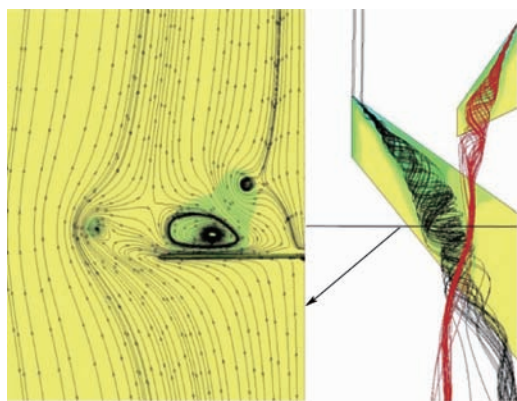


图 16 FC 布局截面及空间流线图($\alpha = 15^\circ$)

Fig. 16 Streamlines over clip plane and flow field of FC configuration ($\alpha = 15^\circ$)

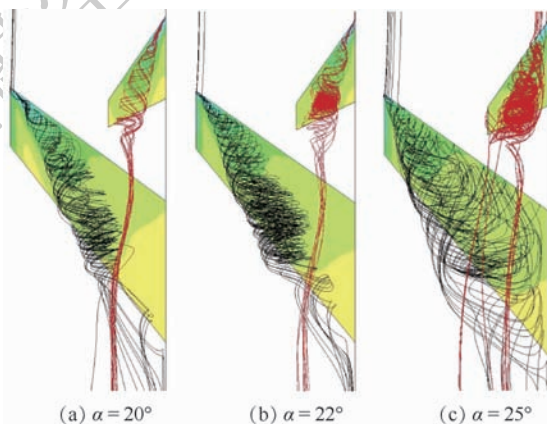


图 17 FC 布局空间流线图

Fig. 17 Streamlines in flow field of FC configuration

2.2.4 FSC 布局的流动分析

FSC 布局中机翼受到机翼前缘涡、边条涡和鸭翼涡的控制,三涡之间的干扰较为复杂。在中

小迎角时,鸭翼涡抑制了边条涡的发展,又因为鸭翼涡和边条涡对机翼前缘涡的外推作用,使得 FSC 布局中机翼的升力系数在中小迎角时较低;在大迎角下,鸭翼涡和边条涡相互耦合,能量得到增强,旋涡变得稳定,但增强后的旋涡对机翼前缘涡的不利干扰更加明显,外推机翼前缘涡并使之加速破裂。图 18 是 $\alpha = 23^\circ$ 时 FS 和 FSC 布局的涡系发展对比。从图中可以看出, FSC 布局中的机翼前缘涡向后发展受到更强的阻挡,破裂较为严重, FSC 布局中的边条涡在前面部分受到鸭翼涡的诱导,变得比较稳定,边条涡发展到后面部分,与机翼前缘涡发生碰撞,后面部分的边条涡随之破散。由于 FSC 布局中机翼上表面受机翼前缘涡和增强的边条涡控制,虽然机翼前缘涡对机翼上表面的控制能力减弱,但鸭翼涡与边条涡耦合之后的旋涡对翼根的控制能力得到了增强,且翼根面积的比重大,对机翼升力的影响更大,所以 FSC 布局在机翼前缘涡破裂时并没有失速。由于边条的后掠角较大,旋涡在大迎角下依然保持稳定,又由于边条涡受到鸭翼涡的下洗作用,边条涡破裂速度变得减缓,所以 FSC 布局在迎角达到 30° 时才开始失速,并且升力系数在失速迎角附近变化缓慢,失速特性良好。

对比 BSC 布局和 FSC 布局可以看到,在 BSC 布局中,鸭翼涡、边条涡和机翼前缘涡三涡可以相互耦合增强,显著提高大迎角时的升力系数和失速迎角,去除鸭翼与边条升力贡献,最大升力系数提高了 0.3,计入鸭翼与边条升力贡献,最大升力系数提高了 0.6。而且与 BS 和 BC 两种布局相比, BSC 布局失速后的升力下降速度也明显变缓,失速特性得到优化。而在 FSC 布局中,鸭翼涡

和边条涡能够相互耦合增强,但两者与机翼前缘涡却相互碰撞、挤压,阻碍了自身涡系的发展,无法像 BSC 布局那样通过三涡相互耦合增强带来明显的升力提高,去除鸭翼与边条升力贡献,最大升力系数仍略有降低,计入鸭翼与边条升力贡献,最大升力系数提高了 0.2。

3 结 论

1) 在前掠机翼加装边条和鸭翼形成的布局中,尽管各布局的升力系数在失速迎角附近变化缓慢,具有缓失速特性,但前掠翼各布局中的机翼前缘涡不能与边条涡和鸭翼涡相互耦合增强,反而受到两涡的不利干扰,大迎角时升力系数明显低于相应的后掠翼布局。

2) 在后掠机翼加装边条和鸭翼形成的布局中,涡系之间通过卷绕和卷并作用,旋涡相互耦合增强,可以提高布局的最大升力系数,并且推迟失速迎角,同时加装边条和鸭翼的 BSC 布局效果最为明显。

3) 大迎角时具备较大的升力系数能够显著提升飞机的机动性,这是飞机气动布局设计时重要的考量指标。后掠翼布局可以有效利用非线性升力以提高大迎角时的升力系数和失速迎角,而前掠翼由于其特殊的流动特点,其机翼前缘涡无法有效耦合于鸭翼涡和边条涡,因而在大迎角时的升力系数较低,这极大地制约了前掠翼气动布局设计在航空领域的广泛运用。

参考文献 (References)

- [1] PUTNAM T W. The X-29 flight-research program; AIAA-1983-2687 [R]. Reston, VA: AIAA, 1983.
- [2] KRONE N J. Divergence elimination with advanced composites; AIAA-1975-1009 [R]. Reston, VA: AIAA, 1975.
- [3] KRONE N. J. Forward swept wing demonstrator; AIAA-1980-1882 [R]. Reston, VA: AIAA, 1980.
- [4] 方宝瑞. 飞机气动布局设计 [M]. 北京: 航空工业出版社, 1997: 779-781.
FANG B R. Aircraft aerodynamic configuration design [M]. Beijing: Aviation Industry Press, 1997: 779-781 (in Chinese).
- [5] 张锡金. 飞机设计手册 [M]. 北京: 航空工业出版社, 2002: 145-165.
ZHANG X J. Aerodynamic design manual [M]. Beijing: Aviation Industry Press, 2002: 145-165 (in Chinese).
- [6] GLOSS B. Effect of canard location and size on canard-wing interference and aerodynamic center shift related to maneuvering aircraft at transonic speeds; NASA TN-D-7505 [R]. Washington, D. C.: NASA, 1974.
- [7] HUMMEL D, OELKER H C. Low-speed characteristics for the wing-canard configuration of the international vertex flow experiment [J]. Journal of Aircraft, 1994, 31 (4): 868-878.

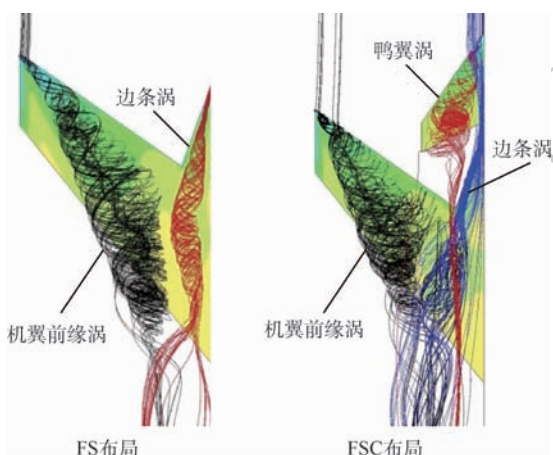


图 18 FS 和 FSC 布局空间流线图 ($\alpha = 23^\circ$)

Fig. 18 Streamlines in flow field of FS and FSC configurations ($\alpha = 23^\circ$)

- [8] OELKER H C, HUMMEL D. Investigation on the vorticity sheets of a close-coupled delta-canard configuration[J]. Journal of Aircraft, 1989, 26(7): 657-666.
- [9] MA B F, LIU P Q, YUAN W. Effects of wing and canard sweep on lift-enhancement of canard-configuration[J]. Journal of Aircraft, 2004, 41(6): 1521-1523.
- [10] 刘沛清, 魏园. 在近距离耦合鸭式布局中的涡系结构[J]. 实验流体力学, 2005, 19(3): 85-89.
- LIU P Q, WEI Y. Vortex structures on close-coupled canard configurations[J]. Journal of Experiments in Fluid Mechanics, 2005, 19(3): 85-89 (in Chinese).
- [11] LIU P Q, WEN R Y, ZHANG G W, et al. Experimental study of canard-spanwise pulsed blowing on a canard configuration[J]. Journal of Aircraft, 2008, 45(5): 1816-1820.
- [12] 马宝峰, 刘沛清, 邓学莹. 大迎角下鸭翼涡与边条涡的干扰特性[J]. 航空学报, 2002, 23(6): 560-563.
- MA B F, LIU P Q, DENG X Y. Characteristics of canard vortex interaction with strake vortex at high incidence[J]. Acta Aeronautica et Astronautica Sinica, 2002, 23(6): 560-563 (in Chinese).
- [13] SPACHT G. The forward swept wing: A unique design challenge: AIAA-1980-1885[R]. Reston, VA: AIAA, 1980.
- [14] MANN M J. A forward-swept wing fighter configuration designed by a transonic computational method[J]. Journal of Aircraft, 1986, 23(6): 506-512.
- [15] MOORE M, FREI D. X-29 forward swept wing aerodynamic overview: AIAA-1983-1834[R]. Reston, VA: AIAA, 1983.
- [16] KEHOE M W, LAURIE E J, BIARKE L J. An in-flight interaction of the X-29A canard and flight control system: NASA TM-101718[R]. Washington, D. C. : NASA, 1990.
- [17] SALTZMAN E J, HICKS J W. In-flight lift-drag characteristics for a forward-swept wing aircraft (and comparisons with contemporary aircraft): NASA T R 3414[R]. Washington, D. C. : NASA, 1994.
- [18] 张彬乾, LASCHKA B. 前掠翼根部流动分离的控制[J]. 航空学报, 1992, 13(5): 241-246.
- ZHANG B Q, LASCHKA B. Flow separation control at the root part offorward-swept wing[J]. Acta Aeronautica et Astronautica Sinica, 1992, 13(5): 241-246 (in Chinese).
- [19] ZHANG B Q, LASCHKA B. Several ways to control the flow separation at the root part of forward-swept wing(FSW) [M]. Munich: Munich Technische University, 2002: 3-51.
- [20] 王晋军, 赵霞, 王双峰, 等. 鸭翼-前掠翼气动布局纵向气动特性实验研究[J]. 空气动力学学报, 2004, 22(2): 237-244.
- WANG J J, ZHAO X, WANG S F, et al. Experimental investigation on longitudinal aerodynamic characteristics of canard-forward-swept wing configuration[J]. Acta Aerodynamica Sinica, 2004, 22(2): 237-244 (in Chinese).
- [21] 展京霞, 王晋军, 赵霞, 等. 近距离鸭翼高度对鸭翼-前掠翼布局纵向气动特性影响的实验研究[J]. 实验流体力学, 2006, 20(2): 50-54.
- ZHAN J X, WANG J J, ZHAO X, et al. Investigation on the effects of the height of close canard wing on the aerodynamic characteristic of a forward-swept configuration[J]. Journal of Experiments in Fluid Mechanics, 2006, 20(2): 50-54 (in Chinese).
- [22] 李岸一, 王旭, 刘文法, 等. 鸭翼涡与边条涡对前掠翼布局的增升研究[J]. 空军工程大学学报(自然科学版), 2010, 11(1): 19-23.
- LI A Y, WANG X, LIU W F, et al. Study on lift-enhancement of canard vortex and strake vortex to configuration with forward-swept wing[J]. Journal of Air Force Engineering University (Natural Science Edition), 2010, 11(1): 19-23 (in Chinese).
- [23] EUGENE L T. Navier-Stokes simulation of a close-coupled canard-wing-body configuration[J]. Journal of Aircraft, 1992, 29(5): 830-838.
- [24] ISMAIL H T, MAX F. Computational study of subsonic flow over a delta canard-wing-body configuration[J]. Journal of Aircraft, 1998, 35(4): 554-560.
- [25] 于冲, 王旭, 董福安, 等. $y +$ 值对翼型气动参数计算精度的影响研究[J]. 空军工程大学学报(自然科学版), 2012, 13(3): 25-29.
- YU C, WANG X, DONG F A, et al. The study of effect on $y +$ for precision of pneumatic parameters about foil[J]. Journal of Air Force Engineering University (Natural Science Edition), 2012, 13(3): 25-29 (in Chinese).
- [26] 王福军. 计算流体动力学分析[M]. 北京: 清华大学出版社, 2004.
- WANG F J. The analysis of computational fluid dynamics[M]. Beijing: Tsinghua University Press, 2004 (in Chinese).
- [27] MENTER F R. Two-equation eddy-viscosity turbulence models for engineering applications[J]. AIAA Journal, 1994, 32(8): 269-289.
- [28] 李人宪. 有限体积法基础[M]. 北京: 国防工业出版社, 2008.
- LI R X. Fundamentals of finite volume method[M]. Beijing: National Defense Industry Press, 2008 (in Chinese).
- [29] BREITSAMTER C, LASCHKA B. Vortical flowfield structure at forward swept wing configurations[J]. Journal of Aircraft, 2001, 38(2): 193-207.
- [30] 安世亚太. ANSYS CFX 对某前掠翼飞机的气动模拟[EB/OL]. (2015-10-29) [2019-01-15]. <http://www.peraglobal.com/>.
- ANSYS. Aerodynamic simulation of a forward-swept wing aircraft by ANSYS CFX[EB/OL]. (2015-10-29) [2019-01-15]. [http://www.peraglobal.com/\(in Chinese\)](http://www.peraglobal.com/(in Chinese)).

作者简介:

张冬 男, 博士研究生。主要研究方向: 空气动力学、飞行力学。

陈勇 男, 博士, 讲师。主要研究方向: 飞行力学、飞行控制。

胡孟权 男, 博士, 副教授。主要研究方向: 飞行力学、飞行控制。

付向恒 男, 博士研究生。主要研究方向: 自动控制、飞行控制。

Effect of strake and canard on aerodynamic characteristics of forward-swept wing and back-swept wing

ZHANG Dong¹, CHEN Yong^{2,*}, HU Mengquan², FU Xiangheng³

(1. Graduate School, Air Force Engineering University, Xi'an 710051, China;

2. Aviation Engineering Institute, Air Force Engineering University, Xi'an 710038, China;

3. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China)

Abstract: In order to analyze the reasons why the design of forward-swept wing aerodynamic configuration cannot be popularized and applied in aviation industry, simplified strake-wing, canard-wing and strake/canard-wing configurations were constituted by fixing strake and canard on forward-swept wing and back-swept wing, so as to deeply understand the flow characteristics and the mechanism of vortices interference between the two different configurations of forward-swept wing and back-swept wing. and First, the reliability and accuracy of numerical computation method were validated by comparing the computing results with experimental data of a standard model. Then, the lift coefficient curves of different configurations were obtained through numerical computation. Finally, the complex vortex interaction mechanism of different configurations were analyzed by pressure contours and streamlines. The results indicate that induction and convolution between vortexes of configurations based on back-swept wing enhance the lift coefficient and increase stalling angle of attack, and the effect was more apparent on the configuration fixed with strake and canard. There is no convolution effect between vortices of configurations based on forward-swept wing, and vortexes of configurations based on forward-swept wing perform an adverse interaction by bumping and squeezing, which makes the lift coefficient of the forward-swept wing much lower than that of the back-swept wing at high angles of attack. The leading-edge vortices of the forward-swept wing cannot be coupled with the canard wing vortices and strake vortices at high angles of attack, and cannot make full use of the non-linear lift force, which is the shortcoming in the aerodynamic layout design of the forward-swept wing.

Keywords: strake; canard; forward-swept wing; back-swept wing; interaction mechanism

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0032

超低信噪比调频连续波引信信号小周期态 Duffing 振子检测



朱志强¹, 侯健², 闫晓鹏^{1,*}, 栗苹¹, 郝新红¹

(1. 北京理工大学 机电动态控制重点实验室, 北京 100081; 2. 北京航计算通讯研究所, 北京 100074)

摘 要: 针对超低信噪比(SNR)下调频连续波(FWCW)引信信号难以检测的问题, 结合 Duffing 振子特性和可停振动系统理论, 建立了小周期态 Duffing 振子检测系统。该系统能够克服传统 Duffing 振子强参考信号检测的固有缺陷, 扩展单个 Duffing 振子的信号频率检测范围, 并降低算法复杂度。在此基础上, 分析 Duffing 振子相轨迹特性, 提出了基于小周期态 Duffing 振子的调频连续波引信信号检测方法。实验结果表明, 小周期态 Duffing 振子检测方法在 -30 dB 的超低信噪比下对真实调频引信辐射信号的平均检测误差小于 1%, 验证了本文方法的有效性。

关 键 词: 调频连续波(FMCW)引信; Duffing 振子; 小周期态; 可停振动系统; 信噪比(SNR)

中图分类号: TJ43+4.1

文献标识码: A

文章编号: 1001-5965(2019)10-2069-10

在对调频连续波(FWCW)引信进行欺骗式干扰时, 检测并获取其参数对于干扰波形的设计具有重要的参考意义^[1]。然而, 由于电子对抗的非合作性, 到达干扰机的调频引信信号往往十分微弱, 且淹没在真实战场环境下的强噪声中, 难以被信号侦察系统检测。因此, 研究低信噪比(SNR)下的调频连续波信号检测具有重要的现实意义。

有关调频连续波引信信号检测问题, 传统的检测方法是基于线性、确定性系统的。文献[2]利用回波信号幅值的变化规律和范围特征进行信号检测, 该时域方法虽然简单, 但提取的信号特征稳定性不高。经典频域方法, 如快速傅里叶变换(FFT)等, 由于无法表述信号的时频局域性质, 其抗噪性能较差且对调频信号的特征检测能力不足^[3-4]。从线性时频域的角度, 文献[1, 5-6]分别

基于改进后的希尔伯特-黄变换(HHT)、Cohen 类模糊函数和 Wigner-Ville 结合 Hough 变换, 利用时频分析方法实现了对单分量和多分量线性调频信号的特征参数识别。文献[7]提出了基于极大 chirplet 变换的调频连续波信号检测方法, 可以用于处理对称三角线性调频连续波信号; 但其处理过程较为繁琐、限制条件较多, 难以实现工程应用。此外, 采用上述检测方法, 可达到的最低检测信噪比门限均不足 -10 dB^[1-8]。可见, 在低信噪比下检测调频连续波信号时, 传统的线性检测方法在信噪比门限或者实时性方面无法满足要求。

随着非线性科学的发展, 混沌理论, 尤其是 Duffing 振子系统, 在微弱信号处理方面表现出独特的优越性^[9-10]。现有的 Duffing 振子检测算法的主要思路为设置强参考信号, 利用混沌系统处于临界状态时对微弱周期信号的极端敏感性及对

收稿日期: 2019-01-22; 录用日期: 2019-04-26; 网络出版时间: 2019-06-11 18:10

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190611.1032.001.html

基金项目: 国家自然科学基金(61673066)

* 通信作者: E-mail: yanxiaopeng@bit.edu.cn

引用格式: 朱志强, 侯健, 闫晓鹏, 等. 超低信噪比调频连续波引信信号小周期态 Duffing 振子检测[J]. 北京航空航天大学学报, 2019, 45(10): 2069-2078. ZHU Z Q, HOU J, YAN X P, et al. Small-scale periodic state Duffing oscillator FMCW fuze signal detection at ultra-low SNR[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2069-2078 (in Chinese).

噪声的免疫特性,实现对微弱周期信号在超低信噪比下的实时检测^[11-12]。由于系统临界阈值受参考信号频率以及噪声强度的影响较大,且单个 Duffing 振子的测频范围有限,因此传统 Duffing 振子对大带宽且频率快速变化的调频信号检测能力严重不足。

针对上述问题,本文构建了基于小周期态 Duffing 振子的微弱信号检测系统模型,并将其应用于低信噪比下的调频连续波引信信号检测方法中。该方法克服了 Duffing 振子强参考系统的固有缺陷,能够实现低信噪比下大带宽和频率快速变化的连续波调频信号的检测,同时降低了计算成本和时间消耗,为调频连续波引信信号的侦察及参数估计提供了技术支撑。

1 调频连续波引信信号基本特征

以三角波调频引信为例,进行调频连续波引信信号特性分析。其发射信号为

$$S_i(t) = \begin{cases} A \cos \{ 2\pi [f_c + (4n+1)\Delta F]t - \pi\beta t^2 \} & nT < t \leq nT + \frac{T}{2} \\ A \cos \{ 2\pi [f_c + (4n-3)\Delta F]t + \pi\beta t^2 \} & nT + \frac{T}{2} < t \leq (n+1)T \end{cases} \quad (1)$$

式中: A 为发射信号幅度; f_c 为信号载频; ΔF 为信号单边调制带宽; T 为调制信号周期,其倒数表示调制信号频率 f_m ; $\beta = 4\Delta F/T = 4\Delta Ff_m$ 为调制斜率; n 为调制周期数。

通常,调频引信发射信号载频 f_c 在 GHz 量级,难以被直接采样和处理。因此侦察设备会对接收到的调频引信信号进行下变频,将其频谱搬移至几十兆赫兹。下变频后信号与式(1)形式基本一致,区别在于幅值 A 、载波 f_c 数值的变化。因此,仍然用式(1)描述侦察设备截获的调频引信信号。

设经下变频三角波调频信号参数为:载频 $f_c = 50$ MHz,单边调制带宽 $\Delta F = 10$ MHz,调制信号频率 $f_m = 100$ kHz。其时频特性如图 1 所示。

可见,三角波调频信号具有频带较宽、信号频率变化较快且频率变化有正负两个不同斜率的特性。这些特性使得非合作条件下的调频引信信号检测方法不仅要具有在噪声下对准周期信号的高

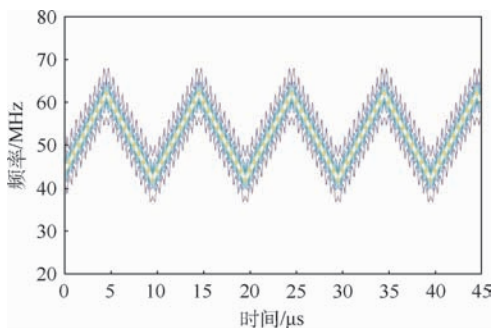


图 1 三角波调频信号时频特性

Fig.1 Time-frequency characteristics of triangular wave frequency-modulated signal

度敏感性,同时还要能保证在大带宽和快频率变化率下的检测精度。

2 小周期态 Duffing 振子检测系统

2.1 可停振动系统

设随机微分方程为

$$\begin{cases} D^q X = F(X, t) + e\xi(t) \\ X(t_0) = x_0 \end{cases} \quad (2)$$

式中: X, F 为二维矢量随机过程; $q \in \mathbf{R}$ 为微分方程的阶数; $e\xi(t)$ 为系统输入的随机扰动; e 为随机扰动幅度; η_0 为微分方程平凡解(即 $F(\eta_0, t) = 0$)。

如果满足以下 2 个条件:

$$1) \lim_{\|x_0 - \eta_0\| \rightarrow 0} P \left\{ \lim_{\substack{t \rightarrow \infty \\ \epsilon \rightarrow \infty}} \|X(t; t_0, x_0) - \eta_0\| \right\} = 1;$$

2) 当式(2)描述的系统输入为周期及准周期信号时,系统有周期或者是准周期解。

则称式(2)所对应的系统为可停振动系统^[13]。其中条件 1) 表示系统在受到微小随机扰动时,以概率 1 渐进稳定,其对应的状态称为可停振动状态。由定义可知,可停振动系统的可停振动状态变化对噪声不敏感,对周期及准周期信号敏感,这为检测未知频率的微弱调频信号提供了一条新的思路。

2.2 Duffing 振子基本特性分析

经典的 Holmes 型 Duffing 振子方程为

$$\ddot{x}(t) + k\dot{x}(t) - x(t) + x^3(t) = \gamma \cos(\omega t) \quad (3)$$

式中: k 为阻尼比; $-x(t) + x^3(t)$ 为非线性恢复力; $\gamma \cos(\omega t)$ 为幅值为 γ ,角频率为 ω 的强参考信号。随着 γ 的增加,Duffing 振子系统依次经历小周期运动、倍周期运动、混沌运动、大尺度周期运动。

设 $k = 0.5$,当 γ 的值大于临界阈值 γ_c (约为 0.826) 时,Duffing 振子系统将从混沌临界运动状态进入到大尺度周期运动状态。这一转变对参数

γ 极其敏感,微小的周期扰动就可能改变系统的动力学行为。因此,设置强参考信号使系统处于临界状态时,当系统中存在微小的周期扰动时,相轨迹会呈现出周期性的在混沌态和大尺度周期态间的转换,如图 2 所示。利用这种转换的周期性实现微弱周期信号检测和频率信息估计的方法称为 Duffing 振子强参考信号检测法。

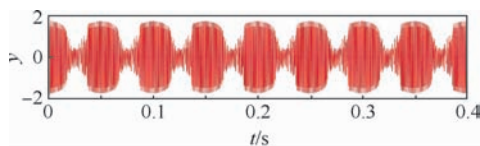


图 2 Duffing 振子系统相轨迹周期性的状态转换

Fig. 2 Periodic state transitions of phase trajectory of Duffing oscillator system

Duffing 振子强参考信号检测系统可以实现低信噪比下的周期信号检测。然而,将其应用于大带宽和频率变化较快的调频引信信号检测时,存在以下固有缺陷:

1) 当强参考信号与待测信号频率非常接近时,系统存在检测盲区,在该区域内系统始终处于大尺度周期态或混沌状态,这意味着无法通过相轨迹的转换来进行信号检测^[14]。

2) 系统的临界阈值对强参考信号频率以及环境噪声强度敏感,因此需要根据它们的变化来设置合适的临界阈值,这会使检测系统复杂化^[15]。

3) 单个 Duffing 振子的频率检测范围较窄,若要实现大带宽频率检测,需设置 Duffing 振子阵列,导致较高的计算成本和时间消耗。

4) 系统相轨迹由混沌态向大尺度周期态转换过程中存在过渡段,对于阵列中的每个振子,都需要较长的时间来确定 Duffing 振子系统当前的状态转换情况,否则可能导致误判^[16]。

2.3 小周期态 Duffing 振子检测系统模型

为避免上述缺陷,本文提出一种新的混沌检测方法来实现超低信噪比下的调频引信信号检测。该方法在 Duffing 振子强参考信号检测系统的基础上,把强参考信号置零,将待测微弱信号直接送入 Duffing 振子系统作为策动力,利用 Duffing 振子处于小周期态时的相轨迹特征实现微弱信号检测。此时,为了使系统能对待测信号足够敏感,仍在系统内部保留参考角频率 ω ,其状态方程为

$$\begin{cases} \dot{x} = \omega y \\ \dot{y} = \omega[-ky + (x+1) - (x+1)^3 + I(t)] \end{cases} \quad (4)$$

式中; $I(t)$ 为待测系统输入信号,在进行调频引信信号检测时,其形式通常为 $A\cos(\omega_0 t + \dot{\omega}t^2) + e\xi(t)$ (ω_0 为初始角频率, $\dot{\omega}$ 为角频率变化率, $e\xi(t)$ 为噪声); $(x+1)$ 是为了保证相轨迹以 $(0, 0)$ 为焦点,尽可能均匀地分布在 4 个象限,以便后续计算。另外,为确保系统工作在小尺度周期态,待测信号幅值 A 不能超过 0.36 ^[17-18]。

基于 Hamilton 系统理论,当 $k > 0$ 且 $I(t)$ 为高斯噪声时, Duffing 振子系统(4)满足式(2)的条件 1); 当 $k > 0$ 且系统中输入微弱的周期信号时, Duffing 振子系统(4)是耗散系统^[19]。所以,当 $k > 0$ 时, Duffing 振子系统(4)是可停振动系统。

根据可停振动系统的定义,可以使用式(4)中 Duffing 系统作为检测模型。因为输入信号微弱且不存在强参考信号,系统处于小周期态,所以将此系统称为小周期态 Duffing 振子检测系统。其相比 Duffing 振子强参考信号系统的主要优势在于:

1) 取消了强参考信号,无需进行临界阈值的复杂设置,也不存在检测盲区。

2) 单个 Duffing 振子的频率检测范围大幅扩展,同时能适应频率的快速变化。

3) 系统结构及算法简单,既无需判别系统的状态转换,也无需设置振子阵列,大幅降低计算成本和时间消耗。

因此,小周期态 Duffing 振子很好地克服了传统 Duffing 振子检测系统的固有缺陷,且在测频范围和算法复杂度等性能上具有明显优势。

小周期态 Duffing 振子系统在不同策动信号下的相轨迹图如图 3 所示。图 3(a)表示在只由高斯白噪声策动时,系统处于可停振动状态,其相轨迹无规律;图 3(b)表示当由微弱调频信号驱动时,系统处于小周期状态;图 3(c)表示在信噪比降低到 -35 dB 时,系统相轨迹虽然变得粗糙,但仍处于小周期态的理想轨迹附近。以上现象验证了本文提出的小周期态 Duffing 振子系统对微弱调频信号敏感,对随机扰动不敏感的特性。因此,利用图 3 所示的相轨迹特征,可以在没有强参考信号和不设置临界阈值的情况下,实现微弱调频信号的检测。

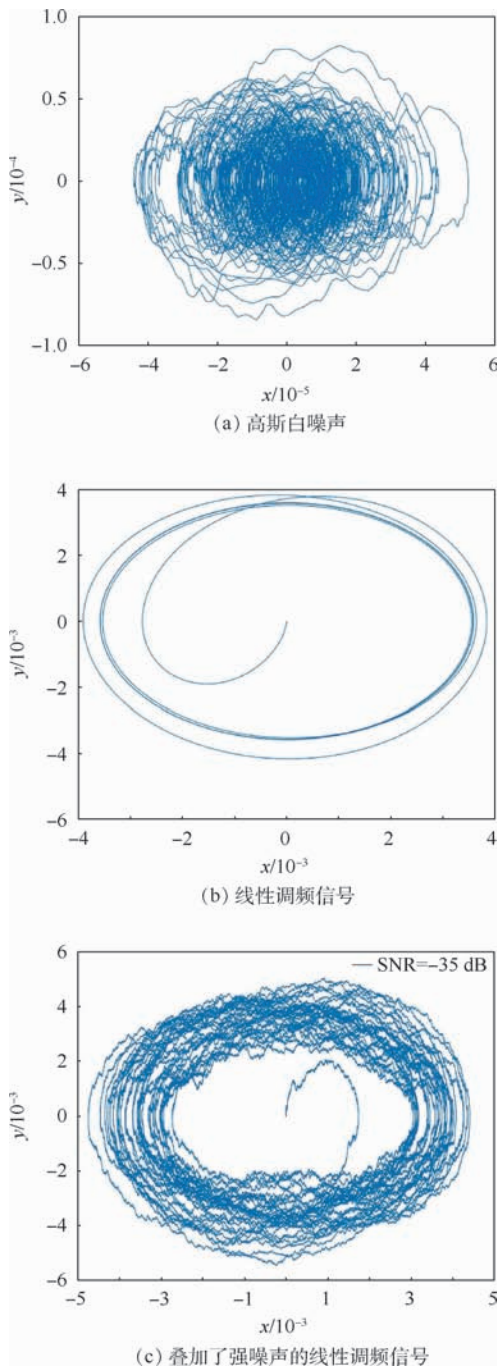


图3 不同输入信号策动下小周期态 Duffing 振荡子系统的相轨迹图

Fig.3 Phase trajectories of small-scale periodic state Duffing oscillator system driven by different input signals

3 小周期态 Duffing 振子微弱信号检测方法

3.1 微弱信号检测方法

从图3可以明显观察到,当系统处于可停振动状态时,相轨迹无规律;当系统处于小周期态时,相轨迹的规律性明显。而且,从统计上来看,系统处于小周期态时相轨迹在分布域内的运动周

期与待测信号周期相同。受此特点的启发,本文提出一种新的微弱调频信号混沌检测方法。

设 φ 为 Duffing 振子系统输出的相点 (x, y) 对应的相空间角,其正弦和余弦函数为

$$\begin{cases} \sin \varphi = \frac{y}{\sqrt{x^2 + y^2}} \\ \cos \varphi = \frac{x}{\sqrt{x^2 + y^2}} \end{cases} \quad (5)$$

式中: x 和 y 分别表示 Duffing 振子的位移和速度。相空间角的正、余弦曲线可以用来描述相轨迹的周期特性。

设置系统初值 $(x, y) = (0, 0)$, 将不同信噪比的线性调频信号 $0.04\cos 2\pi(5 \times 10^7 t + 4 \times 10^{12} t^2)$ 输入小周期态 Duffing 振子检测系统。图4分别给出了不含噪声和信噪比为 -35 dB 时的被测信号时域波形与相空间角正、余弦曲线的对应关系。在没有噪声时,信号为准周期信号,同时相空间角的正余弦波形与被测信号的瞬时频率相同(第1个周期除外),如图4(a)所示;在信噪比达到 -35 dB 时,如图4(b)所示,由于强噪声的影响,已经无法从被测信号时域波形中分辨出信号的特征信息,但正、余弦波形仍然比较清晰并且与图4(a)中周期一致。因此,可以从 Duffing 振子输出信号的相空间正、余弦曲线中得到其时间-频率特性,进而得到待测线性调频信号的时间-频率特性。

在调频连续波引信常用的调制方式中,三角波调制与锯齿波调制本身就是分段的线性调频信号;正弦波调制虽然并非线性调频信号,但由于使用小周期态 Duffing 振子进行频率检测时,只需要

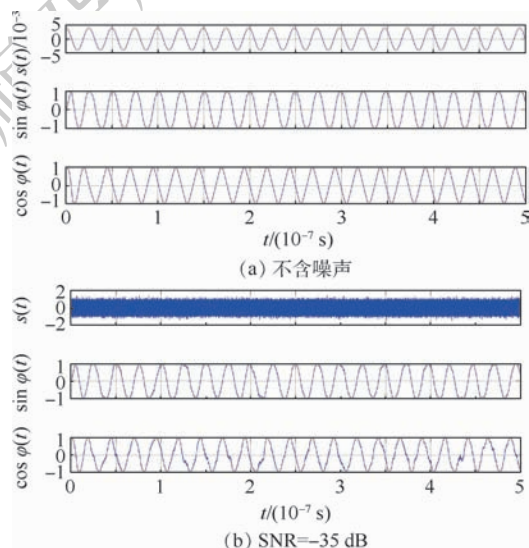


图4 待测信号时域波形和相空间角正弦余弦波形
Fig.4 Time domain waveform of to-be-detected signal and sine cosine waveform in phase space

对很短时间内的几个峰值进行检测,在此时间内正弦波调频信号可视为线性调频信号甚至标准正弦波信号。因而,可以使用小周期态 Duffing 振子实现对连续波调频引信信号在超低信噪比下的检测。

当待测信号频率与参考频率之间的频率差较大时,如图 5 所示,位移 x 和速度 y 均偏离原点 $(0,0)$,且 x 的偏离程度更大,即噪声对位移的影响大于对速度的影响。这是因为较大的频率差使得 Duffing 系统的输出 (x,y) 无法很好地响应待测信号的幅值变化,进而导致小周期态的周期特性被破坏。因此,一方面,此检测方法对于与参考信号频率相差较大的待测信号的测量也会出现偏离,存在一个频率检测范围 $[f_L,f_H]$;另一方面,由于 y 的偏离程度更小,为了更好地克服噪声的影响,文中采用相空间角的余弦函数,即 $\cos \varphi = y/\sqrt{(x^2+y^2)}$,对待测信号的频率参数进行检测。

基于上述分析,基于小周期态 Duffing 振子的微弱调频引信信号检测方法的具体步骤为

- 1) 设 Duffing 振子系统初值 $(0,0)$,阻尼比 $k=0.5$ 。
- 2) 在预估的待测信号频率范围内选择一个适中值作为参考信号频率 f ,以保证被测信号与参考信号频率差不会过大。
- 3) 将含噪待测信号输入小周期态 Duffing 振子检测系统,求解式(4),并计算 Duffing 振子输出相轨迹的余弦函数 $\cos \varphi$ 。

- 4) 利用峰值法找到 $\cos \varphi$ 所有峰值,测量相邻峰值的时间差,将其倒数视为 $\cos \varphi$ 的瞬时频率。
- 5) 根据相轨迹余弦 $\cos \varphi$ 和调频信号同频的原理得到调频信号的瞬时频率。

3.2 性能分析

首先以线性调频引信信号为例,分析此方法的频率检测范围 $[f_L,f_H]$ (将频率检测范围 $[f_L,f_H]$ 定义为一个频率区间,微弱信号检测系统对初始频率在此区间内的线性调频信号的相对检测误差不大于 1%)。设置参考频率 $f=50\text{ MHz}$,固定信噪比和频率变化率,设输入的调频信号为

$$S_i(t)=A\cos(2\pi f_0t+\pi bt^2)$$

(6)

式中: f_0 为待测信号的初始频率; b 为频率变化率。

设置信噪比为 -20 dB ,将频率变化率 $b=0\text{ Hz/s}$ (标准正弦波信号)和 $b=10^{12}\text{ Hz/s}$ 的微弱线性调频信号输入到小周期态检测模型和强参考信号检测模型中,得到单个 Duffing 振子的频率检测范围 $[f_L,f_H]$ 如表 1 所示。

对于强参考信号的 Duffing 振子系统检测方法来说,只有当被测信号与参考信号的相对频率差小于 0.04 时,图 2 中的周期性状态转换才能稳定出现,因此单振子的频率检测范围较小。而小周期态 Duffing 振子检测方法则大大扩展了 Duffing 振子的频率检测区间,仅用单个 Duffing 振子就能有效地检测出淹没在强噪声下的未知周期信号或宽带调频信号。

随后固定 $f_0=50\text{ MHz}$,与参考频率相同,在不同信噪比下对不同频率变化率的线性调频信号分别进行 20 次频率检测仿真,得到的平均检测精度如图 6 所示。从图 6 可以看出,虽然在信噪比过低和频率变化速度超出系统检测能力范围时存在个别异常点,总体来说频率检测精度 η 随频率变化率的增大和信噪比的降低而减小;当频率变化率 b 在 $0\sim5\times10^{12}\text{ Hz/s}$ 的范围内时,信噪比不低于 -25 dB 时,初始频率的检测精度可以达到 99% 以上;当信噪比降低至 $-30\sim-45\text{ dB}$ 时,只

表 1 单个 Duffing 振子的频率检测范围

Table 1 Frequency detection range with single Duffing oscillator

		$[f_L,f_H]/\text{MHz}$		
$b/(\text{Hz}\cdot\text{s}^{-1})$	f/MHz	SNR/dB	小周期态检测模型	强参考信号检测模型
0	50	-20	[0.17,254]	[47.7,52.1]
10^{12}	50	-20	[0.21,212]	[48.2,51.4]

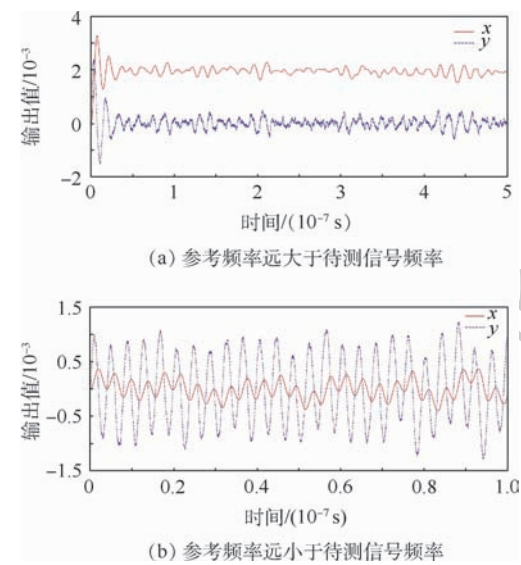


图 5 频率差较大时小周期态 Duffing 振子系统输出的时域波形

Fig.5 Time domain waveform of outputs of small-scale periodic state Duffing oscillator system when frequency difference is large

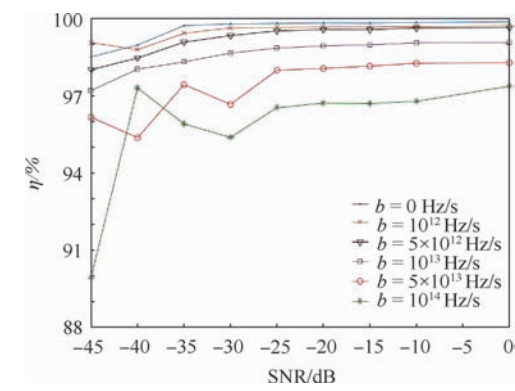


图 6 线性调频信号的频率平均检测精度与信噪比和频率变化率的关系

Fig. 6 Relationship among average frequency detection accuracy, SNR and frequency slope of chirp signals

要 b 的取值仍在 5×10^{12} Hz/s 以下, 依然可以保证较高的频率检测精度。

4 实验验证

为了验证基于小周期态 Duffing 振子的微弱信号检测方法在调频引信信号检测上的效果, 本文使用 3 种不同调制方式的引信仿真模型及某型号调频引信的辐射信号对此方法进行验证。

首先以三角波调频引信模型为例, 设置仿真工作参数如下: 三角波调制频率 $f_m = 100$ kHz, 单边调制带宽 $\Delta F = 10$ MHz, 载波频率 $f_c = 50$ MHz, 信号幅值 $A = 0.004$, 噪声强度为 0.0008 (-20 dB)。

图 7 给出了 -20 dB 信噪比条件下此信号时域波形及采样率为 1 GHz、点数为 $20\,000$ 点时的频谱图。由于噪声的影响, 已经无法获取三角波调频信号的时域波形和频率特性。如果进一步增大采样率到 10 GHz 时 ($200\,000$ 点), 虽然能够获得较高的处理增益, 可以检测到信号的存在, 但是大点数 FFT 存在运算量大、频谱资源浪费的问题, 无法满足实时性检测的需求。

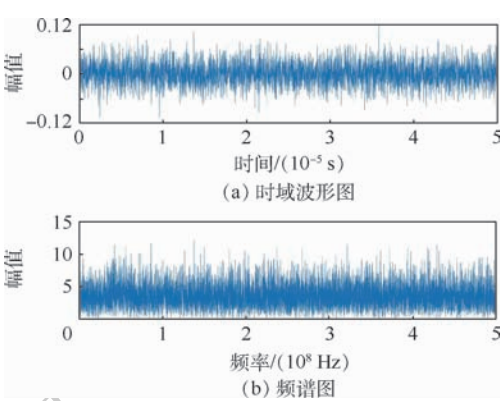


图 7 -20 dB 信噪比下三角波调频信号时域波形和频谱图
Fig. 7 Time domain waveform and spectrogram of triangular wave frequency-modulated signal while SNR is -20 dB

图 8 分别给出了 -20 dB 的信噪比条件下, 采用 Duffing 振子强参考信号阵列检测方法^[20]及本文方法检测得到的三角波调频引信信号的时频特性曲线。根据图 8(a), 在 -20 dB 信噪比下, Duffing 振子强参考信号阵列检测法得到的时频特性曲线中存在多处频率真实值与测量值相差较大的异常点; 根据图 8(b), 本文方法可以很好地对三角波调频信号进行时频分析, 得到的时频特性曲线与真实时频特性曲线几乎重合。因此, 本文方法对于三角波调频引信信号的检测效果优于 Duffing 振子强参考信号阵列检测方法。同时, 根据时频特性曲线对调频信号特征的表征能力, 基于图 8(b) 获得的时频特性曲线, 可以直接实现调频引信信号的特征参数 (如载频、调制频率、最大频偏以及调制类型等) 估计。

表 2 为采用 Duffing 振子强参考信号阵列检测方法和本文方法时, 不同信噪比条件下对三角波调频引信信号的频率检测误差。随着信噪比的

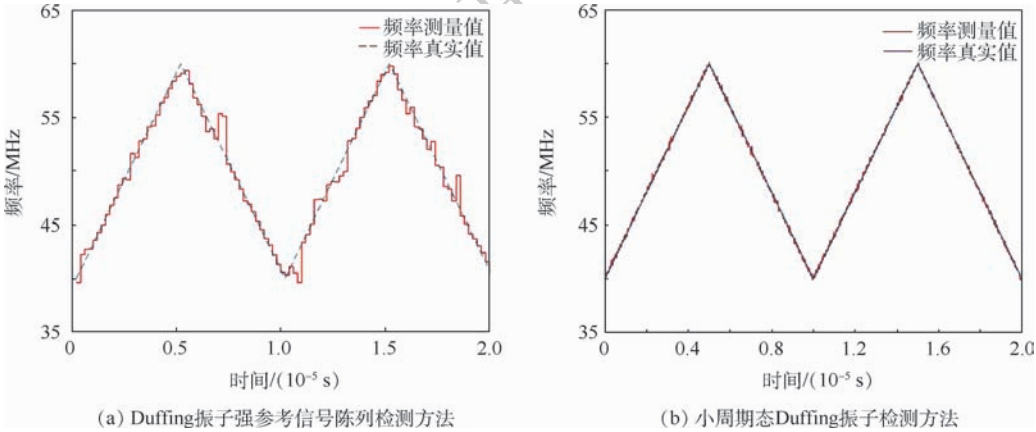


图 8 -20 dB 信噪比下三角波调频信号时频特性

Fig. 8 Time-frequency characteristics of triangular wave frequency-modulated signal while SNR is -20 dB

表 2 不同信噪比条件下三角波调频信号频率测量误差
Table 2 Frequency measurement errors of triangular wave frequency-modulated signal at different SNRs

SNR/dB	强参考信号阵列检测方法测量误差/%	小周期态检测方法测量误差/%
-10	0.2865	0.2801
-20	1.2430	0.2880
-30	3.1455	0.3074
-35	9.0104	0.4644

降低,两种方法的频率检测误差都不断增大,但是当信噪比不小于 -35 dB 时,本文方法均可以得到高于 99% 的检测精度,远高于 Duffing 振子强参考信号阵列推测方法。

为了进一步验证本文方法对不同调制类型的调频连续波信号检测的有效性,在相同的载波频率和调频带宽的基础上,分别对三角波、锯齿波和正弦波调制 3 种不同类型的调频连续波信号进行检测。

图 9 为 -35 dB 信噪比条件下 3 种调频连续

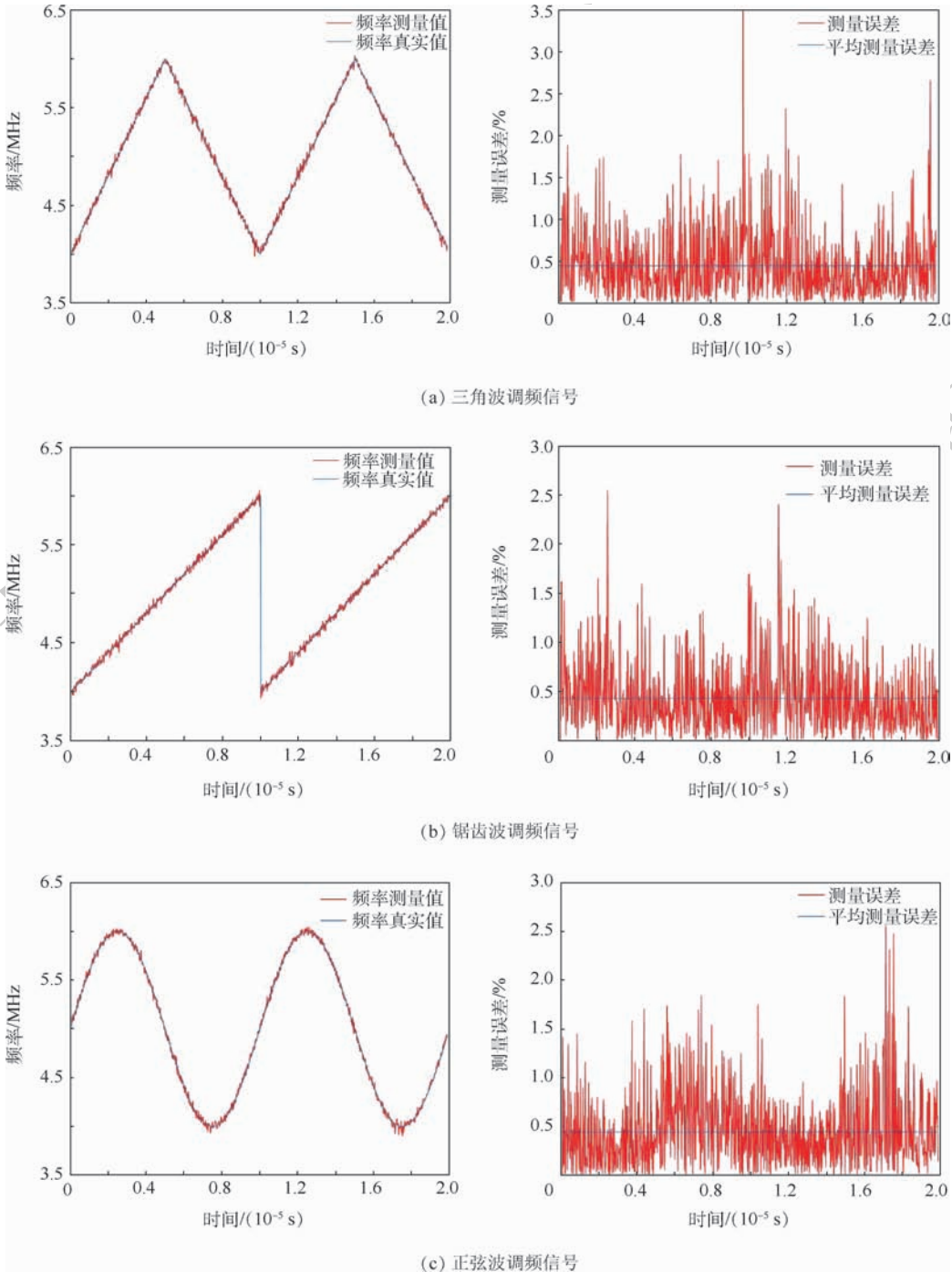


图 9 -35 dB 信噪比下 3 种调频引信信号的时频特性和测量误差

Fig. 9 Time-frequency characteristics and measurement error of three types of frequency-modulated fuze signals while SNR is -35 dB

波引信号的时频特性图,可以看出尽管图像中由于强噪声的影响出现了许多瑕疵点和测量不准确的点,但是频率的平均测量误差均小于1%,证明该方法能实现超低信噪比下对三种调频引信信号的高精度检测。

采集某调频引信真实辐射信号,并加入噪声后,对本文方法进行进一步的半实物仿真验证。此引信的载波频率约为915 MHz,调制频率和最大调制频偏分别约为960 kHz和8 MHz。如图10可知,在-20 dB的信噪比下,利用小周期态Duffing振子可以得到较好的引信检测效果。

为了进行检测能力的量化评估,将该引信信号下变频至50 MHz,以未人为加入噪声时引信信号的频率检测结果作为真实引信频率,计算不同信噪比下的相对检测误差如表3所示。

由于直接对引信辐射信号进行射频采样,采集到的信号存在失真,同时在采集、下变频和滤波过程中也会引入噪声,因而真实引信信号的信噪比检测阈值高于仿真结果。即使如此,本文方法在-30 dB的超低信噪比条件下,对真实调频引信信号依然具有较高的检测精度。至此,小周期态Duffing振子微弱调频引信信号检测方法的超低信噪比和高精度检测性能得到了充分验证。

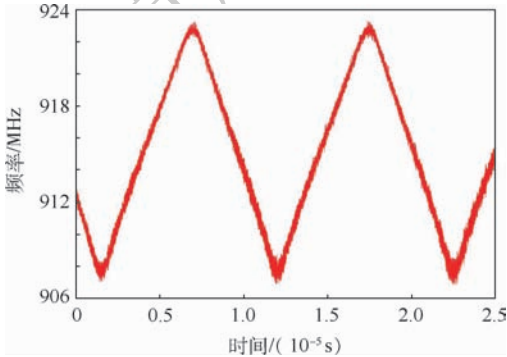


图10 -20 dB信噪比下某调频引信信号时频特性
Fig. 10 Time-frequency characteristics of signal of a FMCW fuze while SNR is -20 dB

表3 不同信噪比条件下某调频引信真实信号
频率测量误差

Table 3 Frequency measurement errors of signal of a FMCW fuze at different SNRs

SNR/dB	测量误差/%
-10	0.2824
-20	0.4945
-30	0.6528
-35	1.1136

5 结 论

本文在分析调频连续波引信辐射信号特征的

基础上,基于可停振动系统理论,提出了一种小周期态Duffing振子混沌检测方法,实现了对微弱调频引信信号的检测。理论分析与实验结果表明,该方法不仅能够克服Duffing振子强参考信号系统的固有缺陷,而且提高了实时性,并扩展了单个Duffing振子的测频范围,具有对调频引信信号的超低信噪比和高精度检测能力,对强噪声环境下非合作辐射源信号侦察具有重要参考意义。

参考文献 (References)

[1] 彭业凌. 线性调频引信信号的分选和识别[D]. 北京: 北京理工大学, 2015: 11, 21-80.
PENG Y L. The signal separation and identification of LFM fuze [D]. Beijing: Beijing Institute of Technology, 2015: 11, 21-80 (in Chinese).

[2] 索中英, 王建民, 吴华, 等. 基于小波变换的连续波多普勒体制无线电引信多普勒信号特征提取方法研究[J]. 探测与控制学报, 2006, 28(3): 29-32.
SUO Z Y, WANG J M, WU H, et al. Research on feature extraction methods for Doppler signal of continuous wave Doppler system radio fuze based on wavelet transform[J]. Journal of Detection and Control, 2006, 28(3): 29-32 (in Chinese).

[3] 黄海燕, 吕乐群, 蒲书缙. 连续波调频引信信号的检测方法[J]. 电子信息对抗技术, 2013, 28(5): 6-10.
HUANG H Y, LV L Q, PU S J. Detection of continuous wave FM radio fuze[J]. Electronic Information Warfare Technology, 2013, 28(5): 6-10 (in Chinese).

[4] 朱文涛, 苏涛, 杨涛, 等. 线性调频连续波信号检测与参数估计算法[J]. 电子与信息学报, 2014, 36(3): 552-558.
ZHU W T, SU T, YANG T, et al. Detection and parameter estimation of linear frequency modulation continuous wave signal [J]. Journal of Electronics and Information Technology, 2014, 36(3): 552-558 (in Chinese).

[5] 雷磊. 基于时频分析的线性调频信号检测与参数估计[D]. 西安: 西安电子科技大学, 2012: 29-52.
LEI L. Detection and parameters estimation of LFM signal based on time-frequency analysis [D]. Xi'an: Xidian University, 2012: 29-52 (in Chinese).

[6] 陈龔豪. 低信噪比雷达辐射源信号的检测[D]. 西安: 西安电子科技大学, 2017: 49-58.
CHEN Y H. The detection of radar emitter signals under conditions of low SNR [D]. Xi'an: Xidian University, 2017: 49-58 (in Chinese).

[7] MILLIOZ F, DAVIES M. Sparse detection in the chirplet transform: Application to FMCW radar signals[J]. IEEE Transactions on Signal Processing, 2012, 60(6): 2800-2813.

[8] ZHAO D, LUO M. A novel weak signal detection method for linear frequency modulation signal based on bistable system and fractional Fourier transform[J]. Optik-International Journal for Light and Electron Optics, 2016, 127(10): 4405-4412.

[9] 唐友福, 刘树林, 雷娜, 等. 基于广义局部频率的Duffing系统频域特征分析[J]. 物理学报, 2012, 61(17): 67-75.
TANG Y F, LIU S L, LEI N, et al. Feature analysis in frequency

- domain of Duffing system based on general local frequency[J]. Acta Physica Sinica, 2012, 61(17): 67-75 (in Chinese).
- [10] JIANG Y, ZHU H, LI Z. A new compound faults detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator[J]. Chaos, Solitons and Fractals, 2015, 1: 1-12.
- [11] 聂春燕. 混沌系统与弱信号检测[M]. 北京: 清华大学出版社, 2009: 21-104.
- NIE C Y. Chaotic system and weak signal detection[M]. Beijing: Tsinghua University Press, 2009: 21-104 (in Chinese).
- [12] 李楠. 水下弱目标信号的 Duffing 振子检测方法研究[D]. 哈尔滨: 哈尔滨工程大学, 2017: 13-42.
- LI N. A detection method of the underwater weak target signal based on Duffing oscillator[D]. Harbin: Harbin Engineering University, 2017: 13-42 (in Chinese).
- [13] 周薛雪, 赖莉, 罗懋康. 基于分数阶可停振动系统的周期未知微弱信号检测方法[J]. 物理学报, 2013, 62(9): 57-69.
- ZHOU X X, LAI L, LUO M K. A new detecting method for periodic weak signals based on fractional order stopping oscillation system[J]. Acta Physica Sinica, 2013, 62(9): 57-69 (in Chinese).
- [14] 李月, 杨宝俊, 林红波, 等. 基于特定混沌系统微弱谐波信号频率检测的理论分析与仿真[J]. 物理学报, 2005, 54(5): 1994-1999.
- LI Y, YANG B J, LIN H B, et al. Simulation and theoretical analysis on detection of the frequency of weak harmonic signals based on a special chaotic system[J]. Acta Physica Sinica, 2005, 54(5): 1994-1999 (in Chinese).
- [15] RASHTCHI V, NOURAZAR M. FPGA implementation of a real-time weak signal detector using a Duffing oscillator[J]. Circuits Systems & Signal Processing, 2015, 34(10): 3101-3119.
- [16] 刘海波, 吴德伟, 金伟, 等. Duffing 振子微弱信号检测方法研究[J]. 物理学报, 2013, 62(5): 42-47.
- LIU H B, WU D W, JIN W, et al. Study on weak signal detection method with Duffing oscillators[J]. Acta Physica Sinica, 2013, 62(5): 42-47 (in Chinese).
- [17] LI M P, XU X M, YANG B C, et al. A circular zone counting method of identifying a Duffing oscillator state transition and determining the critical value in weak signal detection[J]. Chinese Physics B, 2015, 24(6): 196-201.
- [18] 赖志慧, 冷永刚, 孙建桥, 等. 基于 Duffing 振子的变尺度微弱特征信号推测方法研究[J]. 物理学报, 2012, 61(5): 050503.
- LAI Z H, LENG Y G, SUN J Q, et al. Weak characteristic signal detection based on scale transformation of Duffing oscillator[J]. Acta Physica Sinica, 2012, 61(5): 050503 (in Chinese).
- [19] 朱位秋. 非线性随机动力学与控制[M]. 北京: 科学出版社, 2003: 330-335.
- ZHU W Q. Nonlinear stochastic dynamical systems and control[M]. Beijing: Science Press, 2003: 330-335 (in Chinese).
- [20] COSTA A H, ENRÍQUEZ-CALDERA R, TELLO-BELLO M. High resolution time-frequency representation for chirp signals using an adaptive system based on Duffing oscillators[J]. Digital Signal Processing, 2016, 55: 32-43.

作者简介:

朱志强 男, 硕士研究生。主要研究方向: 无线电引信对抗技术。

侯健 女, 博士。主要研究方向: 引信信息对抗技术、微弱信号检测技术。

闫晓鹏 男, 博士, 教授, 博士生导师。主要研究方向: 引信信息对抗技术、引信电磁环境与效应、引信信号处理技术。

Small-scale periodic state Duffing oscillator FMCW fuze
signal detection at ultra-low SNR

ZHU Zhiqiang¹, HOU Jian², YAN Xiaopeng^{1,*}, LI Ping¹, HAO Xinhong¹

(1. Science and Technology on Electromechanical Dynamic Control Laboratory, Beijing Institute of Technology, Beijing 100081, China;

2. Beijing Jinghang Research Institute of Computing and Communication, Beijing 100074, China)

Abstract: Aimed at the problem that the signal of frequency-modulated continuous wave (FMCW) fuze is very difficult to detect at ultra-low signal to noise ratio (SNR), a detection system based on small-scale periodic state Duffing oscillator is established. This system combines Duffing oscillator characteristics with stopping oscillation system theory, eliminating the inherent deficiencies of the traditional transformation-dependent Duffing oscillator detection methods, extending the frequency detection range through a single Duffing oscillator, and reducing the computing cost. On this basis, the phase trajectory characteristics of the small-scale periodic state are analyzed, and then a FMCW fuze signal detection method based on small-scale periodic state Duffing oscillator is proposed. The experimental results show that the small-scale periodic state Duffing oscillator detection method has an average detection error of less than 1% for the real FMCW fuze radiation signal at -30 dB ultra-low SNR, which verifies the effectiveness of the proposed method.

Keywords: frequency-modulated continuous wave (FMCW) fuze; Duffing oscillator; small-scale periodic state; stopping oscillation system; signal to noise ratio (SNR)

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0057

一种区域参考大气密度的建模与应用方法



刘一博, 沈作军*, 张向宇

(北京航空航天大学 航空科学与工程学院, 北京 100083)

摘 要: 为满足空天试验研究中对参考大气模型的实际需求, 即模型应具有跨时空的全球尺度覆盖性以及起降关键航迹点的高精度性, 提出了一种区域参考大气密度的建模与应用方法。首先对实测数据的冬夏季典型月份的大气密度特性, 具体包括月均值与密度扰动进行统计分析, 在此基础上提出了基于探空试验实测数据的全球参考大气模型 (GRAM) 的定量修正方法, 从而构建起了包含有大气扰动与季节变化的区域参考大气密度模型。最后提出了区域参考大气密度模型向 GRAM 的过渡方法与实际应用方法。研究表明, 探测地区的大气密度特性具有明显的冬夏季节性差异, 需要根据季节构建相应的大气密度模型。区域参考大气模型的蒙特卡罗仿真可以有效模拟实测数据的大气密度特性。通过与 GRAM 过渡结合, 参考大气模型既具有全球覆盖性, 又具有很高的局部精度。

关键词: 探空试验; 大气密度; 空天应用; 建模; 仿真

中图分类号: V411.8; V19; P421

文献标识码: A **文章编号:** 1001-5965(2019)10-2079-10

参考大气模型在空天试验中有着重要的应用^[1-3]。它表征了大气的理想气体性质, 包括压力、密度、温度、各元素分子量等随时间、空间的变化情况。模型的输入可以是高度、纬度、或者在一年中所处的天数。不同的参考大气模型有不同的输入和输出。例如美国标准大气模型 (USSA—1962) 以及其后续型号 (USSA—1976), 其输入为海拔高度, 输出内容包括大气温度、压力和密度。不考虑时间因素的一类参考大气模型被称为静态大气模型, 例如中国的参考大气模型 GJB 365.1—87^[4]。在静态模型的基础上, 如果考虑了随时间的变化, 如中国的参考大气模型 GJB 5601—2006^[5], 则模型可以同时反映大气参数随时间和空间变化。

真实大气环境中还存在着大小尺度的扰动, 传统的大气参考模型中缺乏对扰动量的表征。大气环境中的大尺度扰动是大气环境中扰动尺度与

时间尺度较长的扰动现象, 主要包括时间尺度为数天至一个月, 扰动尺度与地球纬度长度相当行星波以及 24 h 周期及其高次谐波的潮汐波等。中小尺度扰动的扰动尺度与时间尺度相对较短, 如几分钟至十几个小时周期的重力波与时间尺度更短的大气湍流扰动等^[6-7]。大气环境中的扰动对于空天试验有着重要的影响, 例如密度的扰动会影响飞行器的热流, 风场的扰动会影响到弹体的落点精度。对大气扰动进行研究并构建合适的仿真模型对于提升空天试验仿真精度具有最重要意义。早在 20 世纪 70 年代, 国外学者 Justus 在文献[8]中对重力波上层大气的混合作用进行了研究, 并在文献[9]中利用蒙特卡罗仿真对大气密度扰动进行了研究。国内学者肖存英等在文献[7]中对临近空间的大气扰动量变化特性进行了定量研究, 并在文献[10]中对临近空间大气密度建模进行了研究, 但由于 2 篇文献所使用的数据

收稿日期: 2019-02-21; 录用日期: 2019-03-30; 网络出版时间: 2019-04-04 16:00

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190403.1446.001.html

* 通信作者。E-mail: 09433@buaa.edu.cn

引用格式: 刘一博, 沈作军, 张向宇. 一种区域参考大气密度的建模与应用方法[J]. 北京航空航天大学学报, 2019, 45(10): 2079-2088. LIU Y B, SHEN Z J, ZHANG X Y. A method for range reference atmospheric density modeling and application[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2079-2088 (in Chinese).

来源为 SABER 探测器,该探测器只可探测到水平尺度大于 100 ~ 200 km 的波动,数据无法反映小尺度扰动的影响^[10]。

空天试验的跨度极广,航程上常常超过数千公里,航线跨越多个气候带甚至南北半球,在导航、制导与控制的设计过程中需要考虑大气环境随时空大幅度改变。传统工程应用中仅仅对起飞点或者降落点的大气环境测量与建模不能完全地满足空天试验需求。为此美国、俄罗斯等国都构建了具有全球覆盖尺度的大气参考模型。例如美国宇航局马歇尔太空飞行中心,利用美国国家环境预测中心(National Centers for Environmental Prediction, NCEP)全球再分析数据库,全球上层大气气候图集(Global Upper Air Climatic Atlas, GUACA)等气候数据,研发了一种涵盖有完整的地面到轨道高度的、且体现完整的全球地理差异变化以及完整的热力学变量随季节以及月度差异变化的全球参考大气模型(Global Reference Atmospheric Model, GRAM),并对其不断进行着更新升级^[11-14]。GRAM 提供了温度、密度、压强、风场、各分子量浓度等大气参数的气候期望值及其扰动量随时空的变化输出。但客观上来自气象卫星、火箭、遥感器、探空器等观测数据并不足以覆盖全部时空,在参考大气模型构建过程中必须使用插值技术进行补充,例如在 GRAM 构建的过程中,使用到了垂直插值、二维插值、过极点插值、数据集合间的过渡插值、季节与月份数据插值等过渡等插值技术^[11-14]。因此对于特定的空域,全球参考大气模型可能不如该地区长时间实际测量所得的结果精度高。一种将基于实测数据的局部参考大气与高覆盖性的大气参考模型相结合的建模方法,对于提高模型的局部精度,更好地满足空天试验对大气模型的需求具有重要意义。

酒泉地区是中国的重要空天试验基地,中国学者对该地区的大气环境特性进行过多方面的研究。杨钧烽等基于 MERRA(the Modern Era Retrospective-analysis for Research and Applications)再分析资料的风场数据,对酒泉地区上空的风切变特性进行了分析^[15]。高振荣等基于酒泉等 6 个地面气象站的数据对河西走廊地区近 58 年的气温变化与突变分析进行了研究^[16]。刘晓云等基于酒泉测站数据,对酒泉地区的气温和降水特征进行了分析^[17]。

本文利用酒泉地区(98.5°E, 39.5°N)的典型气象站 2010—2011 年 2 年的逐日探空试验密度数据,对该地区海拔 1 ~ 23.5 km 的大气密度均值

与扰动的冬夏季节性差异进行了分析。通过将实测数据与 GRAM 进行对比分析,提出了基于实测数据的均值与扰动量参数的模型定量修正方法,从而构建起了区域参考大气模型。在此基础上通过蒙特卡罗仿真与实测数据的对比,说明了该种建模方式的可行性与实用性。最后将局部大气密度模型与 GRAM 相结合,提出了一种由基于实测数据的局部参考大气向高覆盖范围的全球动态参考大气环境过渡的方法,使参考大气模型既具有全球尺度的覆盖性,又具有很高的局部精度,可以更好地满足空天试验需求。

1 大气密度特性冬夏季节差异分析

1.1 数据来源与处理

本文中所使用的实测数据来自于酒泉地区(98.5°E, 39.5°N)典型气象站 2010—2011 年的逐日探空试验(部分日期单日内进行了若干次探空试验),探测信息包括温度、密度、压强、经纬向风速,探测空间包括对流层至平流层下部的部分空域。相较于卫星探测器,探空气球所搭载的探测设备是对密度的直接测量,因此数据中包含了中小尺度扰动的影响。实际测量过程中,受恶劣自然环境影响,部分日期探空气球在进入平流层之前信号丢失,本文在数据处理过程中舍弃了这部分数据。对于顺利进入平流层的试验所测得的各组数据,选取海拔高度范围 1 ~ 23.5 km,将逐日数据中超出 3 倍标准差的异常值剔除,并对逐日数据按月份归类处理,方便分析大气参数随季节变化。经过上述处理后,最终 1 月剩余数据组数为 49 组,7 月为 59 组。

在本文的研究中所应用的另一个研究对象是 GRAM,其数据来源随海拔高度不同而不同^[11-14]:从 0 ~ 27 km,热力学和风的数据来源于全球上层大气气候图集(GUACA);20 ~ 120 km 之间的数据来自中层大气计划(Middle Atmosphere Program, MAP);120 km 以上的热层大气,数据来源可以由用户在马歇尔工程热层模型(the Marshall Engineering Thermosphere model Met-2007, MET-2007)、雅里奇·鲍曼热层模型(Jacchia-Bowman 2006 thermosphere model, JB2006)、海军研究实验室质谱仪非相干散射雷达扩展模型(Naval Research Labs Mass Spectrometer, Incoherent Scatter Radar Extended Model, NRLMSISE-00)以及相关的谐波风模型(Harmonic Wind Model, HWM-93)等数个模型中进行选择。在全球参考大气模型的最新版本 Earth-GRAM2010 中,来自 GUACA 的数据

已被 NCEP 全球再分析数据库的数据所取代^[14]。

1.2 大气密度特性冬夏季差异

本文选取 1 月作为冬季的典型月份,7 月作为夏季典型月份。同时为了更直观地表征密度扰动量,将扰动量无量纲化表示,定义密度扰动为实际值偏离统计月均值地百分比。对实测数据地统计分析表明,冬夏两季酒泉地区海拔 1~23.5 km 的大气密度特性,具体包括大气密度均值与扰动量均有明显差异。实测大气密度数据的均值与扰动量如图 1~图 3 所示。图 2、图 3 中的粗实线表示实测数据的 3 倍标准差范围。

密度均值方面,海拔 9.6 km 是一个分界点。在海拔 1~9.6 km 范围内,冬季密度均值小于夏季,该范围内在海拔 1 km 处两者差值达到最大值 16.56% (相对于冬季均值),差值随海拔增加呈减小趋势。在海拔 9.6~23.5 km 范围内,冬季密度均值大于夏季,两者差值在 16.5 km 处达到最大值 -21.67% (相对于 1 月均值),在这一范

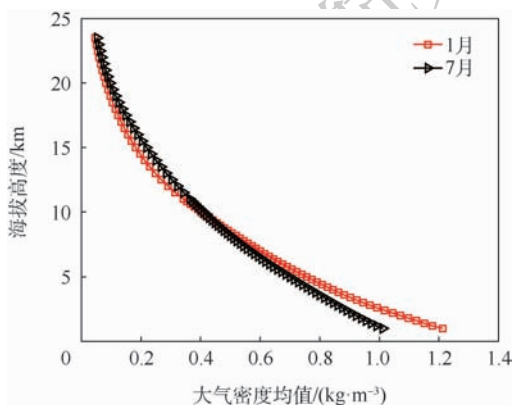


图 1 1 月与 7 月大气密度均值随海拔高度变化
Fig.1 Variation of mean atmospheric density with altitude in January and July

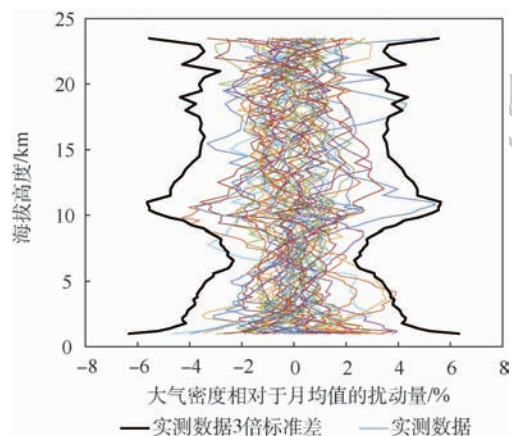


图 2 1 月大气密度扰动量随海拔高度变化
Fig.2 Variation of atmospheric density perturbation with altitude in January

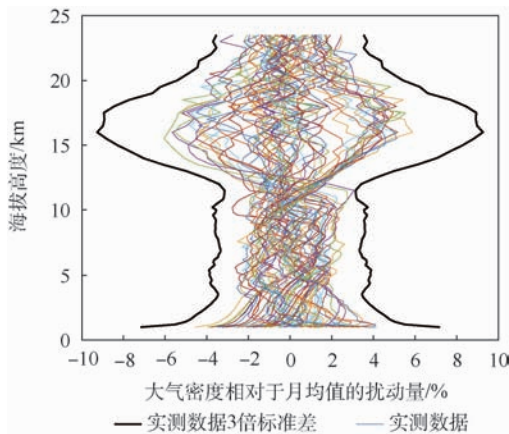


图 3 7 月大气密度扰动量随海拔高度变化

Fig.3 Variation of atmospheric density perturbation with altitude in July

围内两者差值随海拔高度增加呈先增加后减小的趋势。

扰动量方面,1 月与 7 月的数据整体上均呈现随海拔高度增加,扰动量先减小再增大之后继续减小的趋势。为了可以更加直观地分析扰动量的变化并进行对比,对扰动量数据进一步处理。当实测样本足够多时,相对统计月期望值的随机偏差量基本满足具有统计月标准差的高斯分布^[10]:

$$\rho_0 \sim N(\mu, \sigma) \quad (1)$$

式中: ρ_0 为大气密度; μ 为大气参数月均值; σ 为大气参数标准差。

相对于统计月均值的扰动量 ρ_{pm0} 自身的分布,应当是一种期望值为零的高斯分布:

$$\rho_{pm0} \sim N(0, \sigma_p) \quad (2)$$

式中: σ_p 为相对于统计月均值的扰动量的标准差。

因此可以将 σ_p 作为扰动量的一个特征参数来进行定量分析。1 月与 7 月的大气密度相对于统计月均值的扰动量标准差 σ_p 随海拔高度的变化及两者的对比情况如图 4 所示。

随海拔高度变化的趋势方面,1 月与 7 月的大气密度扰动量的标准差 σ_p 在海拔高度 1~23.5 km 范围内整体上均呈现先减小再增大之后继续减小的趋势,2 个月份的数据在对流层内有 2 个明显的拐点。其中 1 月份 σ_p 数据的这 2 个拐点对应的海拔高度分别为 6.6 km 与 11 km,7 月份 σ_p 数据的这 2 个拐点对应的海拔高度分别为 11.5 km,16 km。1 月份 σ_p 数据在 1~17.5 km 范围内随海拔高度升高的单调性有着较为明显的趋势,但在 17.5~23.5 km 范围内的随海拔高度升高出现了增减的反复。7 月份 σ_p 数

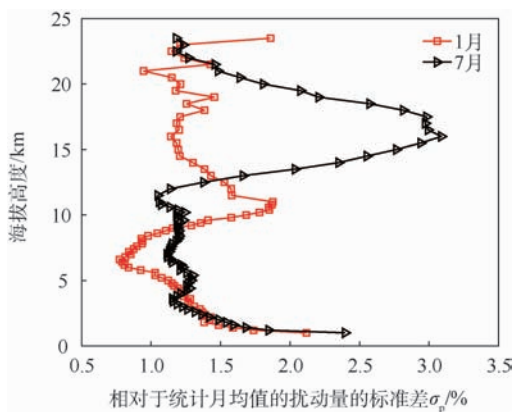


图4 1月与7月大气密度相对月均值的扰动量的标准差随海拔高度变化

Fig.4 Variation of standard deviation of perturbation compared with mean atmospheric density of statistical month with altitude in January and July

据在海拔高度 11.5 km 以下随海拔高度升高存在增减的反复,但在 11.5 ~ 23.5 km 范围内,随海拔高度升高 σ_p 有着明显而平稳的单调性趋势。

数值变化方面,1月的大气密度扰动量的标准差 σ_p 在海拔高度 1 km(接近酒泉地区地面)处达到最大值 2.118%,在海拔高度 6.6 km 处达到最小值 0.773%,此外 11 km 处 σ_p 为 1.876%,在海拔高度 23 km 以上出现了突增,在 23.5 km 处达到了 1.859%;7月的大气密度扰动量的标准差 σ_p 在海拔高度 16 km 处达到了最大值 3.088%,在海拔高度 11.5 km 处达到最小值 1.049%。

通过变化趋势与数值 2 个方面的分析可知,7月大气密度的扰动量的标准差 σ_p 随海拔高度的变化幅度整体上大于 1 月,数值上整体偏大,这意味着酒泉地区海拔高度 1 ~ 23.5 km 范围内的 7 月份大气密度扰动整体上更为剧烈。

2 区域参考大气密度建模

20 世纪 80 年代 Justus 提出了将大气参数表示为均值与密度扰动之和的建模方式^[8],在后续数十年的空天试验当中也依旧采用了这种建模方式^[11-14]。本文中的大气密度建模同样采用均值与随机扰动叠加构成大气参数的方式进行建模,即每一个蒙特卡罗仿真值是按照式(3)产生的:

$$\rho = \bar{\rho} + \rho_p \quad (3)$$

式中: ρ 为大气密度的仿真值; $\bar{\rho}$ 为大气密度的统计月均值的仿真值; ρ_p 为大气密度的扰动值的仿真值。

本文所述扰动量 ρ_{pm} 是一个无量纲化的概念,它是在扰动值的基础上提出的,表示扰动值相对

于月均值的比例:

$$\rho_{pm} = \rho_p / \bar{\rho} \quad (4)$$

区域参考大气密度的建模基于实测数据,通过对 GRAM 均值与扰动量 2 个方面的修正来实现。通过 GRAM 与实测数据的对比分析可以说明进行修正的必要性。将酒泉地区的经纬度坐标及对应统计月份输入 GRAM 后,可得 1 月与 7 月的均值输出与扰动输出。

2.1 基于实测数据的均值修正

由于均值是高斯分布的位置尺度参数,对均值的修正可以提升蒙特卡罗仿真对大气环境的模拟效果。通过对比实测数据与 GRAM 输出的酒泉地区 1 月与 7 月数据密度均值,可以发现 1 月数据海拔高度 6 km 以下 GRAM 数据的均值相较实测数据偏小,6 ~ 23.5 km 范围 GRAM 数据的均值相较实测数据偏大,且在海拔高度 16 ~ 23.5 km 范围内两者误差超过了 5%;7 月全球参考大气模型的数据与实测数据的差距相较于 1 月更大,在海拔高度 10.6 km 以下模型的仿真均值输出小于实测数据,而在 10.6 ~ 23 km 范围内则大于实测数据,并且两者误差也超过了 5%。实测数据与 GRAM 输出的相应月份的大气密度均值具体对比如图 5 和图 6 所示。

通过对比分析可以说明进行基于实测数据的均值修正是十分必要的。通过使用实测数据的均值替代原大气模型中的均值数据来实现对均值的修正,本文中修正量占相对原大气模型输出的百分比来定量描述修正,1 月与 7 月的修正结果分别如表 1 和表 2 所示。

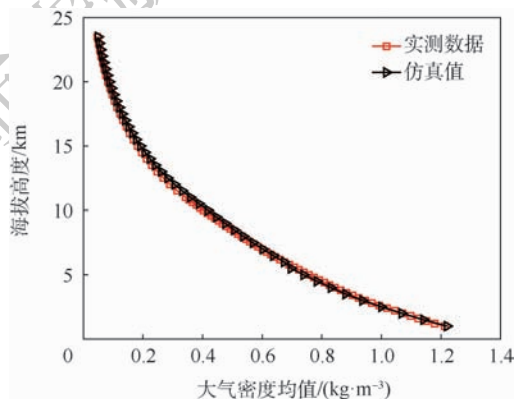


图5 实测数据与 GRAM 输出的 1 月大气密度均值对比
Fig.5 Contrast of mean atmospheric density of measured data to that of output of GRAM in January

2.2 基于实测数据的扰动修正

全球参考大气模型中对密度大小尺度的扰动均有建模^[11-14]。其中大尺度扰动使用了随机的变幅值、变波长、变相位的大尺度扰动建模方法。

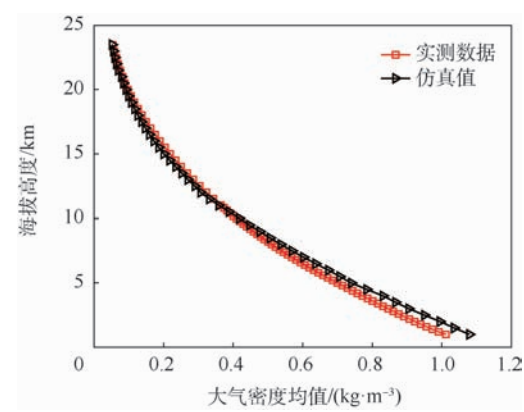


图 6 实测数据与 GRAM 输出的 7 月大气密度均值对比
Fig. 6 Contrast of mean atmospheric density of measured data to that of output of GRAM in July

假设为变幅值的余弦函数,扰动为在一定数值区间内的随机数,构建无量纲化的大尺度扰动函数如下:

$$F(h) = [\text{rand}(0,1) + 0.5] A \cos\left(\frac{2\pi h}{\lambda} + \phi\right) \quad (5)$$

式中: h 为海拔高度; $\text{rand}(0,1)$ 表示在 0 至 1 区间内产生的随机数; A 为幅值; λ 为波长; ϕ 为相位。 A 与 λ 为高度的函数, ϕ 与初始位置选取有关。

小尺度扰动则使用了一阶自回归模型来表示,通过连续的扰动值之间的必要相关性来计算

在新位置处的扰动。该建模方式的主要思路是假设已知 x 方向上某位置 x_1 处的归一化变量 $\alpha(x_1)$ (α 为 x_1 方向上相对均值的偏差值与标准差之比),则下一位置 x_2 的变量 $\alpha(x_2)$ 表达式为

$$\alpha(x_2) = r\alpha(x_1) + (1 - r^2)^{0.5} q(x_1) \quad (6)$$

式中: $q(x)$ 为期望值为 0 以及标准差为 1 的标准高斯分布; r 为一系列归一化变量之间的自相关系数。

但通过实际仿真检验,全球参考大气模型中酒泉地区 1 月与 7 月扰动量的输出与实测数据存在一定差异,对扰动量的修正也是十分必要的。标准差是高斯分布的尺度参数,大气扰动对于大气参数的影响体现在数据标准差上,基于前文对于扰动量自身分布的分析,本文中使用相对于统计月均值的扰动量的标准差 σ_p 来作为修正的依据。由于扰动量修正是在均值修正的基础上进行的,所以计算 σ_p 时所用到的统计月密度均值实际为实测数据的均值。

首先以 1 月数据的修正为例,统计处理全球参考大气模型酒泉地区 1 月大气密度 250 次蒙特卡罗的仿真结果,得到相对于统计月均值的扰动量的标准差 σ_p 。扰动量仿真值与实测数据的扰动量的对比如图 7 所示。

表 1 1 月大气密度均值相对 GRAM 原输出的修正

Table 1 Correction of mean atmospheric density compared to GRAM primary output in January

海拔高度/km	均值相对原输出的修正量 y/%	海拔高度/km	均值相对原输出的修正量 y/%	海拔高度/km	均值相对原输出的修正量 y/%
1	-0.60	9	-2.20	17	-5.13
2	0.22	10	-3.37	18	-5.28
3	0.91	11	-4.61	19	-5.56
4	1.57	12	-5.20	20	-5.57
5	2.26	13	-4.87	21	-5.62
6	0.10	14	-4.85	22	-5.57
7	-0.43	15	-4.83	23	-5.43
8	-1.03	16	-4.91	23.5	-5.57

表 2 7 月大气密度均值相对 GRAM 原输出的修正

Table 2 Correction of mean atmospheric density compared to GRAM primary output in July

海拔高度/km	均值相对原输出的修正量 y/%	海拔高度/km	均值相对原输出的修正量 y/%	海拔高度/km	均值相对原输出的修正量 y/%
1	6.69	9	4.06	17	-8.95
2	7.84	10	1.81	18	-7.97
3	7.05	11	-1.58	19	-6.72
4	7.66	12	-5.07	20	-5.16
5	5.69	13	-5.33	21	-3.87
6	6.67	14	-6.35	22	-6.59
7	6.00	15	-8.18	23	-4.46
8	5.21	16	-9.02	23.5	-3.43

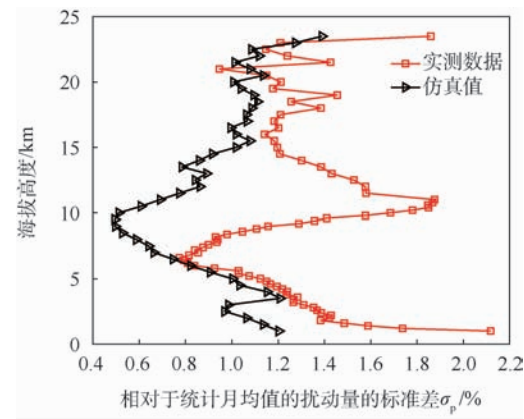


图 7 1 月实测数据与 GRAM 输出的大气密度相对月均值的扰动量的标准差的对比

Fig. 7 Contrast of standard deviation of measured data to that of output of GRAM in terms of perturbation compared with monthly mean atmospheric density in January

与实测数据对比可知,海拔高度 1 ~ 3 km 以及 7 ~ 20 km 范围,全球参考大气模型仿真输出的扰动量相较实际情况偏小,3.5 ~ 6.5 km 以及 20.5 ~ 23.5 km 范围模拟效果相对较好。为了定量说明针对相对于统计月均值的扰动量的标准差 σ_p 的修正,按照式(7)定义修正因子 k :

$$k = \frac{\sigma_{pG}}{\sigma_{pM}} \tag{7}$$

式中: σ_{pG} 为对全球参考大气模型输出统计计算后得到的相对于统计月均值的扰动量的标准差; σ_{pM} 为实测数据相对于统计月均值的扰动量的标准差。

通过调整修正因子使得 σ_{pG} 接近 σ_{pM} ,从而使得蒙特卡罗仿真的扰动输出更加接近实测数据,对于 7 月的扰动量修正过程与此相同。最终 1 月与 7 月的定量修正结果如表 3 和表 4 所示。

2.3 修正的仿真校验

经过 2.1 节中的均值修正与 2.2 节中扰动量的修正,可以通过蒙特卡罗仿真对修正后整体输出结果进行检验,每一个蒙特卡罗的输出值产生过程为

$$\rho_r = (1 + \gamma\%) \bar{\rho}_G + k \rho_{pG} \tag{8}$$

式中: ρ_r 表示修正后的仿真值; $\bar{\rho}_G$ 表示 GRAM 原输出的均值; γ 为 2.1 节表 1 与表 2 中所述修正量; ρ_{pG} 为 GRAM 原输出的扰动值; k 为 2.2 节所述修正因子。

通过大量蒙特卡罗仿真与实测数据的对比可以说明这种建模方式的可行性与实用性。1月与

表 3 1 月大气密度相对月均值的扰动量的标准差相对 GRAM 原输出的修正
Table 3 Correction of standard deviation of atmospheric density perturbation compared to GRAM primary output in January

海拔高度/km	修正因子 k	海拔高度/km	修正因子 k	海拔高度/km	修正因子 k
1	1.76	9	2.85	17	1.11
2	1.33	10	3.28	18	1.27
3	1.31	11	2.70	19	1.33
4	1.06	12	1.82	20	1.20
5	1.12	13	1.61	21	1.07
6	1.02	14	1.51	22	1.11
7	1.29	15	1.17	23	0.95
8	1.60	16	1.12	23.5	1.34

表 4 7 月大气密度相对月均值的扰动量的标准差相对 GRAM 原输出的修正
Table 4 Correction of standard deviation of atmospheric density perturbation compared to GRAM primary output in July

海拔高度/km	修正因子 k	海拔高度/km	修正因子 k	海拔高度/km	修正因子 k
1	1.99	9	2.40	17	2.79
2	1.40	10	2.31	18	2.59
3	1.25	11	1.53	19	2.02
4	1.05	12	1.32	20	1.79
5	1.28	13	1.87	21	1.42
6	1.50	14	2.73	22	1.28
7	1.68	15	2.72	23	1.24
8	2.01	16	3.03	23.5	1.17

7 月各 250 次仿真结果如图 8 与图 9 所示。

由于仿真过程中每一个仿真值是按照均值与扰动值叠加的方法产生的,其中均值本身就是实测数据的统计月均值,所以可以通过检验仿真对实测数据的扰动量的模拟效果来进一步校验区域参考大气模型。统计计算 1 月与 7 月 250 次蒙特卡罗仿真的扰动量,并与 1 月 49 组实测数据和 7 月 59 组实测数据的扰动量进行对比。结果如图 10 ~ 图 13 所示,各图中粗实线为实测数据的大气密度扰动量的 3 倍标准差。

通过扰动量的对比,可以更加直观地看出实测数据落在了仿真值的包络范围内,且仿真值基本都在实测数据的 3 倍标准差范围内。通过相对统计月均值的扰动量的标准差这一特征参数的对比,可以进一步说明区域参考大气对于实测数据的扰动量也有很好的模拟效果。扰动量的标准差的蒙特卡罗仿真结果与实测数据的对比如图 14 和图 15 所示。

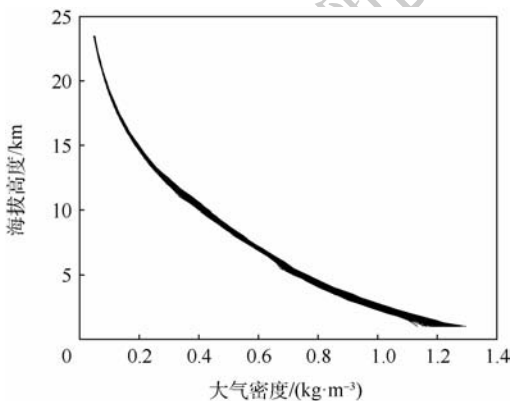


图 8 区域参考大气密度模型 1 月的蒙特卡罗仿真结果
Fig.8 Monte Carlo simulation results of range reference atmospheric density model in January

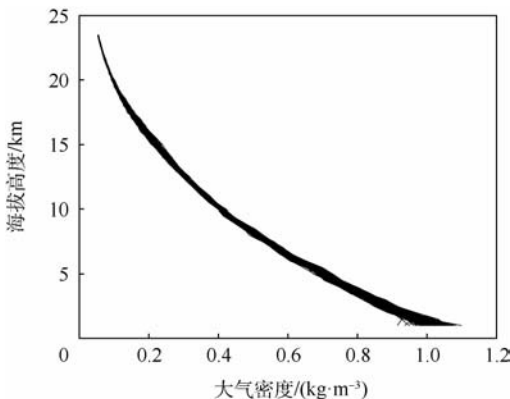


图 9 区域参考大气密度模型 7 月的蒙特卡罗仿真结果
Fig.9 Monte Carlo simulation results of range reference atmospheric density model in July

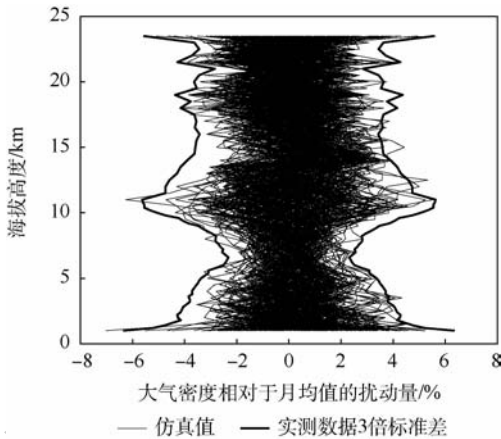


图 10 区域参考大气密度模型 1 月密度扰动量的蒙特卡罗仿真结果
Fig.10 Monte Carlo simulation results of density perturbation of range reference atmospheric density model in January

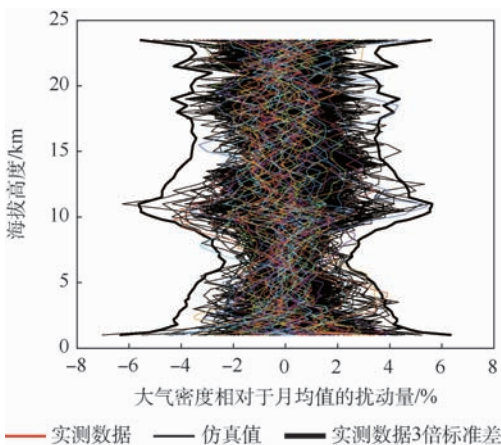


图 11 区域参考大气密度模型 1 月密度扰动量的蒙特卡罗仿真结果与实测数据对比
Fig.11 Contrast of Monte Carlo simulation results of density perturbation of range reference atmospheric density model to that of measured data in January

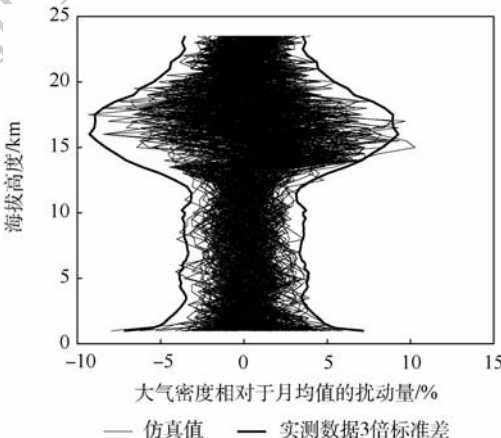


图 12 区域参考大气密度模型 7 月密度扰动量的蒙特卡罗仿真结果
Fig.12 Monte Carlo simulation results of density perturbation of range reference atmospheric density model in July

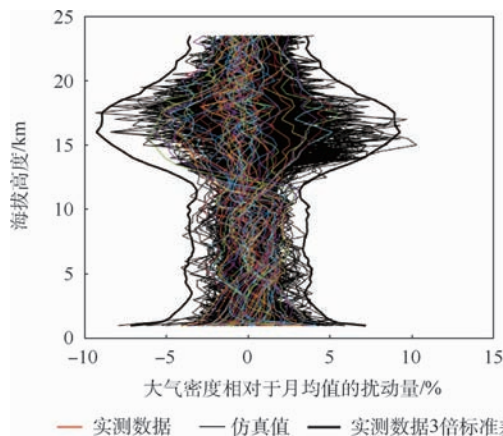


图 13 区域参考大气密度模型 7 月密度扰动量的蒙特卡罗仿真结果与实测数据对比

Fig. 13 Contrast of Monte Carlo simulation results of density perturbation of range reference atmospheric density model to that of measured data in July

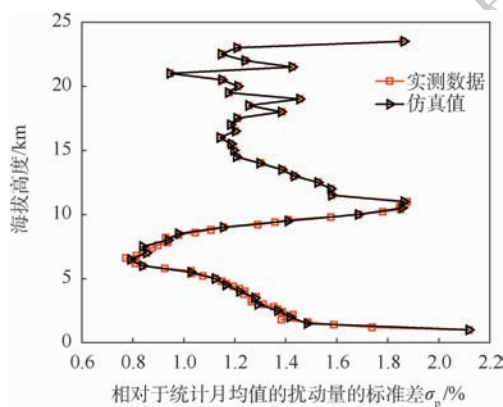


图 14 定量修正后 1 月实测数据与 GRAM 输出的大气密度相对月均值的扰动量的标准差的对比

Fig. 14 Contrast of standard deviation of measured data to that of output of GRAM in terms of perturbation compared with monthly mean atmospheric density after quantitative correction in January

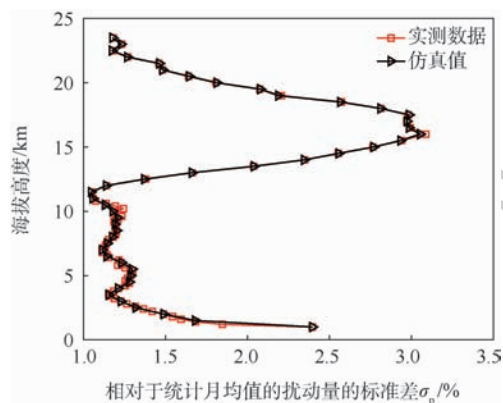


图 15 定量修正后 7 月实测数据与 GRAM 输出的大气密度相对月均值的扰动量的标准差的对比

Fig. 15 Contrast of standard deviation of measured data to that of output of GRAM in terms of perturbation compared with monthly mean atmospheric density after quantitative correction in July

3 区域参考模型向 GRAM 的过渡与应用方法

为了适应跨越时空范围极为广阔的空天试验需求,还需要将修正后的局部大气模型与全球参考大气模型进行过渡。本文中采用一种相对简单的过渡算法处理局部参考大气向全球参考大气过渡的问题。

首先以实测数据的经纬坐标作为中心点,规定经纬度半径值为 R ,当 $R=2^\circ$ 时的半径值为 R_1 ;当 $R=5^\circ$ 时的半径值为 R_2 。对于半径范围在 2° 之内的大气参数,模型输出经过修正的局部参考大气模型的数据;对于半径范围大于 5° 的大气参数,输出全球参考大气模型的数据;对于半径范围在 $2^\circ \sim 5^\circ$ 之间的大气参数,输出 2 种大气模型的过渡数据。即数据输出按照式 (9) 过渡算法执行:

$$M = (1 - w)RA + wG \quad (9)$$

式中: M 为参考大气模型最终输出; RA 为区域参考大气模型的输出; G 为 GRAM 输出; w 为过渡权重数值。 w 具体计算方法如下:

$$w = \begin{cases} 0 & R < R_1 \\ aR^3 + bR^2 + cR + d & R_1 < R < R_2 \\ 1 & R > R_2 \end{cases} \quad (10)$$

为了使过渡平滑光顺,各项参数如下:

$$\begin{cases} a = \frac{2}{(R_2 - R_1)^3} \\ b = \frac{-3R_1R_2}{(R_2 - R_1)^3} \\ c = \frac{6R_1R_2}{(R_2 - R_1)^3} \\ d = \frac{R_1^2(R_2 - 3R_1)}{(R_2 - R_1)^2(R_1 + R_2)} \end{cases} \quad (11)$$

以东经 $86^\circ \sim 94^\circ$,北纬 $36^\circ \sim 42^\circ$ 范围为例,通过海拔高度 21 km 处的大气密度输出情况来表示局部参考大气模型向全球参考大气模型过渡方法,如图 16 所示。

以海拔高度 21 km 为例,区域参考大气向全球参考大气过渡的建模技术的应用方法如图 17 所示。图中黄色实线为假设的飞行器航迹,例如发射点为酒泉地区,降落点为西部某地,红色与蓝色圆环分别表示发射点与降落点的区域参考大气模型范围。在进行飞行仿真的过程中,航迹处于小圆以内的部分使用区域参考大气数据,大圆以外的部分使用全球参考大气模型的数据,内外圆

之间属于过渡范围,按照前文所述的过渡方法使用过渡数据。这样既可以保证关键航迹点的模型精度,同时还使模型具有了全球尺度的覆盖性,并且可以在 2 种模型间平滑过渡。

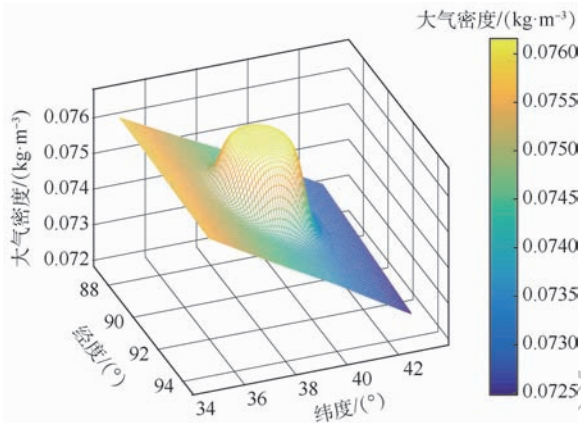


图 16 区域参考大气密度模型向 GRAM 的过渡方法示意
Fig. 16 Illustration of transition method for range reference atmospheric density model to GRAM

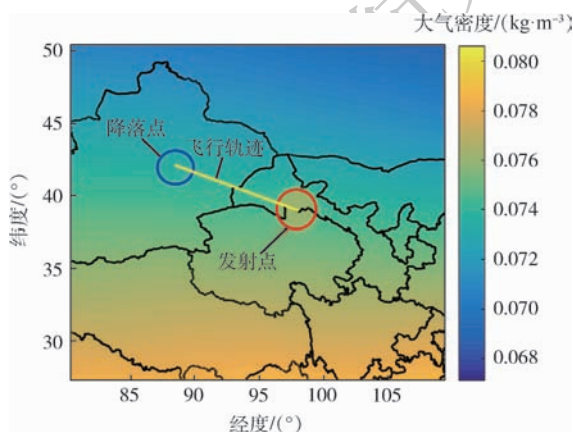


图 17 与 GRAM 过渡结合的区域参考大气密度模型的应用方法示意

Fig. 17 Illustration of application method for range reference atmospheric density model combined with GRAM transition

4 结 论

为满足空天试验对参考大气模型对跨时空的全球覆盖性与局部高精度性的要求,本文探空实测大气密度冬夏典型代表月份的数据分析了探测地区的大气密度特性冬夏季节性差异,通过定量修正 GRAM,构建了区域参考大气密度模型,模型的蒙特卡罗仿真结果可以有效模拟实测数据。最后通过本文所提出的区域参考大气向 GRAM 的过渡与应用方法,使参考大气模型即具有全球尺度的覆盖性,又具有很高的局部精度。受气球探空的试验条件限制,本文仅对海拔高度 1 ~ 23.5 km 空域范围进行了研究,而临近空间范围的大气探

测需要探空火箭与遥感技术来实现。但本文所提出的区域参考大气密度的建模与应用方法在包括临近空间在内的更广阔的空域也同样适用。

参考文献 (References)

- [1] DALE J, BARRY R, WILLIAM V, et al. Atmospheric models for engineering applications: AIAA-2003-0894 [R]. Reston, VA: AIAA, 2003.
- [2] DALE J, WILLIAM V. How atmospheric thermodynamic parameters and model atmospheres have been used to help engineering in aerospace launch vehicle design & development [C] // 50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. Reston, VA: AIAA, 2012.
- [3] ALETA D, JUSTUS C G, VERNON K. Global reference atmospheric model (GRAM) series for aeroassist applications [C] // 43rd AIAA Aerospace Sciences Meeting and Exhibit. Reston, VA: AIAA, 2005.
- [4] 陈昭. 北半球标准大气 (-2 ~ 80 公里): GJB 365. 1—87 [S]. 北京:国防工业技术科技委员会, 1987.
CHEN Z. Standard atmosphere in northern hemisphere (-2 - 80 km): GJB 365. 1—87 [S]. Beijing: Commission of Science Technology and Industry for National Defense, 1987 (in Chinese).
- [5] 李群, 谢志辉, 阎书源, 等. 中国参考大气 (地面 ~ 80 km): GJB 5601—2006 [S]. 北京: 中国人民解放军总装备部, 2006.
LI Q, XIE Z H, YAN S Y, et al. China reference atmosphere (ground ~ 80 km): GJB 5601—2006 [S]. Beijing: The PLA General Armament Department, 2006 (in Chinese).
- [6] 姚志刚, 孙睿, 赵增亮, 等. 风云三号卫星微波观测的临近空间大气扰动特征 [J]. 地球物理学报, 2019, 62 (2): 473-488.
YAO Z G, SUN R, ZHAO Z L, et al. Gravity waves in the near space observed by the microwave temperature sounder of the FY 3C meteorology satellite [J]. Chinese Journal of Geophysics, 2019, 62 (2): 473-488 (in Chinese).
- [7] 肖存英, 胡雄, 王博, 等. 临近空间大气扰动变化特性的定量研究 [J]. 地球物理学报, 2016, 59 (4): 1211-1221.
XIAO C Y, HU X, WANG B, et al. Quantitative studies on the variations of near space atmospheric fluctuation [J]. Chinese Journal of Geophysics, 2016, 59 (4): 1211-1221 (in Chinese).
- [8] JUSTUS C G. Upper atmospheric mixing by gravity waves [C] // International Conference on the Environmental Impact of Aerospace Operations in the High Atmosphere. Reston, VA: AIAA, 1973: 1-4.
- [9] JUSTUS C G. Density perturbation simulation with the global reference atmospheric model [C] // 26th Aerospace Sciences Meeting. Reston, VA: AIAA, 1988: 1-8.
- [10] 肖存英, 胡雄, 杨钧烽, 等. 临近空间 38°N 大气密度特性及建模技术 [J]. 北京航空航天大学学报, 2017, 43 (9): 1757-1765.
XIAO C Y, HU X, YANG J F, et al. Characteristics of atmospheric density at 38°N in near space and its modeling technique [J]. Journal of Beijing University of Aeronautics and Astronau-

- tics, 2017, 43(9): 1757-1765 (in Chinese).
- [11] JUSTUS C G, JOHNSON D L. The GRAM model; Status of development and future aspects[J]. *Advances in Space Research*, 1997, 19(4): 549-558.
- [12] JUSTUS C G, DUVAL A, JOHNSON D L. Earth global reference atmospheric model (GRAM-99) and trace constituents [J]. *Advances in Space Research*, 2004, 34(8): 1731-1735.
- [13] LESLIE F. Earth global reference atmospheric model 2007 (Earth-GRAM07) [C] // 37th COSPAR Scientific Assembly. Washington, D. C.: NASA, 2008.
- [14] LESLIE F W, JUSTUS C G. The NASA marshall space flight center earth global reference atmospheric model—2010 Version; NASA/TM-2011-216467 [R]. Washington, D. C.: NASA Marshall Space Flight Center, 2019.
- [15] 杨钧烽, 肖存英, 胡雄, 等. 临近空间风切变特性及其对飞行器的影响[J]. *北京航空航天大学学报*, 2019, 45(1): 57-65. YANG J F, XIAO C Y, HU X, et al. Wind shear characteristics in near space and their impacts on air vehicle [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2019, 45(1): 57-65 (in Chinese).
- [16] 高振荣, 田庆明, 刘晓云, 等. 近 58 年河西走廊地区气温变化及突变分析[J]. *干旱区研究*, 2010, 27(2): 194-203. GAO Z R, TIAN Q M, LIU X Y, et al. Analysis of temperature change and abrupt change in Hexi Corridor in recent 58 years [J]. *Research in the Arid Areas*, 2010, 27(2): 194-203 (in Chinese).
- [17] 刘晓云, 岳平, 徐殿祥. 酒泉市最近 54a 气温和降水特征分析[J]. *干旱区研究*, 2006, 23(3): 495-499. LIU X Y, YUE P, XU D X. Analysis of temperature and precipitation characteristics in Jiuquan City in recent 54 years [J]. *Research in the Arid Areas*, 2006, 23(3): 495-499 (in Chinese).

作者简介:

刘一博 男, 硕士研究生. 主要研究方向: 空天飞行器制导与控制, 空天飞行器不确定性评估.

沈作军 男, 博士, 教授. 主要研究方向: 飞行力学、制导技术与飞行控制、最优控制、轨迹优化.

张向宇 男, 硕士研究生. 主要研究方向: 空天飞行器仿真环境建模、空天飞行器不确定性评估.

A method for range reference atmospheric density modeling and application

LIU Yibo, SHEN Zuojun*, ZHANG Xiangyu

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: To meet the requirement of reference atmospheric model in aerospace application, that is, the model should have global scale coverage across seasons and space, and high accuracy of key trajectory points, such as the range for takeoff and landing, a method for modeling and application of range reference atmospheric density is proposed. First, the characteristics of atmospheric density in typical months of winter and summer, including monthly mean and density perturbation, are analyzed. On this basis, a quantitative correction method of the global reference atmosphere model (GRAM) based on sounding measurement is proposed in this paper. Then, the range reference atmospheric density model with atmospheric perturbation and seasonal variation is constructed. Finally, the transition method and application method for the transition from the range reference atmospheric density model to GRAM are proposed. The simulation results show that the atmospheric density characteristics of the measured area have obvious seasonal differences in winter and summer, and the atmospheric density model should be built according to the season. The results of Monte Carlo simulation of the range reference atmospheric model show that it can effectively simulate the atmospheric density characteristics of the measured data. Combined with GRAM transition, the reference atmosphere model has both global coverage and high range accuracy.

Keywords: sounding measurement; atmospheric density; aerospace application; modeling; simulation

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2018.0564

基于异步卷积分解与分流结构的单阶段检测器



赵蓬辉¹, 孟春宁^{2,*}, 常胜江¹

(1. 南开大学 电子信息与光学工程学院 现代光学研究所, 天津 300350;

2. 中国人民武装警察部队海警学院 电子技术系, 宁波 315801)

摘 要: 目标检测网络 SSD 的多层回归特征图存在各层回归计算之间相对独立的问题,且基于 SSD 改进的系列算法在提高检测精度的同时难以兼顾实时性。针对上述问题,提出一种基于异步卷积分解与分流(shunt)结构的单阶段目标检测器。基于异步卷积分解算法设计了一种 shunt 结构,交错连接多层特征图,增强了回归计算之间的统一性与协调性。优化了原有高层主流结构,在主流结构与 shunt 结构中分别用最大池化和异步卷积分解 2 种不同的方式对特征图大小进行降维,保留空间相关信息的同时提高了特征的多样性。实验结果表明,将 VOC2007trainval 和 VOC2012trainval 中的图片统一缩小至 300 像素×300 像素进行训练,提出的目标检测器在 VOC2007test 上进行检测时的平均精度均值可达到 80.5%,检测速度超过 30 帧/s。

关 键 词: 目标检测; 卷积神经网络; 异步卷积分解; 分流结构; 结构优化

中图分类号: TP391.4

文献标识码: A

文章编号: 1001-5965(2019)10-2089-10

目标检测作为计算机视觉领域的关键技术,一直是具有挑战性的研究热点。Viola 和 Jones^[1]于 21 世纪初提出了 Viola-Jones 算法首次在计算资源有限的情况下实现了实时的人脸检测。2005 年, Dalal 与 Triggs^[2]提出的 HOG (Histogram of Oriented Gradient) 行人检测器拓展了 Viola-Jones 算法的检测领域。Felzenszwalb 等^[3]提出的可变形部件模型 (Deformable Part based Model, DPM) 及其后续优化算法连续 3 年获得目标类别视觉挑战赛 (The PASCAL Visual Object Classes Challenge)^[4]的冠军,代表了当时基于手工设计特征的目标检测器的最高水平。然而,基于手工设计特征的目标检测器在目标区域选择上策略针对性差,窗口计算冗余量大,尤其针对环境多样性变化和遮挡问题没有很好的鲁棒性^[5]。因此,早期的目标检测算法难以达到实用的要求,陷入发展停滞期。

得益于卷积神经网络的迅速发展, Girshick 等于 2014 年提出了区域卷积神经网络结构 (Regions with Convolutional Neural Networks features, R-CNN)^[6], 标志着目标检测摆脱了缓慢发展的困境,进入了新的发展阶段。此后涌现出的基于深度学习的目标检测算法大致可分为 2 类: 两阶段分类回归系列与单阶段回归系列。

R-CNN 主要计算流程由 2 部分组成, 分别是类别分类和位置回归。此后, 在 R-CNN 基础上提出的 SPP-Net^[7]、Fast R-CNN^[8]、Faster R-CNN^[9]、FPN (Feature Pyramid Networks)^[10] 等算法均将检测任务分成分类问题和位置回归 2 类问题。因此, 这类算法统称为两阶段分类回归系列。随着目标检测数据集的逐渐丰富和网络训练技巧的不断优化, 该系列检测算法的平均检测精度得到迅速提升。两阶段分类回归系列算法虽然能取得

收稿日期: 2018-09-27; 录用日期: 2019-05-18; 网络出版时间: 2019-06-19 08:36

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190617.0931.002.html

基金项目: 公安部技术研究计划 (2017JSYJC10)

* 通信作者. E-mail: mengchunning123@163.com

引用格式: 赵蓬辉, 孟春宁, 常胜江. 基于异步卷积分解与分流结构的单阶段检测器[J]. 北京航空航天大学学报, 2019, 45(10): 2089-2098. ZHAO P H, MENG C N, CHANG S J. Single shot multibox detector based on asynchronous convolution factorization and shunt structure[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2089-2098 (in Chinese).

70%以上的平均精度均值,但其网络结构过于复杂,不仅导致前期的网络训练时间较长,也致使网络参数量较大,难以保证实时性。为获得较高的检测帧率,两阶段分类回归系列算法需要苛刻的GPU硬件配置条件,且不利于将算法移植到移动终端。提高检测精度的同时,保证检测速度日益成为工业界的实际需求。为兼顾精度与速度,单阶段回归算法应运而生。

Redmon等^[11]于2015年提出的端到端一体化网络YOLO v1在VOC07^[4](Pascal VOC 2007)上取得66.4%的平均精度均值,虽然平均检测精度低于大部分两阶段分类回归算法,但检测处理速度最高可达155帧/s。此后,为提高检测精度,陆续出现YOLO9000^[12]、SSD^[13]、YOLOv3^[14]等单阶段回归算法。这类算法将分类问题与位置回归问题统一成一个回归计算问题,与R-CNN等算法形成鲜明对比。为进一步提高平均精度均值,Liu等在SSD的基础上先后提出DSSD^[15]、DSOD^[16]等优化的SSD类算法。在检测流程中,SSD类算法采用锚箱^[8]在多层特征图上以不同比例与尺寸的建议框进行回归计算,一次性检测物体的类别与位置。SSD类算法相对两阶段分类回归算法计算简单且参数量较少,能够在一定程度上兼顾检测精度与速度。

SSD中的多层次回归计算思想优于单一特征图上的目标检测,利于消除YOLO v1中存在的近邻目标检测“竞争”现象^[11],DSSD等算法继承了SSD多层回归计算的特点,通过改进高层框架结构以提升平均精度均值。然而,SSD多层回归计算在结构上存在回归特征图(检测所基于的多层特征图)层与层之间相对独立的情况。虽然通过高层结构改造可在一定程度上提升平均精度,但改造后的复杂结构又影响了检测速度,难以保证实时性^[14,16]。

为解决SSD系列算法存在的层间回归计算相对独立的问题,保证目标检测实时性的同时,进一步提高检测精度,本文基于SSD构造了一种底层结构为卷积神经网络^[17]的基于异步卷积分解与分流结构的单阶段检测器(Single Shot mutibox Detection based on Asynchronous convolution Factorization and shunt structure,FA-SSD)。在底层网络结构中,为避免网络结构过于复杂,以单链无分支的卷积神经网络结构为基础。在高层网络结构中,采用异步卷积分解的两层卷积层与池化层相结合的降维方式,保证空间相关信息的同时,提高特征的多样性。此外,FA-SSD借鉴残差神经网络^[18]的交融结构,提出一种分流(shunt)结构交

错连接多层回归特征图,增强多层回归计算之间的统一性与协调性。

FA-SSD网络的创新点可归纳如下:

1) 借鉴Inception模型^[19]中的卷积分解思想,增加了不同方向不同步长的异步卷积策略,提出一种异步卷积分解算法。采用异步卷积分解算法构造的两层卷积层,与SSD采用一层卷积对特征图降维相比,可在不增加计算量的情况下,提高提取特征的非线性表达能力。

2) 基于异步卷积分解算法构造了一种shunt结构,从原有高层网络结构(主流结构)中分出包含两层采用异步卷积分解的卷积层组成的shunt结构,同时与主流结构实现特征交融,解决了各层回归计算相对独立的问题,增强了多层回归计算之间的统一性与协调性。在shunt结构中采用异步卷积分解算法构造的卷积层可在降低特征图维度的同时,增加特征图之间的交融过程,改善空间相关信息缺失的问题。

3) 优化了高层网络结构中的主流结构。首先使用池化层代替原有结构中的步长为2的卷积层,在主流结构和shunt结构中同时使用池化和卷积两种降维方式,提升特征的多样性;然后在每次交融的特征图后增加一层步长为1大小为 1×1 的卷积层以提升回归特征图之间的卷积深度,扩大各层回归特征图间的特征差异性。

1 SSD网络结构

SSD算法的计算流程如图1所示,其中图1(a)和(b)为底层特征提取阶段,图1(b)和(c)为非极大值抑制(NMS)筛选多个检测结果的阶段,图1(b)表示多层回归特征图把分类任务与位置回归任务统一为回归问题,在高层神经网络中的多层次特征图上分别应用不同尺寸的锚箱生成特定长宽比例的建议框,并在建议框上一次性回归物体类别与位置信息,生成的建议框长宽比有“1、2、3、1/2、1/3”等。

底层结构为VGG结构的SSD网络框架如图2所示,高层结构中的多层回归特征图分别为block4、fc7、block8、block9、block10、block11。其中上述名称分别表示神经网络中每个卷积命名区域中最后的特征图层。黑色圆表示卷积(conv),采用大小为“ 3×3 ”,步长为(1,1)的两层卷积核,即2kernel3×3_s1。红色三角形表示池化(pool),采用大小为“ 2×2 ”,步长为(2,2)的一层最大池化,即pool2×2_s2。上述6层特征图层共生成8732个特定比例与尺寸的建议框,每个

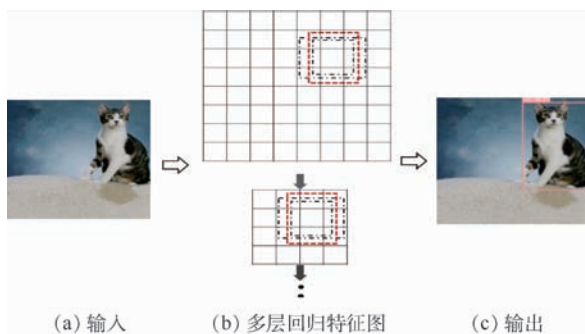


图 1 SSD 算法计算流程

Fig. 1 Computation procedure of SSD algorithm

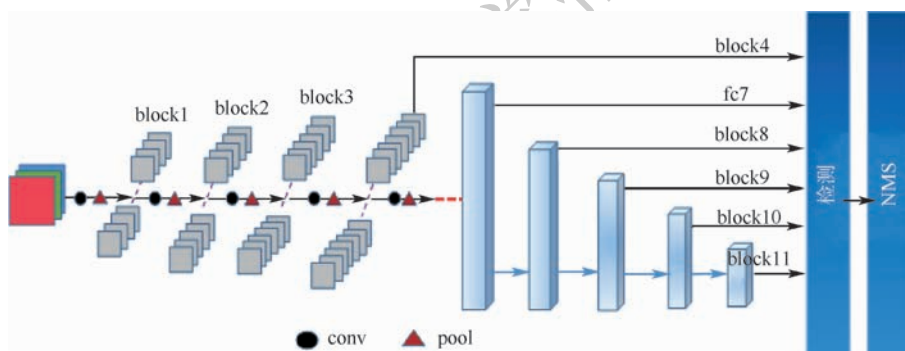


图 2 基于 VGG 前端的 SSD 网络结构

Fig. 2 Network structure of SSD based on VGG front end

框架检测出的结果经过非极大值抑制筛选,得出最终检测的位置和类别。

SSD 计算过程可表示底层与高层两部分。

底层神经网络计算过程可表示为

$$\begin{cases} X_{i'+1} = f(X_{i'}) & i' \in \text{block}'_{1,2,3,4} \\ X_{i''+1} = f_p(X_{i''}) & i'' \in \text{block}''_{1,2,3,4} \\ i' \cup i'' \in \text{block}_{1,2,3,4} \end{cases} \quad (1)$$

高层神经网络计算过程可表示为

$$\begin{cases} X_i = f(X_{i-1}) \\ D = R[f(X_i)] & i \in \text{fc}_7 \cup \text{block}_{4,8,9,10,11} \end{cases} \quad (2)$$

式中: f_p 表示卷积神经网络中的池化计算(下标 p 表示相应特征图层); f 表示卷积运算; fc_7 和 $\text{block}_{4,8,9,10,11}$ 表示回归特征图; $\text{block}''_{1,2,3,4}$ 表示底层网络中每个卷积区域的最后特征图层; $\text{block}'_{1,2,3,4}$ 表示卷积区域其他特征图和池化后的特征图; X_i 表示相应的特征图层; R 表示回归运算; D 表示分类回归与位置定位回归; $\cup$ 表示范畴中的并集和特征图在通道维度上的并列交融。

SSD 可通过各种复杂交叉的底层网络改善处理过程中容易丢失空间相关信息的问题以提高特征的多样性。然而,复杂的高层模型可以提高精度却难以保证实时的处理速度。由图 2 可见,

SSD 高层网络结构的回归特征图中层与层之间相对独立,致使多层回归计算难于统一,例如图 2 中 block_8 与 block_{10} 之间仅用单链的四层卷积层间接地通过 block_9 进行联系,各自通过回归计算出类别与位置,无法直接建立两层之间的有效联系,不利于目标检测在网络中的协调统一。为此,本文提出 FA-SSD 网络,摒弃复杂的底层与高层网络结构的改进方法,仅通过优化主流结构和增加一种轻量级的 shunt 结构增强高层网络多层回归特征图之间的协调性与统一性。

2 FA-SSD 网络结构

FA-SSD 的底层网络基于 VGG,高层网络为多层次的回归计算结构,如图 3 所示,其中红色虚线框为高层主流结构,黑色圆、红色三角形和红色菱形分别表示卷积操作(conv)、池化操作(pool)和回归特征图层之间的网络结构(nn),蓝色菱形和对立三角表示 shunt 结构和特征图层通道方向上的交融操作。图中特征图层的命名方式与图 2 相同,高层网络的回归特征图分别为 block_4 、 fc_7 、 block_8 、 block_9 、 block_{10} 、 block_{11} 。高层结构用两种方式对特征图进行降维,分别是 shunt 结构中的异步卷积分解算法和高层主流结构中的最大池

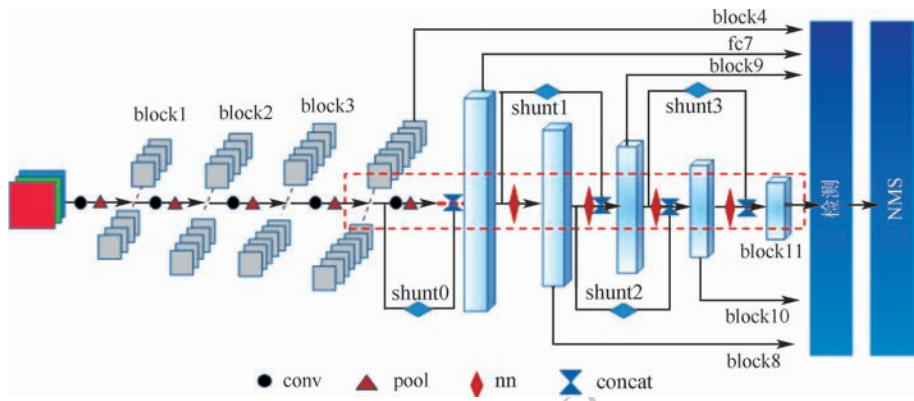


图 3 FA-SSD 网络结构

Fig. 3 Network structure of FA-SSD

化。FA-SSD 中共有 4 个 shunt 结构, 分别是 shunt0、shunt1、shunt2 和 shunt3, 用 2 种连接方式增加回归特征图层之间的联系。4 个 shunt 结构的整体连接方式如图 3 所示。最后在 6 个回归特征图上计算出类别与位置, 通过 NMS 进行一定阈值的筛选, 得出置信度最高的目标。

搭建的网络结构包含以下 3 部分内容。

1) 异步卷积分解算法

SSD 的高层结构完全应用卷积层搭建, 应用步长为 2 的一层卷积层对特征图进行降维。为提高非线性表达能力, 在使用卷积进行降维的同时, 提出一种异步卷积分解算法。该算法借鉴 Inception 模型中的卷积分解思想, 采取不同方向上不同步长的异步卷积策略组合两层异步卷积分解的卷积层, 图 4 展示了异步卷积分解算法中的一层

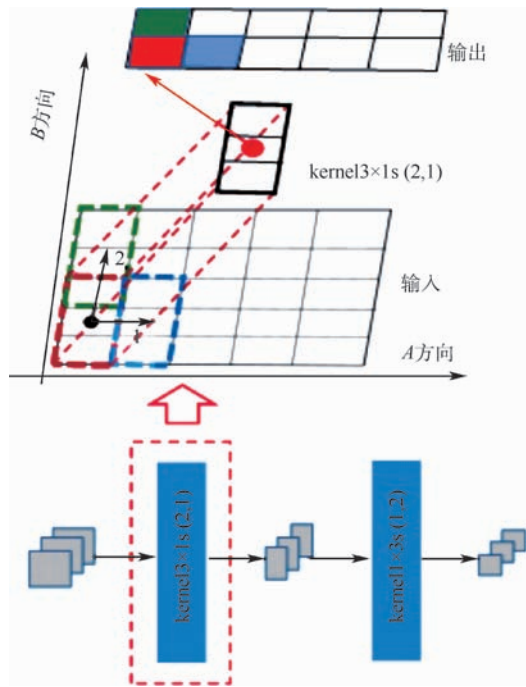


图 4 异步卷积分解算法

Fig. 4 Asynchronous convolution factorization algorithm

的操作过程, 其中 $\text{kernel}3 \times 1s(2,1)$ 表示大小为“ 3×1 ”的卷积核, 在 B 方向上大小与步长分别为 3 和 2, 在 A 方向上的大小和步长分别为 1 和 1。输入特征图在经过一层异步卷积分解的卷积计算后, 仅在 B 方向实现降维。再次经过卷积核 $\text{kernel}1 \times 3s(1,2)$ 的卷积处理后, 特征图在 A 与 B 方向分别降低了维度, 类似 $\text{pool}2 \times 2s2$ 对特征图进行处理的效果。与 SSD 高层网络中的一层 $\text{kernel}3 \times 3s(2,2)$ 卷积结构相比, $\text{kernel}3 \times 1s(2,1)$ 与 $\text{kernel}1 \times 3s(1,2)$ 的结合在未增加计算量的同时, 提高了网络的非线性表达能力。与 $\text{pool}2 \times 2s2$ 的最大池化层相比, 两层异步卷积分解的卷积计算既在卷积核大小为“3”方向上保证了像素之间的重叠, 即在步长为“2”方向上保证了感受野之间的重叠, 保留了足够的空间相关信息。

2) shunt 结构

shunt 中的网络结构即为异步卷积分解算法, shunt 连接方式如图 5 所示, 借鉴残差网络模型中的分支交融结构, 对主流结构采用 2 种方式进行连接。block4 与 fc7 之间存在每层含有 512 个神经元的三层卷积网络 and 一层包含 1024 个神经元的 fc6, 网络层数和神经元个数大于其他回归特征图层之间的网络。为节省计算量、提高回归特征图间的联系, shunt0 起始于回归特征图层 block4, 直接与回归特征图层 block fc7 在通道维度进行并联交融(后续交融或连接均为通道维度的 concat 交融), 与其他 shunt 结构无交错跨越的联系。而 shunt1 与 shunt2 之间以及 shunt2 与 shunt3 之间存在交错连接, 例如 shunt1 的交融端位于 shunt2 的分流(与主流相同的征信息)之后, shunt2 则可跨越 block9 直接连接 block8 与 block10。如果 shunt2 起始端位于 shunt1 交融位置之后, 则会因为特征交融导致特征图通道的倍增, 而增加计算量。在通道维度上进行并联交

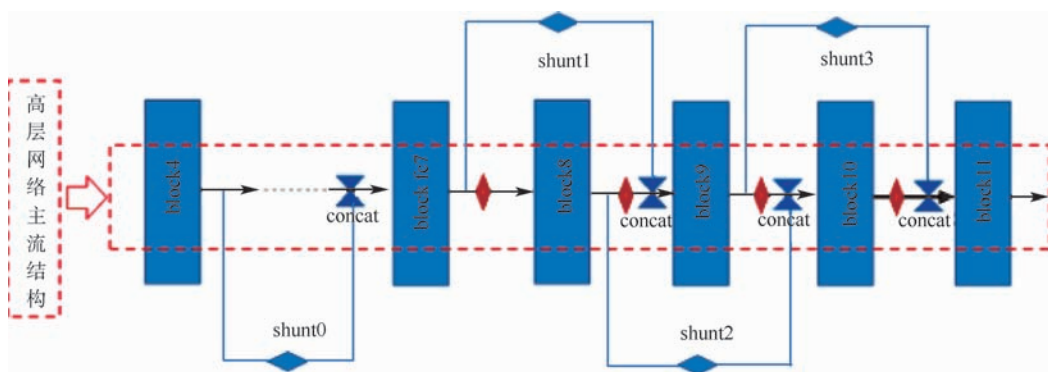


图 5 回归特征图层间的 shunt 结构

Fig. 5 Shunt structure between layers of regressive feature image

融之后,特征图层经过批归一化^[20]处理及回归计算得出类别与位置信息。通过上述 2 种 shunt 连接方式增强高层网络中多层次的回归计算之间的联系,即增强协调性和统一性。

3) 高层网络结构的优化及计算过程

图 6 展示了优化后的高层网络结构中的局部结构,其他局部结构与此类似。实验搭建了多种网络结构并进行了训练及测试,对比实验结果证明图 6 所示的结构取得了最好的测试效果。该结构主要由池化层 $\text{pool}2 \times 2s(2,2)$ 和 shunt 结构组成,池化层和 shunt 结构均可实现特征图的维度降低。其中 $\text{pool}2 \times 2s(2,2)$ 表示大小为“ 2×2 ”,2 个方向步长为“2”的池化, $\text{kernel}3 \times 3s(1,1)$ 表示大小为“ 3×3 ”,2 个方向步长均为 1 的卷积核, $\text{kernel}1 \times 1s(1,1)$ 表示大小为“ 1×1 ”,2 个方向步长均为 1 的卷积核。

高层网络结构的优化后,6 层回归计算过程可分为 3 个部分,即

$$D_1 = R(\text{block}_4) \cup R(\text{block}_8) \quad (3)$$

$$D_2 = R(\text{fc}_7) \cup R(\text{block}_4) \quad (4)$$

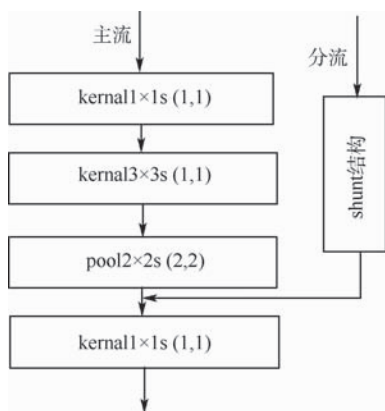


图 6 高层网络局部结构

Fig. 6 Local structure of high-level network

$$\begin{cases} X_{i+2} = f_p(X_{i+1}) \cup f_s(X_i) \\ D_3 = R[f(X_{i+2})] \quad i \in \text{fc}_7 \cup \text{block}_{8,9} \end{cases} \quad (5)$$

式中: f_p 与 f_s 分别表示最大池化和异步卷积分解 2 种降维方式。

多层次的检测结果 D 为 D_1 、 D_2 和 D_3 的并集,即

$$D = D_1 \cup D_2 \cup D_3$$

3 实验与分析

实验所用训练数据集为 VOC2007trainval 与 VOC2012trainval^[4],检测数据集为 VOC2007test^[4]。实验软件配置为 Window10、TensorFlow1.7.0、TensorFlow Layers API、CUDA9.0,硬件配置为 NVIDIA GeForce GTX 1080Ti (一块)、Intel(R) Xeon(R) CPU E5-2609v4 @ 1.70 GHz $\times$ 16。实验检测的基准是算法的百分制(%)平均精度均值 m_AP 和检测速度 v 。

FA-SSD 中参与训练和检测的图像大小为 300 像素 $\times$ 300 像素,因此下文称 FA-SSD300。回归特征图的大小分别是“ 38×38 ”、“ 19×19 ”、“ 10×10 ”、“ 5×5 ”、“ 3×3 ”和“ 1×1 ”,在上述 6 种回归特征图采用锚箱分别生成建议框,具体建议框参数设置和分类定位阈值设定与 SSD 算法相应设置相同,参与对比分析的 SSD300 检测结果如可见对比实验部分内容。

采用批次为 16,共训练 150 000 步。学习率初始设置为 0.01,采用分阶段控制方式间接调整目标函数优化器。衰减步长边界设置为:“20 000、900 000、130 000”,学习率衰减设置为:“1、0.1、0.005、0.001”。学习率终止边界设为 0.000 01,直至收敛。

为增加所用训练样本的价值,在图像预处理阶段对样本进行随机地裁剪,颜色扰动、翻转等数

据增广方式。颜色扰动过程包括图片颜色的“亮度、对比度、饱和度、色相”4 种调整操作,并且随机改变这 4 种图像颜色操作的前后顺序,最后统一压缩为“300 像素×300 像素”的图像进行网络训练。增广产生的数据实例如图 7 所示。

实验对 FA-SSD 的训练网络初试参数设定采用 VGG 参数迁移和参数初始化 2 种方法。其中,迁移 VGG 的参数是经过 ImageNet^[21] 数据训练后获得的权重。训练过程中的损失变化如图 8 所示,其中 loss1 与 loss2 分别是初始化参数和迁移 VGG 参数下的损失。由图 8 可看出,迁移 VGG 参数的训练损失 loss2 收敛速度最快。但 2 种参数情况下的训练最终收敛在近似损失水平。

在迁移 VGG 参数的训练下,检测物体的平均精度均值的变化如图 9 所示。由图 9 可见,除 30 000 步的检测结果外,训练 120 000 步之前平均精度均逐渐提升。在 120 000 步至终止训练阶段,平均精度均值出现振荡,分别于 120 000 步和 140 000 步达到高峰,因此最终检测的平均精度均值为 80.5%。



图 7 数据增广
Fig. 7 Data augmentation

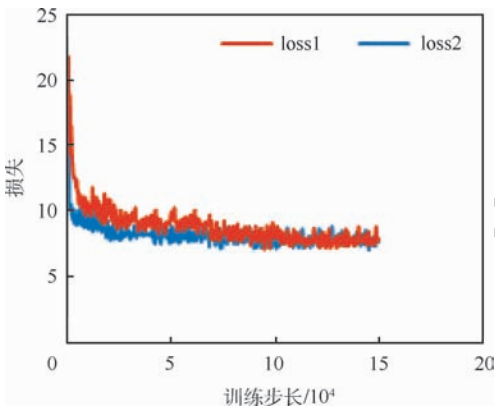


图 8 损失变化曲线
Fig. 8 Loss variation curves

3.1 基于异步卷积分解算法的 shunt 结构对实验结果的影响

FA-SSD 中高层网络共使用了 4 个 shunt 结构,即 shunt0、shunt1、shunt2 和 shunt3,如图 5 所

示。为验证不同的 shunt 连接方式对网络检测结果的影响,实验通过调整 shunt 结构个数,分别搭建了 shunt0000、shunt1000、shunt0101、shunt1110、shunt1111 进行相同的训练与检测。其中,末尾数字串的前后数字顺序表示 shunt0 至 shunt3 结构的位置顺序。“0”表示去除 FA-SSD 中对应位置的 shunt 结构,“1”表示保留所在位置的 shunt 结构,例如 Shunt0101 表示图 6 中仅有 shunt1 与 shunt3 结构的 FA-SSD 网络结构。

不同数目 shunt 结构的检测精度如图 10 所示。由图 10 可见,仅加深高层主流网络深度后,shunt0000 网络检测平均精度均值为 77.8%,相对 SSD 提高 0.6%,说明增加回归特征图之间的卷积层数可提高检测精度。此后,每增加一个 shunt 结构,网络在保证实时性的同时,进一步提高了平均检测精度。最后 shunt1111 (FA-SSD300) 网络的检测平均精度均值达到最高的

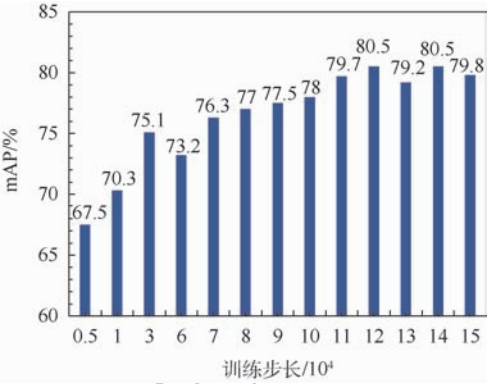


图 9 不同迭代次数下的平均精度均值
Fig. 9 Mean average precision under different numbers of iteration

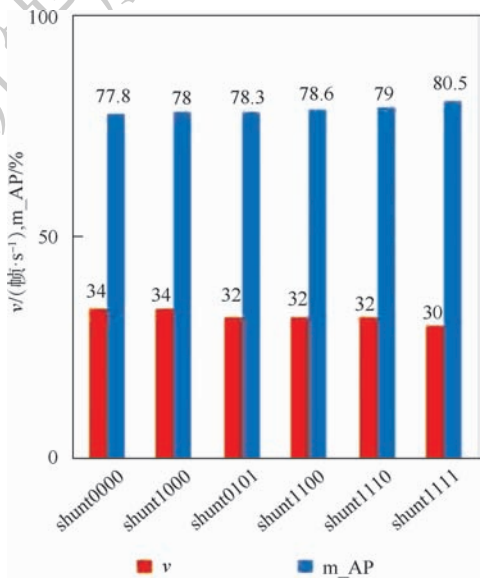


图 10 不同数目的 shunt 结构对检测的影响
Fig. 10 Influence of different numbers of shunt structure on detection

80.5%,相对 SSD 提高了 3.3%。由平均精度均值提高的幅度可推断,仅增加高层主流网络的卷积层数是提高平均精度均值的次要因素,交错连接的 shunt 结构是检测精度提高的主要原因。在增加了 shunt 结构后,FA-SSD300 平均精度均值达到了最高的80.5%,同时保证了检测的实时性。

3.2 高层网络优化对实验的影响

为优化异步卷积分解与 shunt 结构的融合,进一步提高特征的多样性,本文搭建了如图 11 所示的 3 种高层局部网络结构(shunt 并联网络结构)。图 11(a)表示 SSD 原有的 $\text{kernel}3 \times 3s(2,2)$ 卷积层与 shunt 结构的结合,图 11(b)与(c)分别表示异步卷积分解结构、最大池化层和 shunt 结构的搭配。图 11(a)、(b)和(c)分别对应的网络结构为 FA-SSD0、FA-SSD1、FA-SSD2。FA-SSD1 与 FA-SSD2 中的主流结构中的不同部分,在图 11

中标记为红色。为保证 FA-SSD1 与 FS-SSD2 的主流结构深度相同(卷积层数与池化层数),结构(c)相比(b)增加一层 $\text{kernel}1 \times 1s(1,1)$ 的卷积层。在相同条件下对上述 3 种网络进行训练后,在 VOC2007test 上的检测结果如图 12 所示。FA-SSD0 的平均精度均值为 78.3%,相对 SSD 提升 1.1%,检测速度虽然相对 SSD 降低 11 帧/s,但依然能保证实时检测。FASSD1 采用异步卷积分解结构对主流特征图进行降维,平均精度均值达到 80%,相比 SSD 提高 2.8%。主流结构采用最大池化降维时,目标检测的精度与速度都有进一步的提升。FA-SSD2 相比 FA-SSD1 在平均精度均值和检测速度上都提高 0.5%。FA-SSD1 与 FA-SSD2 在结构上的主要区别是主流结构上的降维计算方式。FA-SSD1 采用一种异步卷积分解结构同时进行特征提取与降维,而 FA-SSD2 采用异步

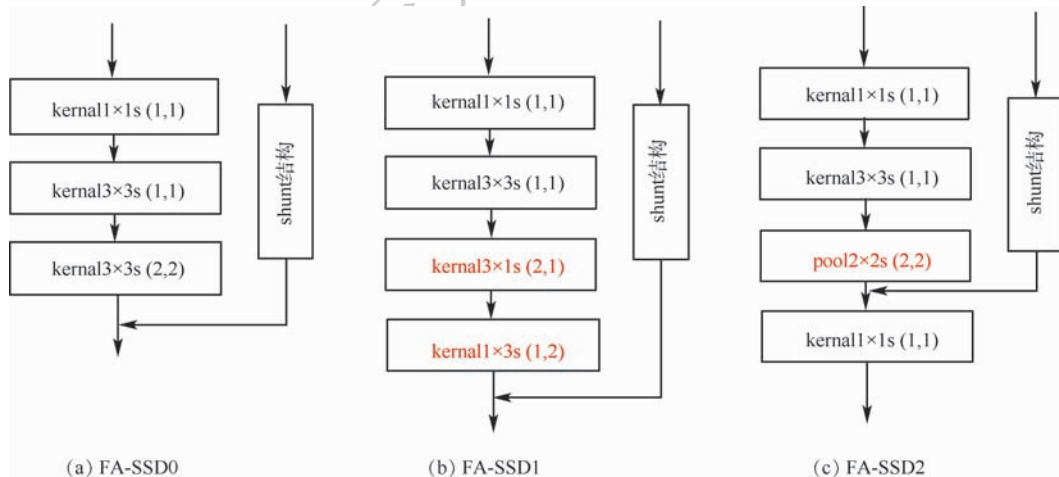


图 11 高层网络的优化

Fig. 11 Optimization of high-level network

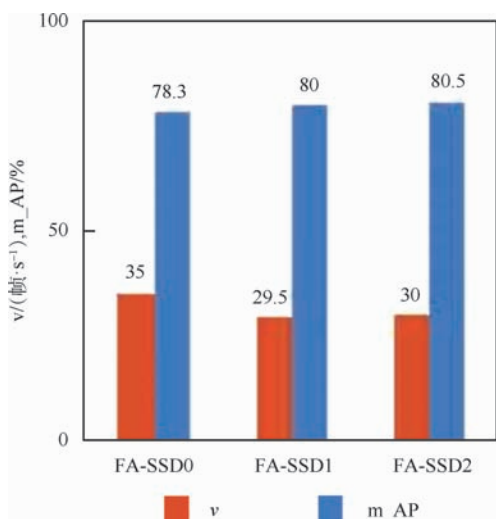


图 12 高层结构优化对检测的影响

Fig. 12 Influence of high-level structure optimization on detection

卷积分解和池化 2 种不同的降维计算方式,提高特征的多样性。由于池化计算相对异步卷积分解结构更简单,FA-SSD2 相对 FA-SSD1 提高了检测速度。异步卷积分解相对池化更有利于传递空间相关信息,同时 2 种降维方式提高了特征的多样性,因此 FA-SSD2 检测精度达到最高的 80.5%。

3.3 对比实验

实验通过端到端的方法训练 FA-SSD300,与近期算法在 VOC2007 和 VOC2012(07+12)中训练,在 VOC2007 上进行了对比,结果如表 1 所示, Faster RCNN 与 R-FCN 属于两阶段分类回归算法,虽然训练用到分辨率最大的图像,但平均精度均值仍然最低,且不具备实时检测能力。先后出现的单阶段回归系列算法 YOLOv2、SSD300、DSOD300 以及 DSSD321 的检测平均精度均值逐步提升。其中, SSD 类算法 DSOD300、DSSD321

虽然较前几种算法取得了更高的检测精度,但不能保证实时性,检测速度分别为 17.4 帧/s 和 9.5 帧/s。FA-SSD300 检测平均精度均值最高,达到 80.5%,且保证了实时检测。为公平对比检测速度,采用与本实验相同的硬件,对 SSD 的检测速度进行测试,检测结果为表 1 中的 SSD300* 所示数据。

8 种算法在 VOC2007test 上针对 20 个常见类别的具体检测结果对比如表 2 所示。其中,最高平均精度均值和每个类最高平均精度已标记为黑色粗体数字。

表 2 中左列 5 种算法为两阶段分类回归系列算法,右列 3 种算法为单阶段回归算法,2 类算法各自平均精度最高的网络分别是 R-FCN 与 FA-

SSD300。虽然与 FA-SSD300 同样达到 80.5% 的平均精度均值,但是 R-FCN 采用信息提取能力更强的复杂网络结构 ResNet-101^[18],处理速度仅为 7 帧/s。除平均检测精度达到最高外,FA-SSD 在 20 类检测中共有 11 类达到最高平均精度,其中“Bus”与“Cat”2 个类的平均检测精度均超过 90%,分别为 90.2% 和 91.7%。

FA-SSD300 在 VOC2007test 上的部分结果如图 13 所示,对难以检测的密集小目标、遮挡目标和局部大目标均取得了较好的检测结果。图 13(e)由于像素分辨率低造成一个小目标漏检,同时(b)和(g)由于目标遮挡过于严重造成次要目标漏检,其他所有位置定位框架均在合理范围内。

表 1 不同算法在 VOC2007test 上的检测结果
Table 1 Detection results of different algorithms on VOC2007test

算法	训练数据	预训练	底层网络	图片大小	建议框数	显卡	速度/(帧·s ⁻¹)	m_AP/%
Fast R-CNN ^[8]	07 + 12	√	VGGNet	600 × 1 000 *	300	K40	3.125	66.9
Faster R-CNN ^[9]	07 + 12	√	VGGNet	600 × 1 000 *	300	K40	5	73.2
R-FCN ^[22]	07 + 12	√	VGGNet	600 × 1 000	300	K40	5.8	75.6
YOLOv2 ^[12]	07 + 12	√	Darknet-19	352 × 352		Titan X	81	73.7
SSD300 ^[13]	07 + 12	×	VGGNet	300 × 300	8 732	Titan X	46	74.3
SSD300 ^[13]	07 + 12	√	VGGNet	300 × 300	8 732	Titan X	46	77.2
SSD300 * ^[13]	07 + 12	×	VGGNet	300 × 300	8 732	1080Ti	43.5	74
DSOD300 ^[16]	07 + 12	×	DS/64-192-48-1	300 × 300	8 732	Titan X	17.4	77.7
DSSD321 ^[14]	07 + 12	√	ResNet	321 × 321	17 080	Titan X	9.5	78.6
FA-SSD300	07 + 12	×	VGGNet	300 × 300	8 732	1080Ti	30	79.0
FA-SSD300	07 + 12	√	VGGNet	300 × 300	8 732	1080Ti	30	80.5

表 2 针对 VOC2007test 具体类别的检测对比
Table 2 Comparison of specific category detections on VOC2007test

类别	Fast R-CNN ^[8]	Faster R-CNN ^[9]	ION ^[22]	R-FCN ^[23]	MR-CNN ^[24]	SSD300 ^[13]	DSSD321 ^[14]	FA-SSD300
Aero	77.0	76.5	79.2	79.9	80.3	79.5	81.9	86.4
Bike	78.1	79	83.1	87.2	84.1	83.9	84.9	85.9
Bird	69.3	70.9	77.6	81.5	78.5	76	80.5	79.6
Boat	59.4	65.5	65.6	72	70.8	69.6	68.4	73.3
Bottle	38.3	52.1	54.9	69.8	68.5	50.4	53.9	53.6
Bus	81.6	83.1	85.4	86.8	88	87	85.6	90.2
Car	78.6	84.7	85.1	88.5	85.9	85.7	86.2	89.2
Cat	86.7	86.4	87	89.8	87.8	88.1	88.9	91.7
Chair	42.8	52	54.4	67	60.3	60.3	61.1	60.0
Cow	78.8	81.9	80.6	88.1	85.2	81.5	83.5	84.3
Table	68.9	65.7	73.8	74.5	73.7	77	78.7	80.9
Dog	84.7	84.8	85.3	89.8	87.2	86.1	86.7	89.1
Horse	82.0	84.6	82.2	90.6	86.5	87.5	88.7	87.4
Mbike	76.6	77.5	82.2	79.9	85	83.97	86.7	86.5
Person	69.9	76.7	74.4	81.2	76.4	79.4	79.7	83.3
Plant	31.8	38.8	47.1	53.7	48.5	52.3	51.7	54.2
Sheep	70.1	73.6	75.8	81.8	76.3	77.9	78	83.2
Sofa	74.8	73.9	72.7	81.5	75.5	79.4	80.9	82.3
Train	80.4	83	84.2	85.9	85	87.6	87.9	89.2
Tv	70.4	72.6	80.4	79.9	81	76.8	79.4	78.5
mAP/%	70.0	73.2	75.6	80.5	78.2	77.2	78.6	80.5

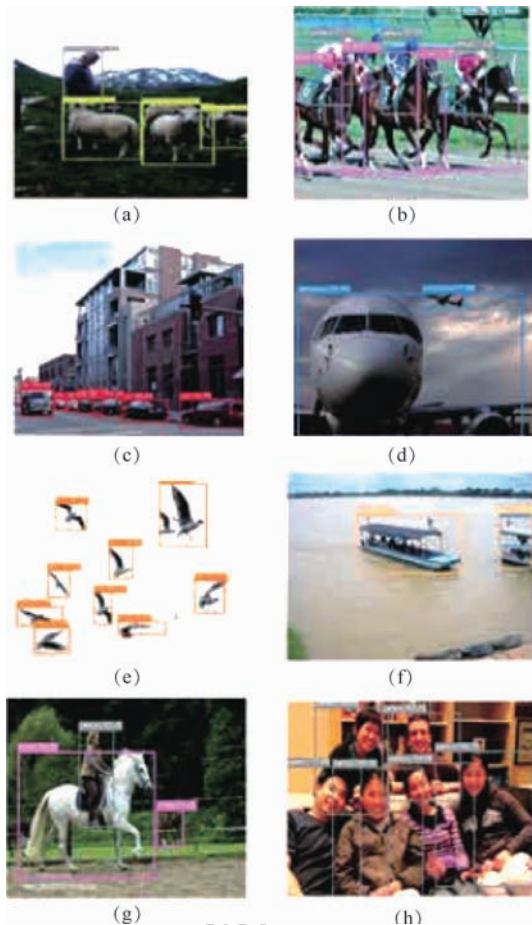


图 13 FA-SSD300 在 VOC2007test 上的部分检测结果

Fig. 13 Partial detection results of FA-SSD300 on VOC2007test

4 结 论

本文基于 SSD 提出一种基于异步卷积分解与 shunt 结构的单阶段目标检测器。采用基于异步卷积分解的 shunt 结构,在没有进行主流结构优化前,检测的平均精度均值便可达到 80.0%,在不增加计算量的同时,提升了提取特征的非线性表达能力,shunt 结构与主流结构的交融提高了回归特征图之间的协调性与统一性。构造了包含池化与异步卷积分解的局部结构,优化原有高层主流网络结构,异步卷积分解与池化两种降维方式的结合,既有利于空间相关信息的传递,也有利于提高特征的多样性。增加 shunt 结构并优化主流结构后,平均精度均值达到 80.5%,相比 SSD 提高 3.3%,比 DSSD321 提高了 1.9%,同时在一块 1080ti 显卡上取得 30 帧/s 的平均处理速度。

FA-SSD 在保证实时性检测的同时,其平均精度在未采用 MS COCO 数据库扩充训练数据的情况下,便超过近期提出的 SSD 类算法。为适用于移动终端的相关应用,下一步研究工作将侧重于

进一步降低网络规模优化网络结构。

参考文献 (References)

- [1] 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, NJ: IEEE Press, 2003: 511-518.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2005: 886-893.
- [3] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2008: 1-8.
- [4] EVERINGHAM M, GOOL L V, WILLIAMS C K I, et al. The pascal, visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [5] 李旭冬, 叶茂, 李涛. 基于卷积神经网络的目标检测研究综述[J]. 计算机应用研究, 2017, 34(10): 2881-2886.
LI X D, YE M, LI T. Review of object detection based on convolutional neural networks[J]. Application Research of Computers, 2017, 34(10): 2881-2886 (in Chinese).
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2014: 580-587.
- [7] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 346-361.
- [8] GIRSHICK R. Fast R-CNN[C] // IEEE International Conference on Computer Vision. Piscataway, NJ: 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[C] // International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.
- [10] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 936-944.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 779-788.
- [12] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 6517-6525.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C] // European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [14] REDMON J, FARHADI A. YOLOv3: An incremental improve-

- ment[EB/OL]. (2018-04-08) [2018-09-21]. <http://cn.arxiv.org/pdf/1804.02767v1>.
- [15] FU C Y, LIU W, RANGA A, et al. DSSD: Deconvolutional single shot detector [EB/OL]. (2017-01-23) [2018-09-21]. <http://cn.arxiv.org/pdf/1701.06659>.
- [16] SHEN Z, LIU Z, LI J, et al. DSOD: Learning deeply supervised object detectors from scratch [C] // IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 1937-1945.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-03-10) [2018-09-21]. <http://cn.arxiv.org/pdf/1409.1556>.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2015: 770-778.
- [19] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 2818-2826.
- [20] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02) [2018-09-26]. <https://arxiv.org/abs/1502.03167>.
- [21] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [22] BELL S, ZITNICK C L, BALA K, et al. Inside-outside Net: Detecting objects in context with skip pooling and recurrent neural networks [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 2874-2883.
- [23] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks [EB/OL]. (2016-06-21) [2018-09-26]. <https://arxiv.org/abs/1605.06409>.
- [24] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C] // IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 1-13.

作者简介:

赵蓬辉 男, 硕士研究生。主要研究方向: 视频目标检测、深度学习。

孟春宁 男, 博士, 副教授。主要研究方向: 图像处理、人工智能、信息安全。

常胜江 男, 博士, 教授, 博士生导师。主要研究方向: 数字图像处理、太赫兹功能器件、深度学习。

Single shot multibox detector based on asynchronous convolution factorization and shunt structure

ZHAO Penghui¹, MENG Chunning^{2,*}, CHANG Shengjiang¹

(1. Institute of Modern Optics, College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China;

2. Department of Electronic Technology, China Coast Guard Academy, Ningbo 315801, China)

Abstract: Single shot multibox detector (SSD) owns the relatively independent regression computations of multi-regressive feature maps, while the object detection algorithms based on SSD cannot make a tradeoff between detection accuracy and real-time speed. To solve the problems above, a single shot mutibox detector based on asynchronous convolution factorization and shunt structure (FA-SSD) is introduced based on asynchronous convolution factorization algorithm and shunt structure. The shunt structure, based on the proposed asynchronous convolution factorization algorithm, is designed to staggerly connect the layers of regression features, enhancing the unity and coordination between regression calculations. In order to optimize the mainstream of high-level structure, the asynchronous convolution factorization algorithm and max pooling are implemented to reduce the dimension of image features in the mainstream and shunt respectively, which can hold the spatial information while improving the diversity of features. According to the experimental results from VOC2007test, FA-SSD achieves a mean average precision of 80.5% after the training of VOC2007trainval and VOC2012trainval with nominal resolution of 300×300 , while the detection speed exceeds 30 frames per second.

Keywords: object detection; convolutional neural networks; asynchronous convolution factorization; shunt structure; structure optimization

Received: 2018-09-27; **Accepted:** 2019-05-18; **Published online:** 2019-06-19 08:36

URL: kns.cnki.net/kcms/detail/11.2625.V.20190617.0931.002.html

Foundation item: Technology Research Project of Public Security Ministry (2017JSYJC10)

* **Corresponding author.** E-mail: mengchunning123@163.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0014

自研激光雷达三维点云配准技术



呼延嘉玥, 徐立军, 李小路*

(北京航空航天大学 仪器科学与光电工程学院, 北京 100083)

摘 要: 为了实现激光雷达点云与图像重建点云的三维空间配准, 基于自研三维扫描激光雷达系统, 提出了新型的快速多尺度因子(FMSR)点云配准算法, 研究了空间点云配准技术。该算法主要包括初始配准和精确配准2个步骤: 初始配准使用基于尺度自适应关键点质量(ASKQ)的点云特征提取算法, 提取关键点的特征匹配对, 求解点云配准初始参数; 精确配准利用K-邻近(KNN)算法全局搜索, 提升计算效率, 多次迭代得到2组点云之间的最优旋转矩阵、最优平移向量和最优尺度因子。仿真和实验结果表明, 所提出的算法对空间目标(尺寸为20.30 m × 7.85 m × 26.56 m)实现空间点云配准, 配准精度达到0.194 m, 运行时间为16.207 s; 与多尺度迭代最近点(S-ICP)算法相比, 配准精度提高了0.131 m, 运行时间提高了30%。所提出的空间点云配准技术可为场景重建和纹理匹配提供算法基础。

关键词: 激光雷达; 图像重建; 空间点云配准; 尺度自适应关键点质量(ASKQ)算法; 快速多尺度因子(FMSR)点云配准算法

中图分类号: TN958.98; V11

文献标识码: A

文章编号: 1001-5965(2019)10-2099-09

近年来, 激光雷达三维扫描技术发展迅速, 是获取空间三维信息最重要的方法之一。因为其便捷快速、测量范围广、精度高的特点, 被广泛应用于数字城市、地形测绘、文物保护、自然灾害监测等领域^[1-2]。传统获取物体表面三维信息的方法, 主要包括基于激光雷达的三维扫描方法^[3]和基于图像的三维重建方法^[4]。

基于TOF(Time of Flight)三维扫描激光雷达, 计算光的飞行时间获取距离信息, 可以解算物体的空间三维坐标。这种方法测距精度高、工作距离远、操作简单, 但无法为重建场景提供纹理信息, 且存在位置限制和遮挡等问题, 通常难以获取完整表面^[5]。基于图像数据三维重建方法, 获取某一物体多视角下的图像, 利用三角测量原理计算图像像素间视差, 从而解算物体的三维几何信

息。这种方法虽然能够获取丰富的纹理信息, 但重建点数量少且精度低, 当图像质量不佳时, 重建效果较差^[6]。激光雷达与图像相结合的三维重建方法兼具两者优势, 因此受到国内外学者关注, 已在纹理映射、导航定位等领域得到广泛应用。实现激光雷达点云与图像重建点云的空间配准技术, 则是三维重建技术研究的关键。

空间点云配准技术分为初始配准和精确配准2个步骤, 初始配准能够初步减小两组点云的旋转和平移错位, 精确配准能够使两组点云之间的配准误差最小^[7]。目前, 最常用的点云配准方法是由Besl和Mckay^[8]在1992年提出的迭代的最近点(Iterative Closest Point, ICP)算法。该算法针对目标数据集中的每个目标点, 寻找参考数据集中与其距离最小的对应参考点, 通过迭代算法估

收稿日期: 2019-01-16; 录用日期: 2019-03-29; 网络出版时间: 2019-04-11 10:16

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190409.1104.003.html

基金项目: 国家重点研发计划(2018YFB0504500); 国家自然科学基金(61671038, 61721091)

*通信作者: E-mail: xiaoluli@buaa.edu.cn

引用格式: 呼延嘉玥, 徐立军, 李小路. 自研激光雷达三维点云配准技术[J]. 北京航空航天大学学报, 2019, 45(10): 2099-2107.
HUYAN J Y, XU L J, LI X L. Three-dimensional point cloud registration technique for self-designed LiDAR[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2099-2107 (in Chinese).

计旋转矩阵、平移向量。然而,该算法存在计算量大、效率低、对初始位置要求高等不足,且忽视了 2 组点云之间的尺度因子^[9]。Chen 和 Medioni^[10]提出点对面搜索最近点的配准方法,降低原始 ICP 算法对初始位置的要求,加快算法收敛速度。Rusinkiewicz 和 Levoy^[11]详细分析 ICP 算法各个步骤,并提出相应的改进算法。戴静兰等^[12]提出基于曲率特征的改进 ICP 算法,邹际祥^[13]提出基于 K-D 树 (K-Dimensional tree) 搜索最近点的方法,加快了最邻近点的查找速度。

本文研究了一种改进的快速多尺度因子 (Fast Multi-Scale Registration, FMSR) 点云配准方法,该方法引入激光雷达点云和图像重建点云之间的尺度因子,提升空间点云配准技术的配准精度和计算效率,并将该方法最终应用于自研激光雷达系统所得点云数据。初始配准使用基于尺度自适应关键点质量 (Adaptive Scale Key Point Quality, ASKQ) 的点云特征提取算法^[14]提取出关键点,实现关键点的特征提取与匹配,利用特征匹配计算初始参数,提升初始配准的配准精度。精确配准采用 K-邻近 (K-Nearest Neighbors, KNN) 算法搜索相邻近点,无需对点云进行网格划分,减少算法的运行时间,实现快速、准确的配准。

1 三维点云数据获取系统

三维点云数据获取系统由激光雷达系统和图像采集系统构成,如图 1 所示。自研三维扫描激光雷达系统包括 4 个组成单元:激光发射与接收单元、扫描单元、控制单元和测距单元。激光发射与接收单元由激光器、望远镜、光电探测器组成,设计同轴光路采集激光发射脉冲、回波脉冲和触发信号。扫描单元用于实现三维扫描功能,垂直方向 360°扫描采用步进电机和 45°扫描转镜实现,水平方向 360°扫描采用高精度数位云台实现。控制单元的核心部件为现场可编程门阵列 (FPGA),用于设置激光器参数、程控增益控制、步进电机控制和数据采集等。测距单元包括自动

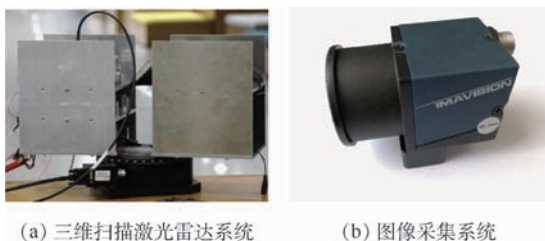


图 1 三维点云数据获取系统

Fig. 1 Three-dimensional point cloud data acquisition system

增益控制电路和时刻鉴别电路。系统具体设计及描述参见笔者团队前期发表的论文^[15-18]。

图像采集系统采用摄像机 (型号 MER-1070-14U3C) 拍摄多视点下的图像。相机镜头 (型号 M7528-MP2) 焦距为 75 mm,工作距离为 0.3 m 到无穷远处。该图像采集系统具有 1070 万像素分辨率,图像数据通过 USB3.0 数据接口传输。

2 基于三维点云的空间配准

基于激光雷达点云与图像重建点云的空间点云配准技术,主要包括点云获取、点云处理和点云融合 3 个步骤,本文实现点云配准的具体流程如图 2 所示。首先,激光雷达系统与图像采集系统采集激光雷达点云和二维图像数据,并利用 SFM (Structure from Motion) 技术从图像中获取三维重建点云。其次,利用 ASKQ 算法选取关键点,提取关键点特征与匹配。然后,利用提取出的匹配对求解初始配准参数,利用基于快速多尺度因子的三维点云配准算法实现精确配准。最后,进行数值仿真和实验,分析验证算法有效性,评价配准精度。

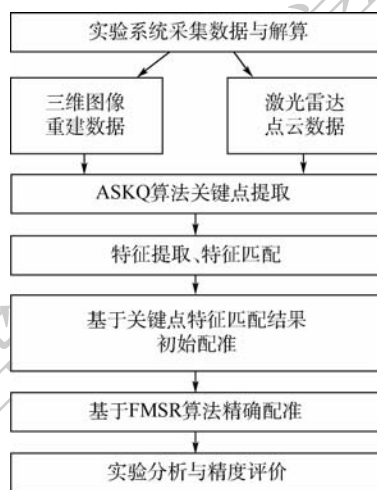


图 2 点云配准流程图

Fig. 2 Flowchart of point cloud registration

2.1 两种点云数据获取

点云数据获取过程包括 2 个部分:基于三维扫描激光雷达的三维坐标解算和基于图像的三维坐标重建。

为了获取三维扫描激光雷达点云,笔者团队设计了点云实时获取与显示的上位机软件,该软件通过 USB 接口实时获取测距和扫描模块相关数据,计算每束激光脉冲对应的水平方向扫描角、垂直方向扫描角和激光脚点的距离信息,最后利用激光雷达方程解算出激光脚点的三维坐标。

为了获取图像三维重建点云,首先利用相机拍摄同一物体在不同视角的二维图像。然后基于 SIFT (Scale-Invariant Feature Transform) 算法提取出图像特征,并基于 RANSAC (RANdom SAMple Consensus) 算法实现特征匹配。最后利用 SFM 技术获取三维重建点云。

2.2 基于 ASKQ 的点云特征提取算法

针对已获取的 2 组点云,本文采用基于 ASKQ 的点云特征提取算法提取关键点,评估关键点质量并且排序,从中选取质量最优关键点。该方法的主要步骤如下:

首先,对 2 组点云中每个采样点 $\mathbf{v}$,选取以 $\mathbf{v}$ 的坐标为球心, r 为半径的球体内的全部点云信息,并由局部表面矩阵 $\mathbf{L}_j = [x_j, y_j, z_j]^T$ ($j = 1, 2, \dots, m$) 表示。

其次,初步选取关键点。利用霍特林变换在局部表面建立局部坐标系,得到 2 个主轴方向。将每个局部表面矩阵 $\mathbf{L}$ 变换到局部坐标系中,使其与主轴方向对齐,并计算出物理量主轴比 δ 为

$$\delta = \frac{\max(\mathbf{X}) - \min(\mathbf{X})}{\max(\mathbf{Y}) - \min(\mathbf{Y})} \quad (1)$$

式中: $\mathbf{X}, \mathbf{Y}$ 分别为变换后局部表面矩阵的 x 分量和 y 分量。如果 $\delta = 1$,则该局部表面为普通对称表面;如果 $\delta > 1$,则为非对称表面。根据采样点 $\mathbf{v}$ 的不同半径 r_k 计算主轴比 δ_k ,当主轴比 δ_k 中存在极大值,对应的采样点 $\mathbf{v}$ 视为关键点,极大值对应半径 r_k 为最佳尺度。

最后,评估提取出的关键点质量并排序,从中选取最优关键点。关键点质量取决于关键点附近局部表面的主曲率。基于变换后的局部表面矩阵所含数据,进行曲面拟合,得到拟合曲面 $\mathbf{C}$,并对曲面 $\mathbf{C}$ 进行 $n \times n$ 的采样,计算关键点质量 Q_k 为

$$Q_k = \frac{1000}{n^2} \sum |K| + \max(100K) + |\min(100K)| + \max(10\kappa_1) + |\min(10\kappa_2)| \quad (2)$$

式中: $K = \kappa_1 \kappa_2$ 为高斯曲率, κ_1 和 κ_2 为拟合曲面 $\mathbf{C}$ 上每点的主曲率。

将关键点根据关键点质量 Q_k 由大到小排序。由于部分距离相近的关键点容易造成数据信息的重复,因此设置最小距离阈值,消除最小距离范围内的其他关键点,仅保留最优关键点。

2.3 关键点特征提取与匹配

由拟合曲面 $\mathbf{C}$ 采样提取的特征向量具有旋转不变性和尺度不变性,本文选择该特征向量作为形状特征。形状特征的向量维数通常较高,存

在占用大量内存、耗费大量时间等问题,因此使用主成分分析 (Principal Component Analysis, PCA) 方法对形状特征进行降维,将形状特征投影到 PCA 子空间中。

利用降维后的 2 组形状特征实现关键点特征匹配。其中,由激光雷达点云获取形状特征为 $\mathbf{f}_p$,由图像重建点云获取形状特征为 $\mathbf{f}_q$,两组形状特征之间的误差 e_r 为

$$e_r = \arccos(\mathbf{f}_p(\mathbf{f}_q)^T) \quad (3)$$

式中: e_r 在 $0 \sim \pi/2$ 之间分布,0 代表两组形状特征向量完全相同, $\pi/2$ 代表两组形状特征向量完全不同或正交。当 e_r 取得最小值时,则为最佳匹配。

为了实现多尺度点云之间的精确配准,本文对 ICP 算法引入尺度因子,研究基于快速多尺度因子的三维点云配准算法。该算法选取目标点云与参考点云之间的误差和作为目标函数,最小化目标函数,求解 2 组点云之间最优尺度因子 S 、最优旋转矩阵 $\mathbf{R}$ 和最优平移向量 $\mathbf{T}$ 。目标函数为

$$e(S, \mathbf{R}, \mathbf{T}) = \sum_{k=1}^N \|\mathbf{S}\mathbf{R}\mathbf{x}_k + \mathbf{T} - \mathbf{y}_k\|^2 \quad (4)$$

式中: $\mathbf{x}_k$ 为目标点云; N 为目标点云数量; $\mathbf{y}_k$ 为参考点云中与目标点云 $\mathbf{x}_k$ 相对应的点。为了保证目标点云中每个点都能在参考点云搜索到对应点,通常选取数量少、精度低的图像重建点云为目标点云,选取数量多、精度高的激光雷达点云为参考点云。

利用 2.2 节和本节所述的 ASKQ 算法获取关键点的特征匹配对,计算初始尺度因子 S_0 、初始旋转矩阵 $\mathbf{R}_0$ 和初始平移向量 $\mathbf{T}_0$ [19]。当确定初始值后,利用多次迭代实现点云的精确配准。迭代过程如下:

1) 对目标点云进行尺度因子为 S_0 、旋转矩阵为 $\mathbf{R}_0$ 和平移向量为 $\mathbf{T}_0$ 的初始变换,得到变换后的目标点云为 $\mathbf{x}_{\text{trans}}$ 。

2) 利用 KNN 算法全局搜索,在参考点云中搜索与变换后的目标点云 $\mathbf{x}_{\text{trans}}$ 最邻近的点 $\mathbf{y}_k$,计算出尺度因子 S_{trans} 、旋转矩阵 $\mathbf{R}_{\text{trans}}$ 、平移向量 $\mathbf{T}_{\text{trans}}$,使得目标函数误差最小:

$$e_{\min} = \sum_{k=1}^N \|\mathbf{S}_{\text{trans}}\mathbf{R}_{\text{trans}}\mathbf{x}_{\text{trans}} + \mathbf{T}_{\text{trans}} - \mathbf{y}_k\|^2 \quad (5)$$

3) 两次迭代间参数的变化关系为

$$\begin{cases} S_{i+1} = S_{\text{trans}} S_i \\ \mathbf{R}_{i+1} = \mathbf{R}_{\text{trans}} \times \mathbf{R}_i \\ \mathbf{T}_{i+1} = S_{\text{trans}} \mathbf{R}_{\text{trans}} \times \mathbf{T}_i + \mathbf{T}_{\text{trans}} \end{cases} \quad (6)$$

将 $S_{i+1}, \mathbf{R}_{i+1}, \mathbf{T}_{i+1}$ 作为初始值重复步骤 1)、步骤 2),当尺度因子变化量 $\Delta S = S_{i+1} - S_i$ 、旋转矩阵变化量 $\Delta \mathbf{R} = \mathbf{R}_{i+1} - \mathbf{R}_i$ 、平移向量变化量 $\Delta \mathbf{T} = \mathbf{T}_{i+1} - \mathbf{T}_i$

均小于设定阈值时,停止迭代过程,输出即为最优尺度因子 S ,最优旋转矩阵 $\mathbf{R}$ 与最优平移向量 $\mathbf{T}$ 。

4) 利用均方根 (Root Mean Square, RMS) 误差分析评价配准精度,计算式为

$$Q_{\text{rms}} = \frac{1}{S} \left(\frac{1}{N} \sum_{k=1}^N \|\mathbf{S}\mathbf{R}\mathbf{x}_k + \mathbf{T} - \mathbf{y}_k\|^2 \right)^{\frac{1}{2}} \quad (7)$$

式中: Q_{rms} 为均方根误差。

3 实验结果与分析

3.1 仿真实验验证与分析

为了验证 ASKQ 算法对于关键点提取、特征提取与匹配的有效性,本文使用 UWA (University of Western Australia) 提供的点云数据库进行仿真实验验证。该点云数据库采用 Minolta Vivid 910 激光扫描仪,获取真实目标点云,空间分辨率为 640×480 。本实验选用的数据集包含点云 30 165 个,点云间隔为 0.366 mm。

关键点质量 Q_k 常作为选取高质量关键点的指标。 Q_k 的值越大,表示关键点的质量越高,但相对应的是关键点的数量减小,平坦目标表面上的关键点将被经验性剔除。本实验将阈值设定为 $Q_k > 4$,将算法中自适应尺度设定为

$$\mathbf{r}_k = [2.5 \quad 5 \quad 7.5 \quad 10 \quad 12.5]$$

本实验设置 4 种尺寸目标 (小鸡模型),分别为 5、10、20、100 cm,关键点提取的结果如图 3 所示。结果显示当目标尺寸越大,关键点数量越多,并且关键点分布均匀。

为了验证不同尺度因子的特征匹配实验结果,4 种目标尺度因子设置为 0.5、1、2、10,尺度因子定义为目标点云与参考点云的比值。基于上述关键点提取结果,进行网格化采样 (20×20) 得到 400 维特征向量,利用 PCA 方法对特征向量降维。匹配结果如图 4 所示,结果显示该算法对不同尺度因子的目标都有良好的匹配结果。

最后,利用 ASKQ 算法提取的关键点的特征匹配对计算初始配准参数,并利用 FMSR 算法实现 4 组目标的精确配准,配准结果如图 5 所示。结果表明该算法对 4 组尺度因子不同的目标均有良好的精确配准结果。

表 1 为 FMSR 算法的精确配准结果,从表中得出以下结论:①对尺度因子为 0.5、1、2、10 的 4 组目标,利用 FMSR 算法计算的最优尺度因子 S 均与理论值 μ 接近;②由于 4 组目标具有固定的相对位置,配准得到的 4 组最优旋转矩阵 $\mathbf{R}$ 一致,最优平移向量 $\mathbf{T}$ 成比例关系,且比例系数与

尺度因子相同,表明 FMSR 算法精确配准结果具

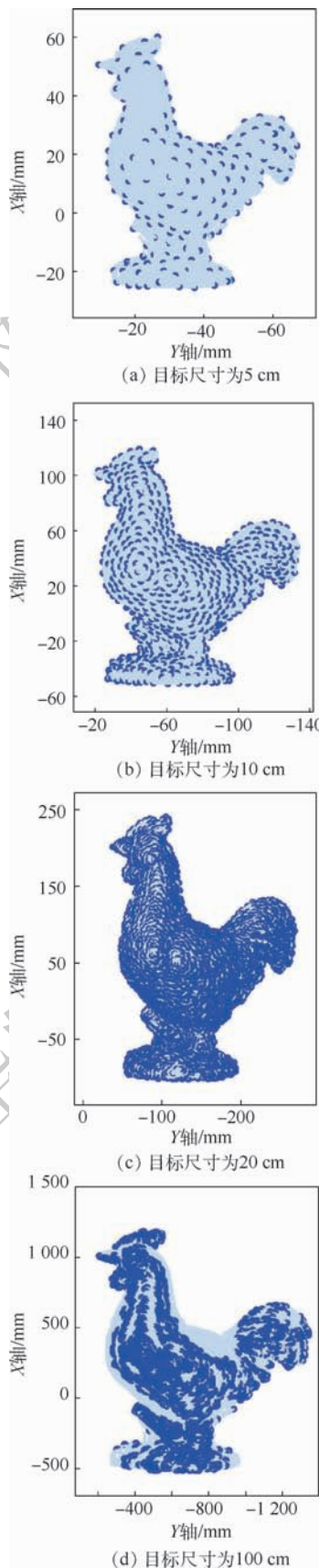


图 3 多尺度小鸡模型的关键点提取结果
Fig. 3 Key point extraction results of multi-scale chicken model

有稳定性;③ 4 组目标的配准精度均为0.002 m,说明FMSR算法配准精度具有稳定性;④4组目

标的算法运行时间不同,算法运行时间与初始配准的精度有关,初始特征匹配对的精度越高,算法

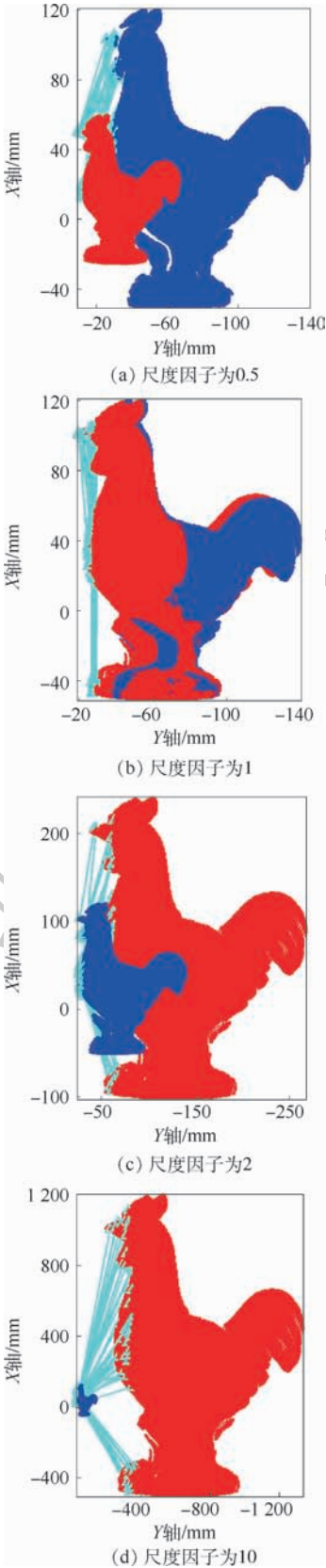


图 4 多尺度小鸡模型的关键点匹配结果

Fig. 4 Key point matching results of multi-scale chicken model

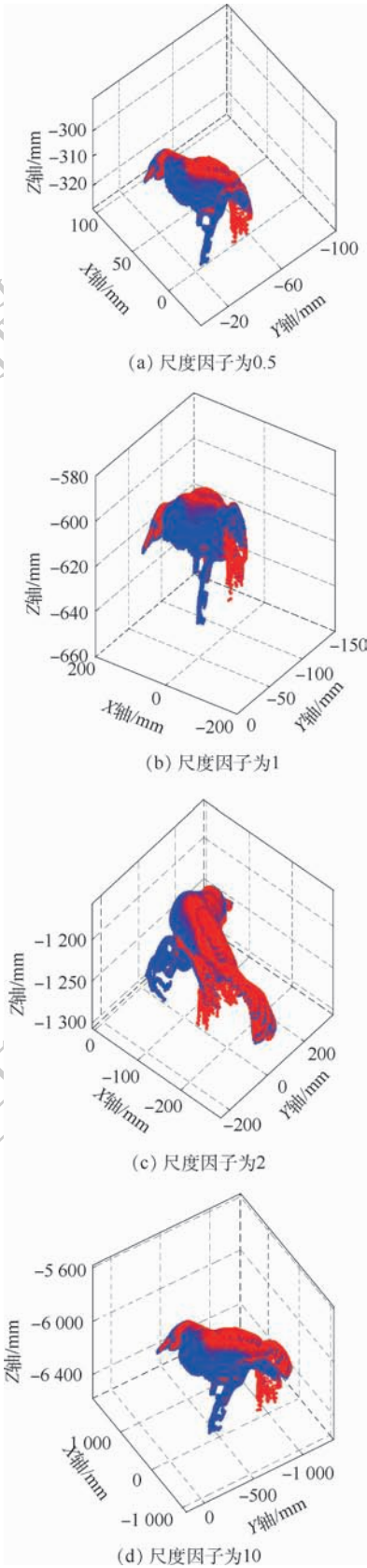


图 5 多尺度小鸡模型精确配准结果

Fig. 5 Fine registration results of multi-scale chicken model

表 1 FMSR 算法精确配准结果

Table 1 Fine registration results of FMSR algorithm

理论值 μ	最优尺度因子 S	最优旋转矩阵 R	最优平移向量 T	均方根误差 Q_{rms}/m	时间/s
0.5	0.487	(-0.483,0.010, -0.049)	(2.659, -145.708, -25.783)	0.002	4.043
1	0.977	(-0.483,0.010, -0.049)	(5.338, -291.396, -51.402)	0.002	2.601
2	1.954	(-0.483,0.010, -0.049)	(10.679, -582.779, -102.807)	0.002	2.509
10	9.768	(-0.483,0.010, -0.049)	(53.439, -2913.906, -513.807)	0.002	2.489

运行越快。

3.2 真实实验验证与分析

3.2.1 实验描述

真实场景实验的目标为北京航空航天大学晨兴音乐厅(尺寸为 20.30 m × 7.85 m × 26.56 m),实验场景如图 6 所示。利用自研三维扫描激光雷达系统采集三维点云,设置激光发射频率为 5 kHz、电机转速为 2 r/s、云台转速为 0.37 (°)/s。获取激光雷达点云 28 136 个,垂直分辨率为 0.144°,水平分辨率为 0.185°。采用摄像机(型号 MER-1070-14U3C)拍摄多视点图像,图像像素为 3 264 × 2 448,并且利用 SFM 技术实现图像三维重建,获取图像重建点云 12 789 个。



图 6 北京航空航天大学晨兴音乐厅建筑的实验现场

Fig. 6 Building experimental site of Chenxing Concert Hall in Beihang University

3.2.2 实验案例分析

本实验中,最小距离阈值设定为 $Q_k > 4$,自适应尺度的选取范围 $r_k = [25 \ 20 \ 15 \ 10 \ 5 \ 1 \ 0.5]$ 。激光雷达点云和图像重建点云的关键点提取结果如图 7(a)和图 7(b)所示。由图可知:①关键点仅取决于局部表面的坐标变化,不受目标整体几何形状的影响,因此 ASKQ 算法提取出的关键点分布均匀。②关键点由局部坐标系提取得出,具有局部不变性,因此关键点鲁棒性佳,不随旋转、平移及尺度因子的变化而改变。③图 7(b)中设定的最小距离阈值小于

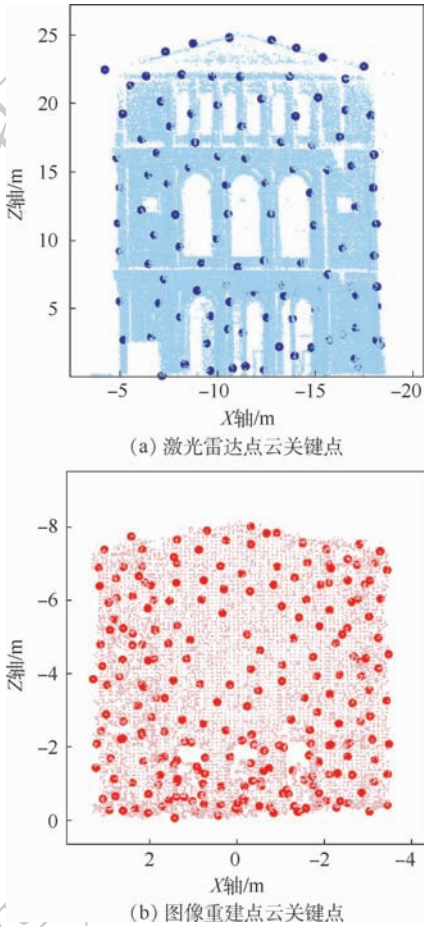


图 7 真实场景关键点提取结果

Fig. 7 Key point extraction results of real scene

图 7(a),因此关键点数量较多。由于图像重建点云密度低于激光雷达点云密度,应从图像重建点云中提取更多关键点,寻找激光雷达点云的最佳匹配。

通过网格化采样提取出关键点的形状特征,使用 PCA 方法降维,并对降维后的形状特征实现匹配。本实验提取出关键点的特征匹配对 22 对,匹配结果如图 8 所示,结果表明所提出算法具有良好匹配效果。

最后,利用关键点的特征匹配对计算初始尺度因子 S_0 、初始旋转矩阵 R_0 和初始平移向量 T_0 ,为 FMSR 算法提供初始参数,最终通过迭代求解实现精确配准。本文选取 S-ICP (Scale-Iterative

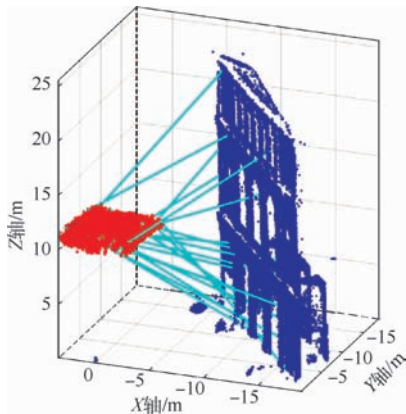


图 8 真实场景的关键点匹配结果

Fig. 8 Key point matching results of real scene

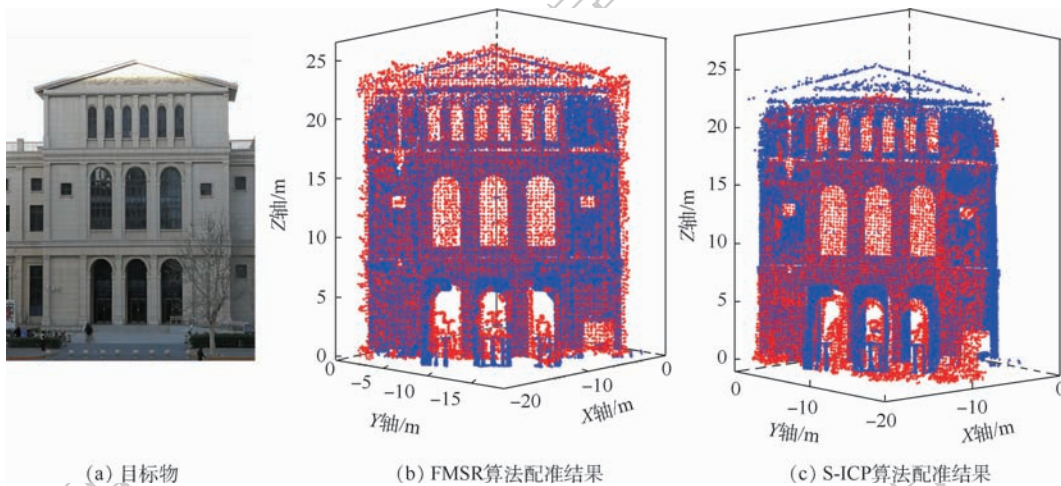


图 9 真实场景的精确配准结果

Fig. 9 Fine registration results of real scene

表 2 S-ICP 算法与 FMSR 算法对比

Table 2 Comparison of S-ICP and FMSR algorithms

参数	S-ICP 算法	FMSR 算法
最优尺度因子 S	3.000	3.300
最优旋转矩阵 R	(1.683, -0.035, 0.935)	(1.561, -0.044, -2.214)
最优平移向量 T	(-39.233, -29.415, 2.564)	(17.394, 12.485, -0.957)
均方根误差 Q_{rms}/m	0.325	0.194
时间/s	23.212	16.207

4 结 论

为了提高激光雷达点云和图像重建点云的空间点云配准精度,自主搭建三维扫描激光雷达系统,主要性能包括:点云垂直角分辨率 0.144° ,水平角分辨率 0.185° ,点云密度 $300\text{ pt}/\text{m}^2$ (pt/m^2 代表每平方米的点数);结合基于 ASKQ 的点云特征提取算法和基于 FMSR 的点云配准算法,实现激光雷达点云和图像重建点云的初始配准和精确配准。

ASKQ 算法为点云配准的初始配准过程提供

Closest Point) 算法为对比算法,利用所提出的 FMSR 算法和 S-ICP 算法配准同一目标。实验目标和 2 种算法的点云配准结果如图 9 所示,2 种算法配准结果的对比如下:在屋顶上方和墙壁右侧位置,应用 S-ICP 算法的配准结果存在部分缺失,而采用所提出的 FMSR 算法则有较优的配准结果。表 2 结果表明,S-ICP 算法的均方根误差为 0.325 m ,FMSR 算法的均方根误差为 0.194 m ,表明 FMSR 算法的配准精度高于 S-ICP 算法,配准精度提高了 0.131 m 。且运行时间由 23.212 s 降低为 16.207 s ,表明 FMSR 算法能够减小运行时间(30%),提高计算效率。

高质量匹配对,从而计算出精确的初始尺度因子 S_0 、初始旋转矩阵 R_0 和初始平移向量 T_0 ,为精确配准提供准确的初始参数。利用 FMSR 点云配准算法进行精确配准,通过 KNN 算法对 2 组点云进行全局搜索,多次迭代计算得到最优尺度因子 S ,最优旋转矩阵 R 和最优平移向量 T 。通过数值仿真和实验结果表明,该算法针对空间目标($20.30\text{ m} \times 7.85\text{ m} \times 26.56\text{ m}$)配准精度达到 0.194 m ,运行时间为 16.207 s 。FMSR 算法有效提升空间点云配准精度和计算效率(30%)。上述方法可以为空间场景重建和纹理匹配提供算法基础。

参考文献 (References)

[1] 骆社周, 习晓环, 王成. 激光雷达遥感在文化遗产保护中的应用[J]. 遥感技术与应用, 2014, 29(6): 1054-1059.
LUO S Z, XI X H, WANG C. The application of LiDAR remote sensing of cultural heritage preservation[J]. Remote Sensing Technology and Application, 2014, 29(6): 1054-1059 (in Chinese).

[2] 赵一鸣, 李艳华, 商雅楠, 等. 激光雷达的应用及发展趋势[J]. 遥测遥控, 2014, 35(5): 4-22.
ZHAO Y M, LI Y H, SHANG Y N, et al. Application and development direction of LiDAR[J]. Journal of Telemetry, Tracking and Command, 2014, 35(5): 4-22 (in Chinese).

[3] 曾齐红. 机载激光雷达点云数据处理与建筑物三维重建[D]. 上海: 上海大学, 2009.
ZENG Q H. Airborne LiDAR point cloud data processing and 3D building reconstruction[D]. Shanghai: Shanghai University, 2009 (in Chinese).

[4] YANG M D, CHAO C F, HUANG K S, et al. Image-based 3D scene reconstruction and exploration in augmented reality[J]. Automation in Construction, 2013, 33: 48-60.

[5] WANG R. 3D building modeling using images and LiDAR: A review[J]. International Journal of Image and Data Fusion, 2013, 4(4): 273-292.

[6] ABAYOWA B O, YILMAZ A, HARDIE R C. Automatic registration of optical aerial imagery to a LiDAR point cloud for generation of city models[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 106: 68-81.

[7] 王欣, 张明明, 于晓, 等. 应用改进迭代最近点方法的点云数据配准[J]. 光学精密工程, 2012, 20(9): 2068-2077.
WANG X, ZHANG M M, YU X, et al. Point cloud registration based on improved iterative closest point method[J]. Optics and Precision Engineering, 2012, 20(9): 2068-2077 (in Chinese).

[8] BESL P J, MCKAY N D. A method for registration of 3-D shapes[J]. Proceedings of SPIE-The International Society for Optical Engineering, 1992, 14(3): 239-256.

[9] YING S, PENG J, DU S, et al. A scale stretch method based on ICP for 3D data registration[J]. IEEE Transactions on Automation Science and Engineering, 2009, 6(3): 559-565.

[10] CHEN Y, MEDIONI G. Object modelling by registration of multiple range images[J]. Image and Vision Computing, 1992, 10

(3): 145-155.

[11] RUSINKIEWICZ S, LEVOY M. Efficient variants of the ICP algorithm[C]//Proceedings 3rd International Conference on 3-D Digital Imaging and Modeling. Piscataway, NJ: IEEE Press, 2001: 145-152.

[12] 戴静兰, 陈志杨, 叶修梓. ICP 算法在点云配准中的应用[J]. 中国图象图形学报, 2007, 12(3): 517-521.
DAI J L, CHEN Z Y, YE X Z. The application of ICP algorithm in point cloud alignment[J]. Journal of Image and Graphics, 2007, 12(3): 517-521 (in Chinese).

[13] 邹际祥. 基于 KD-tree 加速的点云数据配准技术研究[D]. 合肥: 安徽大学, 2013.
ZOU J X. The research of point cloud data registration technique based on KD-tree acceleration[D]. Hefei: Anhui University, 2013 (in Chinese).

[14] MIAN A, BENNAMOUN M, OWENS R. On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes[J]. International Journal of Computer Vision, 2010, 89(2-3): 348-361.

[15] LI X L, LI Y Y, XU L J. Terrestrial laser scanner autonomous self-calibration with no prior knowledge of point-clouds[J]. IEEE Sensors Journal, 2018, 18(22): 9277-9285.

[16] LI X L, LI Y Y, XIE X H, et al. Lab-built terrestrial laser scanner self-calibration using mounting angle error correction[J]. Optics Express, 2018, 26(11): 14444-14460.

[17] LI X L, YANG B W, XIE X H, et al. Influence of waveform characteristics on LiDAR ranging accuracy and precision[J]. Sensors, 2018, 18(4): 1156-1172.

[18] LI X L, WANG H M, YANG B W, et al. Influence of time-pick-off circuit parameters on LiDAR range precision[J]. Sensors, 2017, 17(10): 2369-2389.

[19] XU L J, FENG J, LI X L, et al. Automatic registration method for TLS LiDAR data and image-based reconstructed data[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16(3): 482-486.

作者简介:

呼延嘉玥 女, 硕士研究生. 主要研究方向: 激光雷达及光信号处理。

李小路 女, 博士, 副教授, 硕士生导师. 主要研究方向: 激光雷达及光信号处理。

Three-dimensional point cloud registration technique for self-designed LiDAR

HUYAN Jiayue, XU Lijun, LI Xiaolu *

(School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100083, China)

Abstract: In order to realize the registration of point cloud data respectively obtained from LiDAR and camera, we used a fast multi-scale registration (FMSR) algorithm to register the point cloud data, based on a self-designed three-dimensional scanning laser radar system in our laboratory. The algorithm includes two steps: coarse registration and fine registration. In the coarse registration, an adaptive scale key point quality (ASKQ) algorithm was used to match key points and determine the initial parameters for fine registration. And in the fine registration, K-nearest neighbors (KNN) algorithm was used to simplify the search process and improve the algorithm efficiency. The optimal rotation matrix, translation vector and scale factor between two sets of point cloud data were obtained through many iterations. The simulation verified the stability of FMSR algorithm for multiscale registration. Simulation and experimental results show that the proposed algorithm successfully registers the point cloud data of the self-made LiDAR system and commercial camera. The root-mean-square error of the registration is 0.194 m and the execution time is 16.207 s, for a building with size of 20.30 m $\times$ 7.85 m $\times$ 26.56 m. Compared with an existing scale-iterative closest point (S-ICP) algorithm, the registration accuracy of the proposed algorithm is improved by 0.131 m, and the execution time is reduced by 30%. The proposed point cloud registration method can provide an algorithm basis for scene reconstruction and texture matching.

Keywords: LiDAR; image reconstruction; spatial point cloud registration; adaptive scale key point quality (ASKQ) algorithm; fast multi-scale registration (FMSR) algorithm

Received: 2019-01-16; **Accepted:** 2019-03-29; **Published online:** 2019-04-11 10:16

URL: kns.cnki.net/kcms/detail/11.2625.V.20190409.1104.003.html

Foundation items: National Key R & D Program of China (2018YFB0504500); National Natural Science Foundation of China (61671038, 61721091)

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

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0029

基于飞行数据的无人机平飞动作质量评价模型



滕怀亮¹, 李本威¹, 高永^{1,*}, 杨栋², 张赞¹

(1. 海军航空大学 航空基础学院, 烟台 264001; 2. 中国人民解放军 92074 部队, 宁波 315000)

摘

要: 针对无人机(UAV)操纵质量高低多依赖专家主观评判、不同飞行动作缺乏针对性评判标准的问题,建立了根据飞行数据客观评价无人机平飞动作质量的模型。首先通过建立平飞判别规则识别无人机飞行数据中的平飞动作数据段,然后根据布林通道理论计算各平飞数据段内多个飞行参数的得分值,最后通过“熵权法”确定各参数的权重,进而得到反映不同无人机操控手平飞动作质量的指标数据。将一次四边平飞训练任务中4组不同无人机操控手操纵和1组无人机自主控制的飞行数据输入该模型,计算结果表明,模型可以较好地识别平飞动作并区分不同操控手平飞动作质量的高低,可以为无人机操控手的训练提供参考。

关键词: 质量评价; 无人机; 平飞动作; 布林通道; 熵权法

中图分类号: V279

文献标识码: A

文章编号: 1001-5965(2019)10-2108-07

随着军用无人机大量装备部队以及民用无人机的广泛应用,专业无人机操控手的需求缺口较大,加强对无人机操控手的基础训练,客观合理的评价操控手的飞行技能水平成为无人机发展应用亟待解决的关键问题^[1]。对飞机驾驶人员飞行技能的评价主要有主观法和客观法两种:主观法一般由专家观察飞行过程后对驾驶人员进行打分,如 Kolb-Harper 评价^[2];客观法是选取飞行数据中的参数对驾驶员的飞行技能进行评价的方法。主观评价法具有实际操作简便等优势,应用较为广泛,但评价结果受评价人员情绪、动机、偏好、认知等因素影响较大,有时不能反映飞行操纵人员的实际操作情况。客观评价法克服了主观法的上述缺点,并随着快速数据存取记录器(QAR)等监控技术的进步^[3]越来越受到国内外专家的重视。

根据飞行数据快速、准确地识别飞行动作是利用客观法评估飞行质量的基础。在飞行动作识

别方面,张玉叶等^[4]提出了一种基于主成分分析和距离判别分类的飞行动作识别方法,沈一超等^[5]利用贝叶斯网络实现了飞行动作的识别,但两种方法均需要典型飞行动作数据作为样本数据,在一定程度上限制了其应用。

目前评估无人机自身飞行品质和有人驾驶飞机操纵质量的研究较多,但研究无人机地面操纵人员的操纵质量评估的文献较少。如 Holmberg^[6]、Cotting^[7]等提出了类似有人驾驶飞机的无人机飞行品质评估标准,但无人机本身和无人机地面操控人员属于不同的评价对象,不可混淆;刘莉雯^[8]、张龙^[9]、柳忠起^[10]等采用不同方法对有人驾驶飞机飞行员的操纵质量进行了评价,但有人驾驶飞机和无人机操纵方式和操纵环境差别较大,相关评价体系不能完全适用于无人机。在无人机操纵质量评估方面,夏长俊等^[11]对无人机平飞质量分析的主要内容及计算分析方法做了阐述,给出了相关参数的误差分析,但没有给出具体

收稿日期: 2019-01-22; 录用日期: 2019-04-15; 网络出版时间: 2019-05-21 10:02

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190520.0901.002.html

基金项目: 国家自然科学基金(51505492); 山东省泰山学者建设工程专项

* 通信作者: E-mail: gaoyongbh@sina.com

引用格式: 滕怀亮, 李本威, 高永, 等. 基于飞行数据的无人机平飞动作质量评价模型[J]. 北京航空航天大学学报, 2019, 45(10): 2108-2114. TENG H L, LI B W, GAO Y, et al. Quality evaluation model of unmanned aerial vehicle's horizontal flight maneuver based on flight data[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2108-2114 (in Chinese).

的平飞动作质量评价模型。彭阳和彭晓明^[12]提出了一种基于关键参数偏差的飞行动作质量分析方法,并以某型无人机“坡度 45°~60°盘旋”飞行动作为例进行了分析。Zhou 等^[13]利用 VIKOR 方法对某型无人机的着陆质量进行了评价。但无人机起飞、着陆、平飞、机动等不同飞行阶段,不同飞行动作的特点差异较大,评价指标和参数也应不同。单一的评价体系难以有效评定无人机操控手不同飞行技能的好坏,所以针对不同的飞行动作建立不同的评价体系成为需要解决的重要问题。平飞动作是无人机地面操纵人员首先需要掌握的基础技能,是考验无人机操控手的判断能力、心理素质以及操控能力的最基本的飞行动作,本文针对无人机操控手的平飞技能评估问题,建立了无人机平飞动作质量评价模型。

1 无人机平飞动作质量评价模型的建立

无人机平飞动作质量评价模型主要包含平飞动作识别和平飞动作打分两个环节,具体实现流程如图 1 所示。

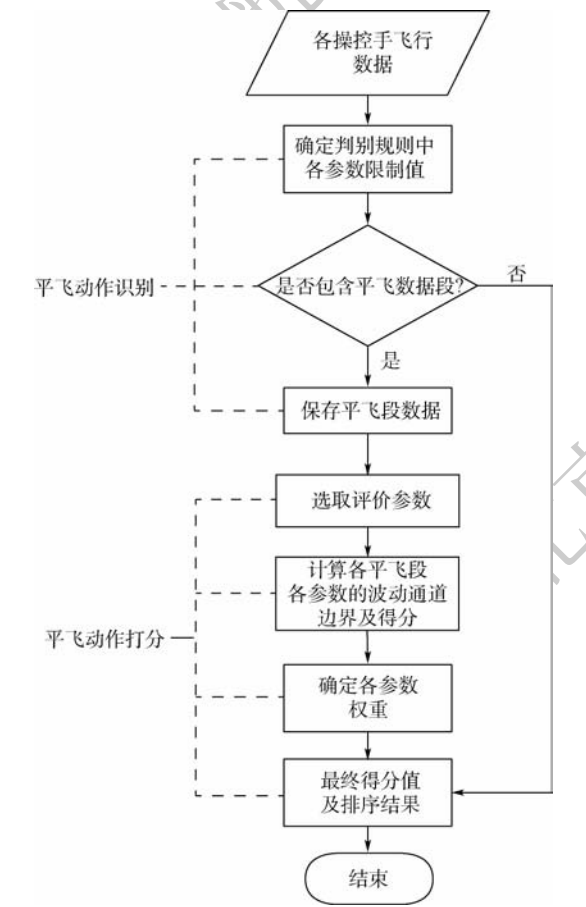


图 1 评价模型实现流程图

Fig. 1 Flowchart for realizing evaluation model

1.1 平飞动作识别

飞行数据记录了无人机在空中的飞行状态及操控手对无人机的操作信息。根据无人机飞行数据可以进行基本飞行动作识别^[14],从无人机飞行历程中提取有代表性的基本飞行动作的时间序列数据。

飞行动作最终体现在飞行参数的响应上,比如平飞就是无人机高度维持在某一稳定范围,爬升即为高度连续正变化,下滑即为高度连续负变化,所以可以根据飞行数据中的高度参数 h 及其变化率 v_h 来判别飞机的平飞动作,根据无人机平飞、爬升和下滑的不同特点,可以构建出下面的判别规则:

$$\begin{cases} |v_h| < d_h, |\Delta h_{\max}| < d_{\text{alt}} \rightarrow \text{平飞} \\ v_h > d_h, \Delta h_{\max} > d_{\text{alt}} \rightarrow \text{爬升} \\ v_h < -d_h, \Delta h_{\max} < -d_{\text{alt}} \rightarrow \text{下滑} \end{cases}$$

其中: d_h 为高度变化率的限制值; $\Delta h_{\max}$ 为平飞段的最大高度差; d_{alt} 为平飞动作的高度判别精度。加州伯克利大学^[15]的无人机自主着陆系统在各坐标轴上距离误差小于 5 cm、角度误差小于 5°;李槲等^[16]研究的无人机航迹跟踪方法的精度能达到 2 m 以内,本文选取 $d_{\text{alt}} = 2 \text{ m}$ 作为平飞动作的高度判别精度,而高度变化率限制值应满足 $d_h < d_{\text{alt}}$,本文取 $d_h = 0.8 \text{ m/s}$ 。即飞行高度变化率稳定在 0.8 m/s 以内且高度稳定在 2 m 以内的连续数据作为平飞动作数据。

平飞精度及高度变化率限制值的大小可以根据操控手不同训练阶段进行设置,同一次评价中采用相同的平飞精度和高度变化率限制值。首先通过高度变化率限制值 d_h 确定可能的平飞范围,即判断无人机飞行是否趋于稳定状态,然后通过平飞动作的高度判别精度 d_{alt} 进一步判断无人机是否处于平飞状态。识别出的平飞数据段为时间段和对应的飞行参数,为直观表示平飞数据段,采用平飞时间段及其对应的高度数据表示,如图 2 所示。

1.2 评价参数选取与波动通道的构建

飞行数据记录了无人机的多种飞行信息,这为评估操控手操作无人机的水平提供了重要依据。目前评估飞行员对各种飞行动作的完成情况,主要通过对某一个或两个飞行参数值的变化范围进行检测,如果变化范围没有超出指定飞行动作规定的参数值要求,那么操作合格。然而实际操作中,因为飞机的飞行动作取决于若干参数,所以要对飞行员操作的飞行动作进行正确评估,

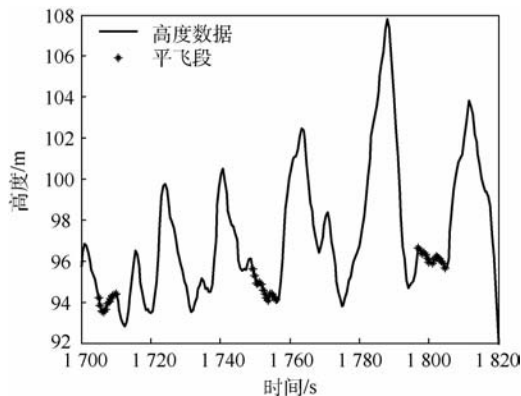


图2 典型平飞数据段

Fig.2 Typical horizontal flight data segment

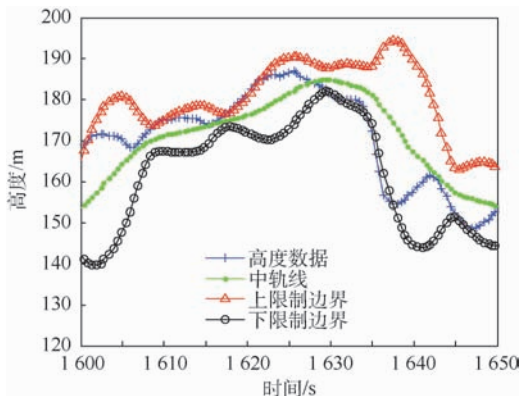


图3 布林线构成的数据带

Fig.3 Data band constituted by Bollinger lines

需要综合考虑多个参数的信息。本文选取对平飞动作较敏感的高度、速度、偏航角、滚转角和俯仰角作为平飞动作的评价参数,这5个参数的稳定性能够反映操控手控制无人机的平稳程度,确定评价参数后,即可构建各参数的波动通道。

数据通道的宽窄与数据本身波动幅度大小的变化直接相关,通道大小随着数据的形态和趋势自动调整,因而也可以反映无人机平飞趋势的变化。参考布林带(Bollinger Band)理论^[17],布林带一共由3条线组成:通道上限制边界线 $U(t)$ 、中轨线 $M(t)$ 和下限制边界线 $D(t)$,其表达式分别为

$$U(t) = \mu(t) + k\sigma(t) \quad (1)$$

$$M(t) = \mu(t) \quad (2)$$

$$D(t) = \mu(t) - k\sigma(t) \quad (3)$$

式中: $\mu(t)$ 为均值函数; $\sigma(t)$ 为标准差函数; k 为通道宽度限制值,可以根据实际情况选取,本文取 $k=2$ 。依据统计学的相关公式,可以通过样本函数 $x(t)$ ($t=1,2,\dots,N$)计算得到时间序列的均值函数 $\mu(t)$ 和标准差函数 $\sigma(t)$,继而确定飞行数据的波动范围。时间序列均值函数的估计值可由下式计算得到:

$$\hat{\mu}(t) = \frac{1}{N_c} \sum_{i=t-\frac{N_c}{2}}^{t+\frac{N_c}{2}} x(i) \quad (4)$$

$$\hat{\sigma}(t)^2 = \frac{1}{N_c} \sum_{i=t-\frac{N_c}{2}}^{t+\frac{N_c}{2}} (x(i) - \hat{\mu}(i))^2 \quad (5)$$

均值估计函数 $\hat{\mu}(t)$ 和标准差估计函数 $\hat{\sigma}(t)$ 在求解时要注意所选数据的窗长度,可选为连续数据点个数 N_c ,本文取 $N_c=20$ 。高度数据的部分布林通道如图3所示。

以高度参数为例,将高度数据 $h(t)$ 分别与下限制边界 $D(t)$ 和中轨线 $M(t)$ 作差,求两个差值

的均值,即可得到高度参数的打分值 s^h 为

$$s^h = \frac{|h(t) - D(t)| + |h(t) - M(t)|}{2} \quad (6)$$

将某操控手的第 i 个平飞段对应的高度数据输入波动通道模型,即可计算得到该操控手第 i 个平飞段的高度参数得分 s_i^h 。因为得分值大小从本质上是反映了飞行数据与均值的偏离程度和方差大小,所以平飞数据段得分值越低,表示操控手的平飞质量越好。

1.3 权重系数的确定

因为不同飞行参数对于评估操控手的权重是不同的,所以选择客观准确的评价指标权重系数对于准确评估飞行成绩起着非常重要的作用。本文采用“熵权法”计算指标权重,熵权法是一种客观的指标权重选取方法。

设某无人机操控手一次飞行任务数据中有 m 个平飞段,选取飞行数据中的 n 个监控参数,以 X_{ij} 表示第 i 个平飞段的第 j 个参数的得分值,则各个平飞段的评价矩阵如表1所示。

表1 无人机操控手飞行参数评分

Table 1 Scoring for flight parameters of UAV manipulators

平飞段	参数1	参数2	...	参数 n
1	X_{11}	X_{12}	...	X_{1n}
2	X_{21}	X_{22}	...	X_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	X_{m1}	X_{m2}	...	X_{mn}

由表1中的评分数据组成多属性决策矩阵:

$$X = [X_1, X_2, \dots, X_m]^T = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix}$$

为消除指标间数量级不同造成的影响,对决策矩阵进行规范化处理,即每个参数向量的单位化:

$$r_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^m X_{ij}^2}} \quad (7)$$

式中: r_{ij} 为单位化后的得分值。由各参数单位向量组成新的决策矩阵 $\mathbf{R} = (r_{ij})_{m \times n}$ 。

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (8)$$

式中: p_{ij} 为第 j 个参数下第 i 个平飞段的贡献度。形成贡献度矩阵 $\mathbf{P} = (p_{ij})_{m \times n}$ 。

$$E_j = -\frac{1}{\ln m} \sum_{i=1}^m (p_{ij} \ln p_{ij}) \quad (9)$$

式中: E_j 为所有平飞段对参数 j 的贡献总量。当 $p_{ij} = 0$, 定义 $p_{ij} \ln p_{ij} = 0$, 即可求得各参数所占的权重:

$$v_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)} \quad (10)$$

从而组成权重向量 $\mathbf{v} = [v_1, v_2, \dots, v_n]$, $v_j > 0$,

$\sum_{j=1}^n v_j = 1$ 。 $\mathbf{R}_i$ 为归一化决策矩阵的第 i 行向量, 于是该次飞行任务第 i 个平飞段所有参数的加权得分值为

$$s_i = \mathbf{R}_i \cdot \mathbf{v}^T \quad (11)$$

该次飞行任务平飞动作得分为

$$s = \frac{\sum_{i=1}^m s_i}{m} \quad (12)$$

至此,无人机平飞动作质量评价模型就构建完成了。

2 平飞动作质量评价的实现

2.1 数据预处理

数据预处理包含伪读数、无效幅值去除和数据长度一致性处理。伪读数是指数据变化超过了该参数变化的合理范围或短时间突变,属于噪声、振动或传感器故障等原因引起的误差,数据中的伪读数可以采用稳健回归算法^[18]进行处理。提取无人机的飞行高度、经纬度数据描述无人机整个飞行历程在空间的航迹示意图,如图 4 所示,操控手操纵无人机升空后在 100 m 高度上下做四边飞行训练。

飞行数据记录的时间单位为天,为便于分析

处理,将其换算为秒。图 5 为某无人机四边飞行训练过程中飞行高度随时间的变化情况,显然无人机升空执行任务是 1 000 ~ 2 400 s 左右,前后两段属于起飞和降落阶段,对平飞动作来说属于无效数据段,应予以去除。对于平飞动作的识别来说,无人机地面停留数据属于干扰识别准确度的无效值,可以通过定义最小高度 $h_{\min}$,删除小于 $h_{\min}$ 的数据的方式进行处理。

无人机飞行数据记录系统采集的参数只有高度和对应的时间信息,无人机高度变化率的近似计算表达式为

$$v_h = \lim_{\Delta t \rightarrow 0} \frac{\Delta h}{\Delta t} \approx \frac{h_{i+1} - h_i}{t_{i+1} - t_i} \quad (13)$$

因各参数传感器采集频率不一致,在数据处理过程中需要进行不同参数的向量长度一致性处理,长度不一致的参数可以通过插值或拟合等方式进行数据的补充或者删除。本文采用了拉格朗日插值的方式进行数据长度一致性处理。拉格朗日插值函数为

$$L_n(x) = \sum_{i=0}^n y_i l_i(x) = \sum_{i=0}^n y_i \left(\prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \right) \quad (14)$$

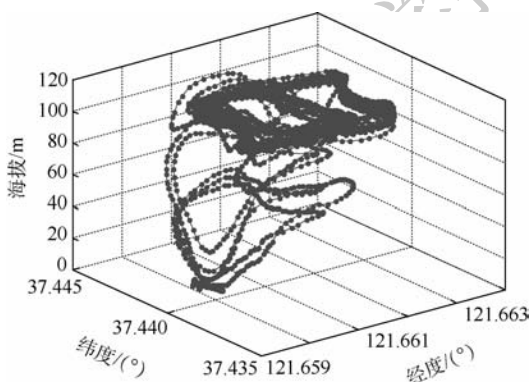


图 4 无人机航迹图

Fig. 4 UAV path map

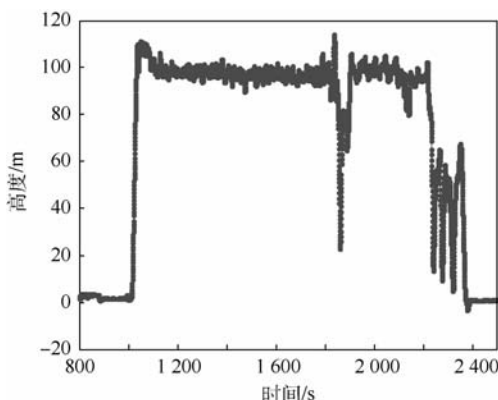


图 5 高度随时间的变化

Fig. 5 Variation of height with time

式中: x_i 和 y_i 分别为第 i 个数据点的横、纵坐标。

插值前后数据对比如图 6 所示,米字符代表插值点数据,空心圆代表原始数据。

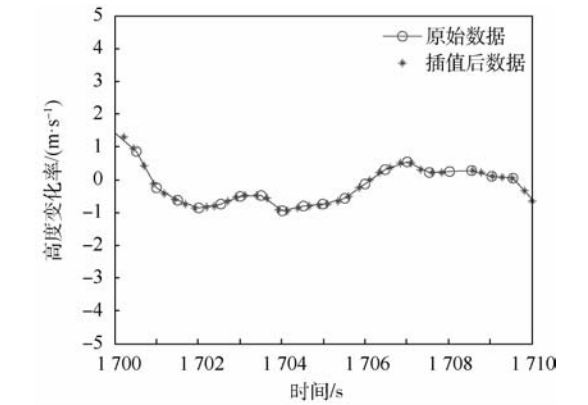


图 6 高度变化率插值结果

Fig. 6 Interpolation results of height change rate

2.2 识别平飞动作

根据构建的平飞判别规则可以识别出符合该规则的平飞时间段。识别出的平飞时间段存在持续时间过短或中间存在间断点的情况,选定连续飞行 5 s 以上的数据段作为具有代表性的平飞数据段,如表 2 所示。

表 2 平飞数据段		
Table 2 Horizontal flight data segments		
序号	起始时间/s	结束时间/s
1	4 666	4 688
2	4 698	4 721
3	4 757	4 783
4	4 849	4 872
5	5 428	5 457
6	5 787	5 808
7	6 172	6 206
8	6 366	6 393
9	6 674	6 699
10	6 787	6 808
11	6 871	6 892
12	7 045	7 074
13	7 237	7 269
14	8 089	8 114
15	8 186	8 210
16	8 266	8 300
17	8 428	8 458
18	8 662	8 698
19	8 852	8 888

2.3 各参数评分值及权重计算

得到平飞段后,计算各平飞时间段对应的各参数的布林带及得分值,分别得到高度、空速、俯仰角、偏航角和滚转角等参数的得分值向量。经过对操控手 1 的 19 个平飞数据段的 5 个飞行参数的打分,最终得到如表 3 所示的评分表。

在参数评分表的基础上,通过熵权法确定各参数权重,通过式 (11) 得到操控手 1 的 19 个平飞数据段得分情况如图 7 所示。

通过式 (12) 即可得到操控手 1 的最终评分值为 0.208 07。

表 3 评分表					
Table 3 Score sheet					
序号	高度得分	速度得分	俯仰得分	偏航得分	滚转得分
1	1.400 4	0.627 5	0.202 2	0.220 6	0.901 7
2	4.208 2	0.672 7	0.339 7	0.275 3	1.118 8
3	1.737 1	0.819 9	0.332 7	0.226 7	0.859
4	1.185 9	0.684 1	0.066 5	0.147 9	0.515 1
5	1.307 8	0.557 7	0.281 8	0.213 3	0.747 4
6	1.006 2	0.450 3	0.120 5	0.183	0.698 8
7	1.735 8	0.784 9	0.132 9	0.197 4	0.746 8
8	4.139 5	0.564	0.152 1	0.165 5	0.674 2
9	1.528 7	0.422 2	0.086 1	0.152 7	0.527
10	1.492	0.462 8	0.174	0.232 8	0.773 3
11	1.009 7	0.445 4	0.059 3	0.163 6	0.540 6
12	0.999 9	0.371 8	0.130 1	0.197 4	0.704 5
13	1.951 1	0.465 5	0.148 8	0.201 5	0.680 9
14	2.379 8	0.641 2	0.065 9	0.152 3	0.483
15	1.930 5	1.171 6	0.301 9	0.367 6	1.366 6
16	1.957 7	0.594 1	0.157 8	0.191 6	0.808 9
17	3.418 5	0.724 9	0.123 6	0.209 6	1.010 2
18	0.876 7	0.595 5	0.213	0.252 8	0.955 1
19	0.653 1	0.556 6	0.118 5	0.165 8	0.646 3

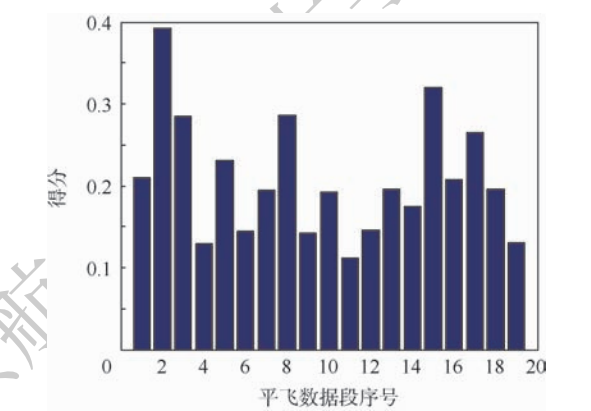


图 7 操控手 1 得分

Fig. 7 Scores of Manipulator 1

3 模型验证与分析

4 名操控手和无人机全自主控制模式执行四边飞行任务均在同一场地环境下进行,风速等天气情况相同,突风一定程度上会影响模型的评价结果,因为无法确定突风发生的时间点,进而修正模型,所以模型暂时无法排除风扰动的影响,利用平飞动作质量评价模型计算得到各操控手的得分及排名情况如表 4 所示。

从各操控手得分情况可以看出,该模型可以较好地区分 4 位操控手平飞动作的好坏。考虑到操控无人机平飞时,主要强调操控的平衡性和稳定性,所以可以采用无人机自主控制执行四边飞行任务的数据来验证算法的准确性。验证采用低空无人机航测系统控制的无人机飞行数据,该系统的航迹跟踪精度可达厘米级,相对 4 名处于训练状态的无人机操控手具有更好的平飞稳定性,通过本文的算法筛选出的平飞数据段得分如图 8 所示。

从图 8 可以看出,无人机自主飞行时满足平飞动作判别规则的数据段有 49 段,明显比 4 位操控手的 19、28、39、5 段更多。而经模型计算,无人机自主控制飞行的平均得分为 0.100 78,明显优于 4 位操控手的得分 0.208 07,0.170 01,0.145 02,0.351 29,从侧面验证了算法的准确性。

表 4 最终得分排名
Table 4 Final score ranking

操控主体	得分	平飞段数	排名
操控手 1	0.208 07	19	第 4 名
操控手 2	0.170 01	28	第 3 名
操控手 3	0.145 02	39	第 2 名
操控手 4	0.351 29	5	第 5 名
自主操纵	0.100 78	49	第 1 名

不同水平操控手平飞操纵质量的高低,可以为无人机操控手的训练提供参考。

后期可以针对不同的飞行动作,建立不同的飞行动作识别和操纵质量评价模型,实现对无人机操控手的综合飞行技能评估。

参考文献 (References)

[1] DAS S, MATTHEWS B L, LAWRENCE R. Fleet level anomaly detection of aviation safety data [C] // IEEE Conference on Prognostics and Health Management. Piscataway, NJ: IEEE Press, 2011: 1-10.

[2] ZHOU K, WANG L X, TAN X S. Flying qualities reduction of fly-by-wire commercial aircraft with reconfiguration flight control laws [J]. Procedia Engineering, 2011, 17: 179-196.

[3] WANG L, WU C X, SUN R S. An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents [J]. Reliability Engineering and System Safety, 2014, 127: 86-96.

[4] 张玉叶, 王颖颖, 王春歆, 等. 分析参数相关和时序特征的飞行动作识别方法 [J]. 计算机工程与应用, 2016, 52 (5): 246-249.

ZHANG Y Y, WANG Y Y, WANG C X, et al. Acrobatic maneuver reorganization method compared with parameters relevance and feature of sequence change [J]. Computer Engineering and Applications, 2016, 52 (5): 246-249 (in Chinese).

[5] 沈一超, 倪世宏, 张鹏. 基于贝叶斯网络的飞行动作识别方法 [J]. 计算机工程与应用, 2017, 53 (24): 161-167.

SHEN Y C, NI S H, ZHANG P. Flight action recognition method based on Bayesian network [J]. Computer Engineering and Applications, 2017, 53 (24): 161-167 (in Chinese).

[6] HOLMBERG J, LEONARD J, KING D, et al. Flying qualities specifications and design standards for unmanned air vehicles [C] // AIAA Atmospheric Flight Mechanics Conference and Exhibit. Reston, VA: AIAA, 2008: 1-14.

[7] COTTING C M. An initial study to categorize unmanned aerial vehicles for flying qualities evaluation [C] // AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. Reston, VA: AIAA, 2009: 1-12.

[8] 刘莉雯, 张天伟, 茹斌. 多参数融合的飞行品质评估模型的建立 [J]. 计算机工程与科学, 2016, 38 (6): 1262-1268.

LIU L W, ZHANG T W, RU B. A flying qualities assessment model based on multi-parameter integration [J]. Computer Engineering and Science, 2016, 38 (6): 1262-1268 (in Chinese).

[9] 张龙, 徐开俊. 采用模糊逻辑理论的飞行操纵品质评估模型 [J]. 中国民航飞行学院学报, 2016, 27 (5): 9-13.

ZHANG L, XU K J. Flight operation quality assessment model using fuzzy logic theory [J]. Journal of Civil Aviation Flight University of China, 2016, 27 (5): 9-13 (in Chinese).

[10] 柳忠起, 袁修干, 樊瑜波. 基于 BP 神经网络的飞行绩效评价模型 [J]. 北京航空航天大学学报, 2010, 36 (4): 403-406.

LIU Z Q, YUAN X G, FAN Y B. Pilot performance evaluation model based on BP neural networks [J]. Journal of Beijing Uni-

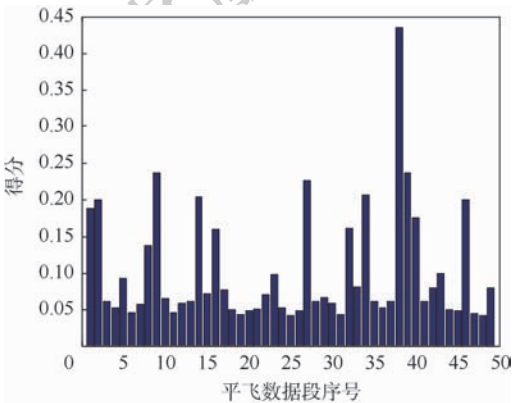


图 8 自主控制模式下的平飞数据段

Fig. 8 Horizontal flight data segments in autonomous control mode

4 结 论

本文通过建立平飞动作判别规则识别无人机平飞动作,通过建立无人机平飞动作质量评价模型评价了无人机操控手平飞动作质量,利用 4 名无人机操控手和自主控制模式飞行的无人机四边飞行数据验证了模型的合理性,结果表明:

- 1) 平飞动作判别规则能够有效识别平飞动作,为平飞动作质量评价提供数据基础。
- 2) 平飞动作质量评价模型能够较好地区分

- versity of Aeronautics and Astronautics, 2010, 36 (4): 403-406 (in Chinese).
- [11] 夏长俊, 孙校书, 李冰. 基于飞行数据的无人机平飞质量分析方法[J]. 海军航空工程学院学报, 2017, 32(5): 478-482. XIA C J, SUN X S, LI B. Analysis method of UAV level flight quality based on flight data[J]. Journal of Naval Aeronautical and Astronautical University, 2017, 32(5): 478-482 (in Chinese).
- [12] 彭阳, 彭晓明. 一种无人机飞行动作质量分析方法[J]. 飞行力学, 2017, 35(6): 65-69. PENG Y, PENG X M. A flight maneuver quality evaluation method for UAVs[J]. Flight Dynamics, 2017, 35(6): 65-69 (in Chinese).
- [13] ZHOU S H, QIAN S L, QIAO X D, et al. The UAV landing quality evaluation research based on combined weight and VIKOR algorithm[J]. Wireless Personal Communications, 2018, 102(2): 2047-2062.
- [14] 杨俊, 谢寿生. 基于模糊支持向量机的飞机飞行动作识别[J]. 航空学报, 2005, 26(6): 738-742. YANG J, XIE S S. Fuzzy support vector machines based recognition for aero-plane flight action[J]. Acta Aeronautica et Astronautica Sinica, 2005, 26(6): 738-742 (in Chinese).
- [15] SHARP C S, SHAKERNIA O, SASTRY S S. A vision system for landing an unmanned aerial vehicle[C] // Proceedings of the 2001 IEEE International Conference on Robotics & Automation. Piscataway, NJ: IEEE Press, 2001: 1720-1727.
- [16] 李隼, 陈清阳, 侯中喜. 自适应引导长度的无人机航迹跟踪方法[J]. 北京航空航天大学学报, 2017, 43(7): 1481-1490. LI Y, CHEN Q Y, HOU Z X. Path following method with adaptive guidance length for unmanned aerial vehicles[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(7): 1481-1490 (in Chinese).
- [17] LIU W, ZHENG W A. Stochastic volatility model and technical analysis of stock price[J]. Acta Mathematica Sinica, English Series, 2011, 27(7): 1283-1296.
- [18] 王玉伟, 高永. 基于稳健回归算法的无人机数据预处理技术研究[J]. 舰船电子工程, 2018, 38(11): 38-41. WANG Y W, GAO Y. Research on UAV data preprocessing technology based on robust regression algorithm[J]. Ship Electronic Engineering, 2018, 38(11): 38-41 (in Chinese).

作者简介:

滕怀亮 男, 硕士研究生。主要研究方向: 航空发动机状态监控与健康评估技术。

李本威 男, 博士, 教授, 博士生导师。主要研究方向: 航空宇航推进理论与工程。

高永 男, 博士, 副教授, 硕士生导师。主要研究方向: 飞行器设计。

Quality evaluation model of unmanned aerial vehicle's horizontal flight maneuver based on flight data

TENG Huailiang¹, LI Benwei¹, GAO Yong^{1,*}, YANG Dong², ZHANG Yun¹

(1. Aviation Foundation College, Naval Aviation University, Yantai 264001, China;

2. Troop 92074 of People's Liberation Army, Ningbo 315000, China)

Abstract: The unmanned aerial vehicle (UAV) manipulator's flying performance mostly relies on experts' subjective evaluation and different flight maneuvers lack pertinent evaluation criteria, so a model which uses flight data to evaluate the horizontal flight quality of UAV is established. Firstly, the flight data segment of UAV's horizontal flight maneuver was identified by the flight discrimination rules. Then, according to Bollinger bands theory, the scores of multiple flight parameters in each flight data segment were calculated. Finally, the weight of each parameter was determined by the entropy weight method and the indexes reflecting the horizontal flight maneuver's quality of different UAV manipulators were obtained. In a quadrilateral flying training mission, four groups of different UAV maneuvers' flight data and one group of flight data under autonomous control were input into the model. The calculation results show that the model can well identify the horizontal flight maneuver and distinguish the horizontal flight maneuver's quality of different manipulators, which can provide advice for the training of UAV manipulators.

Keywords: quality evaluation; unmanned aerial vehicle; horizontal flight maneuver; Bollinger bands; entropy weight method

Received: 2019-01-22; **Accepted:** 2019-04-15; **Published online:** 2019-05-21 10:02

URL: kns.cnki.net/kcms/detail/11.2625.V.20190520.0901.002.html

Foundation items: National Natural Science Foundation of China (51505492); Taishan Scholar Construction Foundation of Shandong Province

* **Corresponding author.** E-mail: gaoyongbh@sina.com

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2019.0038

基于非等维状态的 IMM 混合估计方法

欧能杰, 汪圣利*, 张直

(南京电子技术研究所, 南京 210039)



摘 要: 在目标跟踪领域,交互多模型(IMM)估计器具有良好的性能和较低的复杂度。IMM的成功归因于模式混合,其中各模型输出用于模型条件重初始化。针对IMM算法中存在的非等维状态混合估计问题进行了研究,在总结现有算法的基础上提出了一种最优的IMM混合估计方法。该方法通过将“切换”态的概念引入目标状态,根据当前滤波时刻的模型概率和新息,动态地调整混合策略以实现最优估计。最后,通过仿真实验验证了所提算法在不同模型混合场景中的表现要优于现有的算法。

关 键 词: 目标跟踪;滤波;交互多模型(IMM);混合估计;非等维状态

中图分类号: TN953

文献标识码: A

文章编号: 1001-5965(2019)10-2115-08

由于很难获取目标运动的先验信息,仅采用单模型的跟踪算法在不同机动情形下的跟踪效果往往较差。为了解决这个问题,多模型(Multiple Model, MM)算法^[1]得到了深入研究并被广泛应用于目标跟踪^[2-3]和轨迹预测^[4]等领域。MM算法的核心在于假设模型集包含目标所有可能运动模式的基础上,利用可能模型的概率加权匹配目标运动。MM算法的状态观测一般是模式依赖的,而且量测序列嵌入模式信息,即系统模式序列是Morkov过程。根据模型集设计方法,可以将MM算法分为固定结构和可变结构MM算法^[5]。其中,固定结构MM算法的代表是 n 阶广义伪贝叶斯(n -order Generalized Pseudo-Bayesian, GPB n)算法^[6]和交互多模型(Interacting Multiple Model, IMM)算法^[7]。IMM算法通过应用更好的假设管理技术,具有GPB2的性能和GPB1计算上的优势,逐渐成为研究热点。针对固定结构MM算法高度依赖模型集和模型竞争导致的性能下降等问题^[8],可变结构MM算法^[9-10]通过模型集自适应算法的开发具有较好的跟踪性能和费效比。

IMM算法是在GPB算法的基础上提出的一种具有系统模式Morkov切换的算法。IMM算法进行目标状态估计的计算时,考虑每个模型滤波器都有可能成为当前有效的系统模型滤波器,每个滤波器的初始条件都是基于前一时刻各条件模型滤波器结果的合成,即模型条件重初始化。由于合成过程中存在非等维状态向量的混合,例如当匀速(Constant Velocity, CV)模型和匀加速(Constant Acceleration, CA)模型进行混合估计时,加速度分量即为“额外”分量,而标准IMM算法并未给出关于此问题的描述。IMM混合估计需要匹配状态向量的前两阶矩(均值和协方差),因此低维状态向量需要扩维以匹配高维状态向量,这也是本文讨论的主题。然而,目前此问题只有少量的研究^[11-14]。一种简单的方法^[11]是对低维状态采用零均值和方差进行扩维,但会导致有偏估计。文献[12,14]提出了一种无偏方法,即利用高维状态中的均值和协方差对低维状态进行扩维。该方法通过弹道导弹数据进行了检验并展现了良好的性能。文献[13]认为被扩维的分布应该

收稿日期: 2019-01-29; 录用日期: 2019-05-18; 网络出版时间: 2019-06-11 13:11

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20190610.0911.002.html

* 通信作者。E-mail: slwangbb@sina.com

引用格式: 欧能杰, 汪圣利, 张直. 基于非等维状态的IMM混合估计方法[J]. 北京航空航天大学学报, 2019, 45(10): 2115-2122.
OU N J, WANG S L, ZHANG Z. IMM mixing estimation method based on unequal dimension states[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 2115-2122 (in Chinese).

尽可能反映系统潜在情况(模式切换),采用均匀分布描述“额外”分量,因而提高了模式切换时的跟踪精度。

本文所提方法基于以下思想:当前时刻的最优混合方法应根据时变的目标状态进行动态调整。本文方法在分析现有方法的基础上,将“切换”状态的概念引入目标状态,并利用模型概率和滤波新息进行联合判决,从而选择最优的混合方法,提高跟踪精度。

1 问题描述

对于一个跳变线性系统,其数学描述为

$$\mathbf{x}_{k+1} = \mathbf{F}^j \mathbf{x}_k + \mathbf{G}^j \mathbf{w}_k^j \quad (1)$$

$$\mathbf{z}_k = \mathbf{H}^j \mathbf{x}_k + \mathbf{v}_k^j \quad (2)$$

式中: $\mathbf{x}_k$ 和 $\mathbf{z}_k$ 分别为 k 时刻的状态向量和量测; $\mathbf{F}^j$ 、 $\mathbf{H}^j$ 和 $\mathbf{G}^j$ 分别为模型 j 的状态转移矩阵、量测矩阵和噪声分布矩阵; $\mathbf{w}_k^j \sim \mathcal{N}(0, \mathbf{Q}_k^j)$ 和 $\mathbf{v}_k^j \sim \mathcal{N}(0, \mathbf{R}_k^j)$ 是互相独立的过程噪声和量测噪声序列, $\mathbf{Q}_k^j$ 和 $\mathbf{R}_k^j$ 分别为过程噪声和量测噪声协方差。根据贝叶斯全概率理论,固定模型集的最优估计是全假设树估计,但所需计算资源难以满足^[8]。

因此,IMM 算法利用 GPB 算法进行系统估计时只考虑过去 2 个采样间隔内的历史,并认为系统模式是一阶齐次 Markov 链,其转移概率为

$$\pi_{ij} = P\{m_k^i | m_{k-1}^j\} \quad \forall m^i, m^j \in \Phi \quad (3)$$

式中: Φ 表示模型集; m_k^i 表示 k 时刻模型 j 与系统真实模式匹配。为描述简便起见,假设模型集 M 只包含 2 个非等维状态的模型。事实上,本文算法容易推广到包含任意多个模型的模型集。

假设 k 时刻状态估计的后验概率是高斯分布的,即

$$p(\mathbf{x}_k | m_k^i, \mathbf{Z}^k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^i, \mathbf{P}_k^i) \quad i \in \{1, 2\} \quad (4)$$

式中: $\hat{\mathbf{x}}_k^i$ 和 $\mathbf{P}_k^i$ 分别为系统模式为 m_k^i 时的均值和协方差估计; $\mathbf{Z}^k$ 为从初始时刻到 k 时刻的量测序列集合。与其他 MM 算法不同,IMM 算法在重初始化时考虑到前一时刻模型输出成为当前模型输入的可能性,即通过后向混合概率重初始化状态和协方差。后向混合概率计算式为

$$\mu_k^{i,j} \triangleq p(m_k^i | m_{k+1}^j, \mathbf{Z}^k) = \frac{\pi_{ji} \mu_k^i}{c_j} \quad (5)$$

$$\bar{c}_j = \sum_{i=1}^2 \pi_{ji} \mu_k^i \quad (6)$$

式中: $\mu_k^{i,j}$ 为在给定 $k+1$ 时刻模式为 j 时, k 时刻模型为 i 的概率; μ_k^i 为 k 时刻模型 i 概率; $\bar{c}_j$ 为归

一化因子。混合估计的状态估计及协方差为

$$\hat{\mathbf{x}}_k^{0,j} = E(\mathbf{x}_k | m_{k+1}^j, \mathbf{Z}^k) = \sum_{i=1}^2 \hat{\mathbf{x}}_k^i \mu_k^{i,j} \quad (7)$$

$$\mathbf{P}_k^{0,j} = \sum_{i=1}^2 [\mathbf{P}_k^i + (\hat{\mathbf{x}}_k^i - \hat{\mathbf{x}}_k^{0,j})(\hat{\mathbf{x}}_k^i - \hat{\mathbf{x}}_k^{0,j})^T] \mu_k^{i,j} \quad (8)$$

假设模型 1 状态向量的维度高于模型 2, 即 $n_1^1 > n_2^2$, 且模型 2 的状态分量包含于模型 1, 即

$$\begin{cases} \hat{\mathbf{x}}_k^1 = \begin{bmatrix} \hat{\mathbf{c}}_k^1 \\ \hat{\mathbf{e}}_k^1 \end{bmatrix} \\ \mathbf{P}_k^1 = \begin{bmatrix} \mathbf{P}_k^{c,1} & \mathbf{P}_k^{ce,1} \\ \mathbf{P}_k^{ec,1} & \mathbf{P}_k^{e,1} \end{bmatrix} \end{cases} \quad (9)$$

$$\begin{cases} \hat{\mathbf{x}}_k^2 = \hat{\mathbf{c}}_k^2 \\ \mathbf{P}_k^2 = \mathbf{P}_k^{c,2} \end{cases} \quad (10)$$

式中: $\mathbf{c}_k$ 为模型 1 和 2 的公共分量; $\mathbf{e}_k \in \mathbf{R}^{n_1^1 - n_2^2}$ 为模型 1 的“额外”分量, 在实践应用中可以是加速度、转弯率或推力等。由于模型的状态向量维度不均等, 难以按照式(7)和式(8)实现混合估计。对于模型 2 的重初始化, 可以舍弃“额外”分量, 直接利用 $\hat{\mathbf{c}}_k^1$ 和 $\mathbf{P}_k^{c,1}$ 进行, 即

$$\hat{\mathbf{x}}_k^{0,2} = \mu_k^{1,2} \hat{\mathbf{c}}_k^1 + \mu_k^{2,2} \hat{\mathbf{x}}_k^2 \quad (11)$$

$$\mathbf{P}_k^{0,2} = \mu_k^{1,2} \mathbf{P}_k^{c,1} + \mu_k^{1,2} (\hat{\mathbf{c}}_k^1 - \hat{\mathbf{x}}_k^{0,2})(\hat{\mathbf{c}}_k^1 - \hat{\mathbf{x}}_k^{0,2})^T + \mu_k^{2,2} \mathbf{P}_k^{c,2} + \mu_k^{2,2} (\hat{\mathbf{x}}_k^2 - \hat{\mathbf{x}}_k^{0,2})(\hat{\mathbf{x}}_k^2 - \hat{\mathbf{x}}_k^{0,2})^T \quad (12)$$

然而, 对于模型 1 的重初始化, 低维的状态向量必须按照某种方法进行扩维, 这也正是本文所考虑的问题。

IMM 算法是递推的, 包括模型条件重初始化、模型条件滤波、模型概率更新和估计融合。其中, 式(5)~式(8)给出了模型条件重初始化。

模型条件滤波是在给定重初始化的状态和协方差条件后进行状态估计更新, 包含以下步骤:

1) 状态预测及协方差, 计算式分别为

$$\hat{\mathbf{x}}_{k+1}^{0,i} = \mathbf{F}_k^i \hat{\mathbf{x}}_k^{0,i} \quad (13)$$

$$\mathbf{P}_{k+1}^i = \mathbf{F}_k^i \mathbf{P}_k^{0,i} (\mathbf{F}_k^i)^T + \mathbf{G}_k^i \mathbf{Q}_k^i (\mathbf{G}_k^i)^T \quad (14)$$

2) 新息(滤波残差)及其协方差, 计算式分别为

$$\mathbf{v}_{k+1}^i = \mathbf{z}_{k+1} - \mathbf{H}_{k+1}^i \hat{\mathbf{x}}_{k+1}^{0,i} \quad (15)$$

$$\mathbf{S}_{k+1}^i = \mathbf{H}_{k+1}^i \mathbf{P}_{k+1}^i (\mathbf{H}_{k+1}^i)^T + \mathbf{R}_{k+1}^i \quad (16)$$

3) 滤波更新, 包括滤波增益、状态估计更新及其协方差, 计算式分别为

$$\mathbf{K}_{k+1}^i = \mathbf{P}_{k+1}^i (\mathbf{H}_{k+1}^i)^T (\mathbf{S}_{k+1}^i)^{-1} \quad (17)$$

$$\hat{\mathbf{x}}_{k+1}^i = \hat{\mathbf{x}}_{k+1}^{0,i} + \mathbf{K}_{k+1}^i \mathbf{v}_{k+1}^i \quad (18)$$

$$\mathbf{P}_{k+1}^i = \mathbf{P}_{k+1}^i |_{k+1} - \mathbf{K}_{k+1}^i (\mathbf{S}_{k+1}^i)^{-1} (\mathbf{K}_{k+1}^i)^T \quad (19)$$

模型概率更新, 计算式分别为

$$\mu_{k+1}^i = \Lambda_{k+1}^i \bar{c}_i / \left(\sum_{j=1}^2 \Lambda_{k+1}^j \bar{c}_j \right) \quad (20)$$

$$\Lambda_{k+1}^i = \mathcal{N}(\mathbf{v}_{k+1}^i; 0, \mathbf{S}_{k+1}^i) \quad (21)$$

式中: A_{k+1}^i 为与 m_{k+1}^i 匹配的似然函数,服从高斯分布。

估计融合给出 $k+1$ 时刻的总体估计及其协方差,计算式分别为

$$\hat{\mathbf{x}}_{k+1} = \sum_{i=1}^2 \hat{\mathbf{x}}_{k+1}^i \mu_{k+1}^i \quad (22)$$

$$\mathbf{P}_{k+1} = \sum_{i=1}^2 [\mathbf{P}_k^i + (\hat{\mathbf{x}}_{k+1}^i - \hat{\mathbf{x}}_{k+1}) (\hat{\mathbf{x}}_{k+1}^i - \hat{\mathbf{x}}_{k+1})^T] \mu_k^i \quad (23)$$

2 非等维状态混合估计

如第 1 节所述,本文所研究的焦点是如何将“额外”分量体现到扩维后的状态向量上。假设模型 2 扩维后的状态 $\mathbf{x}_k^{2,a}$ 分布为

$$p(\mathbf{x}_k^{2,a} | \mathbf{Z}^k) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_k^2, \mathbf{P}_k^2) q(\mathbf{e}_k^2) \quad (24)$$

式中: $q(\mathbf{e}_k^2)$ 为模型 2 中“额外”分量 $\mathbf{e}_k$ 的分布。显然,不同混合方法的区别在于 $q(\mathbf{e}_k^2)$ 的选择。本文总结了关于此问题以往的研究^[11-14],下面介绍现有的 3 种方法,分别记为 A_1 、 A_2 和 A_3 。

1) A_1 采取了一种简单的策略,认为由于 $\mathbf{e}_k$ 是模型 1 的分量并不存在于模型 2,故可以用零均值和协方差进行扩维,则相应的分布 $q(\mathbf{e}_k^2)$ 和扩维后的状态 $\mathbf{x}_k^{2,a}$ 为

$$q(\mathbf{e}_k^2) = \delta(\mathbf{e}_k^2) \quad (25)$$

式中: δ 为冲激函数。

$$\begin{cases} \hat{\mathbf{x}}_k^{2,a} = \begin{bmatrix} \hat{\mathbf{c}}_k^2 \\ \mathbf{0} \end{bmatrix} \\ \mathbf{P}_k^{2,a} = \begin{bmatrix} \mathbf{P}_k^{c,2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{cases} \quad (26)$$

2) 文献[12,14]提出了一种无偏方法 A_2 ,认为 A_1 会导致模型 1 状态估计有偏,而 $\mathbf{e}_k^2$ 利用模型 1 的“额外”分量 $\mathbf{e}_k$ 信息进行扩维是无偏的,相应的分布 $q(\mathbf{e}_k^2)$ 和扩维后的状态 $\mathbf{x}_k^{2,a}$ 为

$$q(\mathbf{e}_k^2) = \mathcal{N}(\mathbf{e}_k^2; \hat{\mathbf{e}}_k^1, \mathbf{P}_k^{e,1}) \quad (27)$$

$$\begin{cases} \hat{\mathbf{x}}_k^{2,a} = \begin{bmatrix} \hat{\mathbf{c}}_k^2 \\ \hat{\mathbf{e}}_k^1 \end{bmatrix} \\ \mathbf{P}_k^{2,a} = \begin{bmatrix} \mathbf{P}_k^{c,2} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_k^{e,1} \end{bmatrix} \end{cases} \quad (28)$$

3) A_3 考虑到系统模式切换的可能性,认为 $\mathbf{e}_k^2$ 在其取值区间内服从均匀分布,则相应的分布 $q(\mathbf{e}_k^2)$ 和扩维后的状态 $\mathbf{x}_k^{2,a}$ 为

$$q(\mathbf{e}_k^2) = \prod_{l=1}^{n_e} \mathcal{U}(e_l; a_l, b_l) \quad (29)$$

$$\begin{cases} \hat{\mathbf{x}}_k^{2,a} = \begin{bmatrix} \hat{\mathbf{c}}_k^2 \\ E(\mathbf{e}_k^2) \end{bmatrix} \\ \mathbf{P}_k^{2,a} = \begin{bmatrix} \mathbf{P}_k^{c,2} & \mathbf{0} \\ \mathbf{0} & \text{Cov}(\mathbf{e}_k^2) \end{bmatrix} \end{cases} \quad (30)$$

式中: $n_e = n_x^1 - n_x^2$ 为 $\mathbf{e}_k$ 的维度;参数 e_l 取值区间为 $[a_l, b_l]$; $\text{Cov}(\mathbf{e}_k^2)$ 表示 $\mathbf{e}_k^2$ 的协方差。

事实上,上述 3 种方法是基于不同假设下设计的。 A_1 假设系统模式为 m_k^2 ,因此将“额外”分量 $\mathbf{e}_k$ 置零; A_2 假设系统模式为 m_k^1 ,因此 $\mathbf{e}_k^2$ 直接利用 $\mathbf{e}_k^1$ 信息作为其分布; A_3 假设系统模式可能切换,因此认为 $\mathbf{e}_k^2$ 为均匀分布。显然,在假设成立的条件下,这些算法能够有效提高估计精度,但跳变线形系统的模式是互相转换的,也就是说目标的运动模式是时变的,在没有先验的情况下假设模式固定是不合理的。因此,本文提出方法的核心在于认为更为合理的混合估计方法应当基于目标当前状态而决定。为了实现对目标状态的准确估计,本文引入“切换”状态的概念。对于一个两模型交互的系统,目标状态包括 S_1 、 $S_{1 \rightarrow 2}$ 、 S_2 和 $S_{2 \rightarrow 1}$ 。显然,如果目标真实状态是已知的,则可以选择最佳的混合估计方法。假设目标真实状态为 S_1 ,则 $q(\mathbf{e}_k^2)$ 的最优选择为

$$q(\mathbf{e}_k^2 | S_1) = p(\mathbf{e}_k^2 | m_k^1, \mathbf{Z}^k) = \mathcal{N}(\mathbf{e}_k^2; \hat{\mathbf{e}}_k^1, \mathbf{P}_k^{e,1}) \quad (31)$$

同样地,当目标真实模式为 S_2 时, $q(\mathbf{e}_k^2)$ 的最优选择为

$$q(\mathbf{e}_k^2 | S_2) = p(\mathbf{e}_k^2 | m_k^2, \mathbf{Z}^k) = \delta(\mathbf{e}_k^2) \quad (32)$$

对于“切换”状态 $S_{1 \rightarrow 2}$,由于在此过程中 $\mathbf{e}_k$ 应置零,因此 $q(\mathbf{e}_k^2)$ 的最优选择同式(32)。当目标状态为 $S_{2 \rightarrow 1}$ 时,意味着 $\mathbf{e}_k$ 可能在其取值区间跳变, $q(\mathbf{e}_k^2)$ 的分布应体现出此特性,故采用 A_3 的均匀分布假设,即

$$q(\mathbf{e}_k^2 | S_{2 \rightarrow 1}) = \prod_{l=1}^{n_e} \mathcal{U}(e_l; a_l, b_l) \quad (33)$$

基于上述分析,可以发现 A_1 、 A_2 和 A_3 分别是不同目标状态下的最优方法。因此,在已知目标状态时,3 种方法可以被纳入一个统一的框架,根据当前状态选择最优的一个。然而,跟踪系统很难准确知道目标的真实状态。尽管如此,依然可以利用一些滤波信息,例如模型概率和新息,尽可能地确定目标状态。

相比于状态 S_1 和 S_2 可以通过模型概率 μ_k^i 体现,“切换”状态由于模式切换导致的机动,往往体现在新息变化上。因此,为了检测“切换”状态 $S_{2 \rightarrow 1}$,本文采用一种常用于目标跟踪的机动检测器^[15-17]来确定目标当前状态。显然,当目标从模式 2 切换到模式 1 时,由于模型失配将导致模型 2 的新息增大。因此,令检验统计量为新息的衰减记忆平均值,计算式为

$$\varepsilon_k = \rho(s) \cdot \sum_{i=k-s+1}^k \lambda^{k-i} (\mathbf{v}_i^2)^T (\mathbf{S}_i^2)^{-1} \mathbf{v}_i^2 \quad (34)$$

式中: $\rho(s) = (1 - \lambda) / (1 - \lambda^{s+1})$, $0 < \lambda < 1$ 为衰减因子; $\mathbf{v}_k^2$ 和 $\mathbf{S}_k^2$ 分别为模型 2 的新息及其协方差; s 为滑窗长度; $[k - s + 1, k]$ 为滑窗检测区间。在高斯假设下, 由于 $\mathbf{v}_k^2$ 和 $\mathbf{S}_k^2$ 都服从零均值的高斯分布, 故 $(\mathbf{v}_k^2)^T (\mathbf{S}_k^2)^{-1} \mathbf{v}_k^2$ 服从自由度为 n_z (量测向量维度) 的卡方分布。作为一个独立同分布的高斯变量加权和, 检验统计量 ε_k 虽然不是标准的卡方分布, 但可以通过矩匹配被视作一个经过缩放的卡方变量^[16-17], 有

$$\lim_{k \rightarrow \infty} E(\varepsilon_k) = n_z \quad (35)$$

因此, 检验统计量 ε_k 服从自由度为 n_z 的卡方分布。令假设 H_0 代表目标状态切换未发生, 假设 H_1 则代表模型切换发生, “切换”状态 $S_{2 \rightarrow 1}$ 由式 (36) 检测:

$$\varepsilon_k > \chi_{n_z}^2(\alpha) \quad (36)$$

由式 (36) 可知, 当 ε_k 超过检测阈值 $\chi_{n_z}^2(\alpha)$ 时拒绝假设 H_0 , 否则接受该假设。由概率论可知, 虚警概率 $P_{fa} = \alpha$, 即在目标状态未改变的情况下错误判定目标状态发生变化的概率为 α 。因此, 置信水平 $1 - \alpha$ 通常设的相当高 (例如 95% 或 99%) 以降低虚警概率。当检测门限 $\chi_{n_z}^2(\alpha)$ 过高时, 必然导致漏警概率 P_{in} 增大, 即在目标状态已改变的情况下未成功检测。很显然, 检测器往往需要在虚警概率和检测概率间做性能平衡, 尽量降低误判造成的影响。

当检测器触发阈值, 则状态 $S_{2 \rightarrow 1}$ 被激活, 利用 A_3 重新估计滑窗区间 $[k - s + 1, k]$ 内的目标状态估计。当检验统计量 ε_k 低于式 (36) 阈值时, 则直接利用模型概率确定目标状态, 即

$$q(\mathbf{e}_k^2) = \begin{cases} q(\mathbf{e}_k^2 | S_1) & \mu_k^1 > \mu_k^2 \\ q(\mathbf{e}_k^2 | S_2) & \mu_k^1 \leq \mu_k^2 \end{cases} \quad (37)$$

当 $\mu_k^1 > \mu_k^2$ 时, 认为目标状态为 S_1 , 故采用 A_2 进行混合估计; 否则认为目标状态为 S_2 , 采取 A_1 。图 1 给出了本文方法 (记为 A_4) 的具体流程。

如前所述, 由于门限检测导致的一定的虚警及漏警。下面简要分析各种情况下 A_4 的跟踪性能差异。当目标真实状态未发生变化 (H_0), 则正确接受 H_0 时的跟踪精度趋近于 A_1 或 A_2 , 否则虚警导致 A_4 的跟踪精度趋近于 A_3 。当目标真实状态发生变化 (H_1), 若正确接纳 H_1 , A_4 则采用 A_3 重新估计 $[k - s + 1, k]$ 区间的状态, 其跟踪性能趋近于 A_3 。相比于 A_1 和 A_2 , A_4 能够在成功检测到状态变化后再启用 A_3 , 这意味着 A_4 可以利用量

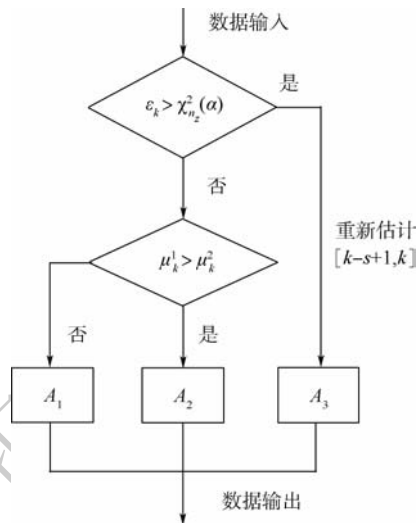


图 1 本文方法原理框图

Fig. 1 Functional block diagram of proposed method

测数据的后验信息来校正目标状态, 从而抑制因模型切换造成的跟踪误差剧增, 极大地降低峰值误差, 其作用类似于 IMM 平滑器^[18], 区别在于前者只在检测器触发时作用。若错误地拒绝接纳 H_1 , 由于漏警发生会导致跟踪误差增大, 跟踪性能接近 A_1 或 A_2 。此外, 滑窗长度 s 的选取同样影响跟踪性能。若滑窗过长会导致检测延迟, 不利于系统快速响应模式切换; 滑窗过短则导致检测的可靠性降低。因此, 在虚警率较低的条件 下尽可能地选择相对较短的滑窗长度以快速响应系统模式切换, 具体设置方法可参考文献 [15-16]。

基于上述分析可以看出, 本文方法并不固定采用某种混合方法, 而是通过新息检验和模型概率确定目标状态以选取最佳方法, 从而将 3 种混合方法的优势结合起来。

3 仿真与分析

在本节中, 对由不同运动模型组成的运动场景进行了仿真实验, 用于比较本文所提方法和现有 3 种方法的性能。在仿真中, 本文考虑了过程噪声的影响, 并给出 2 种评价指标, 最后还讨论了仿真结果。

3.1 轨迹与模型

场景 1: 由 2 个常见的线形运动模型组成。其中, 模型 1 和模型 2 分别为 CA 和 CV 模型, 用于产生一个 2D 运动轨迹 (即 $n_z = 2$)。2 种运动模型的状态向量分别为

$$\mathbf{x}_k^1 = [x_k \quad y_k \quad \dot{x}_k \quad \dot{y}_k \quad \ddot{x}_k \quad \ddot{y}_k]^T \quad (38)$$

$$\mathbf{x}_k^2 = [x_k \quad y_k \quad \dot{x}_k \quad \dot{y}_k]^T \quad (39)$$

式中: x_k 和 y_k 为位置分量; $\dot{x}_k$ 和 $\dot{y}_k$ 为速度分量;

$\ddot{x}_k$ 和 $\ddot{y}_k$ 为加速度分量,即“额外”分量 e_k 。CA 和 CV 模型的状态转移矩阵、量测矩阵和噪声分布矩阵分别为

$$\begin{cases} F_k^1 = \begin{bmatrix} I_2 & TI_2 & T^2 I_2/2 \\ \mathbf{0}_{2 \times 2} & I_2 & TI_2 \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & I_2 \end{bmatrix} \\ F_k^2 = \begin{bmatrix} I_2 & TI_2 \\ \mathbf{0}_{2 \times 2} & I_2 \end{bmatrix} \end{cases} \quad (40)$$

$$\begin{cases} H_k^1 = [I_2 \quad \mathbf{0}_{2 \times 4}] \\ H_k^2 = [I_2 \quad \mathbf{0}_{2 \times 2}] \end{cases} \quad (41)$$

$$\begin{cases} G_k^1 = \begin{bmatrix} T^3 I_2/6 \\ T^2 I_2/2 \\ TI_2 \end{bmatrix} \\ G_k^2 = \begin{bmatrix} T^2 I_2/2 \\ TI_2 \end{bmatrix} \end{cases} \quad (42)$$

式中: I_2 表示 2 阶单位矩阵; $T=1$ s 为采样间隔。

场景 2: 由线形的 CV 模型和非线性的协同转弯 (Coordinated Turn, CT) 模型组成。CT 模型的相应参数设置为

$$\mathbf{x}_k^1 = [x_k \quad y_k \quad \dot{x}_k \quad \dot{y}_k \quad \omega_k]^T \quad (43)$$

$$F_k^1 = \begin{bmatrix} 1 & 0 & \frac{\sin(\omega_k T)}{\omega_k} & \frac{\cos(\omega_k T) - 1}{\omega_k} & 0 \\ 0 & 1 & \frac{1 - \cos(\omega_k T)}{\omega_k} & -\frac{\sin(\omega_k T)}{\omega_k} & 0 \\ 0 & 0 & \cos(\omega_k T) & \sin(\omega_k T) & 0 \\ 0 & 0 & \sin(\omega_k T) & \cos(\omega_k T) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (44)$$

$$G_k^1 = \begin{bmatrix} T^2 I_{2 \times 1}/2 & \mathbf{0}_{2 \times 1} \\ TI_{2 \times 1} & \mathbf{0}_{2 \times 1} \\ \mathbf{0}_{1 \times 2} & 1 \end{bmatrix} \quad (45)$$

$$H_k^1 = [I_2 \quad \mathbf{0}_{2 \times 3}] \quad (46)$$

式中: ω_k 为转弯率。

目标初始状态为 $[1\ 000\ \text{m} \quad 0\ \text{m} \quad 100\ \text{m/s} \quad 0\ \text{m/s}]^T$ 。在 $0 \sim 50\ \text{s}$ 和 $101 \sim 150\ \text{s}$ 内做匀速运动, 在 $51 \sim 100\ \text{s}$ 内, 场景 1 目标做加速度为 $30\ \text{m/s}^2$ 的匀加速运动, 场景 2 做转弯率 $3.6\ (^{\circ})/\text{s}$ 的匀速转弯运动。目标的真实轨迹如图 2 所示。

3.2 IMM 参数设置与性能评价指标

对于上述 2 个场景, 系统量测噪声方差为 $\sigma_m^2 = 50^2\ \text{m}^2$, 其协方差为 $R_k^i = \sigma_m^2 I_2$; 过程噪声协方差为 $Q_k^i = \sigma_a^2 I_2$, 考虑到过程噪声的大小对跟踪性能的影响, 故设为

$$\sigma_a^2 \in \{10^{-2}, 10^{-1.5}, 10^{-1}, \dots, 10^2\} \quad (47)$$

除了混合方法之外, 4 种方法的 IMM 估计器参数设置是相同的, 滤波算法采用卡尔曼滤

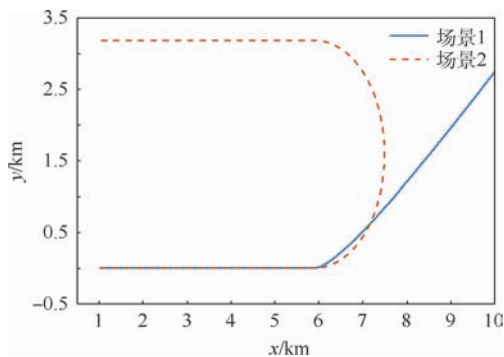


图 2 场景 1 和 2 的真实轨迹

Fig. 2 True trajectories of Scenarios 1 and 2

波^[49]。 A_1 和 A_2 的原理在第 2 节已介绍过。由于 A_3 认为“额外”分量 e_k 服从均匀分布, 式 (29) 中场景 1 的加速度的取值范围为 $[-100, 100]\ \text{m/s}^2$, 场景 2 中转弯率取值范围为 $[-3.6, 3.6]\ (^{\circ})/\text{s}$ 。对于 A_4 , 所需参数设置如下: 衰减因子 $\lambda = 0.95$, 置信度 $\alpha = 0.01$; 滑窗长度 $s = 2$, 阈值 $\chi_2^2(\alpha) = 9.21$ 。系统模式的 Markov 转移矩阵和初始模型概率分别为

$$\pi_{ij} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \quad (48)$$

$$\mu_0 = [0.9 \quad 0.1] \quad (49)$$

本文的性能评价指标如下:

$$\tau_k(\sigma_a) = \sqrt{\sum_{m=1}^M (\mathbf{x}_k - \hat{\mathbf{x}}_k^{(m)})^2 / M} \quad (50)$$

$$\tau^{\sigma_a} = \sum_{k=1}^N \tau_k(\sigma_a) / N \quad (51)$$

式中: $M=500$ 为蒙特卡罗仿真次数; $N=150$ 为目标运动时间; $\tau_k(\sigma_a)$ 是过程噪声方差为 σ_a^2 时的均方根误差 (Root Mean Square Error, RMSE); τ^{σ_a} 为不同 σ_a 下的 RMSE 均值。

3.3 结果与分析

限于篇幅, 本文只选择展示仿真结果中的一部分内容, 但这并不影响 4 种算法间的性能比较。对于场景 1 来说, 图 3 ~ 图 6 分别展示了过程噪声方差 σ_a^2 为 10^{-2} 和 10^{-1} 时的位置 RMSE 以及速度 RMSE。显然, A_1 、 A_2 和 A_4 在目标进行 CV 和 CA 运动时性能良好, 但系统模式从 CV 切换到 CA 时, 前 2 种算法的跟踪误差会急剧增大。 A_3 和 A_4 则能在模型切换时保持较好的跟踪精度。但 A_3 的缺陷在于, 当系统模式不发生切换时, 它的跟踪精度是 4 种方法中最差的。随着过程噪声方差的增大, 4 种算法间的性能差别逐渐变小, 这种趋势如图 7 所示。同样地, 对于场景 2, 图 8 和图 9 给出了过程噪声方差 σ_a^2 为 10^0 时的位置 RMSE 以及速度 RMSE。图 10 给出了不同过程噪

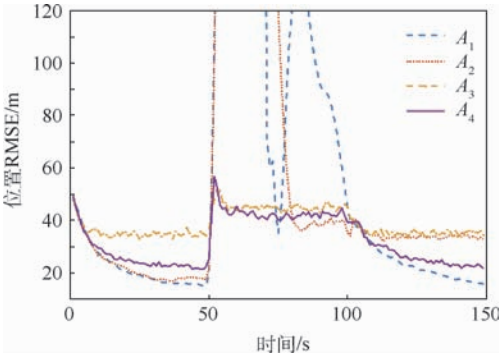


图3 场景1位置均方根误差($\sigma_a^2 = 10^{-2}$)

Fig.3 Position RMSE of Scenario 1 ($\sigma_a^2 = 10^{-2}$)

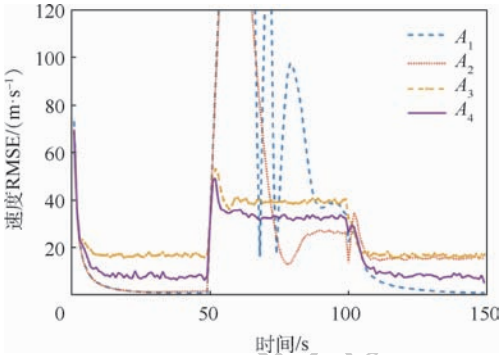


图4 场景1速度均方根误差($\sigma_a^2 = 10^{-2}$)

Fig.4 Velocity RMSE of Scenario 1 ($\sigma_a^2 = 10^{-2}$)

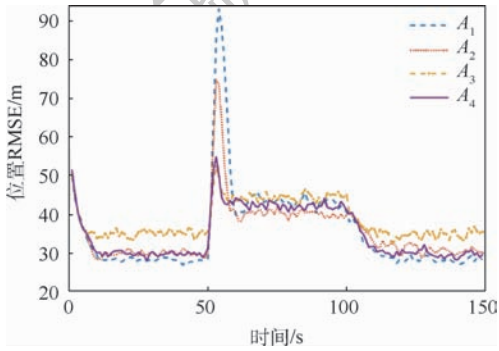


图5 场景1位置均方根误差($\sigma_a^2 = 10^1$)

Fig.5 Position RMSE of Scenario 1 ($\sigma_a^2 = 10^1$)

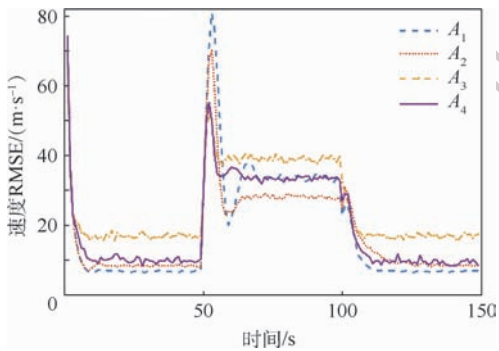


图6 场景1速度均方根误差($\sigma_a^2 = 10^1$)

Fig.6 Velocity RMSE of Scenario 1 ($\sigma_a^2 = 10^1$)

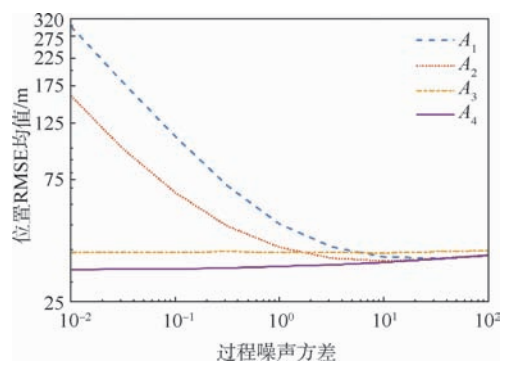


图7 不同过程噪声下场景1位置均方根误差均值

Fig.7 Mean value of position RMSE of Scenario 1 under different process noises

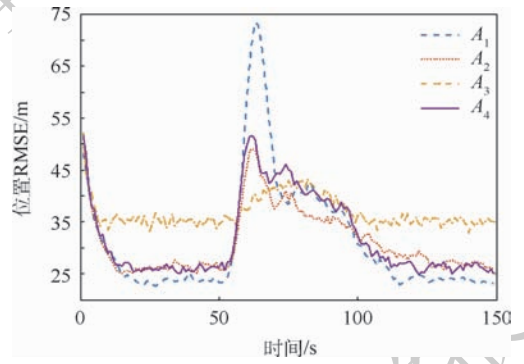


图8 场景2位置均方根误差($\sigma_a^2 = 10^0$)

Fig.8 Position RMSE of Scenario 2 ($\sigma_a^2 = 10^0$)

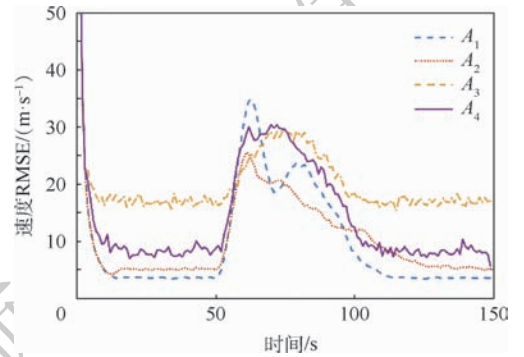


图9 场景2速度均方根误差($\sigma_a^2 = 10^0$)

Fig.9 Velocity RMSE of Scenario 2 ($\sigma_a^2 = 10^0$)

由仿真结果可见,无论目标处于何种状态或过程噪声, A_4 都能保持较好的性能,因此本文所提算法 A_4 的性能要优于其他 3 种算法。基于上述仿真结果,本文总结如下:

1) 过程噪声对不同混合方法有显著影响。当过程噪声较小时,系统模式切换导致的机动要求增大“额外”分量的协方差,这可以通过 $q(\mathbf{e}_k^2)$ 的设计实现。

2) 实际上, A_3 采用的均匀分布始终保持一个较大的协方差,因此对模型切换时的跟踪误差抑制起到了较好的作用,但相比于 A_1 和 A_2 所采

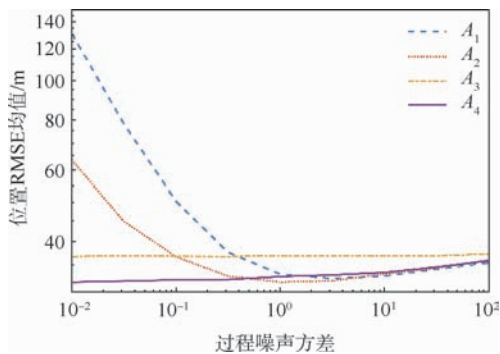


图 10 不同过程噪声下场景 2 位置均方根误差均值

Fig. 10 Mean value of position RMSE of Scenario 2 under different process noises

取的策略, A_3 牺牲了目标状态处于稳态时的跟踪精度。

3) A_4 性能优越的原因在于,通过对目标状态的重新划分,并利用检测器对目标状态进行确认可以灵活地调整当前混合方法,使其具有了 3 种方法的优势。

4) 当过程噪声较大时,其协方差已足够大导致各方法间的差异逐渐消失。

4 结 论

本文提出了一种基于非等维状态的 IMM 混合估计方法,通过仿真结果验证了该方法具有良好的性能,得出以下结论:

1) 过程噪声的大小会影响 IMM 混合方法的性能,较小的过程方差会导致较大的模型切换误差,而较大的过程噪声方差会削弱这种影响。

2) 通过引入“切换”状态实现目标状态的重新划分,结合滤波信息可以实现对目标状态的确定,从而灵活地选取当前最佳的混合方法。

为使本文提出的算法能具有更好的适应性,仍需要对目标状态确认逻辑做更为深入理论分析。

参考文献 (References)

[1] LI X R, JILKOV V P. Survey of maneuvering target tracking. Part V. Multiple-model methods[J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(4): 1255-1321.

[2] MAZOR E, AVERBUCH A, BAR-SHALOM Y, et al. Interacting multiple model methods in target tracking: A survey[J]. IEEE Transactions on Aerospace and Electronic Systems, 1998, 34(1): 103-123.

[3] FOO P H, NG G W. Combining the interacting multiple model method with particle filters for manoeuvring target tracking[J]. IET Radar, Sonar & Navigation, 2011, 5(3): 234-255.

[4] 翟岱亮, 雷虎民, 李炯, 等. 基于自适应 IMM 的高超声速飞

行器轨迹预测[J]. 航空学报, 2016, 37(11): 3466-3475.

ZHAI D L, LEI H M, LI J, et al. Trajectory prediction of hypersonic vehicle based on adaptive IMM[J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(11): 3466-3475 (in Chinese).

[5] LI X R. Multiple-model estimation with variable structure. II. Model-set adaptation[J]. IEEE Transactions on Automatic Control, 2000, 45(11): 2047-2060.

[6] JAFFER A G, GUPTA S C. On estimation of discrete processes under multiplicative and additive noise conditions[J]. Information Sciences, 1971, 3(3): 267-276.

[7] BLOM H A P, BAR-SHALOM Y. The interacting multiple model algorithm for systems with Markovian switching coefficients[J]. IEEE Transactions on Automatic Control, 1988, 33(8): 780-783.

[8] LI X R, BAR-SHALOM Y. Multiple-model estimation with variable structure[J]. IEEE Transactions on Automatic Control, 1996, 41(4): 478-493.

[9] LAN J, LI X R. Equivalent-model augmentation for variable-structure multiple-model estimation[J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(4): 2615-2630.

[10] XU L, LI X R, DUAN Z. Hybrid grid multiple-model estimation with application to maneuvering target tracking[J]. IEEE Transactions on Aerospace and Electronic Systems, 2016, 52(1): 122-136.

[11] BAR-SHALOM Y, WILLETT P K, TIAN X. Tracking and data fusion[M]. Storrs: YBS Publishing, 2011.

[12] YUAN T, BAR-SHALOM Y, WILLETT P, et al. A multiple IMM estimation approach with unbiased mixing for thrusting projectiles[J]. IEEE Transactions on Aerospace and Electronic Systems, 2012, 48(4): 3250-3267.

[13] GRANSTRÖM K, WILLETT P, BAR-SHALOM Y. Systematic approach to IMM mixing for unequal dimension states[J]. IEEE Transactions on Aerospace and Electronic Systems, 2015, 51(4): 2975-2986.

[14] LOPEZ R, DANES P, ROYER F. Extending the IMM filter to heterogeneous-order state space models[C]//49th IEEE Conference on Decision and Control (CDC). Piscataway, NJ: IEEE Press, 2010: 7369-7374.

[15] BAR-SHALOM Y, BIRMIWAL K. Variable dimension filter for maneuvering target tracking[J]. IEEE Transactions on Aerospace and Electronic Systems, 1982, 27(5): 621-629.

[16] JIN B, JIU B, SU T, et al. Switched Kalman filter-interacting multiple model algorithm based on optimal autoregressive model for manoeuvring target tracking[J]. IET Radar, Sonar & Navigation, 2014, 9(2): 199-209.

[17] LI X R, JILKOV V P. Survey of maneuvering target tracking. Part IV: Decision-based methods[C]//SPIE Proceedings Series. Society of Photo-Optical Instrumentation Engineers. Piscataway, NJ: IEEE Press, 2002: 511-534.

[18] LOPEZ R, DANES P. Low-complexity IMM smoothing for jump Markov nonlinear systems[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(3): 1261-1272.

[19] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Journal of basic Engineering, 1960, 82(1): 35-45.

作者简介:

欧能杰 男,硕士研究生。主要研究方向:雷达数据处理。

汪圣利 男,博士,研究员。主要研究方向:信息融合、雷达数据

处理。

张直 男,博士。主要研究方向:雷达数据处理。

IMM mixing estimation method based on unequal dimension states

OU Nengjie *, WANG Shengli, ZHANG Zhi

(Nanjing Research Institute of Electronics Technology, Nanjing 210039, China)

Abstract: The interacting multiple model (IMM) estimator has been proven to be of excellent performance and low complexity in tracking agile targets. The success of IMM attributes to mode mixing, where model outputs are mixed for model-conditional reinitialization. The problem of unequal dimension states mixing in IMM estimation is studied and an optimal method for IMM mixing is proposed based on summarizing the existing methods. By introducing the concept of “switching” state into the target state, the new method dynamically adjusts the hybrid strategy with model probability and innovation to achieve optimal estimation. The simulation results show that the proposed approach outperforms the existing algorithms in the scenarios of mixing different models.

Keywords: target tracking; filtering; interacting multiple model (IMM); mixing estimation; unequal dimension states

《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人,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

(以下按姓氏笔划为序)

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 45 卷 第 10 期 2019 年 10 月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUA)

(Monthly, Started in 1956)

Vol.45 No.10 October 2019

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 北航文化传媒集团

发行范围 国内外发行

联系电话 (010) 82315594 82338922
82314839

电子信箱 jbuua@buaa.edu.cn

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 JBUA

Printed by Beijing Kexin Printing Co., Ltd.

Distributed by BUAA Culture Media Group Limited

Telephone (010) 82315594 82338922
82314839

E-mail jbuua@buaa.edu.cn

http://bhxb.buaa.edu.cn

刊 号 ISSN 1001-5965
CN 11-2625/V

国内定价 50.00 元 / 期

ISSN 1001-5965



9 771001 596199