



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



2022-07

Vol.48 No.7

目次

- 基于任务映射的暗硅芯片功耗预算方法 李鑫, 李智, 周巍, 吴瑞祺, 唐浩然, 陈业航 (1115)
- 部分可观 Petri 网故障的量子贝叶斯诊断
..... 刘久富, 张信哲, 汪恒宇, A. M. TOMAS DIAS, 王志胜, 杨忠 (1125)
- 基于零速修正与姿态自观测的惯性行人导航算法
..... 戴洪德, 张笑宇, 郑百东, 戴邵武, 郑伟伟 (1135)
- 一种多机协同打击的快速航迹规划方法
..... 陈清阳, 辛宏博, 王玉杰, 唐钟南, 贾高伟, 朱炳杰 (1145)
- 基于多分辨率特征自选择的遮挡物识别算法 谢祥颖, 来广志, 那峙雄, 骆欣, 王栋 (1154)
- 气动舵机高压反向直动式减压阀的设计及特性
..... 刘延斌, 王学生, 秦新亚, 王浩, 陈琴珠, 赵赛 (1164)
- 高速干摩擦机械密封的端面变形及摩擦磨损 马润梅, 赵祥, 陈潇竹, 李双喜, 杨海超 (1174)
- 抽吸气流量对催化惰化系统性能影响 王晨臣, 潘俊, 王洋洋, 段伟杰 (1183)
- 二维有限长度柔性膜面上 T-S 波演化的数值研究 洪正, 叶正寅 (1190)
- 基于状态空间网格法的阵风减缓分析 杨澜, 安朝, 谢长川, 杨超 (1200)
- 磁悬浮转子状态反馈解耦自抗扰控制方法
..... 尹增愿, 蔡远文, 任元, 王卫杰, 陈晓岑, 于春森 (1210)
- 基于高效注意力和上下文感知的目标跟踪算法 柏罗, 张宏立, 王聪 (1222)
- 基于 DDQN 的片上网络混合关键性消息调度方法 李国梁, 李峭, 徐亚军, 熊华钢 (1233)
- 基于离散系数改进的 VWKNN 位置指纹定位算法
..... 许钊, 何泾沙, 朱娜斐, 邓万航, 吴霜, 他永君 (1242)
- 大行程解耦三平动微定位平台的设计与优化 林苗, 孟刚, 居勇健, 徐伟胜, 曹毅 (1252)
- 基于 PM 算法的涡旋电磁波引信超分辨率测向方法 孙丽华, 闫晓鹏, 刘强, 郝新红, 章鸿运 (1263)
- 结构引导的图像修复 胡凯, 赵健, 刘昱, 牛余凯, 姬港 (1269)
- 基于 SVC 视频流的低复杂度多播组分解算法 杨静璇, 徐桢 (1278)
- 小天体环的轨道动力学 吴太龙, 王悦 (1287)
- 基于平面变换的高精度相机标定方法 彭妍, 郭君斌, 于传强, 柯冰 (1297)
- 基于注意力机制的光伏组件热斑检测算法 樊涛, 孙涛, 刘虎 (1304)
- 一种新的气动伺服弹性失稳模式的机理分析 姜宇, 杨超, 吴志刚 (1314)
- 面向扑翼翼面运动参数的优化设计 吴越, 谢长川, 杨超, 安朝 (1324)
- 空间光学遥感器高精度 6-DOF 并联平台设计与实现
..... 梁凤超, 谭爽, 黄刚, 范建凯, 林喆, 康晓军 (1332)

CONTENTS

A power budgeting method for dark silicon chips based on task mapping
..... LI Xin, LI Zhi, ZHOU Wei, WU Ruiqi, TANG Haoran, CHEN Yehang (1115)

Partial observable Petri nets fault diagnosis with quantum Bayesian learning
..... LIU Jiufu, ZHANG Xinzhe, WANG Hengyu, A. M. TOMAS DIAS, WANG Zhisheng, YANG Zhong (1125)

Inertial pedestrian navigation algorithm based on zero velocity update and attitude self-observation
..... DAI Hongde, ZHANG Xiaoyu, ZHENG Baidong, DAI Shaowu, ZHENG Weiwei (1135)

A rapid path planning method for multiple UAVs to cooperative strike
..... CHEN Qingyang, XIN Hongbo, WANG Yujie, TANG Zhongnan, JIA Gaowei, ZHU Bingjie (1145)

Occlusion recognition algorithm based on multi-resolution feature auto-selection
..... XIE Xiangying, LAI Guangzhi, NA Zhixiong, LUO Xin, WANG Dong (1154)

Design and characteristics of reverse direct-acting high-pressure reducing valve for pneumatic actuator
..... LIU Yanbin, WANG Xuesheng, QIN Xinya, WANG Hao, CHEN Qinzhu, ZHAO Sai (1164)

End face deformation and friction and wear of high-speed dry friction mechanical seal
..... MA Runmei, ZHAO Xiang, CHEN Xiaozhu, LI Shuangxi, YANG Haichao (1174)

Effect of suction flow rate on performance of catalytic inerting system
..... WANG Chenchen, PAN Jun, WANG Xiangyang, DUAN Weijie (1183)

Numerical investigation on evolution of T-S wave on a two-dimensional compliant wall with finite length
..... HONG Zheng, YE Zhengyin (1190)

Gust load alleviation analysis based on vortex lattice method in state-space form
..... YANG Lan, AN Chao, XIE Changchuan, YANG Chao (1200)

Decoupled active disturbance rejection control method for magnetically suspended rotor based on state feedback
..... YIN Zengyuan, CAI Yuanwen, REN Yuan, WANG Weijie, CHEN Xiaocen, YU Chunmiao (1210)

Target tracking algorithm based on efficient attention and context awareness
..... BAI Luo, ZHANG Hongli, WANG Cong (1222)

A DDQN-based mixed-criticality messages scheduling method for network-on-chip
..... LI Guoliang, LI Qiao, XU Yajun, XIONG Huagang (1233)

VWKNN location fingerprint positioning algorithm based on improved discrete coefficient
..... XU Tian, HE Jingsha, ZHU Nafei, DENG Wanhang, WU Shuang, TA Yongjun (1242)

Design and optimization of large-stroke decoupled three-translational micro-positioning platform
..... LIN Miao, MENG Gang, JU Yongjian, XU Weisheng, CAO Yi (1252)

PM based super-resolution method of azimuth detection for electromagnetic vortex wave fuze
..... SUN Libin, YAN Xiaopeng, LIU Qiang, HAO Xinhong, ZHANG Hongyun (1263)

Images inpainting via structure guidance
..... HU Kai, ZHAO Jian, LIU Yu, NIU Yukai, JI Gang (1269)

Low computational-cost multicast subgrouping for SVC streams
..... YANG Jingxuan, XU Zhen (1278)

Orbital dynamics of rings of small bodies
..... WU Tailong, WANG Yue (1287)

Calibration method for high precision camera based on plane transformation
..... PENG Yan, GUO Junbin, YU Chuanqiang, KE Bing (1297)

Hot spot detection algorithm of photovoltaic module based on attention mechanism
..... FAN Tao, SUN Tao, LIU Hu (1304)

Mechanism analysis of a new aeroservoelastic instability mode
..... JIANG Yu, YANG Chao, WU Zhigang (1314)

Optimal design of motion parameters of flapping wing
..... WU Yue, XIE Changchuan, YANG Chao, AN Chao (1324)

Design and implementation of a high precision 6-DOF parallel platform for a space optical remote sensor
..... LIANG Fengchao, TAN Shuang, HUANG Gang, FAN Jiankai, LIN Zhe, KANG Xiaojun (1332)

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

DOI: 10.13700/j.bh.1001-5965.2021.0011

基于任务映射的暗硅芯片功耗预算方法



李鑫^{1,*}, 李智¹, 周巍¹, 吴瑞祺², 唐浩然³, 陈业航⁴

(1. 西北工业大学 电子信息学院, 西安 710072; 2. 西北工业大学 教育实验学院, 西安 710072;
3. 西北工业大学 软件学院, 西安 710072; 4. 西北工业大学 计算机学院, 西安 710072)

摘 要: 暗硅系统功耗预算问题可被归类为一种 NP-hard 问题, 针对其存在的提高芯片均温与降低通信成本 2 个对立优化目标, 提出了一种基于任务映射的暗硅芯片功耗预算方法。为降低计算复杂度, 基于率先映射高通信量并对后续映射影响较小的任务的规则建立模型, 将任务图转换为最大生成树的形式, 以优先级值的大小决定任务进行映射的先后顺序。在稳态情况下逐个进行核心寻优, 将排序后的任务放置于合适的核心位置, 并以凸二次规划问题形式对已确定映射核心位置的功耗预算进行求解。实验表明: 针对 12 开启核心的 36 核心系统, 与经典的热安全功耗预算方法相比, 所提方法将总功耗预算提高了 11.8%, 通信能耗降低了 38.2%。

关键词: 暗硅; 功耗预算; 任务映射; 多核系统; 动态热管理

中图分类号: TN47

文献标志码: A

文章编号: 1001-5965(2022)07-1115-10

由于登纳德缩放 (Dennard scaling) 定律的失效, 继续缩小芯片特征尺寸会引发严重的高功耗密度问题。因此, 相当一部分芯片核心必须关闭以防止可能的过热和永久损坏, 这种现象被称为暗硅 (dark silicon)^[1]。暗硅现象带来了一种从传统低功耗设计到功耗预算设计的范式转变。目前, 多核系统设计模式的优化目标已转变为: 在热约束条件下, 估计芯片的功耗上限值 (称为功耗预算^[2]), 并将其分配至各个核心或功能单元, 从而使系统的整体性能达到最优^[3-4]。芯片设计人员通常采用的功耗预算是热设计功耗 (thermal design power, TDP), 该方法的目标是确保芯片的绝对热安全, 不能根据芯片的运行状态做出调整, 因而其给出的功耗预算在绝大多数情况下都是极为保守且次优的。考虑到 TDP 的诸多限制, Paganì 等^[5]提出一种取代 TDP 的热安全功耗 (ther-

mal safe power, TSP) 技术。如果预先给定了开启核心的分布, TSP 将根据该分布计算出最大允许功耗预算, 否则, TSP 将给出最坏开启核心分布情况下的最大允许功耗预算。然而, TSP 是一种与任务无关的静态功耗预算计算方法, 其不能对芯片运行过程中开启核心的数量和分布进行动态分配, 导致所有开启核心分配统一的功耗预算。为此, Wang 等^[6]使用模型预测方法, 提出一种适应瞬态温度变化的新型功耗预算技术, 并在随后提出一种基于贪心思想的瞬态功耗 (greedy based dynamic power, GDP)^[7] 预算技术, 该方法通过寻找次优的核心分配以降低计算复杂度, 从而快速获得功耗预算问题较好的次优解, 但未考虑任务间通信延迟等问题, 不足以保证芯片实际运行的性能。

近年来, 多级缓存 (multi-level caches, MLC)

收稿日期: 2021-01-11; 录用日期: 2021-05-09; 网络出版时间: 2021-06-02 16:01

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210602.1053.001.html

基金项目: 国家自然科学基金 (61501377); 陕西省自然科学基金基础研究计划 (2021JM-074)

* 通信作者。E-mail: xinli@nwpu.edu.cn

引用格式: 李鑫, 李智, 周巍, 等. 基于任务映射的暗硅芯片功耗预算方法[J]. 北京航空航天大学学报, 2022, 48(7): 1115-1124.
LI X, LI Z, ZHOU W, et al. A power budgeting method for dark silicon chips based on task mapping[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1115-1124 (in Chinese).

和片上网络(network on chip, NoC)已经分别取代单级缓存和总线,成为暗硅芯片设计的标准组件^[8-10]。然而,上述功耗预算技术均忽视了这些组件,并未考虑 NoC 通信能耗等问题,缺乏合理的任务映射策略^[11-15]来权衡暗硅系统的性能和能耗。目前,大部分工作针对暗硅系统的任务映射和功耗预算两方面问题分别提出了多种优化策略,但少有将这 2 个相关问题结合起来考虑。因此,本文提出了一种基于任务映射的功耗预算方法。首先,对基于 NoC 的暗硅系统模型进行了具体分析和研究。其次,定义了应用于任务放置在不同核心位置的一一映射过程,并综合考虑通信能耗与温度分布的要求建立功耗预算问题的组合优化目标。然后,设计应用任务的优先级排序方法,通过定量计算的优先级值大小决定映射顺序。最后,设计逐个核心寻优的迭代过程,并基于确定的映射核心位置,以凸二次规划问题形式对此时的功耗预算进行求解。

1 模型建立

1.1 结构模型

本文提出的功耗预算方法基于 NoC 架构的暗硅系统,其采用图 1 所示的网格型 2D-Mesh 拓

扑结构形式,主要由资源节点、路由节点和通信链路组成。每个路由节点通过十字网格相连,具有很好的扩展性。除了边界节点,其余路由节点都有 5 个端口,其中,4 个端口和相邻的路由节点相连,1 个端口与可用处理单元(processing element, PE)相连。

由于芯片内部的功耗与每个核心的执行频率紧密相关,为设计功耗预算方法,本文考虑以核心级别来建立暗硅系统版图。一个完整的核心模块主要由缓存(Cache)、可用处理单元(PE)和路由单元(R)组成,如图 2(a)所示,图 2(b)则展示了一个 9 核心(Core)系统的版图设计情况。

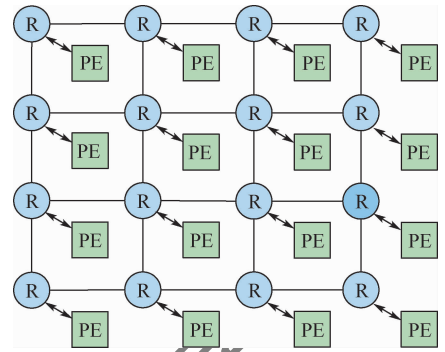


图 1 2D-Mesh 拓扑结构

Fig. 1 2D-Mesh topology structure

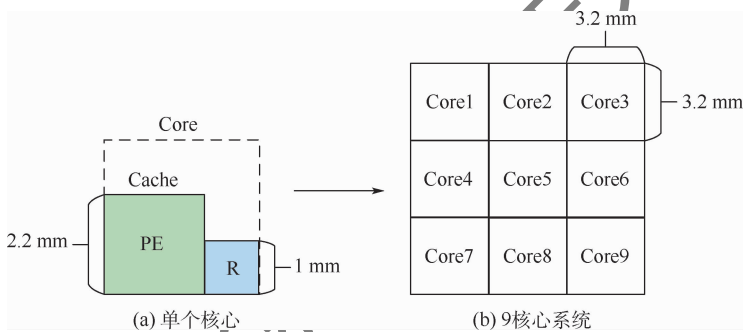


图 2 平面版图设计

Fig. 2 2D layout design

1.2 能耗模型

核心间的通信互连会产生额外的能量消耗。因此,在基于 NoC 的系统中,能量消耗主要由资源模块和网络共同产生,包括计算和通信 2 部分,如下:

$$E_{\text{total}} = E_{\text{comp}} + E_{\text{comm}} \quad (1)$$

式中: E_{comp} 为计算能耗; E_{comm} 为通信能耗。在 NoC 系统中,数据包是网络通信传送的基本单元,由多个微片组成。一个微片从位置 m 传输到位置 n 所需的能耗 $E_{\text{flit}}^{m,n}$ 包括微片在路由器转发所需能耗和链路传输所需能耗,如下:

$$E_{\text{flit}}^{m,n} = (h_{m,n} + 1)E_{\text{router}} + h_{m,n}E_{\text{link}} \quad (2)$$

式中: $h_{m,n}$ 为位置 m 与位置 n 之间的链路距离;

E_{router} 为微片经过路由器所需的平均能耗; E_{link} 为微片通过 2 个路由器间链路所需能耗。

因此,对于全部通信任务集合 C ,所需要消耗的总通信能耗如下:

$$E_{\text{comm}} = \sum_{\omega_{i,j} \in C} \frac{\omega_{i,j}}{B} E_{\text{flit}}^{m,n} \quad (3)$$

式中: $\omega_{i,j}$ 为位置 m 的任务 i 与位置 n 的任务 j 之间的通信量; B 为通信链路的位宽。

1.3 温度模型

热电阻电容网络模型(thermal resistance-capacitance network model)通常被称为 RC 热阻网络,是目前最广泛采用的热模型,其基于电学和热传导之间的二元特性,将芯片和冷却方案(散热

片等)的不同部分以 M 个热节点的方式表示为一阶微分方程系统,如下:

$$AT' + B(T - T_{amb}) = P \quad (4)$$

式中:矩阵 $A = [a_{i,j}]_{M \times M}$ 为热电容值;矩阵 $B = [b_{i,j}]_{M \times M}$ 包含纵向和横向相邻节点之间的热导率值;列向量 $T = [t_i]_{M \times 1}$ 为热节点上的温度;列向量 $T' = [t'_i]_{M \times 1}$ 为每个热节点上的温度相对于时间的一阶导数;列向量 $P = [p_i]_{M \times 1}$ 包含每个热节点上的功耗值; T_{amb} 为环境温度值。

若热节点中包含 N 个核心节点 ($N \leq M$),可将核心温度变化向量定义为 $\tilde{T} = [\tilde{t}_i]_{N \times 1}$,核心功耗向量为 $\tilde{P} = [\tilde{p}_i]_{N \times 1}$,有如式(5)所示的关系:

$$\begin{cases} \tilde{T} = H(T - T_{amb}) \\ H^T \tilde{P} = P \end{cases} \quad (5)$$

式中: $H = [h_{i,j}]_{N \times M}$ 为从全部节点中提取核心节点的选择矩阵。

在稳态情况下,温度模型不存在温度对时间的一阶微分项,即满足 $dT/dt = 0$ 。由于 B 为对称正定矩阵(可逆),结合式(4)和式(5),描述核心温度变化与核心功耗间转换关系的模型如下:

$$\tilde{T} = R\tilde{P} \quad (6)$$

式中: $R = HB^{-1}H^T = [r_1 \ r_2 \ \dots \ r_N]_{N \times N}$ 为稳态情况下温度和功耗间的转换矩阵。

2 基于多目标权衡的映射选择

2.1 任务映射

暗硅时代,芯片上仅有部分核心可以开启,因此需要为运行的应用任务选择特定核心进行映射。如图 3(a)所示,应用任务均以通信任务图(communication task graph, CTG)描述,定义为一

个由三元组 $G = G(K, D, \omega)$ 描述的有向无环图(directed acyclic graph, DAG)。其中, K, D, ω 分别为应用中子任务、有向边缘、边缘权重的向量。具体来说, $k_i \in K (i \in [1, K])$ 为应用中的一个可并行执行的子任务, K 为全部任务数。有向边 $d_{i,j} \in D (i, j \in [1, K])$ 为应用中存在任务 k_i 到任务 k_j 的通信需求。每条边上的权重 $\omega_{i,j} \in \omega$ 为需要传递的通信量。

如图 3(b)所示,基于 NoC 的暗硅系统被定义为架构特征图(architecture characterization graph, ACC) $G' = G'(N, L)$ 。其中, N 和 L 分别为 NoC 系统中核心和核心上路由间的链路向量。以 2D-Mesh 的 NoC 拓扑结构为例,假定每个核心模块包括一个可用处理单元 PE 和一个路由单元 R。若该系统具有 $N_x \times N_y = N$ 个核心,则每个核心 $n_m \in N (m \in [1, N])$ 对应一个二维坐标 (x, y) ,且满足 $x \in N_x, y \in N_y$ 。 $l_{m,n} \in L (m, n \in [1, N])$ 则表示了该 NoC 相邻路由之间的链路连接。

本文中,任务映射被定义为应用于任务 k_i 放置在不同核心 n_m 的一一对应过程,如图 3(c)所示。因此,总核心数目为 N ,任务数即开启核心数目为 K 。映射过程如下:

$$\begin{aligned} \text{map: } K &\rightarrow N, \text{map}(k_i) = n_m \\ \forall k_i \in K, \forall n_m \in N \end{aligned} \quad (7)$$

根据映射位置的远近,任务映射可分为连续映射和非连续映射,如图 4 所示。图 4(a)中,连续映射方法考虑将同一应用的多个子任务映射到彼此相连的核心位置,但这会导致芯片上产生多个热点区域。图 4(b)中,非连续映射方法则放宽了对映射区域的限制,但该方法在优化温度分布的同时,由于子任务间通信距离的延长导致通信能耗增加。

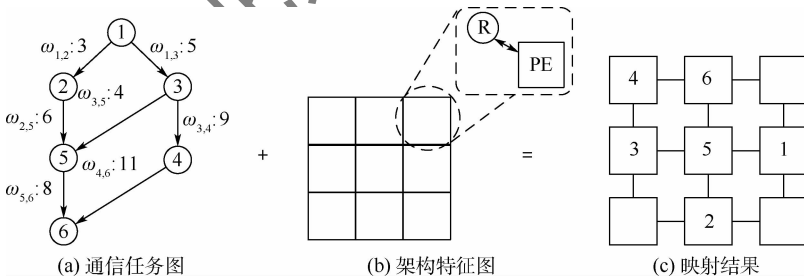


图 3 任务映射示例

Fig. 3 An example of task mapping

2.2 功耗预算的优化目标

任务映射方式与功耗预算问题息息相关,但经典的功耗预算估计方法缺乏核心位置选择能力(如 TSP^[5]),或仅考虑优化温度分布的映射(如 GDP^[7]),均不足以保证系统的最大性能。为此,

本文针对热约束和通信成本两方面对功耗预算估计与分配问题的优化目标进行具体分析。

一方面,由于采用 RC 热阻网络进行温度建模,可以将等效散热电路表示为等效热敏电阻的串联,如图 5 所示。 $R_{chip}, R_{TIM}, R_{spreader}, R_{sink}$ 和 R_{conv}

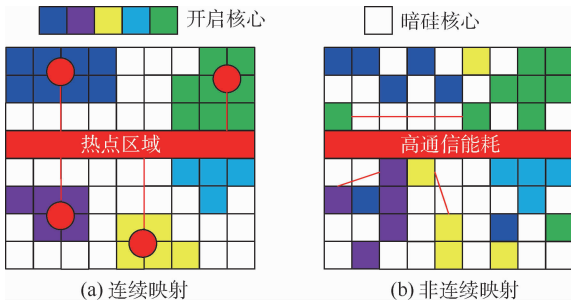


图4 任务映射方式

Fig. 4 Task mapping method

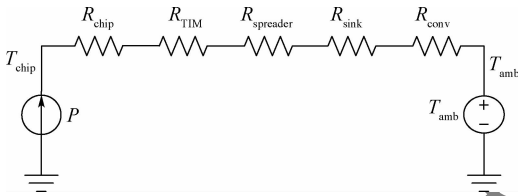


图5 芯片的等效散热电路

Fig. 5 Equivalent heat dissipation circuit of chip

分别为芯片、热界面材料(thermal interface material, TIM)、散热底座、散热片和散热片与空气间对流的热阻。热回路两端的热节点分别表示芯片均温 T_{chip} 和环境温度 T_{amb} 。在 T_{chip} 处连接了一个等效电流源,其电流值为芯片的总功耗 P 。此外,在 T_{amb} 处连接了等效电压源,代表环境温度。由于环境温度 T_{amb} 固定,若要提高芯片的平均温度,则必须增加加热电阻电路中的电流。换句话说,较高的芯片平均温度对应较高的芯片总功耗。因此,将任务分散到相距较远的多个核心上会产生更好的热分布,从而可设置更高的功耗预算。

另一方面,将任务放置到过于分散的核心位置,在提高总功耗预算的同时,通信成本迅速升高导致的通信延迟会令并行计算任务进入等待,反而损害系统性能。由式(2)和式(3)可知,网络上完成一次数据传输所需能耗主要与数据的传输距离和任务本身通信量的大小有关。对于2D-Mesh的NoC拓扑结构,数据传输距离为两位置间的曼哈顿距离。为此,本文参考文献[16]中提出的平均加权曼哈顿距离(average weighted Manhattan distance, AWMD)作为评估通信成本的指标,如下:

$$AWMD = \frac{\sum_{d_{i,j} \in D} \omega_{i,j} h(n_m, n_n)}{\sum \omega_{i,j}} \quad (8)$$

式中:对于一个通信量为 $\omega_{i,j}$ 的有向边 $d_{i,j} \in D$, $h(n_m, n_n)$ 表示其连接的任务 k_i 和 k_j 所对应核心位置 n_m 和 n_n 之间的曼哈顿距离。

综上所述,功耗预算问题在提高芯片均温与

降低通信成本两方面的优化目标均与任务映射到的核心位置距离息息相关,并且相互对立。因此,功耗预算方法应针对应用任务对于计算和通信两方面的不同需求,设计不同的映射选择,对两方面优化目标进行权衡,实现组合优化的目标。如式(6)所示,针对稳态温度情况,有 $\tilde{T} = R\tilde{P}$ 。在温度阈值 T_{th} 下,此稳态系统的功耗预算及对应的映射选择的求解形式如下:

$$\begin{cases} \text{minimize } \|\tilde{T}_{th} - R\tilde{P}\|_2 \quad \&\& \text{AWMD} \\ \text{s.t. } R\tilde{P} \leq \tilde{T}_{th} \end{cases} \quad (9)$$

式中: $\tilde{T}_{th} = T_{th} - T_{amb}$ 为升温阈值向量;“ \leq ”为向量不等号。因此,本文可依据总功耗预算和AWMD指标进行稳态功耗预算方法设计与评估。

暗硅芯片中存在着大量潜在的开启核心分布组合,对于8开启核心的16核心系统而言,存在 $C_{16}^8 = 12\ 870$ 种可能情况,确定每个开启核心和应用子任务的对应关系也存在 $A_{16} = 40\ 320$ 种可能。由此看出,暗硅系统功耗预算估计与分配问题具有 $O[N!/(N-K)!]$ 的计算复杂度,可归类为一种NP-hard问题。因此,本文设计了应用任务优先级排序方法和任务到核心的逐个寻优方法,以降低计算复杂度的同时获取较优的功耗预算估计与分配结果。

3 组合优化功耗预算估计与分配技术

3.1 任务排序

由于确定暗硅芯片开启核心和应用子任务的一一对应关系具有较高的计算复杂度,本文考虑以一定的优先级顺序对任务进行依次放置。为确定任务的优先级顺序,需针对应用任务的子任务数目和通信边缘特征进行具体分析,具体流程如图6所示。图6(a)展示了具有6个子任务和7个通信边缘的通信任务图示例。在进行优先级排序前,考虑将原始通信任务图转换为最大生成树的形式。基于图论中的普里姆(Prim)算法,先选择平均通信需求最大的任务为根进行转换,以 $D(k_i)$ 表示 k_i 的连接边缘数,任务平均通信需求 $A(k_i)$ 为

$$A(k_i) = \sum_{d_{i,j} \in D} \frac{\omega_{i,j}}{D(k_i)} \quad k_i \in K \quad (10)$$

从根任务开始,Prim算法寻找与其相连的具有最大通信需求的任务,并将其和对应边缘添加到树中。以此类推,直到树中包含所有任务为止。

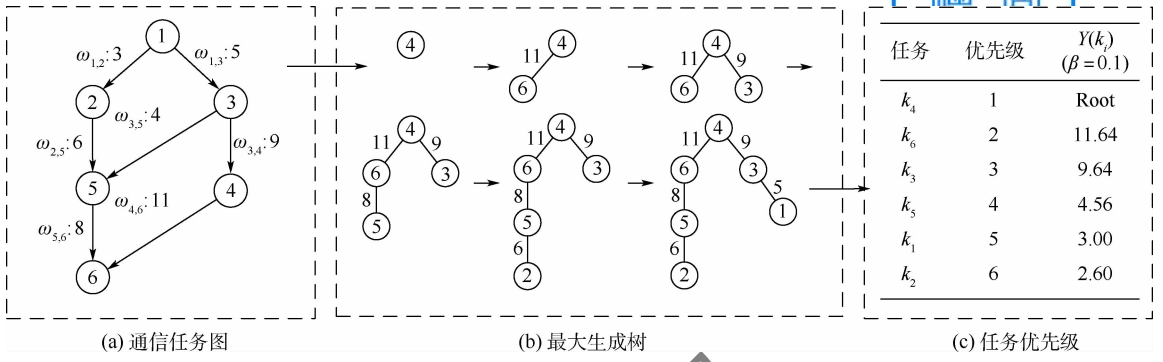


图6 任务优先级排序流程

Fig. 6 Flowchart of task priority ranking

构建最大生成树的流程如图6(b)所示。为减小通信损失,任务优先级模型需要在考虑高通信量的同时选择对后期映射影响较小的任务率先进行映射,因此,需要制定同时考虑树中层数(通信量大小)和后代最大路径长度(对后期映射影响大小)的映射规则。基于该规则建立模型,通过定量计算的优先级值大小决定映射顺序。以 $Y(k_i)$ 表示 k_i 任务基于生成树的任务映射优先级,如下:

$$Y(k_i) = \omega(k_i) \left(\frac{1}{S(k_i)} + \frac{\beta}{\sqrt{1+C(k_i)}} \right) \quad (11)$$

$k_i \in K$

式中: $\omega(k_i)$ 为生成树中该任务与其父任务间的通信量; $S(k_i)$ 为该任务在生成树中的层数; $C(k_i)$ 为该任务到其后代的最大路径长度,并以 β 参数决定其对映射顺序的影响。若令 $\beta=0.1$,应用任务的优先级示例的结果如图6(c)所示。

3.2 核心映射

确定任务的优先级顺序后,还需依据功耗预算问题的优化目标,将任务映射到适当的核心位置。鉴于任务到核心映射的高寻优计算复杂度,考虑建立一种启发式的贪心映射方法。

为此,首先定义 $\text{map}(k_k, n_n^{k_k})$ 为任务到核心的映射过程。其中, k_k 为上述任务排序方法获得的从大到小排序后的第 $k(k \in [1, K])$ 个任务, $n_n^{k_k}$ 为 k_k 任务所对应的具有 N 个核心的NoC系统从上到下、从左到右依次编号的第 $n(n \in [1, N])$ 个核心位置。假定暗硅不分配功耗预算,并且在之前迭代中已经确定位置的开启核心在下次迭代中不再被重复使用。本节提出的稳态逐个核心寻优方法可分为以下3个步骤:

步骤1 寻找第1个开启核心位置。

寻找第1个核心位置情况具有某些特定约束,可将优化目标式(9)改写为

$$\begin{cases} \text{map}(k_1, n_n^{k_1}) : \\ \text{minimize } \|\tilde{T}_{th} - R\tilde{P}\|_2 \\ \text{s. t. } \text{level}(\tilde{P}) = 1, \text{AWMD} = 0, R\tilde{P} \leq \tilde{T}_{th} \end{cases} \quad (12)$$

式中: $\text{level}(\tilde{P})$ 表示 \tilde{P} 中非零值的个数即开启核心数目。由于仅放置根任务,不存在边缘即任务间的通信量,该核心映射的唯一要求是最大化向量 \tilde{P} 中仅存的一个非零值 $\tilde{p}_n (n \in [1, N])$ 。

为简化问题,对图7中所示 $n_p^{k_1}, n_q^{k_1}$ 两种可能位置进行分析。 r_p, r_q 为 R 矩阵中对应这2种可能位置的列向量。在不违反热约束的前提下,可以看出 $n_p^{k_1}, n_q^{k_1}$ 两个位置可以达到的最大温度分别为 $r_p \tilde{p}_p$ 和 $r_q \tilde{p}_q$ 。为提高系统的功耗预算,应选取使整体温度更接近温度阈值的位置,即选取温度向量 $\tilde{T}_{th} - r_p \tilde{p}_p$ 和 $\tilde{T}_{th} - r_q \tilde{p}_q$ 中较小值对应的位置,等价于最小化映射矩阵列向量与升温阈值向量的夹角。对于所有可能位置情况该结论仍保持成立,由于 $\tilde{p}_n > 0$,则 \tilde{T}_{th} 与 r_n 同向,夹角满足 $0^\circ \sim 90^\circ$ 之间。因此,定义 \hat{r}_n 为 r_n 的单位向量,依次计算 \tilde{T}_{th} 与 \hat{r}_n 的内积 $\langle \hat{r}_n, \tilde{T}_{th} \rangle$,其中具有最大内积的核心位置即第1个开启核心位置。假设第1个开启核心为 n_{first} ,计算其对应功耗向量 \tilde{P}_{first} 。

步骤2 确定第2个开启核心位置并计算新的功耗预算。

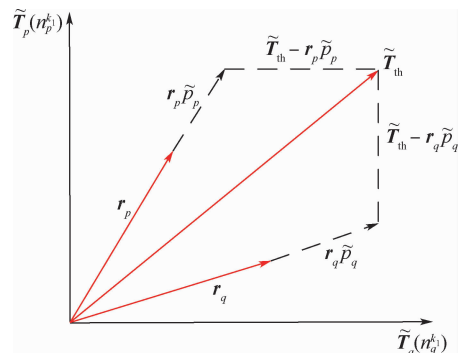


图7 寻找第1个开启核心示意图

Fig. 7 Diagram of finding the first active core

为避免核心的重复选择,确定第2个开启核心位置前,需要从升温阈值 \tilde{T}_{th} 中减去第1个确定开启核心所引起的温度上升,定义剩余升温阈值向量 $T_{rm} = \tilde{T}_{th} - R_{first} \tilde{P}_{first}$,其中, R_{first} 由已选择位置(当前仅有 n_{first})对应的 R 矩阵中列向量组成。同上所述,可通过求解内积 $\langle \hat{r}_n, T_{rm} \rangle$ 获得令系统温度更接近温度阈值的第2个开启核心位置,如下:

$$\begin{cases} \text{map}(k_2, n_n^{k_2}): \\ \text{maximize}(\langle \hat{r}_n, T_{rm} \rangle - \alpha \cdot \text{AWMD}) \\ \text{s. t. AWMD} = h(n_{first}, n_n^{k_2}) \end{cases} \quad (13)$$

此时,依据任务优先级放置的前2个任务间存在通信,因仅存在一条边缘,两者的 AWMD 即为 n_{first} 位置与该位置的曼哈顿距离。寻找第2个开启核心的优化问题如式(13)所示,需要同时考虑提高均温和降低 AWMD 两方面的需求。由于两者不能兼顾,通过设置一个正参数 α 表示方法优化的倾向,如下:

$$\alpha = \frac{\max \langle \hat{r}_n, T_{rm} \rangle - \min \langle \hat{r}_n, T_{rm} \rangle}{\max \text{AWMD} - \min \text{AWMD}} \quad (14)$$

步骤3 将步骤1和步骤2扩展到一般迭代步骤。

为找到第 k 个任务的核心位置,定义 $T_{rm}(k-1)$ 为前 $k-1$ 个任务映射后的剩余升温阈值, \tilde{D} 为所有先行任务的映射核心和目标核心存在的边缘值向量。基于式(8),第 k 个任务的映射寻优迭代过程如下:

$$\begin{cases} \text{map}(k_k, n_n^{k_k}): \\ \text{maximize}(\langle \hat{r}_n, T_{rm}(k-1) \rangle - \alpha \cdot \text{AWMD}) \\ \text{s. t. AWMD} = \frac{\sum_{d_{i,j} \in \tilde{D}} \omega_{i,j} h(n_m, n_n)}{\sum \omega_{i,j}} \end{cases} \quad (15)$$

每次确定核心位置后,还需确定此时的功耗预算。将 R 的相应 k 列收集到矩阵 R_k 中,并把这 k 个核心的功耗预算表示为向量 \tilde{P}_k ,通过求解 $\|\tilde{T}_{th} - R_k \tilde{P}_k\|_2$ 的最小值可以获得当前功耗预算,其展开形式为

$$\begin{cases} \text{minimize} \tilde{P}_k^T R_k^T R_k \tilde{P}_k - 2 \tilde{T}_{th}^T R_k \tilde{P}_k + \tilde{T}_{th}^T \tilde{T}_{th} \\ \text{s. t. } R_k \tilde{P}_k \leq \tilde{T}_{th} \end{cases} \quad (16)$$

为解决该受约束的最优化问题,可令 $\tilde{P}_k = x$, $R_k^T R_k = Q$, $R_k^T \tilde{T}_{th} = c$, $R_k = w$, $\tilde{T}_{th} = b$,则式(16)被转化为一个典型的二次规划问题(quadratic programming, QP)形式,简化后为

$$\begin{cases} \min_x \frac{1}{2} x^T Q x + c^T x \\ \text{s. t. } w x \leq b \end{cases} \quad (17)$$

由于 $R_k^T R_k = Q$ 为对称矩阵,并且如式(18)所示,其对任意非零 x 满足:

$$x^T R_k^T R_k x = (R_k x)^T R_k x = \|R_k x\|_2^2 \geq 0 \quad (18)$$

因此, Q 为对称半正定矩阵,该二次规划问题的目标函数为凸函数,有全局最小值,可在多项式时间内获得当前各开启核心的功耗预算。

本文功耗预算方法的具体流程如图8所示。在通信任务图和架构特征图的基础上,获取任务映射的优先级顺序和典型封装结构下的系统稳态温度模型。以此进行逐个核心寻优的功耗预算方法,并将其计算得到的各核心功耗预算值输入暗硅系统,用以收集通信状态信息和温度状态信息等反馈,以便计算下一次调用时的功耗预算选择。此外,本文功耗预算方法的伪代码如下:

输入: $N, K, G, G', R, \tilde{T}_{th}, \alpha$ 。

输出: $\text{map}(k_k, n_n^{k_k}), \tilde{P}$ 。

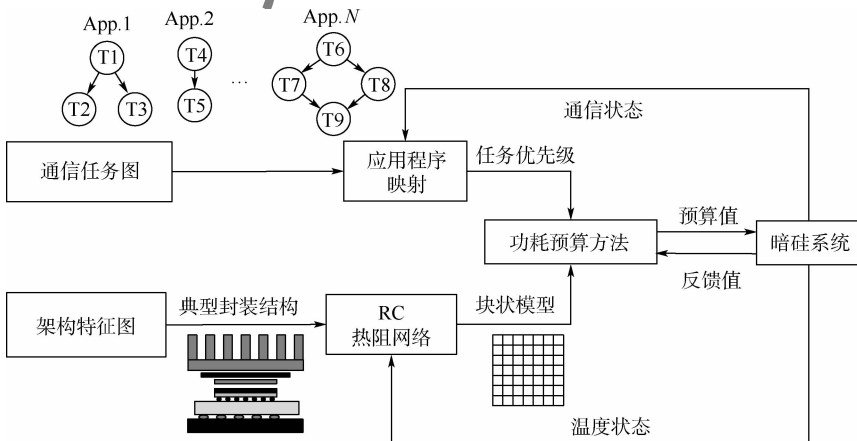


图8 本文功耗预算方法流程

Fig. 8 Flowchart of the proposed power budgeting method

1. 初始化: $\mathbf{K} = [\mathbf{0}]_{K \times 1}$, $\mathbf{N} = [\mathbf{0}]_{K \times 1}$, $\mathbf{R}_k = [\mathbf{0}]_{N \times K}$, $\mathbf{T}_{rm} = \tilde{\mathbf{T}}_{th}$
2. 基于 Prim 算法建立最大生成树
3. for $i = 1$ to K do
4. 根据式(11)计算任务优先级 $Y(k_i)$
5. end for
6. for $k = 1$ to K do
7. 将任务以优先级顺序从大到小放置于 $\mathbf{K}(k)$ 中
8. end for
9. for $k = 1$ to K do
10. for $n = 1$ to N do
11. 根据式(12)~式(15)进行任务到核心的映射
12. if $\text{map}(k_k, n_n^{k_k}) > \text{map}(k_k, N(k))$
13. then $N(k) = n_n^{k_k}$
14. end if
15. end for
16. $\mathbf{R}_k = [\mathbf{R}_{k-1}, r_{N(k)}]_{N \times K}$
17. 根据式(16),对当前功耗预算 $\tilde{\mathbf{P}}_k$ 进行求解
18. $\mathbf{T}_{rm} = \tilde{\mathbf{T}}_{th} - \mathbf{R}_k \tilde{\mathbf{P}}_k$
19. end for
20. 获得最终映射结果和功耗预算

4 实验结果

4.1 实验环境

本文使用的仿真平台是基于 NoC 的 MPSoC, 使用版图绘制工具 ArchFP 构建核心版图,并基于热仿真工具 HotSpot 中的默认配置构建热模型。

从而进行温度分析。依据相关参数配置,采用任务图生成器 TGFF 生成随机任务图序列,作为实验输入的 benchmark 集。

4.2 结果分析

将本文方法与 Pagani 等^[5]提出的 TSP 方法及 Wang 等^[7]提出的 GDP 方法进行对比。由于核心间的通信距离会随着芯片尺寸设置的不同而变化,为便于进行实验对比,默认相邻 2 个核心间的曼哈顿距离为 1,无量纲。由于 TSP 缺少核心位置选择的能力,GDP 方法为仅考虑温度分布的理想模型,二者均未考虑任务本身的影响。因此,在本文实验设置中,2 种对比方法均将应用程序任务按顺序放置于第 1 个可用核心位置。以阈值温度为 80°的条件下,分别将这 3 种方法应用于 9~64 核心的暗硅系统,对得到的功耗预算估计与分配结果进行详细分析。

上述 3 种方法在热约束下总功耗预算和 AWMD 两方面的计算结果如表 1 所示。表中,每种开启核心数量的结果均为 10 个不同 TGFF 生成的随机任务图测试结果的均值。基于表 1 中数据,绘制 3 种方法在相同暗硅比例系统(即开启一半核心情况:8 开启核心的 16 核心系统、18 开启核心的 36 核心系统和 32 开启核心的 64 核心系统)中的 2 种指标对比,如图 9 所示。由表 1 和图 9 可知,相同暗硅比例系统对比时,本文方法始终具有最低的 AWMD,即最低的通信能耗,并且可分配的总功耗预算高于 TSP 方法,与 GDP 方法相比存在一定的不足;随着系统核心数目和开启核心数目的增长,本文方法提供的总功耗预算也随之增多,但仍然处于 TSP 和 GDP 方法之间。同

表 1 三种方法总功耗预算与 AWMD 的计算结果

Table 1 Results of total power budget and AWMD using three methods

核心数	开启核心数	TSP 方法 ^[5]		GDP 方法 ^[7]		本文方法	
		功耗预算/W	AWMD	功耗预算/W	AWMD	功耗预算/W	AWMD
9	2	82.00	1.00	86.01	2.00	83.89	1.00
	4	128.13	1.47	137.69	2.00	134.52	1.02
	8	179.04	2.03	188.53	2.26	187.50	1.30
16	4	127.73	1.47	138.02	2.38	133.70	1.02
	8	179.78	2.17	199.80	2.70	193.94	1.41
	12	210.47	2.20	230.10	2.90	227.24	1.54
25	8	179.17	2.27	198.97	2.82	196.63	1.27
	12	211.70	2.49	239.15	3.61	235.97	1.69
	18	241.39	2.74	268.72	3.60	267.45	1.91
36	12	213.96	2.97	247.25	3.92	238.07	1.80
	18	242.30	3.29	281.56	4.21	276.37	2.27
	24	269.65	3.34	319.59	5.10	316.69	2.58
64	18	254.57	3.68	297.45	5.99	291.76	2.36
	24	265.01	4.33	319.35	6.39	311.45	2.63
	46	313.88	5.02	361.22	6.92	357.12	2.77

时,本文方法的 AWMD 不断增加但一直保持在较低水平。而忽视了应用程序子任务间通信的 TSP 和 GDP 方法,随着可用核心位置的增多通信能耗增长明显。

图 10 展示了 2 种实际多媒体基准程序 MWD 和 MPEG-4 的通信任务图。可以看出,两者具有相同子任务数目但边缘情况不同。针对 36 核心系统,对 TGFF 生成的 12 子任务应用(定义为tgff-12)、MWD 和 MPEG-4 三种不同应用情况进行具体分析。功耗预算对比结果如图 11(a)所示。对于 TSP 方法、GDP 方法而言,这 3 种情况下的核

心映射放置均相同即功耗预算保持不变,显然是不符合实际需求的。本文方法则根据子任务间通信特性获得不同的映射选择。与 MWD 相比,MPEG-4 中存在多个低通信的子任务,因此这些任务可以被映射到更为分散的位置,令其分配的总功耗预算更大。同时,图 11(b)给出了 AWMD 对比结果。由图 11(b)中可知,本文方法有着最低的通信能耗。对于 TSP 和 GDP 方法,虽然任务放置的核心位置相同,但由于通信边缘的差异,导致具有高通信量的 MPEG-4 具有更大的 AWMD。

表 2 给出了本文方法与其他方法在指标变化幅度方面的对比。与经典 TSP 方法相比,本文方法将总功耗预算平均提高 11.8%,且 AWMD 降低 38.2%;与新型 GDP 方法相比,本文方法平均牺牲 3.3%的功耗预算,将 AWMD 降低 56.6%。

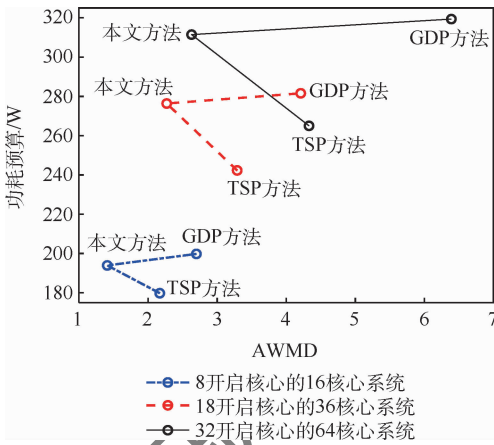
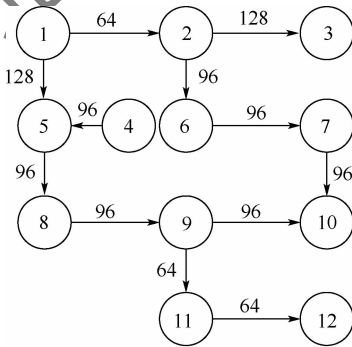
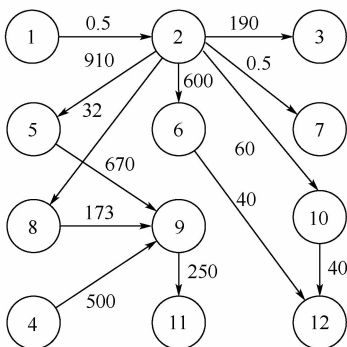


图 9 相同暗硅比例系统功耗预算方法性能对比

Fig. 9 Performance comparison of power budgeting methods under the same system of dark silicon percentages



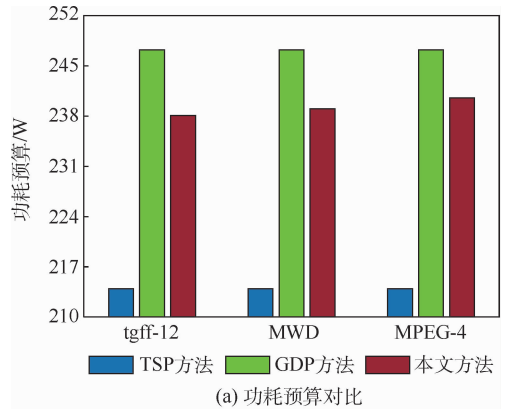
(a) MWD



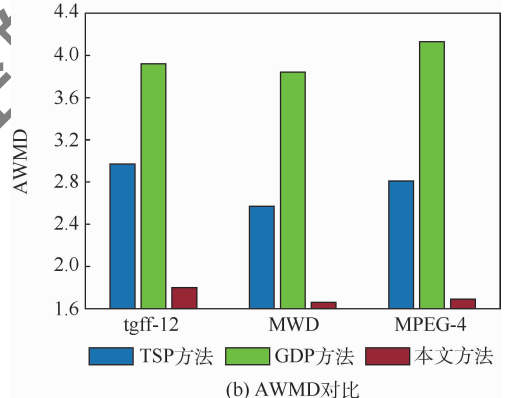
(b) MPEG-4

图 10 两种实际多媒体基准程序的通信任务图

Fig. 10 Communication task graphs of two actual multimedia applications



(a) 功耗预算对比



(b) AWMD对比

图 11 相同子任务数目的 3 种应用结果对比

Fig. 11 Comparison of results of three applications with the same number of subtasks

表 2 本文方法与其他方法提升效果比较 (36 核心系统)
Table 2 Comparison of improved performance between the proposed method and other methods (36 core system)

应用程序	功耗预算提升比例/%		AWMD 减少比例/%	
	本文方法 vs TSP 方法	本文方法 vs GDP 方法	本文方法 vs TSP 方法	本文方法 vs GDP 方法
tgff-12	11.27	-3.71	39.39	54.08
MWD	11.71	-3.33	35.41	56.77
MPEG-4	12.41	-2.72	39.86	59.08

不同方法的代表性核心映射结果及其对应温度分布分别如图 12 和图 13 所示。此外,针对特殊的应用任务需求,本文方法还可以通过调整权重值 α 的大小,决定系统进行连续映射或非连续映射的倾向。 α 越大,则倾向于将开启核心放置于邻近位置,可使通信密集型的应用任务获得更

低的通信能耗;反之,则倾向于将开启核心放置在较远的位置,以使计算密集型的应用任务获得更高的功耗预算。图 14 展示了 MPEG-4 通信任务在不同 α 值下的功耗预算和 AWMD 变化趋势。可以看出,随着 α 的增大,功耗预算和 AWMD 均逐步降低。

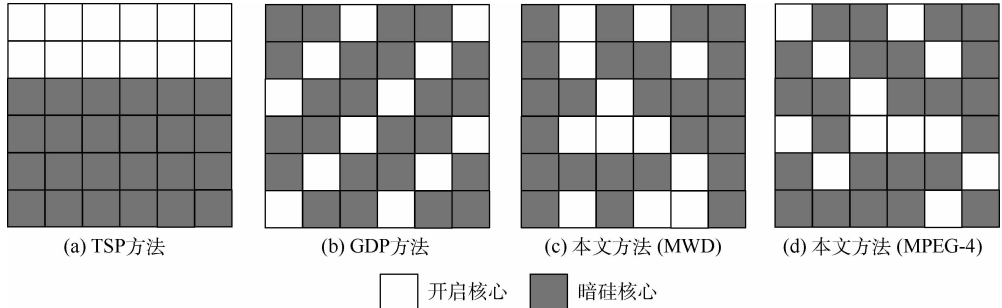


图 12 不同方法的代表性核心映射结果

Fig. 12 Representative core mapping results using different methods

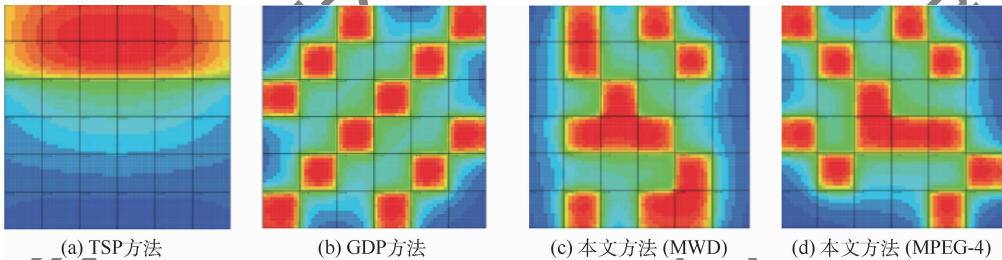


图 13 不同方法的代表性核心映射所对应的温度分布

Fig. 13 Temperature distribution corresponding to representative core mapping using different methods

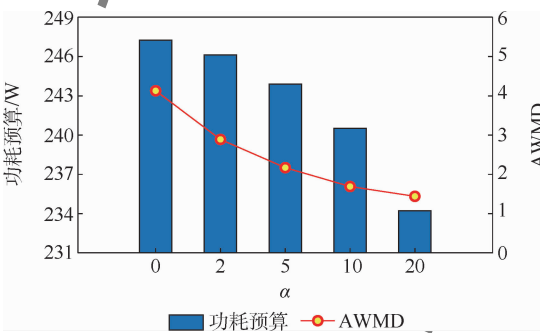


图 14 不同 α 值下 MPEG-4 功耗预算和 AWMD 变化趋势

Fig. 14 Trends of power budget and AWMD for MPEG-4 with different α values

参考文献 (References)

[1] ESMAEILZADEH H, BLEM E, AMANT R S, et al. Dark silicon and the end of multicore scaling [J]. IEEE Micro, 2011, 39 (3): 365-376.

[2] HAGHBAYAN M H, RAHMANI A M, WELDEZION A Y, et al. Dark silicon aware power management for manycore systems under dynamic workloads [C] // Proceedings of the 32nd IEEE International Conference on Computer Design. Piscataway: IEEE Press, 2014: 509-512.

[3] 黄柯衡,张正鸿,王海,等.基于混合优化的多核处理器动态热管理方法 [J]. 电子工艺技术, 2018, 39 (2): 71-75. HUANG K H, ZHANG Z H, WANG H, et al. Hybrid dynamic thermal management method for multi-core microprocessors [J]. Electronics Process Technology, 2018, 39 (2): 71-75 (in Chinese).

[4] 孙奥林,徐奇,陈松.暗硅多核系统芯片资源调度算法 [J]. 计算机辅助设计与图形学学报, 2017, 29 (6): 1145-1154. SUN A L, XU Q, CHEN S. Resource scheduling algorithm for multi-core system chip with dark silicon [J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29 (6): 1145-1154 (in Chinese).

[5] PAGANI S, KHDR H, CHEN J J, et al. Thermal safe power (TSP): Efficient power budgeting for heterogeneous manycore

5 结 论

针对暗硅系统功耗预算问题,本文提出一种基于任务映射的组合优化方法。

1) 将提高芯片均温和降低通信成本的目标进行融合,给出权衡后的核心映射选择和功耗预算。

2) 仿真结果表明,本文方法可以获得优化的开启核心选择,在保证系统通信性能的同时最大限度地提高功耗预算。

- systems in dark silicon[J]. IEEE Transactions on Computers, 2017, 66(1):147-162.
- [6] WANG H, ZHANG M, TAN S X D, et al. New power budgeting and thermal management scheme for multi-core systems in dark silicon [C]// Proceedings of the 29th IEEE International System-on-Chip Conference. Piscataway: IEEE Press, 2016: 344-349.
- [7] WANG H, TANG D Y, ZHANG M, et al. GDP: A greedy based dynamic power budgeting method for multi-/many-core systems in dark silicon[J]. IEEE Transactions on Computers, 2019, 68(4):526-541.
- [8] AGYEMAN M O, AHMADINIA A, BAGHERZADEH N. Performance and energy aware inhomogeneous 3D networks-on-chip architecture generation[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(6):1756-1769.
- [9] AGYEMAN M O, TONG K F, MAK T. Towards reliability and performance-aware wireless network-on-chip design [C]// Proceedings of the 2015 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems. Piscataway: IEEE Press, 2015:205-210.
- [10] RAPP M, SAGI M, PATHANIA A, et al. Power- and cache-aware task mapping with dynamic power budgeting for many-cores [J]. IEEE Transactions on Computers, 2020, 69(1):1-13.
- [11] KANDURI A, HAGHBAYAN M H, RAHMANI A M, et al. Dark silicon aware runtime mapping for many-core systems: A patterning approach [C]// Proceedings of the 33rd IEEE International Conference on Computer Design. Piscataway: IEEE Press, 2015:573-580.
- [12] MAQSOOD T, ALI S, MALIK S U R, et al. Dynamic task mapping for network-on-chip based systems [J]. Journal of Systems Architecture, 2015, 61(7):293-306.
- [13] WANG X H, SINGH A K, LI B, et al. Bubble budgeting: Throughput optimization for dynamic workloads by exploiting dark cores in many core systems [J]. IEEE Transactions on Computers, 2018, 67(2):178-192.
- [14] LI B, WANG X H, SINGH A K, et al. On runtime communication and thermal-aware application mapping and defragmentation in 3D NoC systems [J]. IEEE Transactions on Parallel and Distributed Systems, 2019, 30(12):2775-2789.
- [15] HAGHBAYAN M H, KANDURI A, RAHMANI A M, et al. MapPro: Proactive runtime mapping for dynamic workloads by quantifying ripple effect of applications on networks-on-chip [C]// Proceedings of the 9th International Symposium on Networks-on-Chip. New York: ACM, 2015:1-8.
- [16] FATTAH M, DANESHTALAB M, LIJEBBERG P, et al. Smart hill climbing for agile dynamic mapping in many-core systems [C]// Proceedings of the 50th Annual Design Automation Conference. Piscataway: IEEE Press, 2013:1-6.

A power budgeting method for dark silicon chips based on task mapping

LI Xin^{1,*}, LI Zhi¹, ZHOU Wei¹, WU Ruiqi², TANG Haoran³, CHEN Yehang⁴

(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China;

2. Honors College, Northwestern Polytechnical University, Xi'an 710072, China;

3. School of Software, Northwestern Polytechnical University, Xi'an 710072, China;

4. School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: The power budgeting for dark silicon systems can be regarded as a NP-hard problem. To achieve two opposite optimization objectives of improving chip average temperature and reducing communication cost, a power budgeting method based on task mapping for dark silicon chips is proposed. To reduce the computational complexity, a model is established to transform the task graph into a maximum spanning tree, based on the rule that the task with high throughput and less impact on subsequent mapping is mapped first. The priority value determines the mapping order of tasks. Then, the core-by-core optimization is carried out in a steady state. The sorted tasks are assigned to appropriate active cores. The power budgets of the identified active cores are solved in the form of convex quadratic programming. Experimental results show that compared with the classical thermal safe power budgeting method, the method proposed increases the total power budget by 11.8% and reduces the communication energy consumption by 38.2% for 36-core system with 12 active cores.

Keywords: dark silicon; power budget; task mapping; multicore system; dynamic thermal management

Received: 2021-01-11; **Accepted:** 2021-05-09; **Published online:** 2021-06-02 16:01

URL: kns.cnki.net/kcms/detail/11.2625.V.20210602.1053.001.html

Foundation items: National Natural Science Foundation of China (61501377); Natural Science Basic Research Program of Shaanxi (2021JM-074)

* **Corresponding author.** E-mail: xinli@nwpu.edu.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0010

部分可观 Petri 网故障的量子贝叶斯诊断

刘久富*, 张信哲, 汪恒宇, A. M. TOMAS DIAS, 王志胜, 杨忠

(南京航空航天大学 自动化学院, 南京 211106)

摘 要: 针对故障本身构建量子贝叶斯 Petri 网模型算法, 并利用该子网模型进行 Petri 网系统故障分析。对于部分可观 Petri 网模型中的不可观故障, 根据可达标识图分析变迁点火路径不能判断系统状态, 建立量子贝叶斯子网模型, 通过不确定路径引起的量子干涉重新标定变迁的条件概率表得到量子概率振幅表。根据故障变迁的前置集合并结合量子贝叶斯推理计算变迁触发的先验概率, 由后置集合中的可观变迁修正后验概率, 由最大后验概率估计系统所处状态, 当故障变迁不唯一时, 选取最大概率的故障作为故障源。以实际故障系统建立部分可观 Petri 网模型, 结合可观标签概率序列信息和量子贝叶斯概率估计, 对系统不可观部分进行故障诊断验证算法的有效性。

关键词: 故障诊断; Petri 网; 部分可观; 量子贝叶斯; 量子干涉

中图分类号: TP391

文献标志码: A **文章编号:** 1001-5965(2022)07-1125-10

可靠的健康监测对离散事件系统 (discrete event system, DES) 的安全运行至关重要, 高度约束的系统要求以最小的传感和计算能力运行。在只能获得部分有效信息的情况下, 不可观部分的故障预测和诊断显得尤为重要, 故障的出现会极大地改变系统的行为, 甚至导致整个系统失效^[1-2]。

由于传感器数量等限制, 为克服观测信息不全问题并满足实时性要求, 诊断系统一般采用基于定性模型的诊断方法^[3], Petri 网最早应用于电力^[4]和通信等系统^[5]的故障识别与诊断。文献[5]引入部分可观随机 Petri 网, 系统地计算了与给定时间观测序列相一致的时间和非时间标记轨迹的概率, 根据故障概率与故障置信度比较进行诊断。文献[6]构造基可达图计算基标记来识别故障类的告警等级, 引入 2 种模糊算子计算确定因子值。文献[7]基于增量算法从当前可能状

态及其概率的集合出发, 评估一致轨迹的概率得到状态估计。贝叶斯网络广泛应用于故障预测和诊断, 文献[8]通过 Grover 量子搜索算法快速搜索目标数据, 运用层次分析法删减对诊断故障影响较小的指标, 分析确定变精度粗糙集的错误分类率, 获得最小故障决策表构建贝叶斯网络故障推理模型。

针对系统模型中不可观部分故障, 本文提出一种构建故障的诊断子网, 利用量子贝叶斯估计系统故障概率的方法^[9-13], 引入量子干涉调节量子贝叶斯网络的概率推断^[14-15]。建立对应不可观故障变迁的量子贝叶斯网络模型, 降低模型复杂度^[16-18], 由可观节点触发关系和不可观节点的触发概率, 手动选择量子参数^[19-23]计算故障节点触发概率, 判断系统故障状态。通过对统一故障进行试验分析, 证明本文算法的准确率高于经典贝叶斯理论。

收稿日期: 2021-01-08; 录用日期: 2021-03-20; 网络出版时间: 2021-05-06 16:13

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210506.1303.001.html

基金项目: 国家自然科学基金 (61473144)

*通信作者。E-mail: liujiufu2@126.com

引用格式: 刘久富, 张信哲, 汪恒宇, 等. 部分可观 Petri 网故障的量子贝叶斯诊断[J]. 北京航空航天大学学报, 2022, 48(7): 1125-1134. LIU J F, ZHANG X Z, WANG H Y, et al. Partial observable Petri nets fault diagnosis with quantum Bayesian learning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1125-1134 (in Chinese).

1 部分可观 Petri 网的量子贝叶斯故障诊断问题

1.1 部分可观 Petri 网

部分可观 Petri 网 (partially observed Petri nets, POPN) 定义为一个 6 元组: $POPN = (P, T_o, T_u, \mathbf{Pre}, \mathbf{Post}, \mathbf{M}_0)$, 且 $T = T_o \cup T_u$, T 为 n 维的变迁集, T_o 为可观变迁集, T_u 为不可观变迁集, $|T_u| = n_u \leq n$, n_u 为不可观变迁个数, P 为 m 维的库所集, $\mathbf{Pre} \in \mathbf{N}^{m \times n}$ 和 $\mathbf{Post} \in \mathbf{N}^{m \times n}$ 分别为变迁的前向弧矩阵和后向弧矩阵, 定义矩阵 $\mathbf{C} = \mathbf{Post} - \mathbf{Pre}$, $\mathbf{C} \in \mathbf{N}^{m \times n}$ 为 POPN 的关联矩阵, \mathbf{M}_0 为 POPN 的初始标识。

令库所 p 的标识为 $\mathbf{M}(p)$, 则变迁 $t_i \in T$ 在标识 \mathbf{M} 处逻辑使能的充要条件为: 对于任意的 $p \in \cdot t_i$ ($\cdot t$ 为 t 的前置集合, $\cdot t = \{p \in P: \mathbf{Pre}(p, t) > 0\}$), 都有 $\mathbf{M}(p) \geq \mathbf{Pre}(p, t_i)$

变迁 t_i 在标识 \mathbf{M} 处触发点火记为 $\mathbf{M}[t_i]$ 。如果 t_i 点火后有 $\mathbf{M}' = \mathbf{M} + \mathbf{C}(\cdot, t_i)$, 则记为 $\mathbf{M}[t_i]\mathbf{M}'$, 称其为 POPN 的状态方程, 其中 $\mathbf{C}(\cdot, t_i)$ 为变迁 t_i 触发点火的关联矩阵。

POPN 的库所由库所内 token 资源数表示, 每个库所的 token 集合定义为标识向量 (网络状态) $\mathbf{M} \in \mathbf{N}^m$, \mathbf{M} 的第 i 个位置描述库所 p_i 的 token 资源数。每一步 PN (同一时刻由库所 P 经过变迁 T 到达其他库所 P 的统称) 可能触发一个或多个变迁, 用 $\delta \in \mathbf{N}^n$ 向量识别一组变迁, 其项为 1 或 0。用 $\delta(i)$ 表示向量中的第 i 项, 则变迁 $t \in T$ 在标识 $\mathbf{M}(p)$ 处逻辑使能: $\forall p \in P: \mathbf{M}(p) \geq \sum_{i=1}^n \mathbf{Pre}(p, i)\delta(i)$, 触发后增加 $\sum_{i=1}^n \mathbf{Post}(p, i)\delta(i)$, 得到新标识 \mathbf{M}' :

$$\mathbf{M}'(p) = \mathbf{M}(p) + \sum_{i=1}^n (\mathbf{Post}(p, i) - \mathbf{Pre}(p, i))\delta(i)$$

记为 $\mathbf{M}[\delta]\mathbf{M}'$, $\mathbf{C}(p, i) = \mathbf{Post}(p, i) - \mathbf{Pre}(p, i)$ 。

1.2 观测序列

存在长度为 h 的概率变迁序列集合 (probability transition sequence, PTS) 为 $\sigma = (t_1, \text{Pr}_1)(t_2, \text{Pr}_2) \dots (t_h, \text{Pr}_h) \in (T \times K)^*$, 对应标识为 $\mathbf{M}_0[t_1]\mathbf{M}_1[t_2] \dots [t_h]\mathbf{M}_h$, 简记为 $\mathbf{M}_0[\sigma]\mathbf{M}_h$, 其中, $t_0 < t_1 < \dots < t_h$, t_h 为最后触发的变迁。下标为 i 的变迁 $t_i \in T$ 对应下标为 i 的变迁触发概率 Pr_i 。 $K \in [0, 1]$ 表示概率值范围大小, 但显然为 0 的情况不可能发生, 为 0 时可观概率变迁序列 PTS 不存

在。记 $\log \sigma = t_0 t_1 \dots t_h$ 为忽略触发概率 Pr 的变迁序列。

定义 1 标签 POPN (label POPN, LPOP) 定义为 $\langle \text{POPN}, \mathcal{L} \rangle$, 标签函数定义为 $\mathcal{L}: T \rightarrow L \cup \{\varepsilon\}$, 对于可观变迁有限集 $T_o = \{t \in T | \mathcal{L}(t) = L\}$, 不可观变迁有限集 $T_u = \{t \in T_o | \mathcal{L}(t) = \varepsilon\}$ 。

定义 2 由可观变迁和对应触发概率 Pr_i 组成的集合构成的可观标签概率序列 (observed label probability sequence, OLPS) 为

$\sigma_o = (\gamma_{e_0}, \text{Pr}_{e_0})(\gamma_{e_1}, \text{Pr}_{e_1}) \dots (\gamma_{e_h}, \text{Pr}_{e_h}) \in (L \times K)^*$
 式中: γ_{e_h} 为最后触发的可观变迁; Pr_{e_h} 为最后触发的可观变迁 γ_{e_h} 的触发概率, 那么对于任意 σ_o 。至少存在一个概率变迁序列 PTS σ 与之对应。

1.3 故障诊断问题

定义 3 在 POPN 中, 给定一个变迁 t_i 和其对应的触发条件概率 Pr_i , 组成的变迁概率对是 (t_i, Pr_i) , 其对应的量子概率幅值 φ_i 由式 (1) 确定:

$$\text{Pr}_i = |e^{j\theta_i} \varphi_i|^2 \rightarrow \varphi_i = e^{j\theta_i} \sqrt{\text{Pr}_i} \quad (1)$$

式中: θ_i 为相位角; $\text{Pr}_i \in [0, 1]$ 。则 POPN 中所有变迁的量子概率振幅表为: $\psi = \{(t_i, \varphi_i) | t_i \in T\}$ 。

定义 4 给定一个 $\sigma_o \in (L \times K)^*$, 对应的概率变迁序列集合为

$$Q(\sigma_o) = \{\sigma \in (T \times K)^* | \mathbf{M}_0[\sigma]\mathbf{M}, \mathcal{L}(\sigma) = \sigma_o, \forall t_i \in T: \mathbf{M}[t_i], p_i(\mathbf{M}_0, \sigma) > 0\}$$

式中: $p_i(\mathbf{M}_0, \sigma)$ 表示以 \mathbf{M}_0 为初始标识的概率变迁序列 σ 在不可观变迁 t_i 处的量子触发概率, 标识 \mathbf{M} 可以使其能的条件为变迁可观且触发概率为 p_{e_h} 。

定理 1 给定 OLPS 集合 $\sigma_o \in (L \times K)^*$ 、变迁的条件概率表 ϕ 、故障类 $T_f^i = \{t_i^1, t_i^2, \dots, t_i^k\}$ 和故障概率阈值 ρ , 定义系统故障状态函数 $W: (L \times K) \times K \times \{T_f^1, T_f^2, \dots, T_f^i\} \times K \rightarrow \{N, F\}$ 。 $T_f^i \subseteq T_u$ 可能是执行机构故障、系统结构维护不当、环境干扰等^[24-26], t_i^k 对应具体变迁引起故障。

1) $\exists \sigma \in Q(\sigma_o), \exists t_i^i \in T_f^i$, 满足 $t_i^i \notin \sigma$ 任何一个概率变迁序列 PTS σ 都不包含故障变迁 t_i^i , $\text{Pr}(\sigma_o, \psi, T_f^i) = 0, W(\sigma_o, \phi, T_f^i, \rho) = N$ 。

2) $\exists \sigma \in Q(\sigma_o), \exists t_i^i \in T_f^i$ 满足 $t_i^i \in \sigma$, 又 $\exists \sigma' \in Q(\sigma_o), \exists t_i^i \in T_f^i$ 满足 $t_i^i \notin \sigma'$, 存在包含故障变迁 t_i^i 的 PTS σ , 也存在不包含故障变迁 t_i^i 的 PTS σ' , 系统发生故障概率为

$$\text{Pr}(\sigma_o, \psi, T_f^i) = \max_i \text{Pr}(t_i^i | t_1, t_2, \dots, t_n)$$

如果 $\text{Pr}(\sigma_o, \psi, T_f^i) > \rho, W(\sigma_o, \phi, T_f^i, \rho) = F$; 否则, $W(\sigma_o, \phi, T_f^i, \rho) = N$ 。

3) $\forall \sigma \in Q(\sigma_0), \exists t_f^i \in T_f^i$, 满足 $t_f^i \in \sigma$, 任何一个概率变迁序列 PTS σ 都包含故障变迁 t_f^i , $\Pr(\sigma_0, \psi, T_f^i) = 1, W(\sigma_0, \phi, T_f^i, \rho) = F$.

证明 1) 在条件 $\exists \sigma \in Q(\sigma_0), \exists t_f^i \in T_f^i$ 下, 满足 $t_f^i \notin \sigma$, 即对于故障变迁 t_f^i 不处于变迁序列集合 $Q(\sigma_0)$ 中, 该故障变迁触发概率为 0, 观测结果为系统无故障。

2) 在条件 $\exists \sigma \in Q(\sigma_0) \wedge \exists \sigma' \in Q(\sigma_0)$ 下, 满足 $t_f^i \in \sigma \wedge t_f^i \notin \sigma'$, 即故障变迁 t_f^i 存在于变迁序列 σ 但不在变迁序列 σ' 中, 计算各个故障变迁的触发概率并选取最大概率作为系统故障概率:

$$\Pr(\sigma_0, \psi, T_f^i) = \max_i \Pr(t_f^i | t_1, t_2, \dots, t_n)$$

当满足故障概率大于系统故障阈值时, 系统观测结果为故障; 故障概率小于系统故障阈值时, 系统观测结果为无故障。

3) 在条件 $\forall \sigma \in Q(\sigma_0), \exists t_f^i \in T_f^i$ 下, 满足 $t_f^i \in \sigma$, 即任何一个变迁序列中都存在故障变迁 t_f^i , 该故障变迁触发概率为 1, 观测结果为系统存在故障。 **证毕**

利用变迁序列和量子贝叶斯概率估计, 设计一个状态诊断器对 PORN 模型中不可观部分进行故障诊断, 根据定理 1 确定系统的故障状态。

2 部分可观 Petri 网的量子贝叶斯故障诊断算法

2.1 量子贝叶斯概率估计

贝叶斯网络 $B(G, P_R)$ 的组成包括 n 个节点的有向无环图 G 和各节点对应的条件概率表 P_R [7-8]。论域 $U = \{x_1, x_2, \dots, x_n\}$, U 为 G 的节点变量, $\Pr(x_i | \text{parents}(x_i))$ 表示节点 x_i 与父节点 $\text{parents}(x_i)$ 中变量集合的条件概率。贝叶斯网络中全部变量集合的联合概率分布为

$$\Pr(U) = \Pr(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \Pr(x_i | \text{parents}(x_i)) \quad (2)$$

在贝叶斯网络中, 令 e 为可观变量, Y 为不可观变量, 那么对于变量 x_i 有

$$\Pr(x_i = x_{ij} | e) = \alpha \cdot \Pr(x_{ij}, e) = \alpha \left[\sum_{y \in Y} \Pr(x_{ij}, e, y) \right] \quad (3)$$

式中: 随机变量 x_i 有 m 个基本事件 $\{x_{i1}, x_{i2}, \dots, x_{im}\}$, $\alpha = \frac{1}{\sum_{x_{ij} \in x_i} \Pr(x_i = x_{ij}, e)}$ 。

在量子贝叶斯网络中, 用量子概率振幅代替真实的概率数, 全联合概率分布的量子表示为

$$\Pr(U) = \Pr(x_1, x_2, \dots, x_n) = \left| \prod_{i=1}^n \varphi(x_i | \text{parents}(x_i)) \right|^2 \quad (4)$$

波恩规则对式(3)的应用为

$$\Pr(x_i = x_{ij} | e) = \alpha \left| \sum_y \prod_{i=1}^n \varphi(x_{ij} | \text{parents}(x_{ij}), e, y) \right|^2 \quad (5)$$

展开式(5), 得到量子干涉公式为

$$\Pr(x_i = x_{ij} | e) = \alpha \left(\sum_{k=1}^{|Y|-1} \left| \sum_{i=1}^{|N|} \varphi(x_{ij} | \text{parents}(x_{ij}), e, y = k) \right|^2 + 2 \text{Interference} \right) \quad (6)$$

式中:

$$\text{Interference} = \sum_{k=1}^{|Y|-1} \sum_{m=k+1}^{|Y|} \left| \varphi(x_{ij} | \text{parents}(x_{ij}), e, y = k) \cdot \varphi(x_{ij} | \text{parents}(x_{ij}), e, y = m) \right| \cos(\theta_i - \theta_j)$$

当 $\theta_i - \theta_j = \pi/2$, 即 $\cos(\theta_i - \theta_j) = 0$ 时, 量子贝叶斯网络变为经典贝叶斯网络。由式(4)和式(6)可计算 $x_i = x_{ij}$ 的后验概率为

$$\Pr(x_i = x_{ij} | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{\Pr(x_i = x_{ij}) \left| \prod_{a \neq i} \varphi(x_a | \text{parents}(x_a)) \right|^2}{\sum_{i=1}^m \left(\Pr(x_i = x_{ik}) \left| \prod_{a \neq i} \varphi(x_a | \text{parents}(x_a)) \right|^2 \right)} \quad (7)$$

本文选择手动分配量子干涉参数来获得预测。在量子力学的观点下, 由干涉效应产生的量子参数代表了能量波的位移。在量子认知的视角下, 被解释为事件之间的关联及共享的语义关系。

观测序列中故障集出现的可信度称为置信度 [9], 用 F_b 表示。给定故障变迁 T_f^i 对应的概率序列集合为 $Q(\sigma_0)$, 则

$$F_b = \frac{Q(\sigma_0, T_f^i)}{Q(\sigma_0)} \quad (8)$$

定义 5 量子贝叶斯 Petri 网 (quantum Bayesian Petri net, QBPN) 定义为 $N_{\text{QB}} = (P, T_{\text{QB}_0}, T_{\text{QB}_u}, \mathbf{Pre}, \mathbf{Post}, \psi, \mathbf{M}_0)$, $T_{\text{QB}_0} \subseteq T_0$ 为 QBPN 中可观变迁集, $T_{\text{QB}_0} = \cdot T_{\text{QB}_0} \cup T_{\text{QB}_0} \cdot, \cdot T_{\text{QB}_0}$ 为关于故障变迁的前置可观集合, $T_{\text{QB}_0} \cdot$ 为关于故障变迁的后置可观集合, $T_{\text{QB}_u} \subseteq T_u$ 为 QBPN 中不可观变迁集, 且 $T_{\text{QB}} = T_{\text{QB}_0} \cup T_{\text{QB}_u}$, ψ 为 $t_i \in T_{\text{QB}}$ 的量子概率表。

2.2 部分可观 Petri 网的量子贝叶斯概率故障诊断
构建关于故障变迁的 QBPN 模型诊断系统故障状态, 根据量子贝叶斯估计, 通过故障变迁前置集合中可观变迁触发顺序计算先验概率, 再由后

置集合中可观变迁触发顺序得到故障变迁后验概率,取所有故障变迁中最大触发概率作为系统故障概率判断系统状态^[27-30]。

算法1 基于QBPN模型的故障诊断算法。

输入: POPN模型、量子贝叶斯概率振幅表 ψ 和可能故障变迁 t_i^i 、系统故障概率阈值 ρ 。

输出: 系统故障状态 $W(\sigma_o, \phi, T_f^i, \rho)$ 。

1. 初始化 $\forall t_i^i \in Q(\sigma_o), Pr_i \in [0, 1], \varphi(t_i), \rho \in [0, 1]$

2. for $\forall t_i^i \in \sigma \in Q(\sigma_o)$ do

While $\forall t_k \in T_o, t_k \in M_k, M_k[t_k] \dots M_j[t_j] \dots$

$M_i[t_i^i], k < j \leq i, t_j \in T_u$ do

元素类型转换 $t_k \in T_{QB_u}, t_j \in T_{QB_u}$

While

$\forall t_k \in T_o, t_k \in M_k, M_i[t_i^i] \dots M_j[t_j] \dots M_i[t_k],$

$i < j \leq k, t_j \in T_u$ do

元素类型转换 $t_k \in T_{QB_u}, t_j \in T_{QB_u}$

3. if $\exists t_k \wedge M_i[t_k], t_k$ 不能触发 then

去除 t_k 及其直接相连的流关系

4. for $t_i \in T_{QB}$ do

根据量子概率振幅表 ψ 对 t_i 设置量子概率振幅 $\varphi(t_i)$

5. for $\forall t_k \in T_{QB_u}$ do

根据前向路径可观变迁状态计 $Pr(t_i^i | T_{QB_u})$

6. for $\forall t_k \in T_{QB_u}$ do

根据式(7)修正故障变迁的后验概率

$Pr(t_i^i | T_{QB_u} \cup T_{QB_u}, \cdot)$, 即 $Pr(t_i | T_{QB_u})$

7. 选取最大概率作为系统故障概率 $Pr(\sigma_o, \psi, T_f^i)$

8. if $Pr(\sigma_o, \psi, T_f^i) > \rho$ then

$W(\sigma_o, \phi, T_f^i, \rho) = F$

else $W(\sigma_o, \phi, T_f^i, \rho) = N$

9. 返回 $W(\sigma_o, \phi, T_f^i, \rho)$

针对 POPN 模型中的不可观故障变迁,其前项点火触发关系为父节点,先由父节点向前遍历,当

遇到可观变迁或没有前置约束的不可观变迁时停止搜索,将遍历过程中的变迁,库所和流关系转化成量子贝叶斯节点元素,继续遍历其他父节点分支直到父节点集合遍历完毕;再由后项点火触发关系即子节点向后开始遍历,当遇到可观变迁或没有后置约束的不可观变迁时停止搜索,将遍历过程中的变迁、库所和流关系转化成量子贝叶斯节点元素,继续遍历其他子节点分支直到子节点集合遍历完毕。之后对于由当前标识不能通过变迁 t_k 触发变化到下一变迁的情况,去除该变迁,精简 QBPN 模型结构,通过不确定路径引起的量子干涉重新标定变迁的条件概率表得到量子概率振幅表。根据故障变迁的前置集合并结合量子贝叶斯推理计算变迁触发的先验概率,再由后置集合中的可观变迁修正后验概率,由最大后验概率估计系统所处状态,当故障变迁不唯一时,选取最大概率的故障作为故障源。最终得到状态观测函数输出值。

2.3 算法复杂度分析

以每个故障变迁为开始,分别向前向后遍历变迁序列直到第一个可观变迁停止,最差时每个都要遍历 $l_{max}/2$, 则为 $l_{max} Q_{num}/2$, 此时复杂度为 $l_{max} Q_{num}$; 去掉无关变迁 $l_{max} Q_{num}$; 设置变迁触发概率并标记量子概率幅值, 计算故障概率 l_{max} ; 并计算系统故障率, 故障变迁集中故障变迁数量为 N_f 。综上, 算法的复杂度为 $N_f l_{max}^3 Q_{num}^2$ 。

3 实例分析与验证

给定一个表示液体火箭发动机启动过程的离散事件动态系统的 POPN 模型^[24-25], 如图 1 所示。该系统模型的初始标识为 $M_0 = [1000000000000000]^T$, 可观变迁集 $T_o = \{t_0, t_3, t_4, t_{12}, t_{13}, t_{14}, t_{15}\}$, 不可观变迁集为 $T_u = \{t_1, t_2, t_5, t_6, t_7, t_8, t_9, t_{10}, t_{11}\}$ 。选取系统不可观部分氧化剂气路和燃料气路中的电磁阀门未打开故障 t_{10}

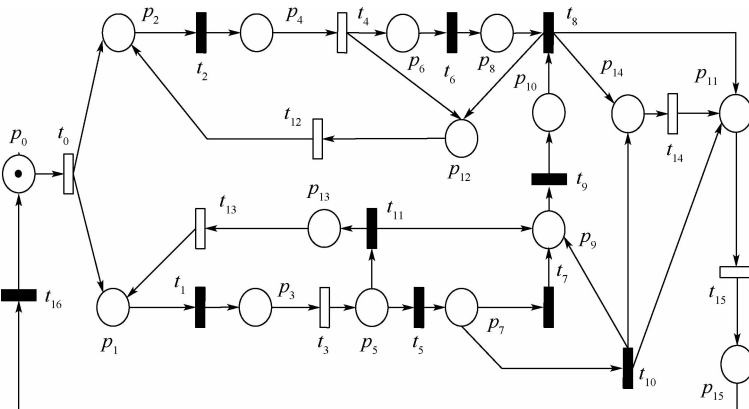


图1 离散事件动态系统的 POPN 模型

Fig. 1 POPN model of discrete event dynamic system

模拟 t_i^{10} , 燃料加注阀泄漏故障 t_{11} 模拟 t_i^{11} , 由此通过仿真验证本文算法的有效性。

3.1 故障变迁的量子贝叶斯 Petri 网

系统的故障状态不确定时,利用 QBPN 故障诊断算法判断系统故障状态。首先,需要根据算法分别构建故障变迁 t_{10} 和故障变迁 t_{11} 对应的 QBPN 模型;然后,计算 2 个故障变迁各自触发概率及系统故障概率,给出系统故障状态。

在 POPN 模型中,给定故障变迁,向前向后搜索其所有路径中的变迁和库所,直到出现可观测变迁为止。对于故障变迁 t_{10} ,其所在路径包括 $t_3 t_5 t_{10} t_{14} t_{12}$, $t_3 t_5 t_{10} t_{15} t_{12}$, $t_3 t_5 t_{10} t_9 t_8 t_{15} t_{12}$, $t_3 t_5 t_{10} t_9 t_8 t_{14} t_{15} t_{12}$ 。其中,路径 $t_3 t_5 t_{10} t_{14} t_{12}$ 中的可观变迁为 t_3 、 t_{12} 、 t_{14} ,不可观变迁为 t_5 、 t_{10} ;路径 $t_3 t_5 t_{10} t_{15} t_{12}$ 中的可观变迁为 t_3 、 t_{12} 、 t_{15} ,不可观变迁为 t_5 、 t_{10} ;路径 $t_3 t_5 t_{10} t_9 t_8 t_{15} t_{12}$ 中的可观变迁为 t_3 、 t_{12} 、 t_{15} ,不可观变迁为 t_5 、 t_8 、 t_9 、 t_{10} ;路径 $t_3 t_5 t_{10} t_9 t_8 t_{14} t_{15} t_{12}$ 中的可观变迁为 t_3 、 t_{12} 、 t_{14} 、 t_{15} ,不可观变迁为 t_5 、 t_8 、 t_9 、 t_{10} 。令变迁 t_5 、 t_8 、 t_9 、 t_{10} 为不可观量子贝叶斯变迁,库所 p_5 、 p_7 、 p_9 、 p_{10} 、 p_{11} 、 p_{12} 、 p_{14} 为量子贝叶斯库所,建立 t_{10} 对应的 QBPN 模型诊断子网,如图 2 所示。

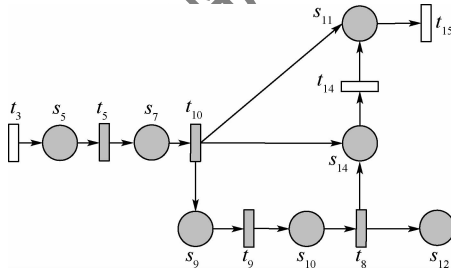


图 2 故障变迁 t_{10} 对应的 QBPN 模型

Fig. 2 QBPN model of fault transition t_{10}

3.2 条件概率表转化成量子概率振幅表

在 POPN 模型中存在不可观变迁,当给定观测点火序列时,触发序列出现不确定的情况,利用式(1)和式(6)得到量子概率振幅表。如图 3 所示, t_9 点火的不确定路径为 $t_{11} \rightarrow p_9 \rightarrow t_9$ 、 $t_7 \rightarrow p_9 \rightarrow t_9$ 和 $t_{10} \rightarrow p_9 \rightarrow t_9$,且由不可观变迁 t_7 、 t_{10} 、 t_{11} 引起。

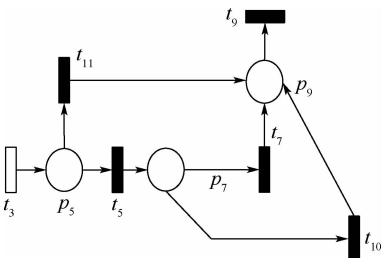


图 3 量子干涉影响的变迁 t_9

Fig. 3 Transition t_9 of the influence of quantum interference

如图 4 所示,变迁 t_{14} 点火的不确定路径包括 $t_5 \rightarrow p_7 \rightarrow t_{10} \rightarrow p_{14} \rightarrow t_{14}$ 、 $t_6 \rightarrow p_8 \rightarrow t_8 \rightarrow p_{14} \rightarrow t_{14}$ 和 $t_5 \rightarrow p_7 \rightarrow t_7 \rightarrow p_9 \rightarrow t_9 \rightarrow p_{10} \rightarrow t_8 \rightarrow p_{14} \rightarrow t_{14}$,通过路径叠加原理, t_8 和 t_{10} 的不确定导致变迁 t_{14} 点火的不确定性。

图 5 中,变迁 t_{15} 点火的不确定路径包括 $t_8 \rightarrow p_{11} \rightarrow t_{15}$ 、 $t_8 \rightarrow p_{14} \rightarrow t_{14} \rightarrow p_{11} \rightarrow t_{15}$ 、 $t_{10} \rightarrow p_{11} \rightarrow t_{15}$ 和 $t_{10} \rightarrow p_{14} \rightarrow t_{14} \rightarrow p_{11} \rightarrow t_{15}$,不确定来源为不可观变迁 t_8 、 t_{10} 和可观变迁 t_{14} 。当可观变迁 t_{14} 点火时,不确定路径会减少到 2 个,路径不确定性降低导致干涉效应的影响降低。

以图 6 所示的变迁 t_{14} 点火概率修正为例,存在不确定路径时,根据量子贝叶斯理论有:表 1

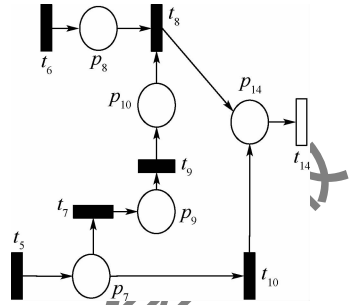


图 4 量子干涉影响的变迁 t_{14}

Fig. 4 Transition t_{14} of the influence of quantum interference

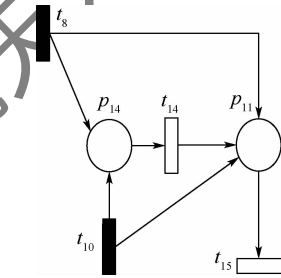
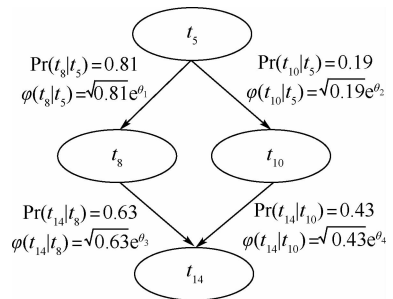


图 5 量子干涉影响的变迁 t_{15}

Fig. 5 Transition t_{15} of the influence of quantum interference



	Pr(t_{14})	$\varphi(t_{14})$
t_8	0.63	$\sqrt{0.63}e^{i\theta_3}$
t_{10}	0.43	$\sqrt{0.43}e^{i\theta_4}$

图 6 经典贝叶斯和量子贝叶斯概率

Fig. 6 Classical Bayesian probability and quantum Bayesian probability

中 t_8 在 t_5 条件下点火概率为 0.81, 量子振幅为 $\sqrt{0.81}e^{\theta_1}$; 而 t_{10} 在 t_5 条件下点火概率为 0.19, 量子振幅为 $\sqrt{0.19}e^{\theta_2}$. t_{14} 在 t_8 条件下点火概率为 0.63, 量子振幅为 $\sqrt{0.63}e^{\theta_3}$; 变迁 t_{14} 在 t_{10} 条件下点火概率为 0.43, 量子振幅为 $\sqrt{0.43}e^{\theta_4}$. 根据图 6 中所示的变迁间的概率信息, 得到如图 7 中量子贝叶斯概率值变化曲线。

表 1 总概率情况和量子贝叶斯情况的概率估计比较
Table 1 Comparison of probability estimates between total probability and quantum Bayesian

概率估计	概率值
$\Pr(t_8 t_5)$	0.81
$\varphi(t_8 t_5)$	$\sqrt{0.81}e^{\theta_1}$
$\Pr(t_{10} t_5)$	0.19
$\varphi(t_{10} t_5)$	$\sqrt{0.19}e^{\theta_2}$
$\Pr(t_{14} t_8)$	0.63
$\varphi(t_{14} t_8)$	$\sqrt{0.63}e^{\theta_3}$
$\Pr(t_{14} t_{10})$	0.43
$\varphi(t_{14} t_{10})$	$\sqrt{0.43}e^{\theta_4}$
总概率下 $\Pr(t_{14} t_5)$	0.592
量子贝叶斯 $\Pr(t_{14} t_5)$	0.795 5
总概率下 $\Pr(\bar{t}_{14} t_5)$	0.408
量子贝叶斯 $\Pr(\bar{t}_{14} t_5)$	0.204 5

式中: \bar{t}_{14} 为非 t_{14} 情况; θ 为干涉角, $\theta = \theta_{t_{14}|t_8} - \theta_{t_{14}|t_{10}}$.

根据式(9)得到图 7 所示概率变化结果。可知, 干涉角 θ 在区间 $[0, \pi]$ 上调节, 对应概率在区间 $[0.565 5, 0.795 5]$ 为增函数。则总概率下 t_{14} 发生概率为 0.592, 由量子贝叶斯理论得到的 t_{14} 触发概率为 0.795 5。

根据式(10)得到图 8 所示概率变化结果。可知, 干涉角 θ 在区间 $[0, \pi]$ 上调节, 对应概率在区间 $[0.435 5, 0.204 5]$ 为减函数。则总概率下 \bar{t}_{14} 发生概率为 0.408, 由量子贝叶斯理论得到的 \bar{t}_{14} 触发概率为 0.204 5。从图 7 和图 8 结合看出, 对应干涉角 θ 取值相同时满足 t_{14} 和 \bar{t}_{14} 发生概率值和为 1, 如表 1 所示。

更加复杂的情况如图 9 中 t_3 条件下 t_8 点火概率, 此时多个干涉角去调节取值以满足需求。

可能路径包括 $t_3 \rightarrow t_5 \rightarrow t_7 \rightarrow t_8$ 、 $t_3 \rightarrow t_5 \rightarrow t_{10} \rightarrow t_8$ 、 $t_3 \rightarrow t_{11} \rightarrow t_7 \rightarrow t_8$ 和 $t_3 \rightarrow t_{11} \rightarrow t_9 \rightarrow t_8$, 不同路径两两间产生叠加效应, 此时产生 6 个干涉角 $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$ 。固定 $\{\theta_3, \theta_4, \theta_5, \theta_6\}$ 得到关于 t_3 条件下 t_8 点火概率和干涉角 $\{\theta_1, \theta_2\}$ 之间的关系 $\Pr(t_8 | t_3) \sim (\theta_1, \theta_2)$ 。如图 10 所示, 概率变换范围为 $[0.83, 0.88]$, 当 $\theta_1 = 0$ 且 $\theta_2 = 0$ 时, 取得最大值, 当 $\theta_1 = \pi$ 且 $\theta_2 = \pi$ 时取得最小值。

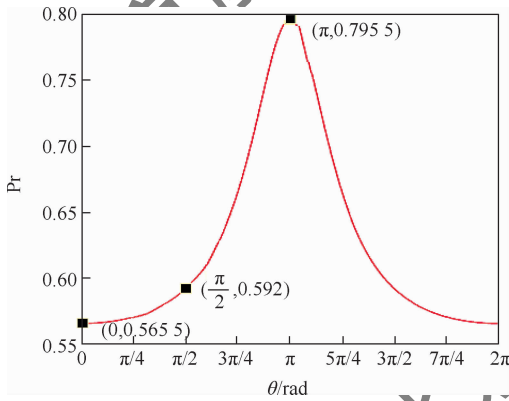


图 7 t_5 条件下 t_{14} 的量子贝叶斯概率值变化曲线

Fig. 7 Change of quantum Bayesian probability value of t_{14} under the condition of t_5

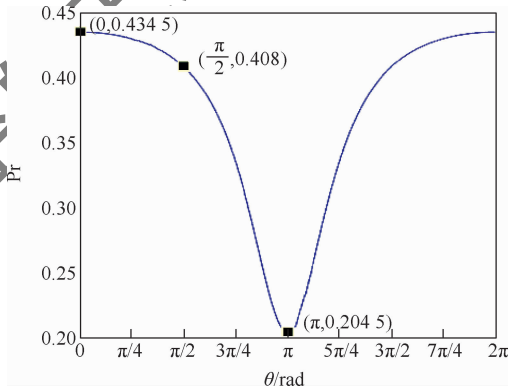


图 8 t_5 条件下 \bar{t}_{14} 的量子贝叶斯概率值变化曲线

Fig. 8 Change of quantum Bayesian probability value of \bar{t}_{14} under the condition of t_5

得到下列算式:

$$\Pr(t_{14} | t_5) = \alpha |\varphi_{t_8} \varphi_{t_{14} | t_8} + \varphi_{t_{10}} \varphi_{t_{14} | t_{10}}|^2 = \alpha (|\varphi_{t_8} \varphi_{t_{14} | t_8}|^2 + |\varphi_{t_{10}} \varphi_{t_{14} | t_{10}}|^2 + 2|\varphi_{t_8} \varphi_{t_{14} | t_8}| |\varphi_{t_{10}} \varphi_{t_{14} | t_{10}}| \cos(\theta_{t_{14} | t_8} - \theta_{t_{14} | t_{10}}))$$

式中:

$$\alpha = \frac{1}{\Pr(t_{14} | t_5) + \Pr(\bar{t}_{14} | t_5)}$$

$$\Pr(t_{14} | t_5) = \frac{0.592 + 0.408 4 \cos \theta}{1 + 0.769 2 \cos \theta} \quad (9)$$

$$\Pr(\bar{t}_{14} | t_5) = \frac{0.408 + 0.360 8 \cos \theta}{1 + 0.769 2 \cos \theta} \quad (10)$$

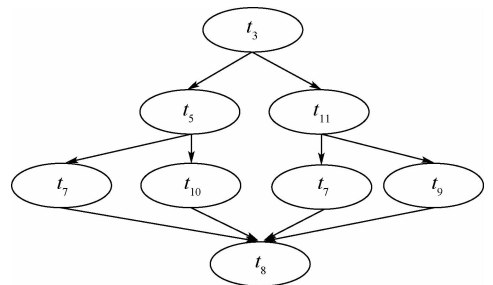


图 9 由变迁 t_3 到 t_8 的不确定路径

Fig. 9 Uncertain path from transition t_3 to t_8

固定 $\{\theta_1, \theta_2, \theta_5, \theta_6\}$ 得到关于 t_3 条件下 t_8 点火概率和干涉角 $\{\theta_3, \theta_4\}$ 之间的关系,如图 11 所示。概率变换范围为 $[0.81, 0.93]$,当 $\theta_3 = \pi$ 且 $\theta_4 = 0$ 时,取得最大值,当 $\theta_3 = 0$ 且 $\theta_4 = \pi$ 时取得最小值。

量子干涉可以是建设性干涉,也可以是破坏性干涉,根据需求需进行选择。根据图 10 和图 11 可知,干涉角 $\{\theta_1, \theta_2, \theta_4\}$ 等于 π 时产生最大建设性干涉效应, $\{\theta_3\}$ 等于 π 时产生最大破坏性干涉效应。对 POPN 模型中所有变迁的不确定性进行分析,修正条件概率。

3.3 基于量子贝叶斯 Petri 网的故障诊断结果

根据变迁条件概率表及可观变迁触发状态,在 QBPN 模型中应用量子贝叶斯推理计算不可观故障节点的故障率,这些信息体现了故障发生的不确定性和量子干涉效应的产生条件。本文基于不确定状态的 POPN 系统模型,利用 QBPN 故障诊断算法估计系统故障概率,表 2 为 QBPN 算法诊断结果。

待诊断故障变迁为 t_{10} 和 t_{11} ,可观变迁为 $t_0, t_3, t_4, t_{12}, t_{14}, t_{15}$ 。当可观标签概率序列 $\sigma_o = (t_0, 1)(t_4, 0.883)(t_3, 0.926)$ 时,变迁 t_{10} 不在变迁概率集合中,故障触发概率为 0,置信度为 0,系统故障概率为 0,系统状态为无故障。当概率变迁序列 $\sigma_o = (t_0, 1)(t_4, 0.883)(t_3, 0.926)(t_{15}, 0.879)$ 时,由可观变迁 t_0, t_3, t_4 计算故障变迁 t_{10} 的故障概率为 $\Pr(t_f^{10} | t_0, t_4, t_3) = 0.1621$,根据可观变迁 t_{15} 触发状态和式(7)计算 QB 函数,得到修正的故障概率为 $\Pr(t_f^{10} | t_0, t_4, t_3, t_{15}) = 0.0471$ 。同理,对故障变迁 t_{11} 进行分析,故障触发概率为 0.0076,系统故障概率为 0.0076,置信度为 0.4,系统故障状态为无故障。当序列 $\sigma_o = (t_0, 1)(t_4, 0.883)(t_3, 0.926)(t_{14}, 0.851)(t_{15}, 0.879)(t_{12}, 0.463)$ 时,先计算故障变迁 t_{10} 触发概率,由可观变迁 t_3 触发,根据式(6)计算故障变迁 t_{10} 触发的条件概率,取 $\theta = \pi/2$ 时概率为 $\Pr(t_f^{10} | t_0, t_4, t_3) = 0.1621$;根据 t_{14} 和式(7)修正故障变迁 t_{10} 的触发概率为 $\Pr(t_f^{10} | t_0, t_4, t_3, t_{14}) = 0.7569$;增加 t_{15} ,修正 t_{10} 的触发概率为 $\Pr(t_f^{10} | t_0, t_4, t_3, t_{14}, t_{15}) = 0.9263$;继续增加 t_{12} , t_{10} 的触发概率为 $\Pr(t_f^{10} | t_0, t_4, t_3, t_{14}, t_{15}, t_{12}) = 0.9668$ 。同理,不断修正故障变迁 t_{11} 触发概率为 $\Pr(t_f^{11} | t_0, t_4, t_3, t_{14}, t_{15}, t_{12}) = 0.0006$,则

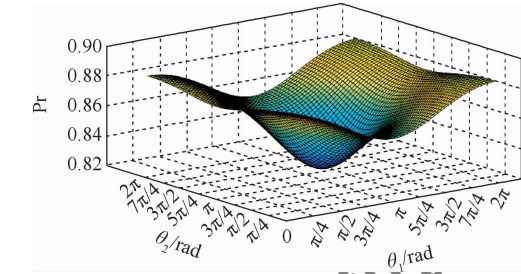


图 10 干涉角 $\{\theta_1, \theta_2\}$ 引起 t_3 条件下 t_8 点火概率变化
Fig. 10 Interference angle $\{\theta_1, \theta_2\}$ causes the change of firing probability of t_8 under the condition of t_3

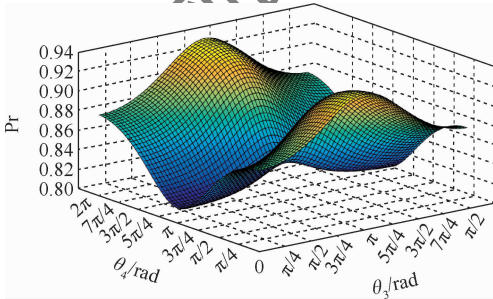


图 11 干涉角 $\{\theta_3, \theta_4\}$ 引起 t_3 条件下 t_8 点火概率变化
Fig. 11 Interference angle $\{\theta_3, \theta_4\}$ causes the change of firing probability of t_8 under the condition of t_3

表 2 基于 POPN 的量子贝叶斯故障算法故障诊断结果

Table 2 Fault diagnosis results of quantum Bayesian fault algorithm based on POPN

可观概率序列 σ_o	故障变迁 T_f^i 触发概率	干涉角 θ/rad	QB 函数结果	置信度 F_b	系统故障概率 $\Pr(\sigma_o, \psi, T_f^i)$	系统状态估计 W	系统实际状态
$(t_0, 1)(t_4, 0.883)(t_3, 0.926)$	$\Pr(t_f^{10} t_0, t_4, t_3)$	$\pi/2$	0	0	0	N	无故障
$(t_0, 1)(t_4, 0.883)(t_3, 0.926)(t_{15}, 0.879)$	$\Pr(t_f^{10} t_0, t_4, t_3)$ $\Pr(t_f^{10} t_0, t_4, t_3, t_{15})$	$\pi/2$ 0	0.1621 0.0471	0.4	0.0471	N	无故障
$(t_0, 1)(t_4, 0.883)(t_3, 0.926)(t_{14}, 0.851)(t_{15}, 0.879)(t_{12}, 0.463)$	$\Pr(t_f^{10} t_0, t_4, t_3)$ $\Pr(t_f^{10} t_0, t_4, t_3, t_{14})$ $\Pr(t_f^{10} t_0, t_4, t_3, t_{14}, t_{15})$ $\Pr(t_f^{10} t_0, t_4, t_3, t_{14}, t_{15}, t_{12})$	$\pi/2$ 0 0 0	0.1621 0.7569 0.9263 0.9668	0.6	0.9668	F	故障
$(t_0, 1)(t_3, 0.907)(t_{13}, 0.736)$	$\Pr(t_f^{11} t_0, t_3, t_{13})$	$\pi/2$	1	1	1	F	故障

系统故障概率为 0.966 8, 系统状态为故障, 置信度为 0.6。当 $\sigma_0 = (t_0, 1)(t_3, 0.907)(t_{13}, 0.736)$ 时, 变迁 t_{11} 存在于所有变迁概率集合, 则 t_{11} 的故障概率为 1, 系统状态为故障, 置信度为 1。

当故障变迁可能存在可观变迁序列对应的变迁序列集合中时, 随着可观变迁的触发增加, 得到不同情况下的故障触发概率并修正。

3.4 算法比较

图 12 中, BPN 算法概率分布区间为 $[0.942 5, 0.975]$, 拟合后为 0.960 7, 期望 $\mu_{\text{BPN}} = 0.960 7$, 标准差 $\sigma_{\text{BPN}} = 0.005 4$ 。图 13 中, QBPN 算法的概率分布区间为 $[0.952 5, 0.980]$, 拟合后为 0.966 3, $\mu_{\text{QBPN}} = 0.966 3$, $\sigma_{\text{QBPN}} = 0.004 3$, 根据期望对比 QBPN 估计的概率更加接近真实概率, 根据标准差对比 QBPN 概率分布更加集中。分别利用 BPN 算法和 QBPN 算法对同一系统状态进行故障诊断, 经过 100 次仿真对比后得到图 14。可以看出, QBPN 算法经过量子干涉修正后能得出更好的概率估计。

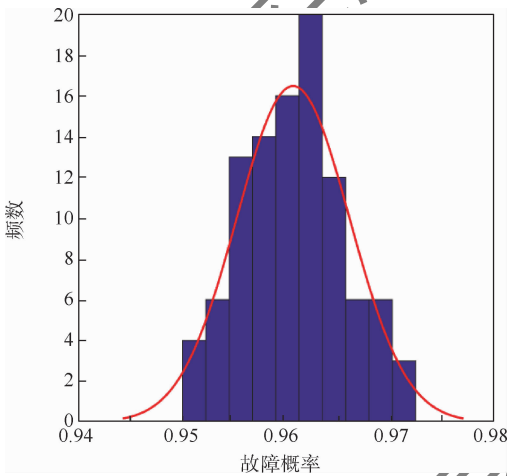


图 12 基于 BPN 算法的概率分布

Fig. 12 Probability distribution based on BPN algorithm

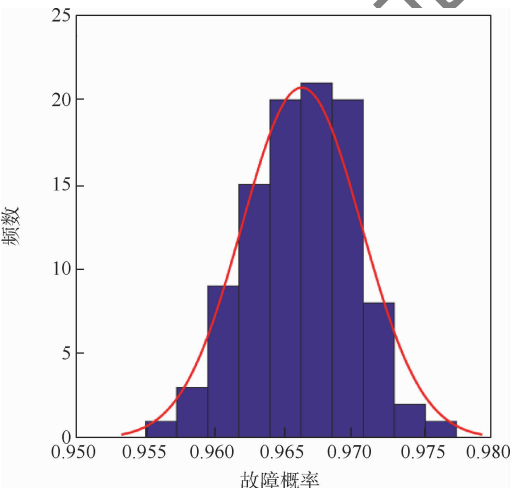


图 13 基于 QBPN 算法的概率分布

Fig. 13 Probability distribution based on QBPN algorithm

表 3 为 BPN 算法与本文算法性能比较。以本文案例模型为基础, 应用 BPN 算法对不可观故障变迁进行诊断, 对于最复杂的可观标签概率序列 OLPS, 经过多次试验后, BPN 算法的响应时间平均为 300 ms, 而应用本文 QBPN 算法, 针对故障变迁建立的子网模型进行故障诊断, 平均响应时间为 230 ms。可以看出, 本文 QBPN 算法减少了计算复杂度, 提高了运算效率。同时在 5 000 次仿真试验中, BPN 算法正确诊断次数为 4 685 次, 诊断正确率为 93.70%; QBPN 算法通过调节量子干涉项, 修正故障触发概率, 正确诊断次数为 4 937, 诊断正确率为 98.74%。本文 QBPN 算法的故障诊断能力更高。

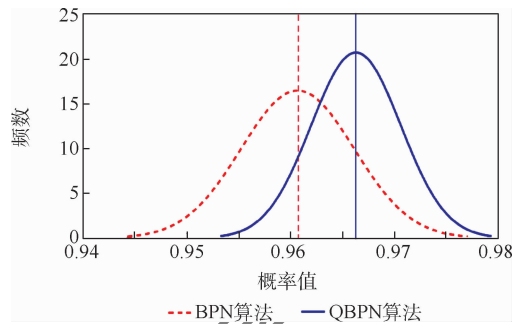


图 14 BPN 算法和 QBPN 算法故障诊断概率比较

Fig. 14 Comparison of fault diagnosis probability between BPN algorithm and QBPN algorithm

表 3 QBPN 算法与 BPN 算法比较

Table 3 Comparison of QBPN algorithm and BPN algorithm

算法	响应时间/ms	诊断正确率/%
BPN	300	93.70
QBPN	230	98.74

4 结论

1) 本文基于 POPN 模型提出了一种 QBPN 故障诊断算法, 结合 Petri 网的直观分析描述方法及量子贝叶斯处理不确定知识表达和推理的特点, 根据可观变迁触发状态, 从不可观变迁引起的量子干涉角度计算不可观变迁触发概率, 对系统模型中不可观节点进行故障诊断, 与系统故障阈值比较估计系统状态。

2) 建立故障变迁对应的 QBPN 模型, 简化模型, 减小算法复杂度, 包含部分信息的子网模型具有更好的容错能力, 对拓扑变化有较好的适应性。

3) 以给定的 POPN 故障诊断模型为基础, 通过仿真平台, 对 QBPN 模型进行仿真验证, 证实了 QBPN 算法的可靠性, 且能减少算法复杂度。

基于 POPN 模型的 QBPN 算法的复杂度依赖

于可能故障变迁的数量,需要大量时间用于QBPN子网中概率数据分析,通过研究变迁广义约束条件的 POPN 模型减少故障变迁集的搜索路径以优化算法速率。本文量子贝叶斯干涉项参数选择是手动进行没有考虑随机性,未来将研究对量子干涉项参数选择的影响,尝试 MCMC 随机采样算法得到更加结合实际变迁概率分布,进一步提高系统故障率的准确性。

参考文献 (References)

- [1] HENRY D, LE PEUVÉDIC C, STRIPPOLI L, et al. Robust model-based fault diagnosis of thruster faults in spacecraft[J]. IFAC-PapersOnLine, 2015, 48(21): 1078-1083.
- [2] DJEBKO K, PUPPE F, KAYAL H. Model-based fault detection and diagnosis for spacecraft with an application for the SONATE triple cube nano-satellite[J]. Aerospace, 2019, 6(10): 105.
- [3] 姜连祥, 李华旺, 杨根庆, 等. 航天器自主故障诊断技术研究进展[J]. 宇航学报, 2009, 30(4): 28-34.
JIANG L X, LI H W, YANG G Q, et al. Research progress of spacecraft autonomous fault diagnosis technology[J]. Acta astronautica Sinica, 2009, 30(4): 28-34 (in Chinese).
- [4] XU B, YIN X, YIN X G, et al. Fault diagnosis of power systems based on temporal constrained fuzzy Petri nets[J]. IEEE Access, 2019, 7: 101896-101904.
- [5] LEFEBVRE D. Fault diagnosis and prognosis with partially observed stochastic Petri nets[J]. Proceedings of the Institution of Mechanical Engineers. Part O: Journal of Risk and Reliability, 2014, 228(4): 382-396.
- [6] AN R M, LIANG W. Unobservable fuzzy Petri net diagnosis technique[J]. Aircraft Engineering and Aerospace Technology, 2013, 8(3): 215-221.
- [7] AMMOUR R, LECLERCQ E, SANLAVILLE E, et al. Faults prognosis using partially observed stochastic Petri-nets: An incremental approach [J]. Discrete Event Dynamic Systems, 2018, 28(2): 247-267.
- [8] 郭栋, 熊文真, 徐建新, 等. 基于变精度粗糙集与量子贝叶斯网络的变压器故障诊断研究[J]. 计算机应用与软件, 2017, 34(2): 93-99.
GUO D, XIONG W Z, XU J, et al. Research on transformer fault diagnosis based on variable precision rough set and quantum Bayesian network [J]. Computer Applications and Software, 2017, 34(2): 93-99 (in Chinese).
- [9] 刘久富, 张治国, 郑锐, 等. 双组元推进系统的部分可观时间 Petri 网故障诊断[J]. 系统工程与电子技术, 2018, 40(6): 1337-1344.
LIU J F, ZHANG Z G, ZHENG R, et al. Fault diagnosis of bi-propellant propulsion system using partially observed time Petri nets[J]. Systems Engineering and Electronics, 2018, 40(6): 1337-1344 (in Chinese).
- [10] LEMM J C. Bayesian approach to inverse time-dependent quantum mechanics[J]. Physics Letters A, 2000, 276(1): 19-24.
- [11] MOREIRA C, WICHERT A. Exploring the relations between quantum-like Bayesian networks and decision-making tasks with regard to face stimuli[J]. Journal of Mathematical Psychology, 2017, 78: 86-95.
- [12] YI Y H. Fault diagnosis based on Bayesian Petri nets[J]. Sensors and Transducers, 2014, 179(9): 114-120.
- [13] CABASINO M P, GIUA A, SEATZU C. Diagnosability of discrete event systems using labeled Petri nets[J]. IEEE Transactions on Automation Science and Engineering, 2014, 11(1): 144-153.
- [14] DON M G, KHAN F. Dynamic process fault detection and diagnosis based on a combined approach of hidden Markov and Bayesian network model [J]. Chemical Engineering Science, 2019, 201: 82-96.
- [15] SAHU A R, PALEI S K. Real-time fault diagnosis of HEMM using Bayesian network: A case study on drag system of dragline [J]. Engineering Failure Analysis, 2020, 118: 104917.
- [16] DECLERCK P, BONHOMME P. State estimation of timed labeled Petri nets with unobservable transitions[J]. IEEE Transactions on Automation Science and Engineering, 2014, 11(1): 103-110.
- [17] BASILE F, CORDONE R, PIRODDI L. A branch and bound approach for the design of decentralized supervisors in Petri net models[J]. Automatica, 2015, 52: 322-333.
- [18] SHEN Q, QIU J, LIU G, et al. Intermittent fault's parameter framework and stochastic Petri net based formalization model [J]. Eksploatacja i Niezawodność-Maintenance and Reliability, 2016, 18(2): 210-217.
- [19] WANG L, CHEN Q, GAO Z, et al. Knowledge representation and general Petri net models for power grid fault diagnosis[J]. IET Generation, Transmission & Distribution, 2015, 9(9): 866-873.
- [20] ZHANG Z D, ZHU J L, PAN F. Fault detection and diagnosis for data incomplete industrial systems with new Bayesian network approach [J]. Journal of Systems Engineering and Electronics, 2013, 24(3): 500-511.
- [21] PENSAR J, NYMAN H, LINTUSAARI J, et al. The role of local partial independence in learning of Bayesian networks[J]. International Journal of Approximate Reasoning, 2016, 69: 91-105.
- [22] YOON S. In-situ sensor calibration in an operational air-handling unit coupling autoencoder and Bayesian inference [J]. Energy and Buildings, 2020, 221: 110026.
- [23] OMOREGBEE H O, HEYNS P S. Fault detection in roller bearing operating at low speed and varying loads using Bayesian robust new hidden Markov model[J]. Journal of Mechanical Science and Technology, 2018, 32(9): 4025-4036.
- [24] 鲁峰, 黄金泉, 吕怡秋, 等. 基于非线性自适应滤波的发动机气路部件健康诊断方法[J]. 航空学报, 2013, 34(11): 2529-2538.
LU F, HUANG J Q, LV Y Q, et al. Aircraft engine gas-path components health diagnosis based on nonlinear adaptive filters [J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(11): 2529-2538 (in Chinese).
- [25] 陈金豹, 翟国富, 王淑娟, 等. 航天电子设备多余物检测信号特征的影响因素分析[J]. 系统工程与电子技术, 2013, 35(4): 889-894.

- CHEN J B, ZHAI G F, WANG S J, et al. Factors affecting characteristics of acoustic signals in particle impact noise detection for aerospace devices [J]. *Systems Engineering and Electronics*, 2013, 35(4):889-894 (in Chinese).
- [26] SHARDA B, BANERJEE A. Robust manufacturing system design using multi objective genetic algorithms, Petri nets and Bayesian uncertainty representation [J]. *Journal of Manufacturing Systems*, 2013, 32(2):315-324.
- [27] ZHU M N, HAN G, GUO H. Optimized fault diagnosis based on FMEA-style CBR and BN for embedded software system [J]. *The International Journal of Advanced Manufacturing Technology*, 2018, 94(9):3441-3453.
- [28] RATH Y, GLIELMO A, BOOTH G H. A Bayesian inference framework for compression and prediction of quantum states [J]. *Journal of Chemical Physics*, 2020, 153(12):124108.
- [29] HUANG Z M, YANG L, JIANG W. Uncertainty measurement with belief entropy on the interference effect in the quantum-like Bayesian networks [J]. *Applied Mathematics and Computation*, 2019, 347:417-428.
- [30] PIENAAR J. Quantum causal models via quantum Bayesianism [J]. *Physical Review A*, 2020, 101(1):012104.

Partial observable Petri nets fault diagnosis with quantum Bayesian learning

LIU Jiufu*, ZHANG Xinzhe, WANG Hengyu, A. M. TOMAS DIAS, WANG Zhisheng, YANG Zhong

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: This paper proposes an algorithm to construct a quantum Bayesian Petri nets model for the fault, and uses the sub net model to analyze the fault of Petri net system. According to the reachability identification diagram, which is the transition firing path can not judge the system state, establish the quantum Bayesian Petri nets subnet model to tackle the unobservable faults in partial observable Petri model. Through the quantum interference caused by the uncertain path, recalibrate the conditional probability table of the transition to obtain the quantum probability amplitude table. According to the pre-set of fault transition and quantum Bayesian reasoning, calculates the firing prior probability of transition. The posterior probability is modified by the observable transition in the post-set, and the state of the system is estimated by the maximum posterior probability. When the fault transition is not unique, the fault with the maximum probability is selected as the fault source. Finally, establishes a partial observable Petri nets model of a real fault system. Combined with the probability sequence information of observable label and quantum Bayesian probability estimation, the fault diagnosis of the unobservable parts of the system is carried out to verify the effectiveness of the algorithm with the data in simulation experiment.

Keywords: fault diagnosis; Petri nets; partially observed; quantum Bayesian; quantum interference

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

DOI: 10.13700/j.bh.1001-5965.2021.0037

基于零速修正与姿态自观测的 惯性行人导航算法



戴洪德*, 张笑宇, 郑百东, 戴邵武, 郑伟伟

(海军航空大学, 烟台 264001)

摘 要: 针对惯性行人导航中航向角发散致使导航精度降低的问题,提出了一种基于零速修正与姿态自观测的惯性行人导航算法。通过四条件零速检测算法对行走步态中的零速区间进行检测。在检测得到的零速区间内,利用零速修正算法原理构造速度误差的观测量;利用零速区间内行人脚部与地面保持静止、只受到重力加速度及姿态角不变的特性,构造姿态角误差的观测量。应用卡尔曼滤波对零速区间内的姿态角、速度及位置的误差进行估计。利用得到的误差状态估计结果对行人导航进行误差校正,提高惯性行人导航的精度。实验表明:小范围矩形路径中,所提算法的导航轨迹相对误差平均值仅占总路程的0.98%,比零速修正算法减小了78.11%;导航轨迹误差标准差仅为0.14 m,比零速修正算法减小了88.62%;400 m标准操场闭合路径中解算终点相对位置误差仅为1.18%。解算轨迹与实际轨迹匹配度较高,具有良好的应用价值。

关键词: 行人导航; 零速修正; 姿态自观测; 卡尔曼滤波; 误差校正

中图分类号: V241.62

文献标志码: A

文章编号: 1001-5965(2022)07-1135-10

基于位置的服务^[1] (location based services, LBS) 逐渐在个人消费领域兴起,行人获取自身精确空间位置的需求正变得越来越迫切,而建立在空间定位基础上的行人导航也成为了当前的研究热点。行人定位与导航算法可以分为有源算法和无源算法2类。有源算法主要有GNSS定位算法^[2]、WIFI定位算法^[3]、蓝牙定位算法^[4]、超声波定位算法^[5]等;而无源算法主要是惯性导航算法^[6]。尽管现在的有源算法(如GNSS定位算法)已经广泛地应用于生产和生活中,但也存在比较明显的局限性,其严重依赖于外部信号,在室内信号受阻及未知环境中无法应用。而无源

算法中的惯性导航算法具有全场景应用、可靠性高等优点,相比于有源算法具有更好的适用性。

利用惯性导航算法对行人进行定位与导航,简称为惯性行人导航。惯性行人导航主要通过将微惯性测量单元(micro inertial measure unit, MIMU)固定于行人足部,采集行人足部运动的加速度和角速度,通过捷联惯导算法解算出行人的位置,从而实现对行人的定位与导航。但由于MIMU存在测量误差,捷联惯导算法的解算精度会随着行走距离的增加而下降,最终会导致导航结果不可用。针对上述问题,Foxlin^[7]利用行人

收稿日期: 2021-01-21; 录用日期: 2021-07-01; 网络出版时间: 2021-07-07 13:48

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210707.1111.001.html

基金项目: 国防科技项目基金(F062102009); 山东省自然科学基金(ZR2017MF036); 山东省高等学校青年创新团队项目(2020KJN003)

* 通信作者. E-mail: 13954559561@126.com

引用格式: 戴洪德, 张笑宇, 郑百东, 等. 基于零速修正与姿态自观测的惯性行人导航算法[J]. 北京航空航天大学学报, 2022, 48(7): 1135-1144. DAI H D, ZHANG X Y, ZHENG B D, et al. Inertial pedestrian navigation algorithm based on zero velocity update and attitude self-observation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1135-1144 (in Chinese).

运动的周期性步态特征,首次将零速修正(zero velocity update, ZUPT)算法引入到惯性行人导航系统捷联解算中,该算法将零速时刻的速度误差作为观测量,利用扩展卡尔曼滤波(extended Kalman filter, EKF)对零速时刻导航误差进行估计与校正,从而实现导航误差的周期性修正,提高了导航精度。尽管ZUPT算法能够有效修正导航误差,但仍存在航向不可观的问题,使得导航精度不高。文献[8]提出一种双惯性测量单元框架,分别在行人足部与肩部固定一个惯性测量单元,通过2个惯性测量单元采集数据,并将两者解算出的航向角之差作为航向角误差的观测量,该算法能够一定程度提升导航精度,但增加惯性测量单元不仅导致系统复杂度提升,也有可能引入新的测量误差。文献[9]提出建筑物航向观测算法,通过预设建筑物的4个方向,利用行人在建筑物内直线行走一定处于4个方向之一的特点,将惯导解算出的航向角与上述4个方向之一做差,作为航向角误差的观测量,从而解决航向不可观的问题,该算法导航结果较好,但实现过程中需要严格规定行人的行走方向及预设建筑物方向,具有较大的应用局限性。文献[10-11]利用磁强计测量行人的磁航向,将磁航向与惯导解算航向之差作为航向角误差的观测量,从而实现航向角误差的观测,该方法由于采用广泛存在的地磁场测量航向,在户外环境具有较好的适用性,但在室内及复杂城市环境中,磁场容易受到干扰,使得磁航向角测量产生较大误差,从而影响导航误差的校正。

本文针对上述问题,提出一种基于零速修正与姿态自观测的惯性行人导航算法。该算法一方面通过零速修正算法解决纯惯导解算误差积累的问题,使惯性行人导航的基本功能得以实现;另一方面,通过姿态自观测算法在零速区间内解决姿态误差增大影响导航精度的问题。首先,应用重力加速度各分量计算得到俯仰角与滚转角,将其与惯导解算出的俯仰角与滚转角求差,作为俯仰角误差与滚转角误差的观测量。然后,在此基础上,利用零速区间内航向角不变的特性,将当前采样点航向角解算值与前2个采样点航向角平均值之差作为航向角误差的观测量。最后,通过卡尔曼滤波对系统误差进行估计,实现对导航误差的校正,从而进一步提高导航精度。本文算法充分利用零速区间的步态特性,无需引入外部设备或预设行走方向,具有较好的实用价值。

1 行人步态特征分析

行人行走过程中,脚部运动具有一定的规律性,并且具有周期性特征。行人的脚部运动从脚跟离地开始,经过脚尖离地、空中摆动、脚跟着地、脚尖着地,最终脚部与地面保持相对静止并持续一段时间,之后又开始下一步的脚跟离地,以此做周期性循环^[12]。从脚跟离地到持续静止结束之间的脚部动作过程称为一个步态周期。步态周期中脚部在空中的运动阶段称为运动区间,而脚部与地面保持静止的阶段称为零速区间^[13]。行人的步态周期如图1所示。

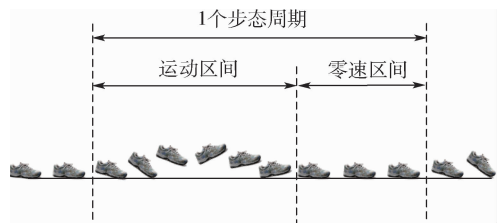


图1 行人步态周期示意图

Fig. 1 Schematic diagram of pedestrian foot gait cycle

2 零速检测算法

行人的步态周期具有运动区间与零速区间,其中零速区间是行人导航的重要特征。零速区间时,脚部与地面保持静止,此时行人脚部受到的加速度模值等于重力加速度,脚部的速度应等于零,并且单个零速区间内脚部的姿态应始终保持不变。根据零速区间的上述特征,就可以构建相应的观测量,利用卡尔曼滤波进行误差估计,实现惯性行人导航的误差校正。而进行误差估计与校正的前提是准确地识别出零速区间,因此需要建立有效的零速检测算法。文献[14]提出一种基于加速度量测幅值的零速检测算法,将三轴加速度的模值作为零速检测的指标,并针对检测过程中的伪零速点进行剔除,从而保证零速检测的准确性与可靠性。文献[15]提出三条件法对零速区间进行检测,并与单条件检测算法进行对比,通过检测到的步数对比,得出三条件检测算法比单条件检测算法更加精确。本文在利用加速度模值、加速度模值标准差、角速度模值的三条件检测算法基础上,增加角速度模值标准差这一条件,采用四条件零速检测算法,应用于本文行走数据的零速区间检测,保证检测结果具有更高的检测精度^[12]。

四条件零速检测算法如下:

1) 加速度模值条件。

$$|a_i| = \sqrt{a_{xi}^2 + a_{yi}^2 + a_{zi}^2} \quad (1)$$

$$C_1(i) = \begin{cases} 1 & h_{|a|_{\min}} < |a_i| < h_{|a|_{\max}} \\ 0 & \text{其他} \end{cases} \quad (2)$$

式中: i 为采样点; a_x 、 a_y 、 a_z 分别为三轴输出加速度信息; $|a_i|$ 为加速度模值; $h_{|a|_{\max}}$ 、 $h_{|a|_{\min}}$ 分别为加速度模值阈值上、下限。

2) 角速度模值条件。

$$|\omega_i| = \sqrt{\omega_{xi}^2 + \omega_{yi}^2 + \omega_{zi}^2} \quad (3)$$

$$C_2(i) = \begin{cases} 1 & |\omega_i| < h_{|\omega|} \\ 0 & \text{其他} \end{cases} \quad (4)$$

式中: ω_x 、 ω_y 、 ω_z 分别为三轴输出角速度信息; $|\omega_i|$ 为角速度模值; $h_{|\omega|}$ 为角速度模值阈值。

3) 加速度模值标准差条件。

$$\sigma_{|a_i|} = \sqrt{\frac{1}{2n+1} \sum_{j=i-n}^{i+n} (|a_j| - \overline{|a_i|})^2} \quad (5)$$

$$C_3(i) = \begin{cases} 1 & \sigma_{|a_i|} < h_{\sigma_{|a|}} \\ 0 & \text{其他} \end{cases} \quad (6)$$

式中: $\sigma_{|a_i|}$ 为以 $|a_i|$ 为采样中点的加速度模值标准差; $2n+1$ 为采样区间长度; $\overline{|a_i|}$ 为加速度模值均值; $|a_j|$ 为采样点 j 的加速度模值; $h_{\sigma_{|a|}}$ 为加速度模值标准差阈值。

4) 角速度模值标准差条件。

$$\sigma_{|\omega_i|} = \sqrt{\frac{1}{2n+1} \sum_{j=i-n}^{i+n} (|\omega_j| - \overline{|\omega_i|})^2} \quad (7)$$

$$C_4(i) = \begin{cases} 1 & \sigma_{|\omega_i|} < h_{\sigma_{|\omega|}} \\ 0 & \text{其他} \end{cases} \quad (8)$$

式中: $\sigma_{|\omega_i|}$ 为以 $|\omega_i|$ 为采样中点的角速度模值标准差; $2n+1$ 为采样区间长度; $\overline{|\omega_i|}$ 为角速度模值均值; $|\omega_j|$ 为采样点 j 的角速度模值; $h_{\sigma_{|\omega|}}$ 为角速度模值标准差阈值。阈值设置和采样窗口的选取参考文献[12]。

当采样点 i 同时满足4个条件时,即

$$C_1(i) \& C_2(i) \& C_3(i) \& C_4(i) = 1 \quad (9)$$

则将采样点 i 判定为零速点,连续零速点则构成行走时的零速区间。

3 惯性行人导航系统

3.1 惯性行人导航系统实现原理

本文设计的惯性行人导航系统由MIMU惯性传感器、信号采集传输和导航计算机共同构成。其中,MIMU惯性传感器安装于行人的左脚后跟位置,通过数据线与导航计算机相连,将采集到的行走数据传输到导航计算机中。传感器安装实物

图及惯性行人导航系统构成分别如图2(a)和图2(b)所示。



(a) 传感器安装实物图



(b) 惯性行人导航系统构成

图2 传感器安装实物图及惯性行人导航系统构成
Fig.2 Physical map of sensor installation positions and composition of inertial pedestrian navigation system

本文设计的惯性行人导航系统工作原理为:行人运动过程中,MIMU惯性传感器输出的数据被采集到导航计算机中,导航计算机对采集的数据进行捷联惯导解算,得到行人运动的速度、位置及传感器姿态信息。由于采用纯惯导解算方法时,导航误差会随时间逐渐积累,需要对导航误差进行修正;对采集的数据同时进行零速检测,检测出行走时的零速区间,在零速区间内运用零速修正与姿态自观测算法估计出导航姿态、位置及速度误差,并对惯导解算结果进行误差修正,从而实现对行人的精确定位与导航。基于零速修正与姿态自观测的惯性行人导航算法原理如图3所示。

图3中,零速检测之后,在检测到的零速区间内,利用重力加速度分量可以计算出俯仰角 θ_{ACC} 和滚转角 γ_{ACC} ,而捷联惯导解算之后可以获得解算俯仰角 θ_{INS} 、滚转角 γ_{INS} 及航向角 ϕ , θ_{ACC} 和 γ_{ACC} 分别与 θ_{INS} 和 γ_{INS} 作差可以获得俯仰角和滚转角误差观测量,而当前时刻的航向角 ϕ 与前2个时刻的航向角平均值作差可以得到航向角误差观测量,俯仰角、滚转角和航向角误差观测量共同构成姿态误差观测量;同时,在零速区间内,惯导解算速度 v_{INS} 就作为速度误差观测量。以上述速度误差与姿态误差作为观测量,运用卡尔曼滤波进行

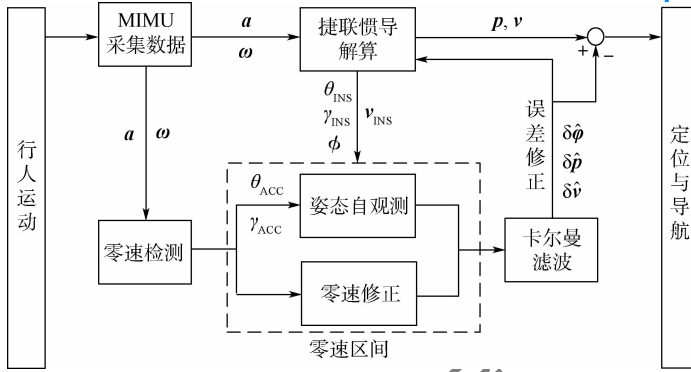


图3 基于零速修正与姿态自观测的惯性行人导航算法原理

Fig.3 Working principle of zero velocity update and attitude self-observation algorithm of inertial pedestrian navigation

状态估计,就可以得到姿态、位置和速度误差的估计值,利用误差估计值就可以实现惯性行人导航系统的误差修正。

3.2 惯性行人导航系统模型

惯性行人导航系统采用捷联惯导算法进行导航解算,因此先分析捷联惯导系统误差模型。在小扰动情况下,捷联惯导的系统误差模型表示如下^[16]:

$$\begin{cases} \delta \dot{\varphi} = -\omega_{in}^n \times \delta \varphi + \delta \omega_{in}^n - C_b^n \delta \omega_{ib}^b \\ \delta \dot{v}^n = (C_{bf}^n \times) \delta \varphi - (2\omega_{ie}^n + \omega_{en}^n) \times \delta v^n + v^n \times (2\delta \omega_{ie}^n + \delta \omega_{en}^n) + C_b^n \delta f_{sf}^b + \delta g^n \\ \delta \dot{p}^n = \delta v^n \end{cases} \quad (10)$$

式中: $\delta \varphi = [\delta \varphi_E \ \delta \varphi_N \ \delta \varphi_U]^T$ 为捷联惯导的姿态角误差; $\delta v = [\delta v_E^n \ \delta v_N^n \ \delta v_U^n]^T$ 为捷联惯导的速度误差; $\delta p = [\delta x \ \delta y \ \delta z]^T$ 为捷联惯导的位置误差; $\delta \omega_{ib}^b$ 为陀螺仪测量误差; δf_{sf}^b 为加速度计测量误差; f_{sf}^b 为载体坐标系中的加速度计测量值; ω 为角速度; C 为姿态转换矩阵; g 为重力加速度。

由于行人速度相对较低且 MIMU 惯性传感器中的噪声较大,地球自转角速度在测量过程中会被噪声淹没,上述模型中的 ω_{en}^n 、 ω_{in}^n 、 ω_{ie}^n 及 δg^n 可以忽略不计^[16]。针对惯性行人导航系统简化后的捷联惯导误差模型可以表示为

$$\begin{cases} \delta \dot{\varphi} = -C_b^n \delta \omega_{ib}^b \\ \delta \dot{v}^n = (C_{bf}^n \times) \delta \varphi + C_b^n \delta f_{sf}^b \\ \delta \dot{p}^n = \delta v^n \end{cases} \quad (11)$$

3.3 系统状态方程与观测方程

根据误差模型,选取上述姿态角、位置及速度误差作为卡尔曼滤波器的状态变量,建立连续系统状态方程如下:

$$\dot{x}(t) = F(t)x(t) + G(t)w(t) \quad (12)$$

式中:

$$x(t) = [(\delta \varphi)^T \ (\delta p^n)^T \ (\delta v^n)^T]^T$$

$$F(t) = \begin{bmatrix} \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} \\ (C_{bf}^n \times) & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix}$$

$$G(t) = \begin{bmatrix} -C_b^n & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & C_b^n \end{bmatrix}$$

$$w(t) = [(\sigma_{\omega}^b)^T \ (\sigma_f^b)^T]^T$$

其中: σ_{ω}^b 和 σ_f^b 分别为陀螺仪和加速度计的噪声方差阵。

系统观测方程为

$$z(t) = Hx(t) + v(t) \quad (13)$$

式中: H 为观测矩阵; $v(t)$ 为观测噪声。

对上述连续系统进行离散化可得

$$X_k = \Phi_{k/k-1} X_{k-1} + \Gamma_{k/k-1} W_{k-1} \quad (14)$$

$$Z_k = HX_k + V_k \quad (15)$$

式中: X_k 为被估计状态; Z_k 表示对 X_k 的观测量; W_{k-1} 为离散化后的观测噪声。

$$\Phi_{k/k-1} = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} & \Delta t \mathbf{I}_{3 \times 3} \\ \Delta t (C_{bf}^n \times) & \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} \end{bmatrix}$$

$$\Gamma_{k/k-1} = \begin{bmatrix} -\Delta t C_b^n & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \Delta t C_b^n \end{bmatrix}$$

4 零速修正与姿态自观测算法

4.1 零速修正算法

当行人处于零速区间时,脚部的理论速度应该等于零,但由于存在测量误差等因素,导致实际计算得到的速度并不等于零。因此,将零速时刻解算得到的速度作为系统速度误差的观测量,零速修正算法的观测方程为

$$Z_{vk} = \delta V_k^n = V_k^n - \mathbf{0} = H_v X_k + V_{vk} \quad (16)$$

式中: $V_k^n = [V_{xk}^n \ V_{yk}^n \ V_{zk}^n]^T$; $H_v = [\mathbf{0}_{3 \times 3} \ \mathbf{0}_{3 \times 3} \ \mathbf{I}_{3 \times 3}]$ 。

4.2 姿态自观测算法

当行人处于零速区间时,由于脚部与地面保持静止,MIMU 惯性传感器的输入信息只有重力加速度,则如下等式在零速区间成立:

$$\begin{bmatrix} f_x^b \\ f_y^b \\ f_z^b \end{bmatrix} = C_n^b \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix} \quad (17)$$

在本文中,依照航向角、俯仰角和滚转角的次序将地理坐标系变换到载体坐标系,以此建立数据模型,即可推出 $C_n^{b[17]}$,进而可得

$$\begin{bmatrix} f_x^b \\ f_y^b \\ f_z^b \end{bmatrix} = \begin{bmatrix} g \sin \gamma \cos \theta \\ -g \sin \theta \\ -g \cos \gamma \cos \theta \end{bmatrix}$$

可以通过三轴加速度计输出确定出 MIMU 惯性传感器的俯仰角和滚转角:

$$\begin{cases} \theta_{ACC} = \arcsin(-f_y^b/g) \\ \gamma_{ACC} = \arctan(-f_x^b/f_z^b) \end{cases} \quad (18)$$

此时,可以利用惯导解算出的俯仰角 θ_{INS} 和滚转角 γ_{INS} 与通过加速度确定的俯仰角 θ_{ACC} 和滚转角 γ_{ACC} 作差,构造俯仰角和滚转角误差观测量,但由于无法通过加速度计算航向角,基于上述条件无法构造航向角误差观测量,无法通过卡尔曼滤波修正航向角误差。而通过进一步分析零速区间可以发现,脚部与地面处于相对静止时,同一零速区间内的航向角也应具有不变的特性,因此,可以将当前采样点航向角解算值与前2个采样点航向角平均值之差作为航向角误差的观测值:

$$\delta\phi_k = \phi_k - \frac{\phi_{k-1} + \phi_{k-2}}{2} \quad (19)$$

由此,可以得到姿态自观测算法的观测方程为

$$Z_{\varphi k} = \delta\psi_k = \begin{bmatrix} \delta\theta \\ \delta\gamma \\ \delta\phi \end{bmatrix} = H_{\varphi} X_k + V_{\varphi k} \quad (20)$$

式中:

$$\begin{cases} \delta\theta = \theta_{INS} - \theta_{ACC} \\ \delta\gamma = \gamma_{INS} - \gamma_{ACC} \end{cases}$$

$$H_{\varphi} = [I_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3}]$$

将零速修正算法与姿态自观测算法的观测方程进行组合,得到系统观测方程为

$$Z_k = \begin{bmatrix} Z_{vk} \\ Z_{\varphi k} \end{bmatrix} = \begin{bmatrix} \delta V_k^n \\ \delta\psi_k \end{bmatrix} = H X_k + V_k \quad (21)$$

式中:

$$H = \begin{bmatrix} H_v \\ H_{\varphi} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & I_{3 \times 3} \\ I_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix}$$

$$V_k = \begin{bmatrix} V_{vk} \\ V_{\varphi k} \end{bmatrix}$$

根据上述零速修正与姿态自观测算法,通过卡尔曼滤波就可以得到系统误差的估计值 $\delta\hat{\varphi}_k^n$ 、 $\delta\hat{p}_k^n$ 、 $\delta\hat{v}_k^n$,从而实现对捷联惯导解算参数的误差修正:

$$\begin{cases} \hat{\varphi}_k^n = \varphi_k^n \oplus \delta\hat{\varphi}_k^n \\ \hat{p}_k^n = p_k^n + \delta\hat{p}_k^n \\ \hat{v}_k^n = v_k^n + \delta\hat{v}_k^n \end{cases} \quad (22)$$

5 实验验证

本文采用的 MIMU 惯性传感器型号为荷兰 Xsens 公司生产的 MTi-G-710 产品,其能够实时测量载体的加速度及角速度。加速度计的量程为 $\pm 50 \text{ m/s}^2$,陀螺仪的量程为 $\pm 450^\circ/\text{s}$,MIMU 惯性传感器的工作频率设置为 50 Hz。利用本文设计的惯性行人导航系统进行实际行走实验,先进行 $9 \text{ m} \times 17 \text{ m}$ 小范围矩形路径的行走。实验得到的结果如图 4~图 6 所示。

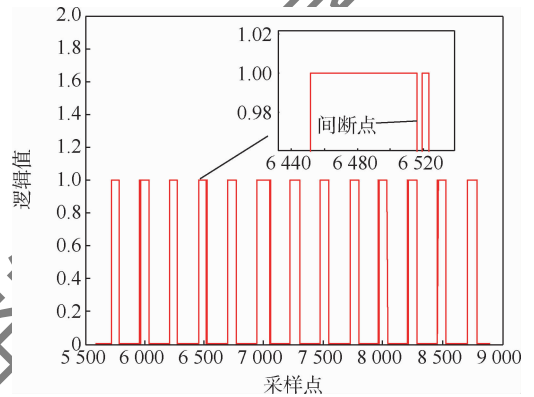
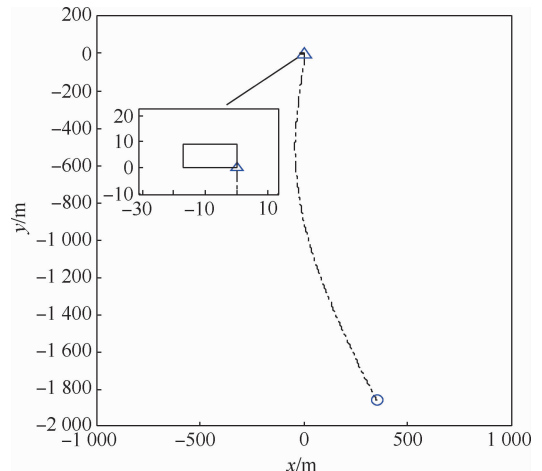


图4 四条件零速检测算法零速检测结果

Fig.4 Results of zero velocity test of four-condition zero velocity detection algorithm



— 实际行走轨迹 - - - 纯惯导解算轨迹
△ 起点 ○ 终点

图5 纯惯导算法轨迹

Fig.5 Trajectory of pure inertial navigation algorithm

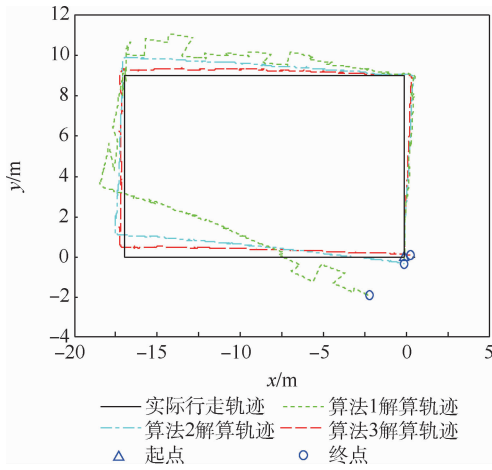


图6 不同算法轨迹对比

Fig. 6 Comparison of different algorithm trajectories

图4为采用四条件零速检测算法得到的零速检测结果。当逻辑值为1时,则代表相应的采样点处于零速时刻,逻辑值为0时,代表相应的采样点处于非零速时刻。从图4中可以看到,四条件零速检测算法能够较好地检测出行人运动的零速时刻。但对于理想的步态周期,零速区间内的零速时刻点应该是连续的,中间不会出现逻辑值为0的间断点。实际情况下,由于存在数据测量误差,零速区间内出现间断点不可避免,这些间断点即为误检点。而四条件零速检测算法能够充分利用采集到的数据信息,使零速区间内出现误检点的数量尽可能得少,使零速检测具有较高的精确度。为评价零速检测算法的精确度,采用零速区间平均单步误检率指标^[12]对本文的零速检测算法精确度进行评估。零速区间平均单步误检率计算方法如下:

$$\eta = \left(\frac{1}{N} \sum_{k=1}^N \frac{m_k}{M_k} \right) \times 100\% \quad (23)$$

式中: N 为零速区间个数; M_k 为第 k 个零速区间内零速时刻点个数; m_k 为第 k 个零速区间内误检点个数。

式(23)的含义即先计算每个零速区间的误检率,再对所有零速区间的误检率求平均值,得到平均单步误检率。对本文零速检测结果计算平均单步误检率,得到结果为 $\eta = 2.38\%$ 。零速区间内零速时刻点约为70个,平均每个零速区间会出现1~2个间断点,这对于系统解算影响很小,因此,四条件零速检测算法能够适用于惯性行人导航零速区间的检测,具有较高的检测精确度。

图5为纯惯导算法解算得到的行走轨迹。由于纯惯导算法无法对误差进行修正,误差会随时间不断积累,得到的解算轨迹是一条发散的曲线。

为便于比较,在图中同时画出实际行走的轨迹,由于实际行走轨迹与纯惯导解算的轨迹在量级上存在较大差距,因此在图5中显示位于起点附近。图5轨迹证明,行人导航中无法利用纯惯导算法进行导航。

图6为采用不同修正算法得到的轨迹对比。图中,黑色曲线为实际行走轨迹;绿色曲线为仅采用零速修正算法(算法1)进行误差修正得到的轨迹;蓝色曲线为采用零速修正与俯仰角、滚转角观测算法(算法2)进行误差修正得到的轨迹;红色曲线为采用本文提出的零速修正与姿态自观测算法(算法3)进行误差修正得到的轨迹。算法2相比于算法1加入了俯仰角、滚转角误差观测测量,但未加入航向角误差观测测量;算法3相比于算法2增加了航向角误差观测测量,与算法2中的俯仰角、滚转角误差观测测量构成完整的姿态误差观测测量。

从图6可以得到,算法1解算轨迹由于存在零速修正的作用,不再呈现为发散的曲线,而是能够大致刻画出实际的行走形状,且起点与终点位置较为接近。但由于只有速度误差的观测测量,缺少姿态误差的观测,算法1对于姿态误差的修正效果较差,导致轨迹中出现部分曲折形状,曲折处轨迹与实际行走轨迹存在较大的偏离,且整个轨迹的航向角也存在一定程度的偏差,算法解算精度较低。

算法2解算轨迹由于在零速修正基础上加入了俯仰角、滚转角误差的观测测量,成功消除了算法1解算轨迹中存在的曲折形状,使曲折轨迹变成直线轨迹,且终点与起点位置非常接近,使得整个解算轨迹呈现出较为理想的矩形路径形状。但由于缺少航向角误差的观测测量,算法2对于航向角误差的修正效果较差,使得解算得到的轨迹与实际行走轨迹存在较大的航向角误差,2条轨迹的匹配度较低。

由于算法3在零速修正与俯仰角、滚转角观测算法基础上又加入了航向角误差的观测测量,使得算法2中存在的航向角误差得到一定的消除,得到的轨迹与实际行走轨迹更加接近,导航误差得到进一步减小。

从图6可以直观地看出3种算法导航误差的大小,算法3导航误差最小且解算轨迹与实际行走轨迹最接近,算法1导航误差最大且解算轨迹与实际行走轨迹差异最大。

为定量对比3种算法的导航精度,选取实际行走轨迹中的4个点作为参考点,分别为转折点1(tp1)、转折点2(tp2)、转折点3(tp3)及终点(fp),如图7所示。

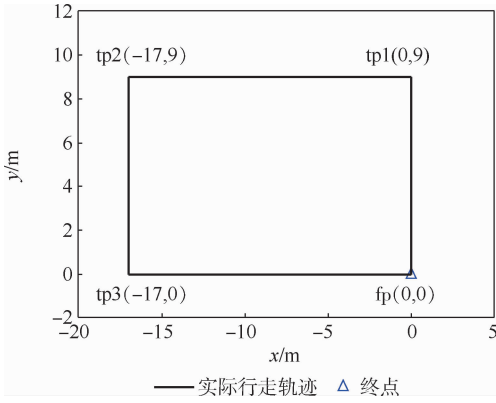


图 7 实际行走轨迹参考点示意图

Fig. 7 Schematic diagram of actual walking track reference point

在 3 种算法的解算轨迹中,同时取相应的转折点及终点,作为实际行走轨迹参考点在解算轨迹中的对应点。对应点 o 与参考点 re 的位置误差定义为

$$\Delta\gamma = \gamma_o - \gamma_{re} = \sqrt{(x_o - x_{re})^2 + (y_o - y_{re})^2} \quad (24)$$

分别计算 4 个对应点与 4 个参考点的位置误差,将 4 个位置误差取平均值,作为解算轨迹与实际轨迹的导航轨迹误差平均值;对 4 个位置误差求标准差,作为解算轨迹与实际轨迹的导航轨迹误差标准差,计算方法为

$$\Delta R = \frac{\Delta\gamma_{tp1} + \Delta\gamma_{tp2} + \Delta\gamma_{tp3} + \Delta\gamma_{fp}}{4} \quad (25)$$

表 1 三种算法位置误差与导航轨迹误差

Table 1 Position error and navigation track error of three kinds of algorithm

算法	参考点坐标	对应点坐标	$\Delta\gamma$	ΔR	ΔS
零速修正算法(算法 1)	(0,9)	(0.68,8.86)	0.69	2.33	1.23
	(-17,9)	(-16.63,10.71)	1.75		
	(-17,0)	(-18.52,3.69)	3.99		
	(0,0)	(-2.12,-1.93)	2.87		
零速修正与俯仰角、滚转角观测算法(算法 2)	(0,9)	(0.62,8.83)	0.64	0.86	0.41
	(-17,9)	(-16.84,10.00)	1.01		
	(-17,0)	(-17.64,1.29)	1.44		
零速修正与姿态自观测算法(算法 3)	(0,0)	(0.04,-0.35)	0.35	0.51	0.14
	(0,9)	(0.50,8.84)	0.52		
	(-17,9)	(-17.00,9.39)	0.39		
	(-17,0)	(-17.36,0.65)	0.74		
	(0,0)	(0.38,0.09)	0.39		

为验证长距离行走下本文算法的实用性,选择学校操场进行闭合圈行走实验,操场卫星图如图 8 所示,图中五角星处为行走起点。

采用本文设计的惯性行人导航系统,沿操场内道行走一圈并回到起点,总路程为 400 m,记录实验数据,分别利用 3 种算法进行导航解算,解算得到的操场行走轨迹如图 9 所示,图中三角形处

$$\Delta S = \sqrt{\frac{\sum_{i=1}^4 (\Delta\gamma_i - \Delta R)^2}{4}} \quad i = 1, 2, 3, 4 \quad (26)$$

式中: $i = 1, 2, 3, 4$ 分别代表转折点 1、转折点 2、转折点 3、终点。

通过导航轨迹误差平均值及标准差 2 个定量指标,即可量化 3 种算法解算轨迹与实际轨迹的接近程度,从而评估 3 种算法的导航精度。

由于矩形实验中涉及的行走轨迹为标准的矩形,解算得到的轨迹也较为接近矩形,可以通过上述 4 个参考点的位置误差平均值及标准差来描述整个轨迹的导航误差,对于其他较为复杂的行走路径,则应该选择轨迹中多个具有代表性的点来描述导航轨迹误差。

3 种算法的 4 个位置误差及导航轨迹误差计算结果如表 1 所示。

从表 1 中的计算结果可以得到,零速修正与姿态自观测算法具有最小的导航轨迹误差平均值, ΔR 仅为 0.51 m,占总路程的 0.98%,相对零速修正算法减小了 78.11%,表明零速修正与姿态自观测算法导航的精度最高。而在具有最小导航轨迹误差平均值 ΔR 的基础上,该算法又具有最小的导航轨迹误差标准差, ΔS 仅为 0.14 m,相对零速修正算法减小了 88.62%,表明该算法解算得到的轨迹与实际轨迹的误差均维持在较小水平且误差变化较小,不存在误差积累或者突变的情况。

为行走起点,圆形处为各轨迹解算终点。

从图 9 可以得到,在操场闭合圈行走实验中,由于距离增加,算法 1 解算轨迹与实际行走轨迹差距较大,解算终点离起点较远,由于缺少姿态误差修正,轨迹中仍然存在曲折形状,航向角偏差也很大,算法 1 在长距离行走中已无法胜任导航解算。

算法2由于加入了俯仰角、滚转角误差的观测测量,解算轨迹较为平滑,解算轨迹与操场的形状较为接近,解算终点离起点的距离也有所减小。在



图8 操场卫星图

Fig.8 Satellite image of playground

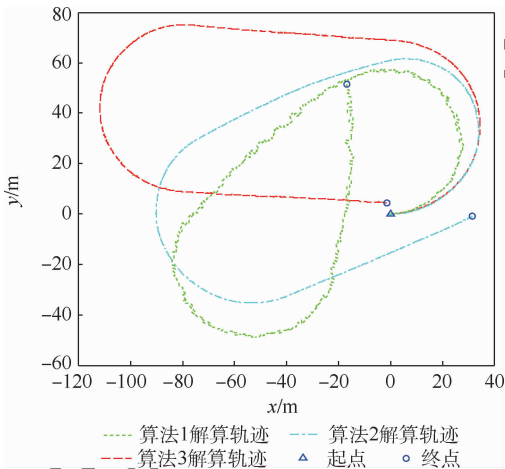


图9 操场行走轨迹

Fig.9 Trajectory of walking on playground

距离较短情况下,解算轨迹与实际轨迹较为接近,但由于算法2缺少对航向角误差的修正,随着行走距离的增加,解算的航向角误差逐渐变大,最终使得解算轨迹与实际行走轨迹存在较大的航向角偏差。

算法3解算轨迹较为平滑,与操场内圈形状最为接近。在此基础上,由于算法3相比于算法2具有航向角误差观测测量,成功修正了解算轨迹中的航向角误差,使得解算终点与起点较接近,导航误差最小,达到了较为理想的解算精度。

为定量对比3种算法在操场闭合圈实验中的导航精度,以解算终点的相对位置误差为指标,对三者进行对比,计算结果如表2所示。

从表2可以得到,算法3具有最小的位置误差4.72 m,在400 m的操场闭合圈行走轨迹中,解算终点相对位置误差为1.18%,导航精度比算法1和算法2有较大的提高,达到了理想的导航精度。

上述矩形和操场闭合圈行走实验结果共同表明,利用本文算法导航解算得到的轨迹与实际轨迹的吻合度较高,轨迹形状误差较小,解算得到的轨迹终点与实际终点的相对位置误差也较小,导航准确性较高。在长距离行走中,算法能够对航向角误差进行自观测的优势则体现的更为明显。

表2 三种算法解算终点的相对位置误差

Table 2 Relative position error of the end point of three kinds of algorithm

算法	起点坐标/m	终点坐标/m	位置误差/m	相对位置误差/%
零速修正算法(算法1)	(0,0)	(-16.80,51.33)	54.00	13.50
零速修正与俯仰角、滚转角观测算法(算法2)	(0,0)	(31.50,-0.86)	31.51	7.88
零速修正与姿态自观测算法(算法3)	(0,0)	(-1.50,4.48)	4.72	1.18

6 结论

1) 本文提出的基于零速修正与姿态自观测的惯性行人导航算法无需增加传感器且不需要对行人行走轨迹进行约束,就能实现对航向角误差的自观测,具有较好的实用价值。

2) 本文算法将零速修正与姿态自观测进行结合,与零速修正算法相比增加了姿态角误差的观测测量,提高了误差估计的精度。通过实际行走实验,矩形路径中,本文算法的导航轨迹误差平均值仅为0.51 m,只占总路程的0.98%,相对零速修正算法减小了78.11%,导航轨迹误差标准差仅为0.14 m,相对零速修正算法减小了88.62%;操场闭合圈路径中,本文算法的解算终点相对位

置误差仅为1.18%。

3) 本文算法相比于零速修正算法,有效减小了航向角误差的发散,进一步提高了惯性行人导航的精度。

参考文献 (References)

- [1] HUANG H, GARTNER G, KRISP J M, et al. Location based services: Ongoing evolution and research agenda[J]. Journal of Location Based Services, 2018, 12(2): 63-93.
- [2] XIN S, LI X, GENG J, et al. The implementation of multi-GNSS real-time precise point positioning service system: GPS/GLO-NASS [C]//2017 Forum on Cooperative Positioning and Service (CPGPS). Piscataway: IEEE Press, 2017: 202-207.
- [3] SUN W, XUE M, YU H, et al. Augmentation of fingerprints for indoor WiFi localization based on Gaussian process regression [J]. IEEE Transactions on Vehicular Technology, 2018, 67

- (11):10896-10905.
- [4] ZHUANG Y, YANG J, LI Y, et al. Smartphone-based indoor localization with bluetooth low energy beacons [J]. *Sensors*, 2016, 16(5):596.
- [5] QI J, LIU G P. A robust high-accuracy ultrasound indoor positioning system based on a wireless sensor network[J]. *Sensors*, 2017, 17(11):2554.
- [6] SUN W, DING W, YAN H F, et al. Zero velocity interval detection based on a continuous hidden Markov model in micro inertial pedestrian navigation[J]. *Measurement Science and Technology*, 2018, 29(6):7.
- [7] FOXLIN E. Pedestrian tracking with shoe-mounted inertial sensors[J]. *IEEE Computer Graphics and Applications*, 2005, 25(6):38-46.
- [8] 徐元, 陈熙源, 李庆华, 等. 一种基于双IMU框架的室内个人导航方法[J]. *中国惯性技术学报*, 2015, 23(6):714-717.
XU Y, CHEN X Y, LI Q H, et al. An indoor personal navigation method based on double IMU framework[J]. *Chinese Journal of Inertial Technology*, 2015, 23(6):714-717 (in Chinese).
- [9] ABDULRAHIM K, HIDE C, MOORE T, et al. Aiding low cost inertial navigation with building heading for pedestrian navigation[J]. *Journal of Navigation*, 2011, 64(2):219-233.
- [10] WANG Q, YIN J, NOURELDIN A, et al. Research on an improved method for foot-mounted inertial/magnetometer pedestrian-positioning based on the adaptive gradient descent algorithm [J]. *Sensors*, 2018, 18(12):4105.
- [11] QIU S, WANG Z L, ZHAO H Y, et al. Inertial/magnetic sensors based pedestrian dead reckoning by means of multi-sensor fusion[J]. *Information Fusion*, 2018, 39:108-119.
- [12] 郑伟伟, 戴邵武, 王希彬, 等. 惯性行人导航系统传感器安装位置研究[J]. *传感技术学报*, 2020, 33(6):815-824.
- ZHENG W W, DAI S W, WANG X B, et al. Research on sensor installation position of inertial pedestrian navigation system[J]. *Chinese Journal of Sensors and Actuators*, 2020, 33(6):815-824 (in Chinese).
- [13] 戴洪德, 李松林, 周绍磊, 等. 基于伪标准差和N-P准则的行人导航零速检测[J]. *中国惯性技术学报*, 2018, 26(6):701-707.
DAI H D, LI S L, ZHOU S L, et al. Zero speed detection of pedestrian navigation based on pseudo standard deviation and N-P criterion[J]. *Chinese Journal of Inertial Technology*, 2018, 26(6):701-707 (in Chinese).
- [14] 陈国良, 杨洲. 基于加速度量测幅值零速检测的计步算法研究[J]. *武汉大学学报(信息科学版)*, 2017, 42(6):726-730.
CHEN G L, YANG Z. Research on step-counting algorithm based on accelerated measurement amplitude zero velocity detection[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(6):726-730 (in Chinese).
- [15] ZHENG L, ZHOU W, TANG W, et al. A 3D indoor positioning system based on low-cost MEMS sensors[J]. *Simulation Modelling Practice and Theory*, 2016, 65:45-56.
- [16] 周绍磊, 李松林, 戴洪德, 等. 一种优化高度通道的行人导航算法[J]. *海军航空工程学院学报*, 2018, 33(5):60-66.
ZHOU S L, LI S L, DAI H D, et al. A pedestrian navigation algorithm for optimizing elevation channel[J]. *Journal of Naval Aeronautical Engineering Institute*, 2018, 33(5):60-66 (in Chinese).
- [17] 秦永元. 惯性导航[M]. 北京: 科学出版, 2014:4-7.
QIN Y Y. *Inertial navigation* [M]. Beijing: Science Press, 2014:4-7 (in Chinese).

Inertial pedestrian navigation algorithm based on zero velocity update and attitude self-observation

DAI Hongde*, ZHANG Xiaoyu, ZHENG Baidong, DAI Shaowu, ZHENG Weiwei

(Naval Aviation University, Yantai 264001, China)

Abstract: Aiming at the problem of reduced navigation accuracy caused by the divergence of the heading angle in inertial pedestrian navigation, an inertial pedestrian navigation algorithm based on zero velocity correction and attitude self-observation is proposed. A four-condition zero velocity detection algorithm is used to detect the zero velocity interval in the walking gait. In the detected zero velocity interval, the principle of the zero velocity update is used to construct the observation of the velocity error; the characteristic that only gravity acts and the heading angle remains unchanged in the zero velocity intervals is used to construct the observation of the attitude angle error. Then, the attitude angle, velocity and position error in the zero velocity interval are estimated by Kalman filtering. The error correction of pedestrian navigation is carried out using the obtained state estimation to further improve the accuracy of inertial pedestrian navigation. Actual walking experiments show that in the rectangular path, the average value of navigation trajectory relative error of this algorithm is only 0.98%, which is reduced by 78.14% compared with the zero velocity update algorithm and the standard deviation of navigation trajectory error of this algorithm is only 0.14 m, which is reduced by 88.62% compared with the zero velocity update algorithm. In the closed loop path of the classical 400 m playground, the relative position error of the solution end point is only 1.18%. The solved trajectory has a high degree of matching with the actual trajectory, which has good application value.

Keywords: pedestrian navigation; zero velocity update; attitude self-observation; Kalman filtering; error correction

Received: 2021-01-21; **Accepted:** 2021-07-01; **Published online:** 2021-07-07 13:48

URL: kns.cnki.net/kcms/detail/11.2625.V.20210707.1111.001.html

Foundation items: Defense Science and Technology Project Foundation of China (F062102009); Shandong Provincial Natural Science Foundation (ZR2017MF036); Young Innovation Team of Colleges and Universities in Shandong Province (2020KJN003)

* **Corresponding author.** E-mail: 13954559561@126.com

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

DOI: 10.13700/j.bh.1001-5965.2021.0022

一种多机协同打击的快速航迹规划方法

陈清阳*, 辛宏博, 王玉杰, 唐钟南, 贾高伟, 朱炳杰

(国防科技大学 空天科学学院, 长沙 410073)



摘 要: 针对多架自杀式无人机对重要目标进行协同打击的问题,在飞行器运动学、飞行器碰撞、时空协同等约束条件下,提出了一种空间分层分布的协同打击策略,在满足飞行器碰撞约束的前提下,有效提高了对重要目标防御系统的抵抗能力与飞行器自身的生存率。在此基础上,进一步提出了一种空间协同的多机打击快速航迹规划方法,结合 Dubins 曲线,将飞行器数目增加带来的计算量指数增长的问题,转化为多项式乘积形式的计算量,实时生成满足时空协同要求的次优航迹。通过仿真实验与实际飞行试验验证了所提方法的有效性,无人机可以在生成航迹的引导下,有效到达打击目标。

关键词: 自杀式无人机; 协同打击; 空间分层; 碰撞约束; 航迹规划; Dubins 曲线

中图分类号: V249.122-3

文献标志码: A **文章编号:** 1001-5965(2022)07-1145-09

随着通信、电子、控制等技术的飞速发展,无人机越来越广泛地出现在人们的日常生活中。从民用的无人机航拍、航测、农业植保运用、电力巡线、货物运输等,到军用的情报收集、区域侦察、目标监视等运用,无人机正以一种快速发展的态势,渗透到人们生活的方方面面。尤其是随着美国、俄罗斯、以色列等军事强国在军用无人机方面的大量投入,高空长航时无人机、察打一体无人机、穿透型隐身无人机也在飞速发展。其中,小型的自杀式巡飞弹或无人机,以其低成本、方便携带等特征,逐渐成为人们关注的焦点。

在自杀式无人机的广泛应用中,采用多机协同进行饱和和攻击,可以以较低的成本、较高的成功率实现对高价值目标的有效打击,降低不断增强的区域协同防空系统带来的影响,具有重要的应用前景^[1-4]。协同的过程,可以从空间和时间 2 个维度进行定义。时间上的协同,期望多个自杀式无人机能同时到达打击目标,或按照预定的时间顺序与间隔到达打击目标;空间上的协同,则期望

多个无人机以空间均匀分布的方式,到达打击目标,或按照期望的间距和队形到达打击目标。目前关于时间协同的研究较多^[5-8],关于空间协同的研究较少。同时考虑时间与空间的多维约束,会给问题求解的过程增加更多的约束条件,大大增加问题的复杂度与所需的计算时间。基于 Dubins 曲线的多机协同航迹规划方法,是一种相对简单且能满足一定实时性要求的常用方法。文献[9-10]提出了基于 Dubins 曲线的多机协同航迹规划方法,多架飞机期望能同时到达预定的多个目标,考虑了多机之间的冲突、路径长度的调整等因素,并在文献[11]中拓展到三维空间中。文献[12-13]为了提高航迹的可跟踪性,采用毕达哥拉斯曲线代替 Dubins 曲线,保证了曲率的连续性,同时基于所设计的毕达哥拉斯曲线,路径长度的调整可以通过曲率的实时搜索来实现,具有一定的可行性,但没有针对路径之间的冲突所带来的反复搜索问题进行详细讨论,难以保证在存在冲突的情况下,实时计算得到每架飞行的期望航

收稿日期: 2021-01-15; 录用日期: 2021-05-07; 网络出版时间: 2021-05-17 15:47

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210517.1350.001.html

基金项目: 国家自然科学基金(61801495)

* 通信作者. E-mail: chy1982_008@nudt.edu.cn

引用格式: 陈清阳, 辛宏博, 王玉杰, 等. 一种多机协同打击的快速航迹规划方法[J]. 北京航空航天大学学报, 2022, 48(7): 1145-1153. CHEN Q Y, XIN H B, WANG Y J, et al. A rapid path planning method for multiple UAVs to cooperative strike[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1145-1153 (in Chinese).

迹。文献[14]引入了 Bezier 曲线进行多机协同航迹规划,同时通过对形状参数的搜索,对多机之间的路径进行调整,保证多机之间的时间同步性。文献[2]将 Dubins 曲线运用于多弹协同打击的中段航迹生成过程,同时针对末段打击的制导方法进行研究,提出了一种多弹协同对目标进行时空一致饱和攻击的方法,并进行了仿真验证。文献[15]针对给定时间约束下的多无人机最优集结问题进行研究,提出了一种基于时域映射的分布式优化框架,对无人机的状态与指令进行决策与优化,最终保证多架无人机在给定时间内集结到给定位置,并通过仿真实验进行了验证。文献[16]提出了一种多智能体信息协同的一致性算法,不管在有或无外部基准输入的情况下,都能保证多个智能体之间的信息与行为的一致性,并通过仿真实验进行了验证。上述方法都对多架无人机或多个智能体之间的时间协同与一致性问题进行了研究,并针对曲率连续性、路径之间的冲突等问题进行了分析。但是,面向多机协同对高价值目标进行饱和攻击的需求,仍存在一定的局限性:

1) 在存在路径冲突的情况下,上述方法都是通过反复迭代搜索来实现的,迭代的过程难以保证在有限的时间内实时得到可行的解。尤其是在多机饱和攻击的需求下,飞行器数目的增加更会导致计算量与计算时间指数增长。

2) 对于多机饱和攻击的需求,为了有效降低敌方防空系统的影响,一般是期望多机以空间均匀的方式到达打击目标,因此,对空间上的协同也具有较高的要求。如何将上述方法拓展到时间与空间的多维约束下,也是需要解决的问题。

针对上述两方面的难点,本文提出一种空间协同约束下的多机打击快速航迹规划方法。与其他方法相比,具有以下创新点:

1) 为了有效降低敌方防空系统的影响(避免同一高度的无人机,容易被敌方的转动式防御武器顺序各个击破),同时解决多机航迹规划中的冲突判断与消解问题,提出一种空间分层分布的安全高效协同打击策略,以空间的分层分布,解决多机之间可能存在的航迹冲突与迭代搜索。

2) 为了解决飞行器数目增加带来的计算量指数增长的问题,提出一种空间协同约束下的快速航迹规划方法,通过降低对路径最短(时间最短)的要求,来顺序搜索满足空间均匀分布的各个无人机与过渡目标状态之间的对应关系,将指

数搜索的过程转化成多项式搜索的过程,有效提高问题求解效率。

3) 基于 Dubins 曲线的航迹表示方法,可以通过半径参数的优化来对多机的航迹进行同步,以某一飞机的参考航迹为基准,对其余飞机的半径进行求解,得到满足同步要求的多机航迹。

1 问题描述

多机协同打击任务定义如下:沿着初始期望航线飞行的多架自杀式无人机,在某一时刻接收到地面控制站的指令,向地面某一高价值目标进行打击。考虑敌方对高价值目标的防御,将多个无人机对目标打击的过程分成 2 段:

1) 在中制导段,为了提高多机协同打击的效果,多架无人机沿着地面控制站规划的期望航迹飞行,期望在相同的时间到达敌方防御圈。在向目标防御圈飞行的过程中,允许多机之间存在一定的通信,通过相互之间的信息共享进行速度/航线的调节,以提高多机之间的同步性。

2) 在末制导段,多架无人机启动末端打击制导模式,相互之间不再通信。机载控制系统以打击过程设置的预定打击时间为优化目标,对无人机的飞行速度/航迹进行调节,以实现多机同时到达与打击目标的效果。

目前针对末制导的研究较多,都围绕着地面控制站发现目标后,实时规划多架自杀式无人机向目标防御圈飞行的期望航迹的问题进行研究,示意图如图 1 所示。

针对图 1,对本文涉及的一些变量进行定义。无人机的初始状态为 $p_i = (x_i, y_i, \theta_i)$,下标 i 表示无人机的初始状态, (x_i, y_i) 为无人机的初始位置信息, θ_i 为无人机的初始航向角信息。对于 N 架无人机对目标打击的问题来说,第 j 个无人机的初

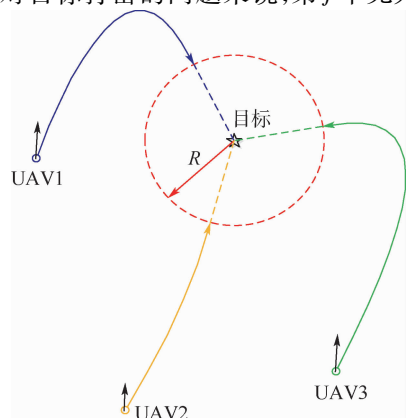


图 1 多机协同打击航迹规划问题示意图

Fig. 1 Demonstration of path planning problem for cooperative strike with multiple UAVs

始状态表示为 $p_{ij} = (x_{ij}, y_{ij}, \theta_{ij})$, $1 \leq j \leq N$, 下标 j 表示无人机的编号。

为了计算无人机向目标威胁圆飞行的期望航迹,定义无人机到达目标威胁圈的状态为过渡目标状态,无人机在目标威胁圆(假设重要目标的防御圈为一个圆柱形的区域,投影到平面上为一个半径为 R 的圆)上的可选过渡目标状态为 $p_f = (x_f, y_f, \theta_f)$ 。考虑到飞行器控制系统的感知精度、控制精度等因素,将 $[0, 2\pi)$ 范围的圆心角进行离散化,离散分辨率为 $\Delta\varphi$,则可选的过渡目标状态数目为 $M = \left\lceil \frac{2\pi}{\Delta\varphi} \right\rceil$, 可选过渡目标状态的集合为 $p_{fg} = (x_{fg}, y_{fg}, \theta_{fg})$, $1 \leq g \leq M$, 下标 f 表示无人机过渡目标状态,下标 g 表示过渡目标状态序号。

根据上述状态定义,中制导段的协同航迹规划问题,先要解决的是通过算法生成期望航迹 $f(j, g) = (f_x, f_y)$, 建立从 p_{ij} 到 p_{fg} 的映射关系:

$$p_{ij} \xrightarrow{f(j, g)} p_{fg} \quad 1 \leq j \leq N, 1 \leq g \leq M \quad (1)$$

同时满足以下要求:

- 1) p_{fg} 在威胁圆上均匀分布,即 $\|\theta_{fg_a} - \theta_{fg_b}\| = l \frac{2\pi}{N} (1 \leq g_a \leq M, 1 \leq g_b \leq M, g_a \neq g_b, 1 \leq l \leq N-1)$ 。
- 2) $\frac{\|f(1, g_1)\|}{v_1} = \frac{\|f(2, g_2)\|}{v_2} = \dots = \frac{\|f(N, g_N)\|}{v_N}$,

其中, $\|f(j, g_j)\|$ 为对应第 j 架无人机到达第 g_j 个过渡目标状态的期望路径的长度, v_j 为第 j 架无人机的飞行速度(地速)。由于本文考虑的自杀式无人机是属于同一构型的飞行器,忽略风速的影响(风速对飞行器之间同步性的影响,在飞行器沿着规划航迹飞行的过程中,通过一定范围内的速度调节来解决),在航迹规划阶段可以认为 $v_1 = v_2 = \dots = v_N$, 因此,该条件可以简化为 $\|f(1, g_1)\| = \|f(2, g_2)\| = \dots = \|f(N, g_N)\|$ 。

除了满足上述时间与空间的约束,在航迹规划的过程中,还必须考虑以下约束条件:

1) 飞行器的运动学约束。本文主要考虑无人机的最小转弯半径,即 $|k_{f(j, g)}| \leq k_{\max}$, 其中 $k_{f(j, g)}$ 为 $f(j, g)$ 路径上任意一点的曲率, k_{\max} 为无人机飞行过程中所允许的最大曲率。

2) 飞行器之间的碰撞约束。为了保证无人机自身的安全性,必须保证飞行器之间在任意时刻不会发生碰撞,即

$$\|f(j, g_r)_t - f(j_q, g_q)_t\| \geq D \quad (2)$$

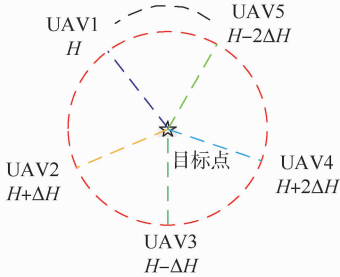
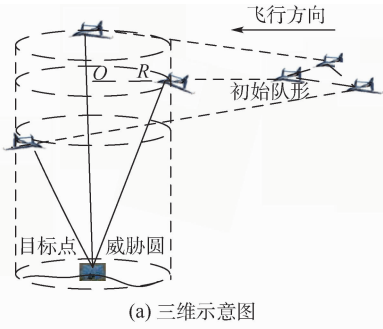
$$1 \leq r \leq N, 1 \leq q \leq N, r \neq q$$

式中: $f(j_r, g_r)_t$ 为第 r 架无人机沿着期望航迹飞行,在 t 时刻预期到达的位置; $f(j_q, g_q)_t$ 为第 q 架无人机沿着期望航迹飞行,在 t 时刻预期到达的位置; D 为无人机飞行过程中所允许的最小安全间距。

2 空间分层分布的协同打击策略

针对上述多约束条件下的协同航迹规划问题,本节采用一种空间分层分布的协同打击策略与顺序搜索的次优路径快速生成方法相结合的模式,对各个约束条件“分而治之”。首先,针对飞行器之间的碰撞约束,规划航迹之间冲突检测与消解是一个复杂的迭代过程,尤其是在考虑时间属性的情况下,导致计算的维度从单纯的空间位置判断,升级到含有时间属性的序列点的比较与不断优化,计算量大大增加。目前,相关的方法难以在有效时间内判断出可能存在的冲突,并计算出有效的应对策略。因此,本文采用图2所示的空间分层分布的协同打击策略。无人机在初始时按照一定的队形,沿着一定的航线和高度飞行,在发现打击目标后,地面控制站根据每架无人机的状态与后续空间约束下的多无人机协同打击快速航迹规划方法得到的结果,进行一种高度上的有效分布,得到类似图2所示的效果(定义无人机到达过渡目标状态的高度以 H 为中间位置,往上按照 ΔH 的间隔逐层递加,往下按照 ΔH 的间隔逐层递减)。以选定的 UAV1 为基准(沿着高度 H 的平面飞行),沿着逆时针方向,随着方位角的变化,按照 $H + \Delta H, H - \Delta H, H + 2\Delta H, H - 2\Delta H, \dots$ 的顺序,将对应不同方位角的无人机分配到不同的高度平面。为了保证无人机在向规划航迹飞行的过程中不会发生相互碰撞,不同的高度层所对应的无人机,即按照初始状态下 N 架无人机的高度顺序进行分配,从而有效降低计算量,同时保证无人机之间的安全性。

通过上述空间分层进行协同航迹规划的思想,一方面可以解决飞行器之间的碰撞约束,减轻后续航迹搜索的计算压力,保证算法的实时性;另一方面,针对敌方防空武器系统来说,为了防御上述饱和和攻击的自杀式无人机,转动式的防御系统需要实现一种方位和俯仰无规则运动的控制模式,尤其是俯仰方向的频繁加减速运动,对敌方控制系统提出了更高的要求,可以有效提高自杀式无人机的生存率。基于该空间分层的协同打击策略,在后续的快速航迹规划方法设计中,就不再考虑飞行器之间的碰撞约束。



(a) 三维示意图

(b) 俯视图 (无人机下方的符号代表无人机的飞行高度)

图2 空间分层分布的协同打击策略示意图

Fig.2 Demonstration of cooperative strike strategy based on hierarchical space distribution

3 空间协同的多机协同打击快速航迹规划方法

空间分层分布的协同打击策略,只是考虑了无人机航迹搜索的平面,但是距离完成空间和时间协同的打击要求,还存在较大的差距。本节提出一种顺序搜索的次优路径快速生成方法,在保证规划路径接近最优性的前提下,将飞行器数目增加带来的计算量指数增长的问题,转化为飞行器数目与可选过渡目标状态数目乘积的形式,有效提高搜索效率,保证方法的实时性。在所设计的方法中,考虑 Dubins 曲线所具有的路径最短、位置/航向连续、圆形曲线半径与路径长度的直接映射关系等特性,采用 Dubins 曲线作为基本的路径曲线。

3.1 Dubins 曲线

为了解决空间机器人在 2 个构型(早期主要考虑地面的移动机器人,主要考虑的状态信息包括位置和方位)之间运动的路径生成问题, Dubins^[17] 于 1957 年提出了由 2 个最小转弯半径的圆弧和 1 段直线连接所构成的最短路径,并命名为 Dubins 曲线^[18]。Dubins 曲线的 4 种基本形式如图 3 所示。

由图 3 可见, Dubins 曲线一般由 3 段基本曲线组成。其中,初始段和终止段由满足机器人运动学约束的最大曲率的圆弧组成,初始段圆弧同时与机器人的初始方位角相切,终止段圆弧同时

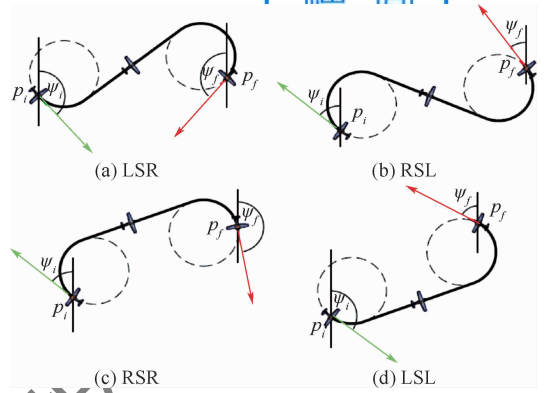


图3 Dubins 曲线的 4 种构型

Fig.3 Four different types of Dubins curves

与机器人的期望方位角相切。在初始段圆弧与终止段圆弧之间,由一段与 2 个圆弧相切的直线构成。根据该设计, Dubins 曲线一般主要包括 4 类,分别为 LSL、LSR、RSR、RSL (L 代表逆时针盘旋的圆弧, R 代表顺时针盘旋的圆弧, S 代表直线段)。该曲线能保证所生成的航迹满足机器人的位置和方位角连续的要求。同时, Dubins^[17] 也证明了,最小转弯半径的圆弧与中间的切线所构成的曲线,是满足位置/方位约束的最短路径。

在应用 Dubins 曲线解决移动机器人的路径规划问题时,一般采用的是先计算满足初始状态和目标状态(位置/方位)的 4 条 Dubins 曲线,对 4 条 Dubins 曲线进行比较,再选择其中最短的曲线,用于移动机器人的控制。4 种类型的 Dubins 曲线的长度计算分别如下所示。

本文中讨论的无人机的初始状态为 $p_i = (x_i, y_i, \theta_i)$, 过渡目标状态为 $p_f = (x_f, y_f, \theta_f)$, 对应 RSR 类型的 Dubins 曲线的初始圆心为(方位角定义顺时针为正, 范围为 $[-\pi, \pi)$, 无人机的最小转弯半径为 R_{min})

$$\begin{cases} x_{oi} = x_i + R_{min} \sin(\theta_i + \pi/2) \\ y_{oi} = y_i + R_{min} \cos(\theta_i + \pi/2) \end{cases} \quad (3)$$

式中: (x_{oi}, y_{oi}) 为针对初始状态的圆心的坐标。

针对目标状态的圆心为

$$\begin{cases} x_{of} = x_f + R_{min} \sin(\theta_f + \pi/2) \\ y_{of} = y_f + R_{min} \cos(\theta_f + \pi/2) \end{cases} \quad (4)$$

式中: (x_{of}, y_{of}) 为针对目标状态的圆心的坐标。

从而 RSR 类型的 Dubins 曲线的路径长度 d 为

$$d = \sqrt{(x_{of} - x_{oi})(x_{of} - x_{oi}) + (y_{of} - y_{oi})(y_{of} - y_{oi})} + R_{min}(\theta_f - \theta_i) \quad (5)$$

其他 3 种类型 Dubins 曲线的路径长度可以类似推导得到。可以发现,路径的长度与圆的半径具有直接的映射关系,既可以根据半径计算相

应路径的长度,又可以在多机协同航迹规划的过程中,根据期望的路径长度,搜索相应的圆形半径。因此,采用 Dubins 曲线作为无人机的基本路径曲线,用于实时的多机协同航迹规划。

3.2 多机协同打击的快速航迹规划方法

以上述 Dubins 曲线为基础,提出一种空间协同的多机协同打击快速航迹规划方法,基本思路如下:

1) 对于 N 个初始状态, $\mathbf{p}_{ij} = (x_{ij}, y_{ij}, \theta_{ij})$, $1 \leq j \leq N, M$ 个可选过渡目标状态 $\mathbf{p}_{fg} = (x_{fg}, y_{fg}, \theta_{fg})$, $1 \leq g \leq M$ 。根据上述 Dubins 曲线计算出连接所有初始状态与目标状态的最短路径,得到一个 $N \times M$ 的数字矩阵 w_{jg} ($1 \leq j \leq N, 1 \leq g \leq M$) 以及相应的 $N \times M$ 条 Dubins 曲线 $f_{ij \rightarrow fg}$ 。为了后续搜索方便,不失一般性,在计算与存储 $N \times M$ 数字矩阵的过程中,假设每一列过渡目标状态的方位角是一致的,第 g 列的过渡目标状态的方位角为 $(g-1)\Delta\varphi$ 。

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1M} \\ w_{21} & w_{22} & & w_{2M} \\ \vdots & \vdots & & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NM} \end{bmatrix} \quad (6)$$

2) 对数字矩阵进行搜索,得到 $N \times M$ 个数据中的最小值 w_{nm} ,从而确定对应第 n 个无人机的过渡目标状态为 \mathbf{p}_{jm} (\mathbf{p}_{jm} 的方位角为 $(m-1)\Delta\varphi$),得到数据结构 $[\mathbf{p}_{in}, \mathbf{p}_{jm}]$,并计算由 \mathbf{p}_{in} 和 \mathbf{p}_{jm} 所确定的最短可行路径 $f_{in \rightarrow jm}$,重新标记为 $[\mathbf{p}_{i1}, \mathbf{p}_{f1}, f_{i1 \rightarrow f1}]$ 。

3) 将数字矩阵中的第 n 行和第 m 列删除,得到一个 $(N-1) \times (M-1)$ 的矩阵。

4) 以 $(m-1)\Delta\varphi$ 为起始状态,相隔 $\Delta\theta = 2\pi/N$ 确定一个过渡目标状态的方位角 $((m-1)\Delta\varphi + \Delta\theta)$,在 $(N-1) \times (M-1)$ 的矩阵中对应 $\left[\frac{((m-1)\Delta\varphi + \Delta\theta)}{\Delta\varphi} \right]$ 所在的列 ($[\cdot]$ 是对数值取整的操作),寻找列中最小的数值,根据最小数值所在的行数确定对应 $((m-1)\Delta\varphi + \Delta\theta)$ 方位角的无人机,得到 $[\mathbf{p}_{i2}, \mathbf{p}_{f2}, f_{i2 \rightarrow f2}]$ 。

5) 重复步骤 3) 和步骤 4),直到找到 N 个过渡目标状态所对应的 N 架无人机,得到新的结构数组 $[\mathbf{p}_{ij}, \mathbf{p}_{fj}, f_{ij \rightarrow fj}]$ ($1 \leq j \leq N$)。

6) 对 N 条 Dubins 曲线 $f_{ij \rightarrow fj}$ ($1 \leq j \leq N$) 的长度进行排序,确定其中最长的路径,定义为参考航线。参考航线的长度 L 为

$$L = \max_{1 \leq j \leq N} |f_{ij \rightarrow fj}| \quad (7)$$

7) 以参考航线的长度 L 为基准,根据 3.1 节讨论的 Dubins 曲线的长度与圆弧半径的关系,对半径参数进行计算,使得所有的路径长度都与参考航线的长度一致。由于 Dubins 曲线的长度与半径的关系式存在三角函数等非线性函数,在计算的过程中可以采用二分法等数值方法进行搜索。

需要说明的是,本文所提出的空间分层分布的协同打击策略中,无人机在接收到规划航迹,向目标高度飞行的过程,存在一个由于高度变化,而导致式(5)的期望航线长度的计算存在一定偏差的问题。为了提高飞行器的安全性,降低多架无人机往目标飞行的过程中,相互碰撞的风险,目前采用的策略是在生成 Dubins 曲线之后,各个飞行器以最大的爬升角或者下滑角,快速地到达期望的高度层,再沿着生成的 Dubins 曲线向目标威胁圆飞行。因此,三维空间中的航线长度计算可以在式(5)的基础上,采用式(8)进行更准确计算,从而进一步提高多机协同打击的同步性。

$$d = \sqrt{(x_{of} - x_{oi})(x_{of} - x_{oi}) + (y_{of} - y_{oi})(y_{of} - y_{oi})} + R_{\min}(\theta_f - \theta_i) + \Delta h \left(\frac{1}{\sin \gamma_{\max}} - \frac{1}{\tan \gamma_{\max}} \right) \quad (8)$$

式中: Δh 为无人机向期望高度飞行需要变化的高度差; γ_{\max} 为无人机的最大爬升角或者下滑角。从而可以根据式(8)进行半径参数的搜索与优化。

根据以上方法,一方面确定了 N 架无人机与对应过渡目标状态之间的映射关系,实现了 N 架无人机在威胁圆上的均匀分布;另一方面也生成了用于引导无人机飞行的期望航线。由于是在 Dubins 曲线的基础上进行轨迹延伸,以保证多架无人机之间的时间同步,生成的航线能保证无人机在较短的时间内到达威胁圆,是一种次优的航线。

4 仿真与飞行试验结果

为了验证本文方法,开展了一系列的仿真与实际飞行试验,下面对仿真与试验的情况进行详细介绍。

4.1 仿真结果

实验场景设置如下:

1) 8 架无人机起飞后,沿着预设的航线飞行,仿真飞机的飞行速度为 19 m/s。

2) 在某一时刻,地面控制站软件设置了任意的打击目标点与预设的防御圈(防御圈半径设置为 300 m)。

3) 地面控制站软件根据打击目标点进行航

迹规划,并根据生成的航线引导无人机飞行。仿真结果如图4所示。

从仿真结果和表1的规划结果可以看到,在8架无人机飞行并进行协同打击的情况下,所设计的航迹规划方法能实时给出期望的航线(所使用的电脑为 Intel i9-9900 八核处理器,16 GB 内存,所需的计算时间为 5 ms)。同时,8架飞机沿着生

成的航线飞行,根据表1的规划结果与图4(d)、(e)所示的剩余路径的长度变化,在第8 295与8 296帧的瞬间(飞行时间大约为 145 s),多架飞机几乎同时到达威胁圆,实现了时间上的协同(路径长度的保存是以离散的点来表示的,因此剩余路径的曲线都是一个一个的台阶,也导致了最终到达时存在一个数据离散化导致的偏差)。

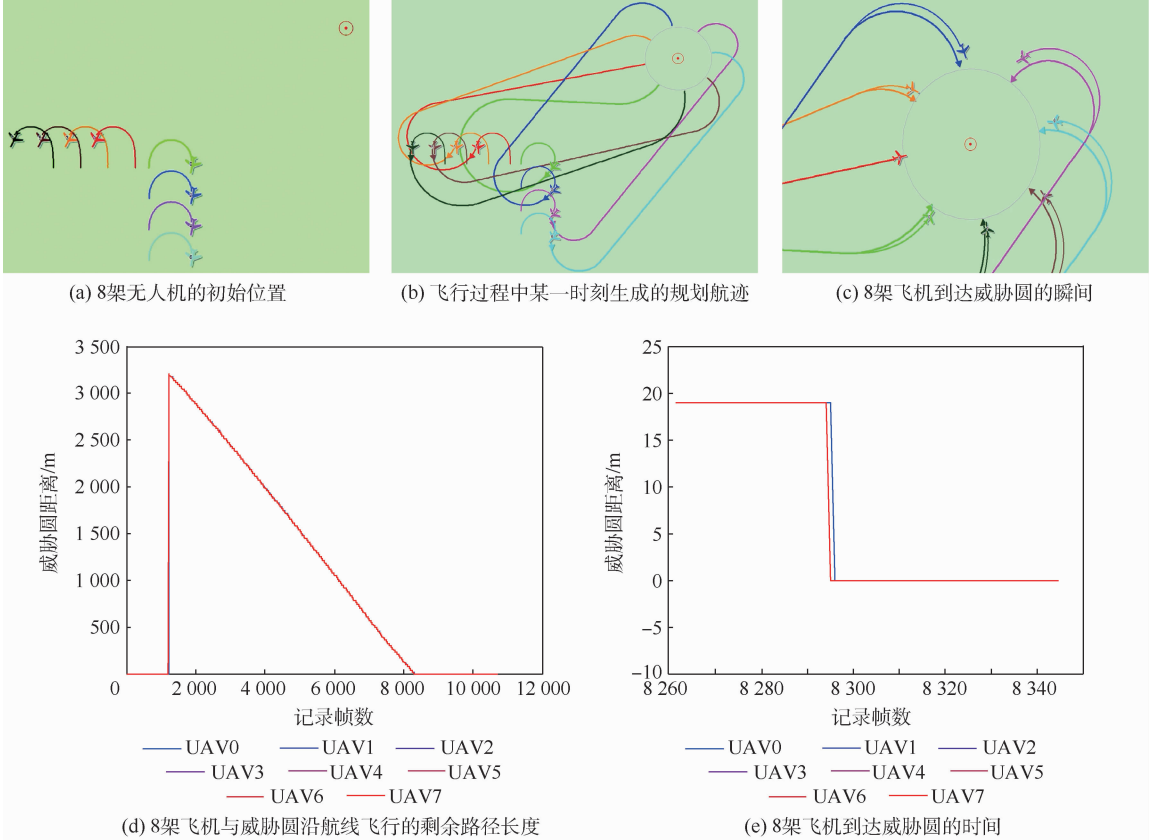


图4 模拟打击结果

Fig. 4 Simulation results of cooperative strike

表1 仿真实验的规划结果

Table 1 Planned results during simulation

无人机序号	规划航线长度/m
UAV0	3 232.8
UAV1	3 205.4
UAV2	3 206.1
UAV3	3 197.7
UAV4	3 218.5
UAV5	3 204.6
UAV6	3 218.1
UAV7	3 218.2

4.2 飞行试验

为了进一步对本文方法进行验证,在仿真实验的基础上,笔者所在团队开展了相关的飞行试验。采用如图5所示的3架泡沫模型机,对地面设定的目标进行模拟打击,该模型机采用电池作为唯一能源,通过电调驱动电机,带动螺旋桨旋转

进行飞行,飞行的速度约为18 m/s。同时,平台具有较好的俯冲特性(大约能达到 20°的俯冲角),适合用于开展模拟打击飞行试验。

无人机飞行控制系统采用笔者所在团队自研的 RX-0701 自驾仪(见图6),自驾仪上集成了三轴陀螺仪、三轴加速度计、三轴磁力计、空速、气压高度、GNSS 导航模块等传感器,通过机载处理器进行传感器融合与导航解算,进一步通过打击制导与控制算法^[19],输出相应的脉冲宽度调制信号驱动机载舵机,控制飞机飞行。同时,为了实时监控飞行器的状态,在必要时给无人机发送一定的控制指令,在控制系统与地面控制站之间配置了 900 MHz的数传链路,基本的系统框架如图7所示。

3架无人机进行协同打击的试验过程如下:

1) 3架无人机手抛起飞,在天上沿着预先设定的四边形航线飞行,正常飞行的过程采用文



图 5 飞行试验平台

Fig. 5 Experimental flight platforms



图 6 飞行控制系统

Fig. 6 Flight control system

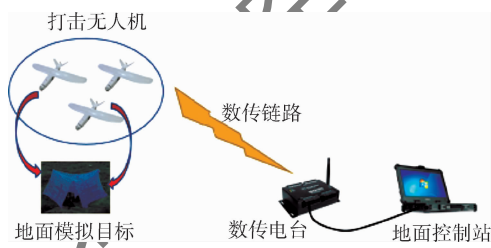


图 7 模拟打击过程系统框架

Fig. 7 Architecture of flight striking system

献[20]中的航迹跟踪算法。

2) 在无人机沿着四边形航线飞行的过程中,地面控制站根据预先布置的地面装备(如图8中的伪装网所示)设定打击的目标点,并将打击目标点的位置信息,通过数传链路发送给3架无人机(自杀式无人机采用GNSS导引的方式,代替可见光导引的制导方式)。

3) 在某一时刻,地面控制站向3架无人机发送打击指令,同时将生成的向过渡目标状态飞行的航线发送给无人机上的控制系统。

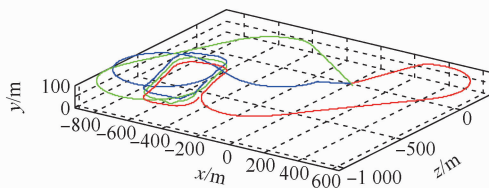
4) 3架无人机沿着规划航迹飞行。在飞行的过程中,3架无人机按照预先设定的飞行时间与距离过渡目标点的距离,进行速度的调整^[21],保证3架无人机尽量同步到达过渡目标状态。

5) 在到达过渡目标状态后,切换到末端打击模式,采用文献[16]中的打击制导算法进行飞行,并最终实现对模拟目标的打击,试验结果如图8所示。

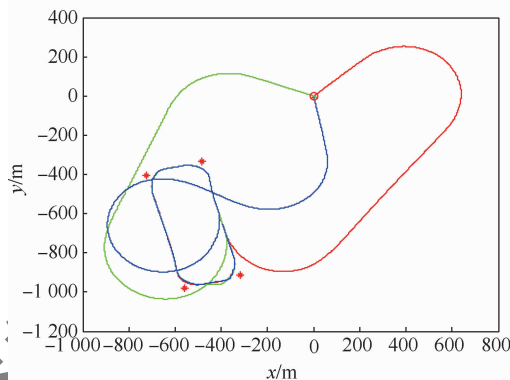
从图8和表2可以看到,本文方法在无人机飞行过程中能实时生成引导无人机到威胁圆的期望航线。无人机在跟踪航线并进行模拟打击的过程中,能实现基本的时空同步,3架无人机接近同一时刻到达模拟打击目标(最终到达模拟目标的最大时间差在1.3s),充分验证了本文方法的有效性。



(a) 打击前一时刻各机位置



(b) 三机飞行航迹



(c) 三机飞行航迹在水平面的投影

图 8 三机协同打击试验结果

Fig. 8 Experimental results of cooperative strike with three UAVs

表 2 飞行试验的规划结果

Table 2 Planned results during flight experiment

无人机序号	规划航线长度/m
UAV0	1 876.4
UAV1	1 913.1
UAV2	1 865.8

5 结 论

针对多架自杀式无人机对重要目标进行打击的问题,本文提出了一种空间分层分布的协同打击策略,并进一步提出了一种多机协同打击的快速航迹规划方法。

1) 空间分层分布的协同打击策略,能有效解决多飞行器航迹规划的碰撞问题,实现对问题的

降维,提高方法的效率与实时性。

2) 空间分层分布的协同打击策略,能有效提高敌方防御系统应对多架自杀式无人机攻击所带来的防御难度,增强无人机自身的生存率。

3) 所提出的快速航迹规划方法,将飞行器数目增加所带来的计算量指数增长的问题,转化为飞行器数目与可选过渡目标状态数目乘积的形式,有效降低计算复杂度,能解决更多飞行器进行饱和攻击所带来的实时性问题。

如何对方法进一步优化,将生成的次优航迹进一步提升为路径最短的最优航迹,是下一步需要研究的内容。同时,如何克服外界风场等扰动的影响,进一步提高无人机之间的时空同步性,也是值得研究的内容。

参考文献 (References)

- [1] KIM H G, KIM H J. Backstepping-based impact time control guidance law for missiles with reduced seeker field-of-view[J]. IEEE Transactions on Aerospace and Electronic Systems, 2019, 55(1): 82-94.
- [2] ZENG J, DOU L H, XIN B. A joint mid-course and terminal course cooperative guidance law for multi-missile salvo attack [J]. Chinese Journal of Aeronautics, 2018, 31(6): 1311-1326.
- [3] ZHEN Z Y, XING D J, CAO C. Cooperative search-attack mission planning for multi-UAV based on intelligent self-organized algorithm[J]. Aerospace Science and Technology, 2018, 76: 402-411.
- [4] DUAN H B, ZHAO J X, DENG Y M, et al. Dynamic discrete pigeon-inspired optimization for multi-UAV cooperative search-attack mission planning[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(1): 706-720.
- [5] LIU X, LIU L, WANG Y J. Minimum time state consensus for cooperative attack of multi-missile systems[J]. Aerospace Science and Technology, 2017, 69: 87-96.
- [6] HE S M, KIM M G, SONG T, et al. Three-dimensional salvo attack guidance considering communication delay[J]. Aerospace Science and Technology, 2018, 73: 1-9.
- [7] KANG S, WANG J N, LI G, et al. Optimal cooperative guidance law for salvo attack: An MPC-based consensus perspective[J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 54(5): 2397-2410.
- [8] 彭志红, 孙琳, 陈杰. 基于改进差分进化算法的无人机在线低空突防航迹规划[J]. 北京科技大学学报, 2012, 34(1): 96-101.
- PENG Z H, SUN L, CHEN J. Online path planning for UAV low-altitude penetration based on an improved differential evolution algorithm[J]. Journal of University of Science and Technology Beijing, 2012, 34(1): 96-101 (in Chinese).
- [9] SHANMUGAVEL M, TSOURDOSY A, ZBIKOWSKIZ R, et al. Path planning of multiple UAVs using Dubins sets[C]//AIAA Guidance, Navigation, and Control Conference and Exhibit. Reston: AIAA, 2005: 1-17.
- [10] SHANMUGAVEL M, TSOURDOSY A, WHITE B, et al. Cooperative path planning of multiple UAVs using Dubins paths with clothoid arcs[J]. Control Engineering Practice, 2010, 18(9): 1084-1092.
- [11] SHANMUGAVEL M, TSOURDOSY A, ZBIKOWSKIZ R, et al. 3D Dubins sets based coordinated path planning for swarm of UAVs[C]//AIAA Guidance, Navigation, and Control Conference and Exhibit. Reston: AIAA, 2006: 1-20.
- [12] SHANMUGAVEL M, TSOURDOSY A, ZBIKOWSKIZ R, et al. A solution to simultaneous arrival of multiple UAVs using Pythagorean Hodograph curves [C] // Proceedings of the 2006 American Control Conference. Piscataway: IEEE Press, 2006: 2813-2818.
- [13] SHANMUGAVEL M, TSOURDOSY A, ZBIKOWSKIZ R, et al. 3D path planning for multiple UAVs using Pythagorean Hodograph curves [C] // AIAA Guidance, Navigation, and Control Conference and Exhibit. Reston: AIAA, 2007: 1-14.
- [14] ASKARI A, MORTAZAVI M, TALEBI H, et al. A new approach in UAV path planning using Bezier-Dubins continuous curvature path[J]. Part G: Journal of Aerospace Engineering, 2016, 230(6): 1103-1113.
- [15] 丁超, 魏瑞轩, 周凯. 基于时域映射的多无人机系统给定时间分布式最优集结[J]. 北京航空航天大学学报, 2021, 47(2): 315-322.
- DING C, WEI R X, ZHOU K. Distributed optimal rendezvous of multi-UAV systems in prescribed time based on time-domain transformation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(2): 315-322 (in Chinese).
- [16] 陈晔, 张鑫, 金鑫, 等. 一种多智能体协同信息一致性算法[J]. 航空学报, 2017, 38(12): 3212-22.
- CHEN W, ZHANG X, JIN X, et al. A cooperative information consensus algorithm for multi-agent system[J]. Acta Aeronauticae Astronautica Sinica, 2017, 38(12): 3212-22 (in Chinese).
- [17] DUBINS L E. On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents[J]. American Journal of Mathematics, 1957, 79(3): 497-516.
- [18] CHEN Q Y, LU Y F, JIA G W, et al. Path planning for UAVs formation reconfiguration based on Dubins trajectory[J]. Journal of Central South University, 2018, 25(11): 2664-2676.
- [19] XIN H B, CHEN Q Y, WANG Y J, et al. Terminal guidance simulation and flight test for small UCAV[C]//International Conference on Control, Automation and Diagnosis. Piscataway: IEEE Press, 2019: 1-6.
- [20] 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).
- [21] CHEN Q Y, LI Y. UAVs formation flight control based on following of the guidance points[C]//Proceedings of 2016 IEEE Chinese Guidance, Navigation and Control Conference. Piscataway: IEEE Press, 2016: 730-735.

A rapid path planning method for multiple UAVs to cooperative strike

CHEN Qingyang*, XIN Hongbo, WANG Yujie, TANG Zhongnan, JIA Gaowei, ZHU Bingjie

(College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Focusing on the cooperative strike problem to important with multiple suicide UAVs, a cooperative striking strategy based on hierarchical space distribution is proposed. The strategy is proposed with the vehicle kinematic constraint, the collision constraint of UAVs, and the space-time cooperative constraint. With the proposed strategy, the collision constraint can be solved. What's more, with the strategy, the resistibility of the UAVs to the recovery system of the target can be improved, and the chance of survival can be increased. A rapid path planning method for multiple UAVs with space cooperative requirements is proposed. The Dubins curve is combined in the method, and the exponential increment of computation with the number of UAVs is transformed into a polynomial form. The real-time requirement can be satisfied with the method, and sub-optimal trajectories can be generated. Simulation and flight experiments are carried out, and the results show that the UAVs can be guided to the target with the generated paths effectively, and the effectiveness of the proposed method is verified.

Keywords: suicide UAV; cooperative strike; hierarchical space; collision constraint; path planning; Dubins curve

北京航空航天大学学报

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

DOI: 10.13700/j.bh.1001-5965.2021.0289

基于多分辨率特征自选择的遮挡物识别算法

谢祥颖, 来广志, 那峙雄, 骆欣*, 王栋

(国网数字科技控股有限公司, 北京 100053)



摘 要: 光伏组件的遮挡物识别是光伏运维系统中不可或缺的一环,传统识别算法多依赖人工巡检,成本高昂且效率低下。基于卷积神经网络,提出了一种面向光伏组件的遮挡物识别算法 PORNet。通过引入特征金字塔,构建多个分辨率下具有丰富语义信息的图像特征,提升对遮挡物尺度和密度的敏感性。通过特征自选择,筛选出语义最具代表性的特征图,以加强物体环境的语义信息表达。用筛选出的特征图完成遮挡物识别,从而提升识别准确率。在自建光伏组件落叶遮挡数据集上进行了实验比较和分析,并对识别性能进行了评估,通过与现有物体识别算法相比,所提算法的准确率和召回率分别提升了 9.21% 和 15.79%。

关键词: 光伏组件; 遮挡物识别; 卷积神经网络; 特征金字塔; 特征自选择

中图分类号: TP391

文献标志码: A **文章编号:** 1001-5965(2022)07-1154-10

随着中国社会及经济的发展,人们在电力能源方面的需求与日俱增,在这种形势下,传统火力发电技术在满足大部分用电需求的同时对环境造成了污染。为了降低碳排放,光伏发电这一新型清洁能源的装机量与日俱增。在光伏电站中,光伏组件是最重要的部件,负责将太阳能转化为电能。然而,由于光伏电站选址的多样性,其可能建造在山区、屋顶等位置。在外部自然因素的影响下,光伏组件上会出现落叶、鸟粪和灰尘等遮挡物,引发热斑效应,降低光电转换效率甚至影响整组光伏阵列。因此,光伏组件的遮挡物识别是光伏电站运维中亟待解决的问题之一。

传统的光伏组件遮挡物识别依赖人工巡检的方式,即通过工作人员在光伏组件处实地观察来判断是否存在遮挡物。这种方式需要耗费大量的人力及时间成本,检测效率低下;此外,对于特殊场景(如屋顶、墙壁等),人工巡检实现难度较大。

近年来,随着计算机视觉领域和人工智能技术的发展,越来越多的研究者开始关注物体识别问题,并将其应用于现实问题中,如图像分类^[1-3]、人脸识别^[4-5]、行人重识别^[6-7]等。现有工作大多利用卷积神经网络进行图像语义信息的提取,并基于该信息完成物体识别。众多实际问题的成功应用体现了卷积神经网络针对不同场景的强大的特征表示学习能力。因此,本文旨在利用物体识别技术,实现对光伏组件遮挡物的精确识别。

然而,基于现有技术完成光伏组件遮挡物识别任务面临着 3 项挑战:①受自然环境影响,光伏组件上的遮挡物分布稀疏程度不同,遮挡物既可能呈聚集趋势,也可能散乱分布。例如,在无风的天气下,光伏组件上可能存在具有较多落叶遮挡的区域,分布密集;而在有风的天气下,光伏组件则可能会存在单点分布的落叶遮挡情况。因此,

收稿日期: 2021-06-02; 录用日期: 2021-07-04; 网络出版时间: 2021-07-23 17:05

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210723.1245.001.html

基金项目: 国家重点研发计划(2018YFB1500800); 国家电网有限公司科技项目(SGTJDK00DYJS2000148)

*通信作者. E-mail: lx@ustc.edu.cn

引用格式: 谢祥颖, 来广志, 那峙雄, 等. 基于多分辨率特征自选择的遮挡物识别算法[J]. 北京航空航天大学学报, 2022, 48(7): 1154-1163. XIE X Y, LAI G Z, NA Z X, et al. Occlusion recognition algorithm based on multi-resolution feature auto-selection [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1154-1163 (in Chinese).

如何准确识别各种分布下的遮挡物是一个难点。

②同一遮挡物存在的位置与识别结果强相关。例如,当落叶存在于光伏组件旁的地上时,其显然并未对光伏组件造成遮挡。因此,如何避免光伏组件周围环境中存在物体的干扰也是一个挑战。

③在光伏组件图像采集的过程中,受拍摄距离和角度等因素的影响,同一种遮挡物可能以不同形态和大小出现在图像中。因此,如何自适应地对不同尺度的遮挡物进行识别同样是一个需要解决的问题。

为了解决上述问题,本文提出了一种面向光伏组件的遮挡物识别算法 PORNet。该算法将图像多个分辨率的特征引入到整体光伏组件的遮挡物识别任务中,并自动选择语义信息最具代表性的特征完成识别,从而全面增强对光伏组件遮挡物的密度变化、尺度变化及背景变化的敏感性,提高识别的准确性。本文提出的 PORNet 分为图像特征提取、多分辨率特征提取和特征自选择 3 个模块。对于图像特征提取模块,采用浅层残差网络,在网络的不同深度处获得各分辨率的特征图。对于多分辨率特征提取模块,引入特征金字塔,利用自深向浅的特征融合机制,构建具有丰富语义信息的多尺度特征。对于特征自选择模块,根据最大激活,自动选择语义最具代表性的特征,并舍弃其他特征,最终直接使用该特征完成遮挡物识别。

1) 本文提出了一种面向光伏组件的遮挡物识别算法 PORNet,该算法基于多尺度特征自选择,实现了在不同密度、不同尺度及不同背景下的遮挡物识别。

2) 本文提出的 PORNet 将特征金字塔引入识别问题,利用原始特征生成对光伏组件图像不同分辨率的特征表示,并根据最大激活,完成特征的自动选择,让 PORNet 能使用语义最具代表性的特征完成识别,提高识别效果。

3) 在自建光伏组件落叶遮挡数据集上,针对遮挡物识别问题开展实验,结果证明了本文算法相对于已有基准算法的优越性。

1 相关工作

物体识别的研究主要集中于 2 方面,即网络结构和数据增强。对于网络结构,残差网络 ResNet^[8]将学习目标从目标输出改变为目标输出与

输入的残差,成功缓解了深层网络中梯度消失问题。随后的许多工作^[9-12]基于 ResNet,开始探究如何更好地利用各种尺度的特征,并在网络中进行融合,如 DenseNet^[10]、ResNeXt^[11]等。除了对整体网络结构的改进,一些工作从算子层面进行了创新^[13-15]。空洞卷积提出对卷积核进行填充^[13];可变形卷积提出对卷积的位置进行合理变换^[14];Involution 提出一种利用自注意力机制生成卷积核再进行卷积计算的新算子^[15]。

数据增强同样是辅助完成识别任务的一类方法。Mixup^[16]提出使用按比例混合的成对样本及其标签进行训练以提高网络的鲁棒性。Cutout^[17]提出在训练期间,随机擦除图像中的部分像素,促使网络考虑环境信息。CutMix^[18]提出将一个样本的随机区域裁剪贴至另一个样本的方法以提升样本丰富性。除了这种固定策略的数据增强方式外,还有许多工作利用强化学习的思想,自动搜索最合适的数据增强策略^[19-22]。

传统物体识别算法大多只关注物体类别特征,而忽略物体本身的密度、尺度及背景信息,因此直接应用在光伏组件的遮挡物识别上难以获得理想效果。数据增强方法对于物体背景强相关的识别问题同样不适用。因此,本文提出了特征自选择光伏组件遮挡物识别网络,旨在选择语义最具代表性的特征完成识别,从而解决光伏组件遮挡物识别中的物体状态变化问题。

2 多分辨率特征自选择算法

在光伏组件的遮挡物识别问题中,受自然因素、图像采集等因素影响,遮挡物的密度和尺度都会有很大变化。此外,仅在物体本身能够对光伏组件构成遮挡时才需要被识别,因此识别不仅与物体类别有关,还与物体背景信息有关。传统识别算法中常用的单一分辨率特征对小尺度和低密度物体不够敏感,且易淡化物体背景信息,因此并不适用于此场景。在本文提出的多分辨率特征自选择算法中,将特征金字塔引入到遮挡物识别过程中,以加强对各个尺度、各个密度下的遮挡物的语义表达;并通过对多分辨率特征的自动选择,全面提升对光伏组件遮挡物背景信息的敏感性。算法整体流程分为图像特征提取、多分辨率特征提取及特征自选择 3 个模块,如图 1 所示。

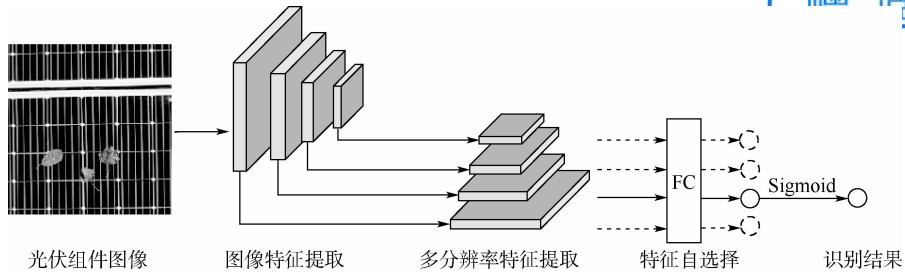


图1 本文算法整体流程示意图

Fig. 1 Overall process of the proposed algorithm

在图像特征提取模块中,使用在物体识别中常用的残差网络对全局特征进行提取。大部分物体识别算法只保留采样倍数最大的特征图。然而,在光伏组件的遮挡物识别中,为了关注多个尺度的遮挡物信息,会在网络的不同阶段保留不同下采样倍数的特征图,以用于后续识别过程。

在从浅到深获得了不同分辨率的特征图后,此时浅层的特征分辨率较高,具有较多的纹理信息;深层的特征分辨率较低,具有丰富的语义信息。传统物体识别算法中大多只关注物体类别信息,因此直接使用语义信息较丰富的最深层特征进行分类。然而,光伏组件图像中,无论是尺度较小的遮挡物识别还是遮挡物背景信息的获取,都需要一定的纹理信息,因此需要考虑将纹理信息与语义信息结合起来,生成更有效的特征。在多分辨率特征提取模块中,引入特征金字塔机制,采用从深到浅的方式对特征进行逐层融合,将语义信息逐步传递至各尺度的特征图,从而完成与纹理信息的结合。

在获得多分辨率特征图后,需要对其进行合理融合以用于识别。由于在一般的识别任务中,待识别物体的类别信息更为重要,而此时各分辨率的特征对于物体识别都有贡献,且贡献难以衡量,可以直接为每个分辨率的特征设置相同的权重,并按权重进行融合。然而,在光伏组件的物体识别任务中,单纯的特征融合有2个问题:①可能将错误的背景信息引入遮挡物;②可能存在密度较低、尺度较小的遮挡物,多个分辨率特征直接融合会淡化其语义,减弱对遮挡物的识别能力。因此要根据语义信息,选择对特征表达最好的特征用于识别。本文的特征自选择模块是以最大激活为依据来选择语义最具代表性的特征,将其用于最终的识别,同时舍弃其他特征。

值得注意的是,本文提出的基于多分辨率特征自选择算法具有可扩展性,可以扩展至多个类别,或迁移至其他场景的识别任务中。

3 光伏组件遮挡物识别网络

3.1 问题定义

本文中的相关符号及含义如表1所示。假设有 N 张标注好的光伏组件图像 $\{I_1, I_2, \dots, I_N\}$, 第 i 张图像的标签为 $y_i \in \{0, 1\}$, 其中, $y_i = 1$ 表示第 i 张图像存在遮挡物, $y_i = 0$ 表示第 i 张图像不存在遮挡物。本文使用标注图像训练模型 F , 训练结束后,可以用模型对任意一张光伏组件图像 I' 进行识别,判断是否存在遮挡物。模型输出 $s' = (I' | \theta_F)$ 表示图像 I' 中存在遮挡物的概率, θ_F 表示模型 F 的参数。

表1 符号表示

Table 1 Summary of main notations' representation

符号	含义
I_i	第 i 张图片类别
ReLU	ReLU 函数
Sigmoid	Sigmoid 函数
BN	批归一化层
FC	全连接层
GAP	全局平均池化
L_{cls}	分类损失函数
Conv	卷积层

3.2 图像特征提取模块

在图像特征提取模块中,基于残差网络对整体光伏组件图像进行多个尺度的全局特征提取。残差网络的基本单元是残差单元,如图2所示。对于输入信号 X , 相比于直接学习到输出信号 Y 的函数 S , 残差单元学习输出信号与输入信号的残差 $Y - X$, 即

$$Y = X + S(X) \quad (1)$$

此时梯度能直接从输出传递至输入,避免了网络层数过深导致梯度消失的问题。

本模块中的整个特征提取网络由4个大残差块构成,每个大残差块又由2个残差单元组成。在残差单元的具体实现中,采用了如图3所示的结构,对于输入信号 X 、输出信号 Y 可以表示为

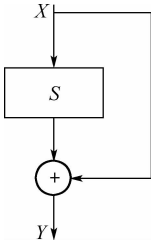


图 2 残差单元结构

Fig. 2 Structure of residual unit

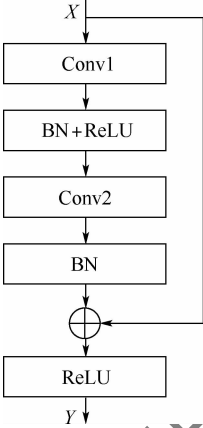


图 3 实际实现中残差单元结构

Fig. 3 Structure of residual unit in actual implementations

$$Y = \text{ReLU}(\text{BN}(\text{Conv2}(\text{ReLU}(\text{BN}(\text{Conv1}(X))))) + X \quad (2)$$

特征提取网络结构如图 4 所示。首先,图像输入到一个卷积核大小为 7×7 、步长为 2、卷积填充大小为 3 的卷积层中进行降采样;然后,经过步长为 2 的最大池化层进一步降采样;最后,输入到 4 个残差块中。这 4 个残差块中除第一个残差块外,每个残差块的第一个残差单元的卷积层的步长都为 2,以将特征图的尺度减小为输入的一半。设第 i 个残差块的输出为 C_i ,输入图像的高为 H 、宽为 W ,则在通过图像特征提取模块后可以得到 C_1 、 C_2 、 C_3 和 C_4 共 4 个不同尺度的特征,且有 $C_i \in \mathbf{R}^{\lfloor \frac{H_i}{2^{i+1}} \rfloor \times \lfloor \frac{W_i}{2^{i+1}} \rfloor \times C_i}$, $i = 1, 2, 3, 4$, $C_i' = 2^{i+5}$ 表示第 i 个特征图的通道数。

3.3 多分辨率特征提取模块

多分辨率特征提取模块结构如图 5 所示。在本模块中,从 C_4 开始,将深层特征的语义信息逐步向浅层融合。由于不同分辨率的特征间的通道数不同,需要先将每个模块用 1×1 卷积统一至相同通道数。设第 i 个特征图 C_i 所用的 1×1 卷积为 T_i ,则统一通道数后的特征图 H_i 可以表示为 $H_i = T_i(C_i)$ (3)

此处 1×1 卷积起到了梯度缓冲的作用,防止后续过程中自浅向深的特征融合操作影响主干网络。

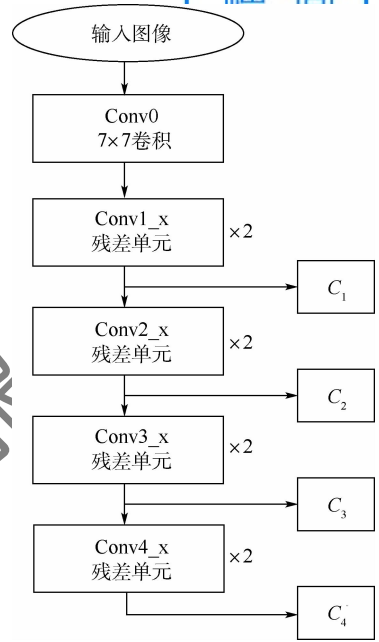


图 4 特征提取网络结构

Fig. 4 Structure of feature extraction network

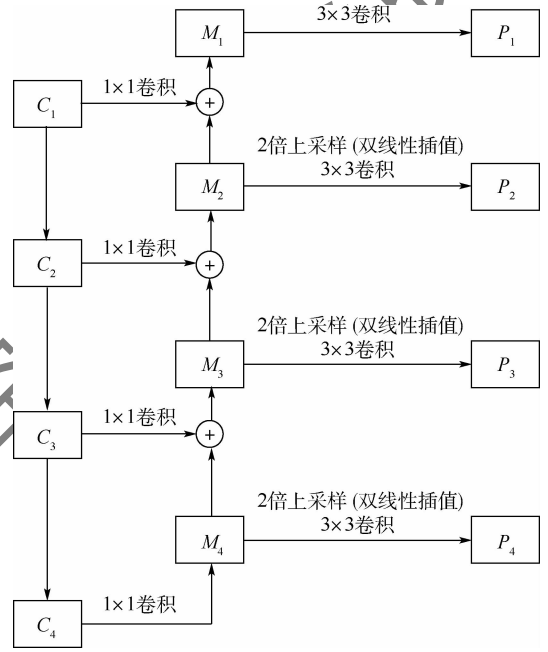


图 5 多分辨率特征提取网络结构

Fig. 5 Structure of multiple resolution feature extraction network

得到相同通道数的特征图 H_i 后,从 H_4 开始逐步向浅层融合,融合方式如下:设 M_i 为第 i 层对应尺度在融合后获得的特征,则要获得 M_{i-1} ,先将 M_i 使用双线性插值的方法,将其上采样至原尺度的 2 倍。此时尺度和通道数均与 H_i 相同,再将其与 H_i 逐个元素地相加。为消除上采样带来的混叠效应,同时提高感受野,使用 3×3 卷积对 M_i 进行处理。设 M_i 使用的 3×3 卷积操作为 G_i ,则最终输出为 $P_i = G_i(M_i)$ (4)

综上,从 C_i 得到 P_i 可表示为

$$P_i = \begin{cases} P_4 & i = 4 \\ G_i(T_i(C_i)) + \text{Up}_{2\times}(T_{i+1}(C_{i+1})) & i = 1, 2, 3 \end{cases} \quad (5)$$

式中: $\text{Up}_{2\times}$ 表示 2 倍上采样操作。

3.4 特征自选择模块

在一般的识别任务中,特征融合的常用方式是直接融合。然而,特征间的尺度不同,不能直接相加来完成融合。一个简单的特征融合方案是:利用特征间通道数一致的特点,先使用全局平均池化,将不同特征转换为与通道数相同维度的向量;再将各个分辨率对应的特征向量逐个元素直接相加,获得融合后的向量。上述操作可表示为

$$P = \sum_{i=1}^4 \text{GAP}(P_i) \quad (6)$$

而在本模块中,没有选择融合后再进行识别,而是先在各尺度同时进行预测,如图 6 所示。每个特征图在进行全局平均池化后,直接输入到全连接层获得激活结果,之后选择能得到最大激活的特征图作为最终使用的特征图。上述操作可表示为

$$P = \arg \max_p \text{FC}(\text{GAP}(P_i)) \quad (7)$$

该操作也可以看作将被选择的特征图权重设为 1,其余分辨率特征的权重设为 0 后,按权重进行融合的结果。

在训练过程中,各尺度特征间使用的全连接层是共享参数的,原因如下:最大激活输出的操作会将除最大激活以外的输出反传的梯度置零。在训练初期,由于各层的参数都是随机的,正样本也会在随机的一层特征上进行训练。假如全连接层不共享参数,此后,对于正样本而言,该层的激活会比其他层更大,最终导致只有一层特征得到训练的情况。

此外,特征自选择的作用同时体现在了正负样本上。对于正样本,由于识别网络参数共享的

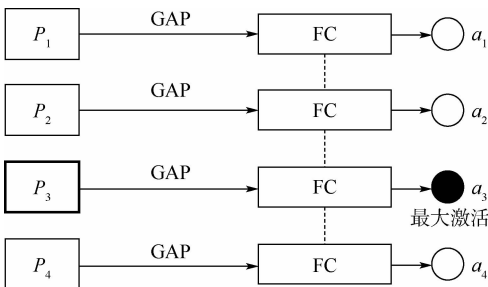


图 6 特征自选择模块结构

Fig. 6 Structure of feature auto-selection module

机制,训练过程中会自动选择最合适分辨率的特征来对网络进行训练,且保证各个分辨率的特征都有机会得到训练;对于负样本,由于最终输出的是最大激活,训练时会形成互相监督的机制,一旦某层对负样本的预测概率过高,就会产生反传梯度,从而进一步压低对负样本的预测概率。

3.5 识别结果生成

在完成特征选择后,直接使用计算最大激活的全连接层来得到激活值,再使用 Sigmoid 函数得到最终分数 s ,上述过程可表示为

$$s = \text{Sigmoid}(\text{FC}(P)) \quad (8)$$

在训练过程中,对于输出的最终分数 s ,使用交叉熵进行训练。设训练样本的标签为 y , $y = 1$ 表示为正样本, $y = 0$ 表示负样本,则训练阶段的损失 L_{cls} 为

$$L_{\text{cls}} = y \ln s + (1 - y) \ln(1 - s) \quad (9)$$

本文提出的特征自选择光伏组件遮挡物识别算法具体过程如算法 1 所示。

算法 1 特征自选择光伏组件遮挡物识别算法。

输入:光伏组件图像 I_{test} 、有标注的光伏组件图像训练集 I 。

输出:光伏组件图像 I_{test} 最终的遮挡物识别得分 s 。

1) 初始化网络参数 θ_F 。

2) REPEAT:

3) 随机选择一张图像训练集 I 中的光伏组件图像。

4) 通过图 4 所示网络提取多尺度全局特征 C_1 、 C_2 、 C_3 和 C_4 。

5) 通过图 5 所示网络提取多分辨率特征 P_1 、 P_2 、 P_3 和 P_4 。

6) 对多分辨率特征 P_1 、 P_2 、 P_3 和 P_4 做全局平均池化,并输入到全连接层中,获得对应激活 a_1 、 a_2 、 a_3 和 a_4 。

7) 使用 $\text{Sigmoid}(\max\{a_1, a_2, a_3, a_4\})$ 计算出识别得分。

8) 根据损失函数 L_{cls} 及输入图像标签优化全局网络参数。

9) UNTIL:损失函数收敛。

10) 将 I_{test} 输入到训练完毕的网络中,按照与训练网络单次迭代相同方式,获得最大激活值 a 。

11) 通过 Sigmoid 函数获得最终光伏组件遮挡物得分 s 。

4 实验结果及分析

4.1 数据集

本文实验数据采集自真实无人机所拍摄的无遮挡光伏板图像及带落叶遮挡的光伏板样本,并通过手动粘贴、裁剪等方式,获得用于训练的正负样本各 3 800 张,以及用于测试的正负样本各 380 张。在生成测试正样本时,考虑到现实情况中

的落叶密度变化,特别生成了落叶密度较为稀疏的正样本。在生成测试负样本时,为了体现光伏组件遮挡物识别的遮挡物背景强相关性,也生成了图像中包含落叶但未对光伏组件形成遮挡的负样本。部分训练及测试样本如图 7 所示,其中第 1 行为正样本,第 2 行为同列第 1 行正样本的局部放大图,第 3 行为同列第 1 行正样本所对应负样本,同时,最后 1 列为测试时所用的其中一对稀疏样本。

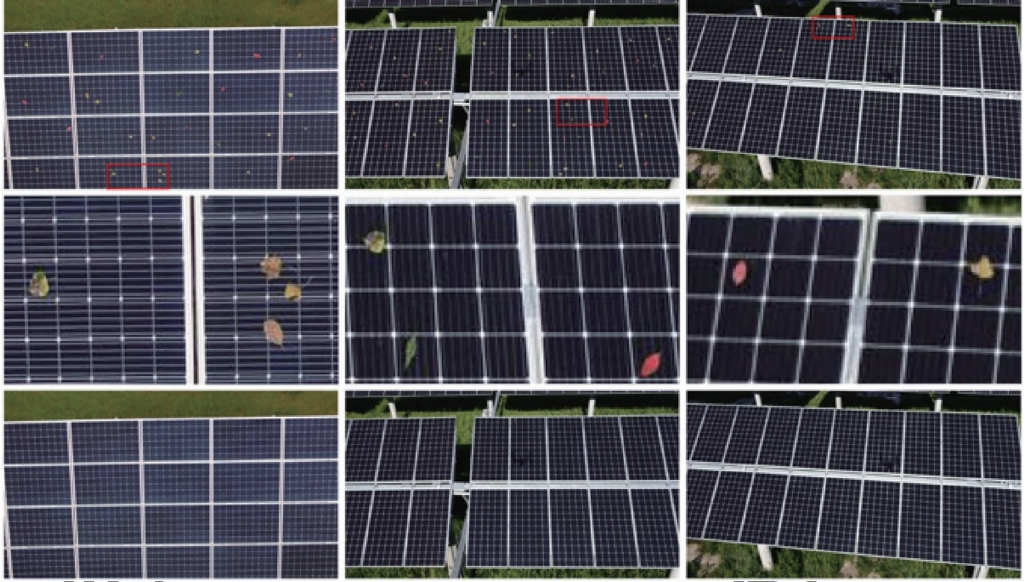


图 7 训练及测试样本示意图

Fig. 7 Illustration of partial training and test samples

4.2 对比算法及实验设置

为表明本文算法 PORNet 在光伏组件遮挡物识别上精度的优越性,以及验证特征自选择的必要性,对比了参数量相近的 VGG 网络(包括 VGG11、VGG13 和 VGG16)、残差网络(Res18)以及与本文算法网络结构相同但仅使用直接融合后的特征完成识别任务(FuseRes18)的多种基准算法。

在训练及测试阶段,防止引入过多噪声,除了将输入图像重新缩放至 224×224 外,本文算法中未引入其他图像增强方法。针对 224×224 的输入,各模块处理过程中所输出的特征图信息如表 2 所示。

实验在一台具有 NVIDIA GTX 1080ti 显卡的服务器上进行。本文算法 PORNet 及基准算法均基于 Pytorch 框架实现。训练过程中,使用 Adam 优化器对网络参数进行优化,学习率设定为 0.000 1,批大小设定为 64。Adam 优化器中的 beta1 和 beta2 分别设定为 0.9 和 0.999, weight_decay 设置为 0.000 1。每个模型迭代训练 300 次,保留最终的模型进行评估指标的计算。

表 2 各模块特征图信息

Table 2. Feature map information of different modules

特征图名称	尺度	通道数
C_1	56	64
C_2	28	128
C_3	14	256
C_4	7	512
P_1	56	256
P_2	28	256
P_3	14	256
P_4	7	256

4.3 评估方法

本文选用了 3 个指标进行评估:准确率、召回率及 AUC(area under curve)。准确率能很好地衡量模型整体预测的正确性,但不能体现针对不同类别样本的预测准确性,尤其在正负样本比例相差较大的情况下。召回率能体现网络对遮挡识别的召回能力,但在分类阈值设定得比较低时无法体现模型真实性能。AUC 表示 ROC 曲线下的面积,对于识别问题是一个公正而全面的评价指标,而且对正负样本的比例不敏感,适用于测试样本中正负样本相差比较悬殊的评估场景。

4.4 实验结果及分析

将本文算法及其他基准算法在自建光伏组件落叶遮挡数据集上进行训练和测试,测试结果如表3所示。

从表3中结果可以看出,本文算法对比基础的 Res18,在准确率、召回率和 AUC 上分别提升了 9.21%、15.79% 和 0.026 1。同时,与本文在图像特征提取模块使用的基础网络结构不同的 VGG 网络进行对比,在准确率、召回率和 AUC 上也均有不同程度的提升。实验结果说明,本文算法可以提高光伏组件的遮挡物识别性能。从运行时参数上看(见表4),本文算法 PORNet 相比于 Res18,在增加了多分辨率特征提取模块后,参数量增加了 2.49×10^6 ,MAC 增加了 2.38×10^9 。尽管如此,速度仅降低了 9.55 帧/s,相比于识别精度的提高,在运行时增加的这些损耗是完全可以接受的。此外,值得注意的是,PORNet 比 VGG16 的参数量和计算量都要小,但其精度要优于 VGG16,进一步证明了 PORNet 的优越性。

表3 不同算法测试结果

Table 3 Test results of different algorithms

算法	准确率/%	召回率/%	AUC
VGG11	84.21	68.42	0.916 2
VGG13	86.84	73.68	0.941 8
VGG16	92.11	84.21	0.981 3
Res18	89.47	81.58	0.965 4
FuseRes18	89.47	78.95	0.943 2
PORNet	98.68	97.37	0.991 5

在表3中,值得注意的是,FuseRes18 相比于 Res18,虽然准确率一致,但召回的正样本数更少,AUC 也更低。在光伏组件的遮挡物识别场景下,漏放显然比误召所造成的后果更加严重。召回率降低的原因在于:直接融合所有分辨率的特征会引入过多的无用信息,甚至引入错误的背景信息,尤其是在池化后,因此理论上,此时使用融合后特征进行识别,必然不如直接使用 C_4 进行识别的精度更高。Res18 和 FuseRes18 的结果证明了这一结论。

为了证明特征自选择的有效性,在测试集中分别挑出图中遮挡物尺度较大和尺度较小的图像,使用 PORNet 进行测试,测试图如图8所示。

结果表明,针对第1行遮挡叶片尺度较小的图像,分辨率较大的 P_1 ,相比于其余的 P_2 、 P_3 和 P_4 ,激活更大;而第2行中遮挡叶片尺度较大的图像, P_3 上的激活大于其余特征图。这间接证明了不同尺度遮挡物存在最适合识别的分辨率的差异,尺度大的物体适合用分辨率小的特征图进行识别,而尺度小的物体则适合用分辨率大的特征图进行识别。

表4 不同算法运行时参数

Table 4 Runtime parameters of different algorithms

算法	参数量/ 10^6	MAC/ 10^9	速度/(帧·s ⁻¹)
VGG11	8.79	6.98	201.79
VGG13	8.97	10.43	180.74
VGG16	14.03	14.31	156.14
Res18	10.66	1.69	125.38
FuseRes18	13.15	4.07	120.08
PORNet	13.15	4.07	115.83



图8 不同尺度测试正样本示意图

Fig. 8 Illustration of test positive samples with different scales

为了观察光伏组件遮挡物识别的背景强相关性带来的影响,可视化了部分被 Res18 及 FuseRes18 误召、而 PORNet 识别正确的测试负样本,如图 9 所示。可以看出,易误召样本图像中的落叶分布在了未对光伏组件造成遮挡的位置。对于基准算法 Res18,其过于关注图像中遮挡物的类别信息,因此当图像中存在遮挡物时,无论是否真造成遮挡,都会倾向于将其误判为正样本。而 PORNet 通过特征自选择筛选出语义最具代表性的特征,从而更好地保留了遮挡物的背景信息。此外,在去掉特征自选择机制(FuseRes18)后,依旧会产生误召情况,进一步证明了特征自选择的有效性和必要性。



图 9 易误召测试负样本示意图

Fig. 9 Illustration of easy recalled test negative samples

为了确定 PORNet 学到的信息与遮挡物本身有关,本文使用 Grad-CAM^[23] 算法,对 PORNet 的激活特征图做了可视化,可视化样本选择了 Res18 不能召回而 PORNet 可以召回的正样本,可视化结果如图 10 所示。可以看出,高激活区域的确是遮挡物所在区域,证明 PORNet 是根据落叶的存在来判断遮挡信息,而不是根据光伏板组件本身的特点这种与遮挡物无关的信息进行识别,具有良好的可解释性。同时,这也证明了 PORNet 对多分辨率特征的使用是有效的,弥补了 Res18 对在遮挡物尺度较小、分布较稀疏场景下识别精度低的缺陷。

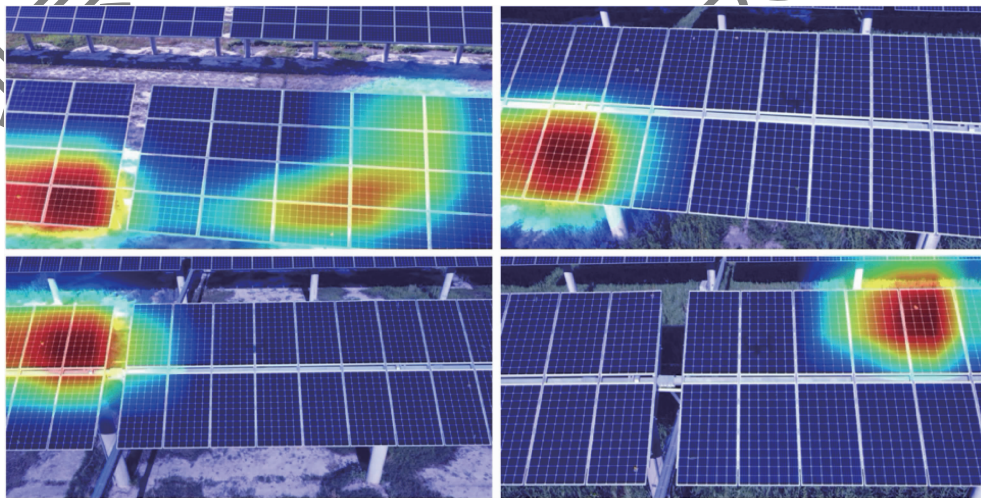


图 10 难召回测试正样本高激活区域可视化

Fig. 10 Visualization of high response regions for hard test samples

5 结 论

本文提出了面向光伏组件的遮挡物识别算法 PORNet,结论如下:

1) PORNet 可以实现比现有算法更为优秀的光伏组件遮挡物识别性能,在自建光伏组件落叶遮挡数据集上,准确率和召回率分别能提升 9.21% 和 15.79%,优于其他遮挡物识别基准算法。

2) PORNet 通过引入特征金字塔和特征自选择机制,有利于提高光伏组件上遮挡物的识别效果,尤其针对遮挡物密度较低、尺度较小场景下的识别。

3) PORNet 可以通过查看激活区域对遮挡物位置进行粗略定位,具有良好的可解释性。

PORNet 主要应用于 RGB 图像中,在未来的工作中,将尝试将其应用于红外图像的光伏组件热斑识别。此外,利用 PORNet 的良好可解释性,

可以尝试进一步检测遮挡物的位置,辅助光伏运维系统对光伏组件遮挡进行更精准的定位。

参考文献 (References)

- [1] JIA D, WEI D, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 248-255.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems, 2012: 1106-1114.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C] // International Conference on Learning Representations, 2015: 1-14.
- [4] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 815-823.
- [5] LIU W, WEN Y, YU Z, et al. SphereFace: Deep hypersphere embedding for face recognition [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6738-6746.
- [6] LI W, ZHU X, GONG S. Person re-identification by deep joint learning of multi-loss classification [C] // Proceedings of International Joint Conference on Artificial Intelligence. New York: ACM, 2017: 2194-2200.
- [7] ZHONG Z, LIANG Z, CAO D, et al. Re-ranking person re-identification with k-reciprocal encoding [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 3652-3661.
- [8] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
- [9] ZAGORUYKO S, KOMODAKIS N. Wide residual networks [C] // Proceedings of the British Machine Vision Conference, 2016: 1-12.
- [10] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 2261-2269.
- [11] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5987-5995.
- [12] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5693-5703.
- [13] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1-9.
- [14] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C] // IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 764-773.
- [15] LI D, HU J, WANG C, et al. Involution: Inverting the inference of convolution for visual recognition [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021.
- [16] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization [C] // International Conference on Learning Representations, 2018.
- [17] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout [EB/OL]. (2017-08-15) [2021-06-01]. <http://arxiv.org/abs/1708.04552>.
- [18] YUN S, HAN D, OH S J, et al. CutMix: Regularization strategy to train strong classifiers with localizable features [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 6023-6032.
- [19] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: Learning augmentation policies from data [EB/OL]. (2018-05-24) [2021-06-01]. <https://arxiv.org/abs/1805.09501>.
- [20] HO B, LIANG E, STOICA I, et al. Population based augmentation: Efficient learning of augmentation policy schedules [C] // Proceedings of the 36th International Conference on Machine Learning, 2019: 2731-2741.
- [21] LIM S, KIM I, KIM T, et al. Fast AutoAugment [EB/OL]. (2019-05-01) [2021-06-01]. <http://arxiv.org/abs/1905.00397>.
- [22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of IEEE Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.
- [23] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [J] International Journal of Computer Vision, 2020, 128(2): 336-359.

Occlusion recognition algorithm based on multi-resolution feature auto-selection

XIE Xiangying, LAI Guangzhi, NA Zhixiong, LUO Xin*, WANG Dong

(State Grid Digital Technology Holding Co., Ltd., Beijing 100053, China)

Abstract: The identification of obstructions of photovoltaic modules is an indispensable link in modern photovoltaic operation and maintenance systems. Traditional identification methods mostly rely on manual inspections, but they are costly and inefficient. Therefore, based on the convolutional neural network, PORNet, an occlusion recognition algorithm for photovoltaic modules, is proposed. By introducing feature pyramids, image features with rich semantic information at multiple resolutions are constructed, enhancing the sensitivity to the scale and density of occlusions. Through feature auto-selection, the most representative feature maps are screened out to strengthen the semantic information expression of the object contexts. Finally, the screened feature map is used to complete the occlusion recognition, improving the recognition accuracy. Experimental comparison and analysis are carried out on the self-built photovoltaic module falling leaf occlusion dataset, and the recognition performance is evaluated. Compared with existing object recognition methods, the accuracy and recall rate of the proposed method are increased by 9.21% and 15.79%, respectively.

Keywords: photovoltaic module; occlusion recognition; convolutional neural network; feature pyramid; feature auto-selection

Received: 2021-06-02; **Accepted:** 2021-07-04; **Published online:** 2021-07-23 17:05

URL: kns.cnki.net/kcms/detail/11.2625.V.20210723.1245.001.html

Foundation items: National Key R & D Program of China (2018YFB1500800); Technology Project of State Grid Corporation of China (SGTJDK00DYJS2000148)

* **Corresponding author.** E-mail: lx@ustc.edu.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0292

气动舵机高压反向直动式减压阀的设计及特性

刘延斌¹, 王学生^{1,*}, 秦新亚¹, 王浩¹, 陈琴珠¹, 赵赛²

¹ 华东理工大学 承压系统与安全教育国家重点实验室, 上海 200237; ² 上海航天精密机械研究所, 上海 201600)

摘 要: 气动舵机在防空导弹的飞行姿态控制系统中是关键部件, 直接影响导弹的飞行稳定性及姿态控制能力。弹载高压气体经过减压为气动舵机提供动力, 可以减小气源系统所占的空间且尽可能多地携带气体以利于提高导弹射程。针对某型号导弹, 设计了一种具有锥形阀瓣的高压反向非平衡直动式减压阀, 建立了该类减压阀热力学及静力分析的数学模型, 并在此基础上开发了其设计校核软件。建立了基于 AMESim 软件的稳态及非稳态进口压力下的仿真模型, 对其压力、流量特性及阀芯位移特性进行了模拟分析。结果表明: 该减压阀在设计参数下具有较好的压力及流量特性, 且理论计算结果与仿真分析结果吻合较好。

关键词: 气动舵机; 减压阀; 设计方法; AMESim; 特性分析

中图分类号: V245.4; TH138.5

文献标志码: A **文章编号:** 1001-5965(2022)07-1164-10

气动舵机是防空导弹飞行姿态控制系统的重要组成部分, 在气动舵机的气源系统中, 减压阀的作用是为系统执行机构提供满足压力、流量及稳定性要求的冷气来源。因此, 减压阀在气动舵机系统中具有十分重要的作用。并且, 在满足上述基本功能要求的前提下, 减压阀的质量应该尽可能轻, 可靠性高, 与此同时结构尽可能简单^[1-3]。

减压阀的减压作用是: 利用节流效应, 即气流通过节流口时将会伴随一定的能量损失, 进而降低气流的压力。当输入压力及输出流量存在波动时, 节流口后的压力也会波动, 因此, 需要一定的调节手段才能确保减压阀的出口压力维持在一定的范围。减压阀在结构设计上一般包括阀芯、阀座、敏感元件, 如活塞、调节弹簧、复位弹簧及调节组件和密封结构等。减压阀可以分为以下种类: ①根据输出压力的调节方式分为直动式和先导式; ②根据输出压力对阀芯的作用分为平衡式和非平衡式; ③根据按气体流动方向与阀芯开启方

向的不同分为正向式和反向式^[4-5]。

国内外针对气体减压模块及气体减压阀已有一定研究。孙冰等^[6]对飞行器增压系统中的两型双级气体减压器进行建模, 并对不同结构参数开展了仿真分析, 根据仿真结果, 提出可以优先减小减压器的二级阀芯质量或增大膜片刚度来提高稳定性。Chen 等^[7]通过建立不同开度、孔板直径及孔板倒角半径、套管直径及减压级数的减压阀数值计算模型, 研究了可压缩湍流工况下这些参数对噪声及能耗的影响规律, 提出控制噪声及能耗的最佳匹配参数。Casari 等^[8]开展了以甲烷为燃料的车载减压阀流道内热力学特性的数值计算和试验研究, 并且对比分析了恒物性参数与变物性参数对流道内压力和温度分布的影响。Chen 等^[9]提出了用于燃料电池电动汽车上氢气减压的多级高压减压阀, 并研究了不同结构参数对其内部流动特性的影响, 以获得更好的氢气减压效果, 并指出需要考虑制造难度及成本进而选择适合的结构参数。Fil 等^[10]用 AMESim 软件建立了

收稿日期: 2021-06-02; 录用日期: 2021-08-29; 网络出版时间: 2021-09-29 09:18

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210927.2043.001.html

基金项目: 中国航天科技集团公司第八研究院产学研合作基金 (SAST2020-111)

* 通信作者。E-mail: wangxs@ecust.edu.cn

引用格式: 刘延斌, 王学生, 秦新亚, 等. 气动舵机高压反向直动式减压阀的设计及特性[J]. 北京航空航天大学学报, 2022, 48(7): 1164-1173. LIU Y B, WANG X S, QIN X Y, et al. Design and characteristics of reverse direct-acting high-pressure reducing valve for pneumatic actuator[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1164-1173 (in Chinese).

气体压力调节器的仿真模型,研究了脉宽调制载波频率对开关阀控制特性的影响,通过设定压力和用户负荷值的阶跃变化,获得了调节器的瞬态特性。Shirato等^[11]针对气动执行机器中阀门存在的非线性问题,引入了质量流量反馈,改进了系统线性度,从而提高气动执行器的控制精度。在气动或液动控制系统的仿真方面,郭嘉博和肖敏^[12]基于AMESim软件建立了带安全阀功能的减压阀仿真模型,并进行仿真验证和性能分析,研究表明,弹簧刚度、进油孔直径、弹簧预紧力和反馈作用面直径等对减压阀的动态响应具有重要影响。滕浩等^[13]针对运载火箭高压小流量减压阀建立了AMESim仿真模型,并进行了动态仿真和试验研究,研究表明,起动段压力长时间下降及压力爬行是因摩擦力所致并提出了改进措施。此外,还有研究也通过采用AMESim与其他软件(如MATLAB/Simulink)联合建模来对减压系统或电液控制系统进行仿真并指导提升控制系统的性能^[14-16]。

按照现行国家标准GB/T 12244—2006^[17],减压阀进口的压力变化一般控制在进口压力的80%~105%,出口压力要求保持稳定。减压阀阀后压力一般控制在阀前压力的0.5倍左右,超过该范围减压阀的性能会受影响^[18]。而本文所研究的气动舵机用减压阀,进口压力波动范围为51.0~10.5 MPa,出口压力约为9.5 MPa,远远超出标准规定的范围。因此,需要针对其特殊工况开展专门的设计和分析研究,进而能够满足某型号导弹气动舵机对气源系统的特殊使用需求,并且达到使用精度及稳定性要求。

综合考虑该减压阀的特殊工况、结构的复杂程度、密封性能及稳定性等因素,本文设计了一种高压反向非平衡直动式减压阀。首先,建立该类减压阀的静态数学模型,开展热力学计算及阀芯受力分析,得到该类减压阀的设计计算流程与方法,并编制计算软件。然后,建立该减压阀在稳态与非稳态进口压力下的AMESim模型,分析其工作特性。

1 数学模型

在气动系统中,能量的传递及转化通过流体介质的一系列状态变化得以实现,在实际过程中往往是很复杂的过程,且都是程度不同的不可逆过程。在分析过程中,往往需要进行合理简化,把复杂过程适度简单化,进而转变为基本热力过程。因此,可以采用热力学第一定律和理想气体状态

方程等进行研究。本文研究的气动舵机系统,考虑到气流速度远大于与外界进行热交换的速度,近似看作等熵流动,但考虑到流体流经节流口的收缩与扩张引发明显的机械能损失,以及实际流动中流道面积收缩等引起的流量变化,需要通过增加流量系数的方法进行修正^[4]。

氮气在气动舵机系统中广泛采用,其是双原子气体,分子量为28,气体常数为296.93 J/(kg·K),等熵指数 $\kappa = 1.40$,临界压力比 $\varepsilon_{cr} = 0.528$ 。

1.1 流量计算方程

在气动系统中,根据开口系统稳定流动能量方程,并考虑到气流对外不做功,多数情况下,气体与外界的热量交换可以忽略,能量方程可以转化为

$$h_i + \frac{V_i^2}{2} = h_o + \frac{V_o^2}{2} = C \quad (1)$$

对于定值比热容的理想气体:

$$\begin{cases} h_i = c_p T_i \\ h_o = c_p T_o \end{cases} \quad (2)$$

$$T = \frac{p}{R\rho} \quad (3)$$

$$R = c_p - c_v \quad (4)$$

$$\kappa = \frac{c_p}{c_v} \quad (5)$$

式中: h_i 为进口流体的比焓,J/kg; h_o 为出口流体的比焓,J/kg; V_i 为流体进口速度,m/s; V_o 为流体出口速度,m/s; c_p 为流体定压比热容,J/(kg·K); c_v 为流体定容比热容,J/(kg·K); T 为流体温度,K; p 为流体压力,Pa; R 为气体常数; ρ 为流体密度,kg/m³; T_i 为进口流体温度,K; T_o 为出口流体温度,K。

将式(2)~式(5)代入式(1),可以得到

$$\frac{\kappa}{\kappa-1} \cdot \frac{p_i}{\rho_i} + \frac{V_i^2}{2} = \frac{\kappa}{\kappa-1} \cdot \frac{p_o}{\rho_o} + \frac{V_o^2}{2} = C \quad (6)$$

或

$$\frac{\kappa}{\kappa-1} RT_i + \frac{V_i^2}{2} = \frac{\kappa}{\kappa-1} RT_o + \frac{V_o^2}{2} = C \quad (7)$$

式中: p_i 为减压阀进口压力,Pa; p_o 为减压阀出口压力,Pa; ρ_i 为进口流体密度,kg/m³; ρ_o 为出口流体密度,kg/m³。

流动是等熵过程,因此可以得出如下关系:

$$\frac{T_o}{T_i} = \left(\frac{p_o}{p_i} \right)^{\frac{\kappa-1}{\kappa}} \quad (8)$$

$$\frac{\rho_o}{\rho_i} = \left(\frac{p_o}{p_i} \right)^{\frac{1}{\kappa}} \quad (9)$$

考虑到节流口后的流速要比节流口前大很

多,因此可以推导:

$$V_o = \sqrt{2c_p(T_i - T_o)} = \sqrt{\frac{2k}{k-1}RT_i\left(1 - \frac{T_o}{T_i}\right)} = \sqrt{\frac{2k}{k-1}RT_i\left[1 - \left(\frac{p_o}{p_i}\right)^{\frac{k-1}{k}}\right]} \quad (10)$$

利用连续性方程,可以得到质量流量:

$$q_m = \rho_o V_o A_o = A_o p_i \sqrt{\frac{2k}{RT_i(k-1)}} \sqrt{\left(\frac{p_o}{p_i}\right)^{\frac{2}{k}} - \left(\frac{p_o}{p_i}\right)^{\frac{k+1}{k}}} \quad (11)$$

式中: A_o 为截面流通面积, m^2 。

在工程上考虑到影响气体流动的因素很多,因此,通常在式(11)前增加一个流量系数 μ 来进行修正。

1) 当 $p_o/p_i > \varepsilon_{cr}$ 时,为亚临界流动,出口截面压力等于反压,流动受反压控制,流量通过式(12)计算:

$$q_m = \mu A_o p_i \sqrt{\frac{2k}{RT_i(k-1)}} \sqrt{\left(\frac{p_o}{p_i}\right)^{\frac{2}{k}} - \left(\frac{p_o}{p_i}\right)^{\frac{k+1}{k}}} \quad (12)$$

2) 当 $p_o/p_i \leq \varepsilon_{cr}$ 时,为超临界流动,出口截面压力等于临界压力,流量不受反压控制,保持最大值不变:

$$p_o/p_i = \left(\frac{2}{1+k}\right)^{\frac{k}{k-1}} \quad (13)$$

因此式(13)可以改写成

$$q_m = \mu A_o p_i \sqrt{\frac{2k}{RT_i(k+1)}} \left(\frac{2}{1+k}\right)^{\frac{1}{k-1}} \quad (14)$$

1.2 节流口面积计算

考虑到锥形阀瓣在密封性能上具有一定的优势^[19],因此本文设计采用锥形阀瓣结构。图1为气动舵机减压阀结构参数。

阀座通道面积计算公式为

$$A_f = \frac{\pi}{4}(D_2^2 - D_3^2) \quad (15)$$

式中: A_f 为阀座通道面积, m^2 ; D_2 为阀座孔径, m ; D_3 为阀杆直径, m 。

锥形阀瓣节流面积计算公式为

$$A_i = \pi D_2 h \sin \alpha - \pi h^2 \sin^2 \alpha \cos \alpha \quad (16)$$

式中: A_i 为锥形阀瓣节流口面积, m^2 ; h 为稳定工作时阀瓣开口量, m ; α 为阀瓣半角, $(^\circ)$ 。

1.3 阀芯静力分析

图2为气动舵机减压阀阀芯受力分析。在对阀芯进行静力学分析前,需要对一些次要因素进行合理的假设^[20-22],从而便于对主要影响因素进行把握和分析:

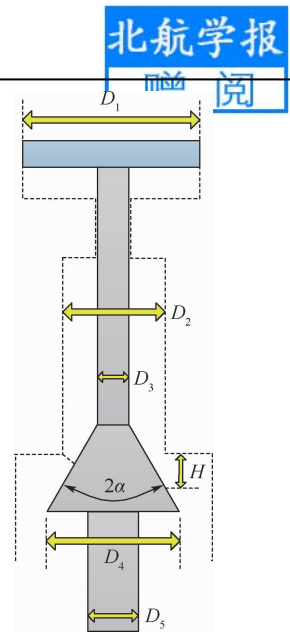


图1 气动舵机减压阀结构参数

Fig. 1 Structure parameters of pressure reducing valve for pneumatic actuator

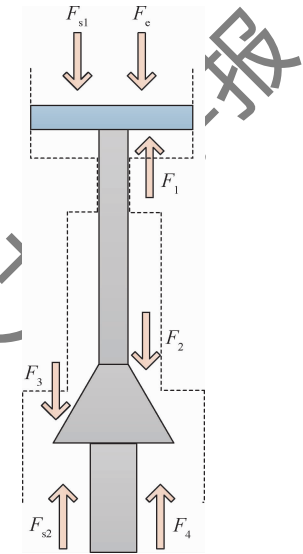


图2 气动舵机减压阀阀芯受力分析

Fig. 2 Spool stress analysis of pressure reducing valve for pneumatic actuator

- 1) 经过阀节流面积的流动为等熵流动。
- 2) 阀芯所承受的摩擦力可以忽略不计。
- 3) 氮气流过节流口时对阀芯的气流反力可以忽略不计。

- 4) 阀芯自重可忽略不计。

稳定工作状态时阀芯的受力平衡方程为

$$F_e + F_{s1} + F_2 + F_3 = F_1 + F_{s2} + F_4 \quad (17)$$

F_e 为环境大气压强作用于活塞上的力,即

$$F_e = p_e A_o \quad (18)$$

$$A_o = \frac{\pi}{4} D_1^2 \quad (19)$$

式中: p_e 为环境大气压强, $p_e = 0.1 \times 10^6$ Pa; D_1 为反馈腔活塞直径, m 。

F_{s1} 为调压弹簧施加于活塞上的力,即

$$F_{s1} = K_1 X_1 + K_1 (H - h) \quad (20)$$

式中: K_1 为调压弹簧的刚度系数, N/m ; X_1 为调压弹簧的预压缩量, m ; H 为初始阀瓣开口量, m 。

F_1 为出口压力施加于活塞上的力,即

$$F_1 = p_o A_1 \quad (21)$$

式中:

$$A_1 = \frac{\pi}{4} (D_1^2 - D_3^2) \quad (22)$$

F_2 为出口压力施加于节流口后阀瓣上部有效面积上的力,即

$$F_2 = p_o A_2 \quad (23)$$

式中:

$$A_2 = \frac{\pi}{4} [(D_2 - 2h \sin \alpha \cos \alpha)^2 - D_3^2] \quad (24)$$

F_3 为入口压力施加于节流口前阀瓣上部有效面积上的力,即

$$F_3 = p_i A_3 \quad (25)$$

式中:

$$A_3 = \frac{\pi}{4} [D_4^2 - (D_2 - 2h \sin \alpha \cos \alpha)^2] \quad (26)$$

式中: D_4 为锥形阀瓣下端直径, m 。

F_4 为入口压力施加于阀瓣下部有效面积上的力,即

$$F_4 = p_i A_4 \quad (27)$$

式中:

$$A_4 = \frac{\pi}{4} (D_4^2 - D_5^2) \quad (28)$$

式中: D_5 为锥形阀瓣下端圆柱直径, m 。

F_{s2} 为复位弹簧施加于阀瓣上的力,即

$$F_{s2} = K_2 X_2 - K_2 (H - h) \quad (29)$$

式中: A_0 为环境大气压强作用在活塞上的有效面积, m^2 ; A_1 为反馈压力在活塞上的有效作用面积, m^2 ; A_2 为出口压力施加于节流口后阀瓣上部的有效面积, m^2 ; A_3 为入口压力施加于节流口前阀瓣上部的有效面积, m^2 ; A_4 为入口压力施加于阀瓣下部有效面积, m^2 ; K_2 为复位弹簧的刚度系数, N/m ; X_2 为复位弹簧的预压缩量, m 。

因此,式(17)可以改写为

$$F_e + K_1 X_1 + K_1 (H - h) + p_o A_2 + p_i A_3 = p_o A_1 + K_2 X_2 - K_2 (H - h) + p_i A_4 \quad (30)$$

整理可以得到

$$p_o = [F_e + K_1 X_1 - K_2 X_2 + (K_1 + K_2)(H - h) + p_i (A_3 - A_4)] / (A_1 - A_2) \quad (31)$$

因此,通过节流口面积计算公式及对应的流量方程,进行联立求解,可以推导得出稳定工作状态时阀瓣的开启高度 h 的表达式(以 $p_o/p_i < \varepsilon_{cr}$

为例):

$$h = \frac{D_2}{\sin(2\alpha)} - \left[\left(\frac{D_2}{\sin(2\alpha)} \right)^2 - \frac{q_m}{\mu \pi p_i \sin^2 \alpha \cos \alpha \sqrt{\frac{2k}{RT_i(k-1)} \left(\frac{2}{1+k} \right)^{\frac{1}{k-1}}}} \right]^{\frac{1}{2}} \quad (32)$$

进一步的,可以得出出口压力的表达式为

$$p_o = \frac{F_e + K_1 X_1 - K_2 X_2 + p_i (A_3 - A_4)}{A_1 - A_2} + (K_1 + K_2) \left\{ H - \frac{D_2}{\sin(2\alpha)} + \left[\left(\frac{D_2}{\sin(2\alpha)} \right)^2 - \frac{q_m}{\mu \pi p_i \sin^2 \alpha \cos \alpha \sqrt{\frac{2k}{RT_i(k-1)} \left(\frac{2}{1+k} \right)^{\frac{1}{k-1}}}} \right]^{\frac{1}{2}} \right\} / (A_1 - A_2) \quad (33)$$

2 设计计算方法

2.1 主要参数

对于该种气动舵机用高压反向非平衡直动式减压阀的设计,一般需要考虑以下设计参数^[12,23]:①锥形阀瓣的角度、阀瓣下端直径;②节流口面积;③阀杆直径、阀座孔径、反馈腔活塞直径;④弹簧刚度系数及压缩量;⑤压力偏差等。

2.2 设计校核方法及流程

1) 根据减压阀进口压力 p_i 、出口压力 p_o 、进口温度 T_i 及出口温度 T_o 计算流体密度 ρ 及流量 q_m 。

2) 选择合适的流量系数 μ , 计算对应的减压阀节流口面积 A_1 。

3) 针对所选的锥形阀瓣结构,初步选定阀瓣半角、反馈腔活塞直径 D_1 、阀座孔径 D_2 、阀杆直径 D_3 、锥形阀瓣下端直径 D_4 、锥形阀瓣下端圆柱直径 D_5 、初始阀瓣开口量 H 。

4) 根据计算得到的减压阀节流口面积 A_1 ,进一步计算得出稳定工作时阀瓣开口量 h_1 。

5) 针对稳定工作点,初步选定调压弹簧和复位弹簧的刚度系数 K_1 与 K_2 , 设定其中一个弹簧的压缩量 X_1 , 并计算另一个弹簧的压缩量 X_2 。

6) 选择其他工况点,在相同的其他参数情况下,重新计算减压阀节流口面积 A_1 ,并进一步得出该工况下稳定工作时阀瓣开口量 h_2 。

7) 根据上述已经计算得出的调压弹簧和复

位弹簧的刚度系数 K_1 与 K_2 及对应的压缩量 X_1 与 X_2 , 并根据阀芯受力平衡方程, 计算得出该工况下减压阀出口压力 $p_{o,2}$, 同时, 比较 $p_{o,2}$ 与 p_o 的偏差是否处在允许的范围内。

8) 如压力偏差满足设计要求, 设计校核过程结束; 如不满足要求, 则反复进行前述步骤, 重新调整相关参数, 重新设计至满足不同工况的设计要求。

详细的气动舵机减压阀设计校核计算流程如图 3 所示。

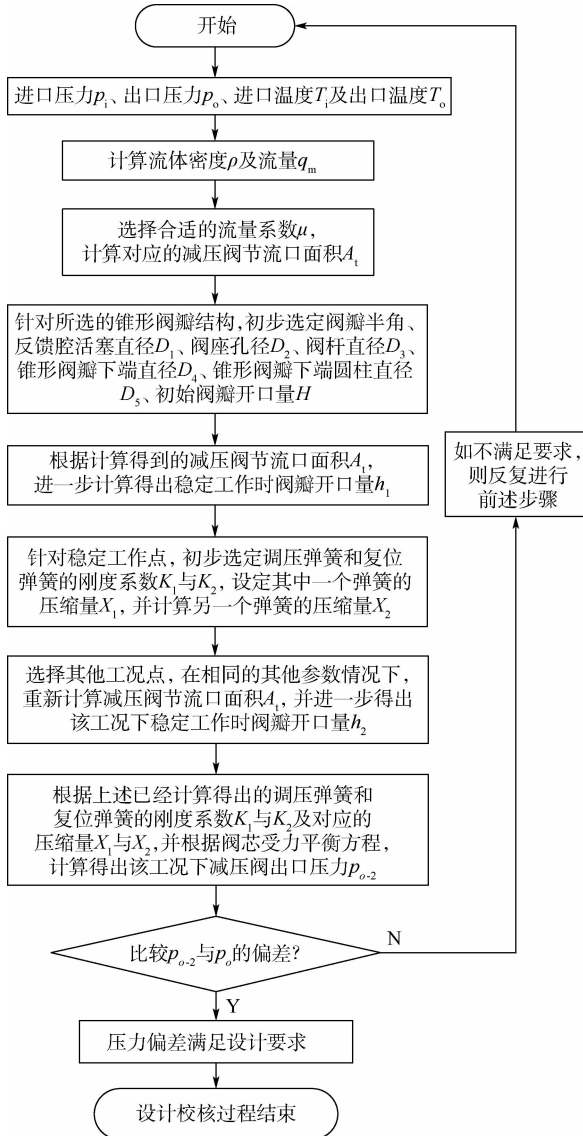


图 3 气动舵机减压阀设计校核计算流程

Fig. 3 Design and check calculation flow chart of pressure reducing valve for pneumatic actuator

2.3 设计校核软件开发

针对该种气动舵机减压阀, 为便于设计计算及校核不同操作工况下的压力变化, 开发了一种能够实现上述功能的软件。在该软件中, 主要分为设计模式和校核模式。

1) 在设计模式下, 根据初始设计要求, 选择软件数据库中合适的流体介质, 并录入主要设计工况参数及减压阀主要结构参数等, 根据气体状态方程及阀芯受力分析等公式, 完成主要过程参数的计算及弹簧的初步设计。

2) 在校核模式下, 以设计模式工况计算得出的弹簧刚度系数及预压缩量等参数作为初始条件, 根据设定的其他工况进出口压力开展校核计算, 并对比出口压力偏差情况, 从而判断是否满足设计要求。

3) 通过使用该减压阀设计校核软件, 可以大幅提高减压阀设计效率, 实现减压阀的初步设计, 并利用校核模块实现不同工况的校核。在上述工作基础上, 通过模拟软件进行减压阀不同工况下的静态特性仿真分析, 根据仿真结果可以对其他影响因素进一步优化, 使得设计的减压阀在满足使用需求的前提下, 具有良好的静态和动态特性。

3 减压阀结构参数及 AMESim 模型建立

针对某型气动舵机减压阀的设计参数要求, 通过以上方法开展设计, 得出减压阀的主要结构参数, 并在此基础上建立该减压阀的 AMESim 模型^[13, 21-25]。表 1 为某型气动舵机减压阀的设计参数, 图 4 和图 5 分别为稳态及非稳态进口压力下减压阀 AMESim 模型。

表 1 某型气动舵机减压阀的设计参数

Table 1 Design parameters of pressure reducing valve for a certain pneumatic actuator

参数	数值
气源气体	氮气
气源压力/MPa	0 ~ 51.0
额定出口压力/MPa	9.8
气源温度/K	293.15
阀座孔径/mm	6.0
阀杆直径/mm	4.0
锥形阀瓣半角/(°)	30.0
锥形阀瓣下端直径/mm	12.0
高压腔容积/L	0.1
反馈腔容积/L	0.1
反馈腔活塞直径/mm	20.0
初始阀瓣开口量/mm	0.5
阀瓣质量块/kg	0.1
质量块限位/mm	0 ~ 0.5
流量系数 μ	0.7
调压弹簧刚度/(N · mm ⁻¹)	800.0
调压弹簧预压缩量/mm	2.7
复位弹簧刚度/(N · mm ⁻¹)	500.0
复位弹簧预压缩量/mm	0
出口节流孔直径/mm	1.2

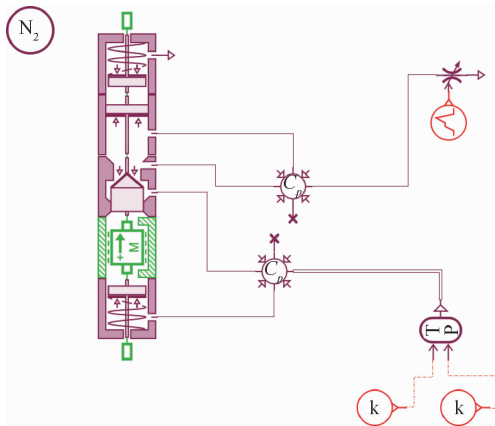


图 4 稳态进口压力下减压阀 AMESim 模型
Fig. 4 AMESim model of pressure reducing valve under steady inlet pressure

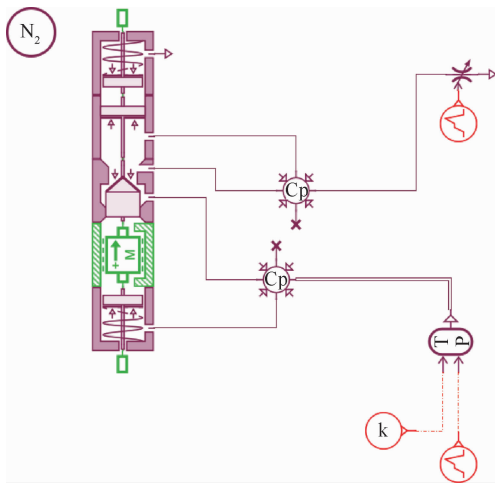


图 5 非稳态进口压力下减压阀 AMESim 模型
Fig. 5 AMESim model of pressure reducing valve under unsteady inlet pressure

4 减压阀特性仿真分析

在建立的 AMESim 仿真模型基础上,开展模拟分析。压力特性曲线反映出口压力与进口压力的关系,流量特性曲线反映出口压力与流量的关系^[26-28]。分别针对稳态进口压力及非稳态进口压力 2 种不同工况,进行该类气动舵机减压阀的压力与流量特性分析,并对阀芯位移的变化趋势展开探讨。

4.1 稳态进口压力工况特性

在稳态进口压力工况下,进口压力始终保持固定值,本节分别模拟了进口压力为 10.5 MPa、30.0 MPa 及 51.0 MPa 对应工况下的出口压力、流量、阀芯位移变化趋势。模拟时长均为 2 s,根据模拟结果可以看出已经达到较好的稳定状态。

4.1.1 出口压力变化趋势

图 6 为稳态进口压力下出口压力变化趋势。

可以看出,随着时间的延长,出口压力均很快达到稳定状态,其中进口压力在 51.0 MPa 时,出口压力达到稳定状态所需时间最短,进口压力在 10.5 MPa 时,出口压力达到稳定状态所需时间最长,但不超过 0.2 s。另外,在相同时刻,进口压力越高,则对应的出口压力也越高。在 3 种不同进口压力下,该减压阀出口稳定状态下的压力处于 9.22 MPa 和 9.85 MPa 之间,与设计出口压力 9.8 MPa 相比,最大偏差为 5.9%。因此,该减压阀的出口压力特性良好。

在上述 3 种不同进口压力下,通过自开发软件理论计算的出口压力与 AMESim 模拟的出口压力对比可以发现(见表 2),最大偏差为 2.4%。

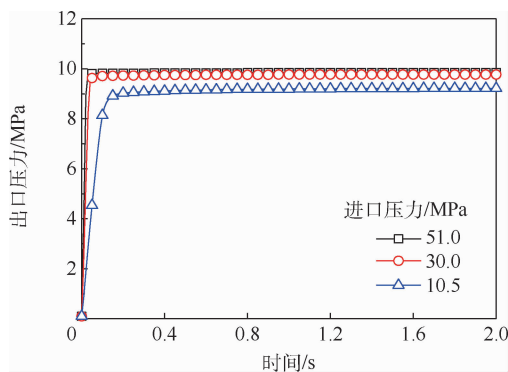


图 6 稳态进口压力下出口压力变化趋势

Fig. 6 Changing tendency of outlet pressure under steady inlet pressure

表 2 出口压力偏差对比

Table 2 Comparison of outlet pressure deviation

进口压力/ MPa	出口压力/MPa		偏差绝对 值/%
	理论计算	模拟	
51.0	9.87	9.85	0.2
30.0	9.70	9.77	0.7
10.5	9.45	9.22	2.4

4.1.2 出口流量变化趋势

图 7 为稳态进口压力下出口流量变化趋势。减压阀出口节流口面积保持不变,出口流量随着时间的延长均很快达到稳定状态,大约在 1 s 以后,所有流量趋于稳定值。其中进口压力在 51.0 MPa 和 30.0 MPa 时,出口流量变化趋势比较接近,而进口压力在 10.5 MPa 时,出口流量的变化趋势更加缓慢。最终,出口流量稳定在 16.7 ~ 17.4 g/s 之间,与设计出口流量 17.8 g/s 的最大偏差为 6.2%。

4.1.3 阀芯位移变化趋势

图 8 为稳态进口压力下阀芯位移变化趋势。随着时间的延长,阀芯位移在进口压力作用下均很快达到稳定状态,并且随进口压力的降低呈现

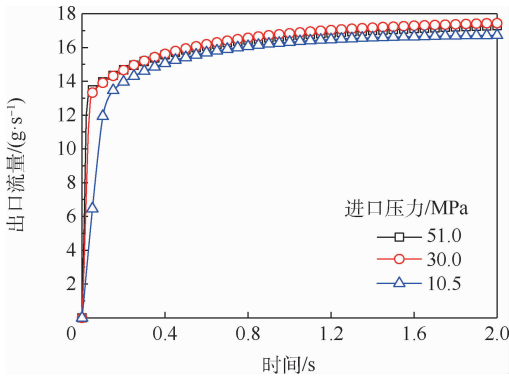


图7 稳态进口压力下出口流量变化趋势

Fig.7 Changing tendency of outlet flow under steady inlet pressure

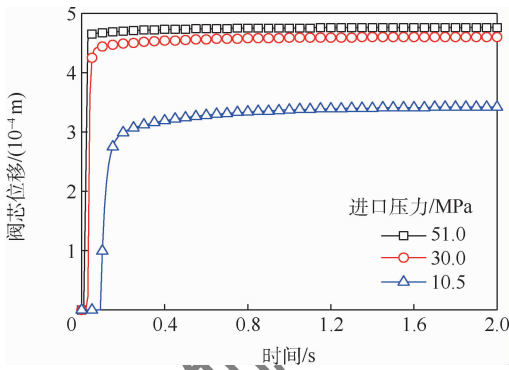


图8 稳态进口压力下阀芯位移变化趋势

Fig.8 Changing tendency of spool displacement under steady inlet pressure

出逐渐减小的趋势。其中,进口压力在 51.0 MPa 时,阀芯位移达到稳定状态所需时间最短且位移量最大,进口压力在 10.5 MPa 时,阀芯位移达到稳定状态所需时间最长且位移量最小。也就是说,为了保证出口压力稳定,进口压力越高,对应的阀瓣与阀座间开口量越小,进口压力越低,对应的阀瓣与阀座间开口量越大,这与理论计算结果具有很好的一致性。

4.2 非稳态进口压力工况特性

在非稳态进口压力工况下,进口压力始终变化,本节模拟了进口压力从 0 MPa 线性增加至 51.0 MPa 工况下(见图9)的出口压力、流量、阀芯位移变化趋势。模拟时长同样为 2 s,根据模拟结果可以看出图 10 已经达到较好的稳定状态。

4.2.1 出口压力变化趋势

图 10 为非稳态进口压力下出口压力变化趋势。随着进口压力的不断升高,出口压力逐步上升并趋于平稳。在起初阶段,随着进口压力的变化,出口压力增速很快,当进口压力达到 8.5 MPa 左右,出口压力增长速度开始放缓并最终稳定在 9.8 MPa 左右,基本与设计出口压力

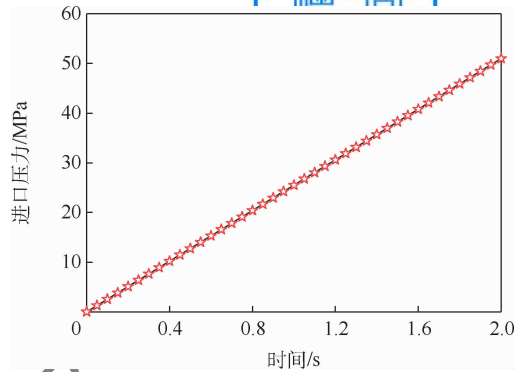


图9 非稳态进口压力变化趋势

Fig.9 Changing tendency of unsteady inlet pressure

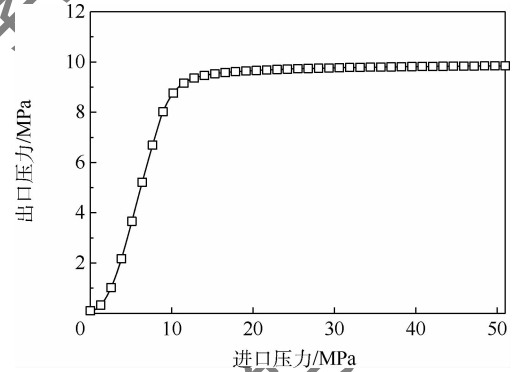


图10 非稳态进口压力下出口压力变化趋势

Fig.10 Changing tendency of outlet pressure under unsteady inlet pressure

相当,说明非稳态进口压力工况下出口压力特性良好。

4.2.2 出口流量变化趋势

图 11 为非稳态进口压力下出口流量变化趋势。与非稳态进口压力下出口压力变化趋势类似,随着进口压力的不断升高,出口流量逐步上升并最终趋于平稳。在起初阶段(约 8.5 MPa 之前),出口流量增速很快,之后出口流量增长速度开始放缓并最终稳定在 17.1 g/s 左右,与设计出口流量偏差约为 3.9%,说明非稳态进口压力工况下出口流量特性良好。

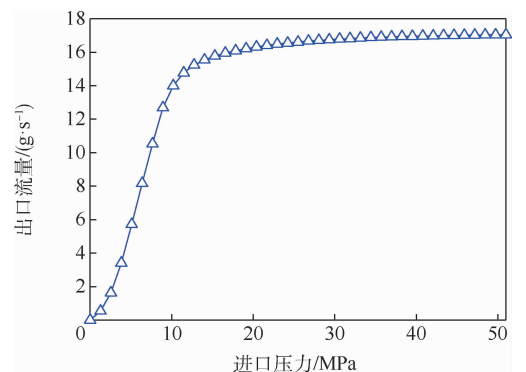


图11 非稳态进口压力下出口流量变化趋势

Fig.11 Changing tendency of outlet flow under unsteady inlet pressure

4.2.3 阀芯位移变化趋势

图12为非稳态进口压力下阀芯位移变化趋势。在起初阶段,进口压力约达到8.5 MPa之前,阀芯基本不移动,这说明该阶段在进出口压力作用下,阀芯受力平衡并未被打破,进而阀芯不产生位移量。之后,阀芯受力不再保持平衡,并开始向上移动,随着出口压力和流量的趋于稳定,阀芯位移也趋于平稳值 4.8×10^{-4} m左右。

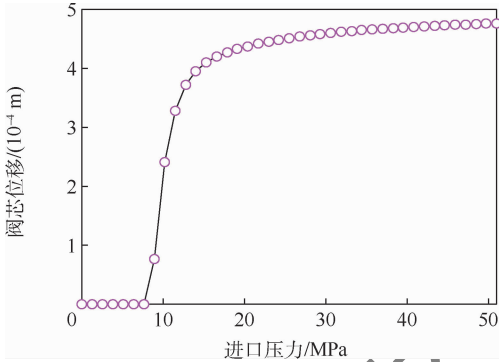


图12 非稳态进口压力下阀芯位移变化趋势

Fig. 12 Changing tendency of spool displacement under unsteady inlet pressure

5 结论

在理论计算的基础上,本文设计了某型导弹气动舵机减压阀,并进行多工况校核,开发了该类减压阀设计校核软件。基于AMESim软件建立了该类减压阀的稳态及非稳态进口压力下的仿真模型,开展了压力、流量及阀芯位移的仿真分析。得出以下结论:

1) 建立了该类减压阀静态特性数学模型,在确定主要设计参数基础上,明确了该类减压阀的设计思路和方法。

2) 开发了该类减压阀静态特性设计校核软件,建立了可以实现该类减压阀稳态及非稳态进口压力工况下的AMESim仿真模型,与AMESim仿真结果对比发现,不同稳态进口压力下该软件计算的出口压力值与AMESim模拟值吻合较好,在3种不同工况下出口压力的最大偏差为2.4%。

3) 在稳态进口压力下,在相同时刻,进口压力越高,则对应的出口压力也越高。在3种不同进口压力下,出口稳定状态下的压力与设计出口压力的最大偏差为5.9%;出口稳定状态下的流量与设计出口流量的最大偏差为6.2%;随着进口压力的升高,对应的阀瓣与阀座间开口量逐渐减小,这与理论计算结果具有很好的一致性。

4) 在非稳态进口压力下,出口压力迅速增加并趋于平稳,最终稳定在9.8 MPa左右,说明非稳

态进口压力工况下出口压力特性良好,出口流量逐步上升并最终稳定在17.1 g/s左右,与设计值偏差约为3.9%,说明非稳态进口压力工况下出口流量特性良好;阀芯受力平衡随进口压力的变化而发生变化,随着出口压力和流量的趋于稳定,阀芯位移最终也趋于稳定。

参考文献 (References)

- [1] 陈汉超. 气动舵机高压减压阀的分析与设计[J]. 北京理工大学学报(自然科学版), 1980, 9(2): 57-62.
CHEN H C. Analysis and design of a high pressure reducer valve for pneumatic actuator[J]. Transaction of Beijing Institute of Technology (Natural Science), 1980, 9(2): 57-62 (in Chinese).
- [2] POOLE H A. A stored gas actuator unit for the TOW missile; 650338 [R]. Washington, D. C.: SAE International, 1965: 361-376.
- [3] 陈青山. 某型近距防空导弹弹载气路系统设计[J]. 航空兵器, 2007(6): 16-19.
CHEN Q S. A kind of close quarters aerial defence missile gas supply system design[J]. Aero Weaponry, 2007(6): 16-19 (in Chinese).
- [4] 陈汉超, 盛永才. 气压传动与控制[M]. 北京: 北京工业学院出版社, 1987.
CHEN H C, SHENG Y C. Pneumatic transmission and control [M]. Beijing: Beijing Institute of Technology Press, 1987 (in Chinese).
- [5] 周铭杰. 浅谈减压阀的结构设计[J]. 液压与气动, 2005(3): 62-63.
ZHOU M J. Discuss of structural design of the relief pressure valve lightly[J]. Chinese Hydraulics & Pneumatics, 2005(3): 62-63 (in Chinese).
- [6] 孙冰, 许琪, 陈阳, 等. 双级气体减压器稳定性影响因素数值分析[J]. 北京航空航天大学学报, 2014, 40(12): 1660-1665.
SUN B, XU Q, CHEN Y, et al. Numerical analysis of influence factors on stability for dual-stage gas pressure reducing regulator [J]. Journal of Beijing University of Aeronautics and Astronautics, 2014, 40(12): 1660-1665 (in Chinese).
- [7] CHEN F Q, QIAN J Y, CHEN M R, et al. Turbulent compressible flow analysis on multi-stage high pressure reducing valve [J]. Flow Measurement and Instrumentation, 2018, 61: 26-37.
- [8] CASARI N, PINELLI M, SUMAN A, et al. Reducing pressure valve with real gases: An integrated approach for the design [J]. Energy Procedia, 2018, 148: 607-614.
- [9] CHEN F Q, REN X D, HU B, et al. Parametric analysis on multi-stage high pressure reducing valve for hydrogen decompression [J]. International Journal of Hydrogen Energy, 2019, 44(59): 31263-31274.
- [10] FIL K U, ILIUKHIN V N, GRESHNIAKOV P I. Digital control system of gas pressure regulator [C] // 2020 International Conference on Dynamics and Vibroacoustics of Machines (DVM). Piscataway: IEEE Press, 2020: 1-5.

- [11] SHIRATO Y, OHNISHI W, FUJIMOTO H, et al. Controller design of mass flow rate loop for high-precision pneumatic actuator [C] // 2020 IEEE 16th International Workshop on Advanced Motion Control (AMC). Piscataway: IEEE Press, 2020: 40-45.
- [12] 郭嘉博, 肖敏. 带安全阀功能的减压阀动态特性仿真与分析[J]. 机械工程与自动化, 2016(3): 58-60.
GUO J B, XIAO M. Dynamic characteristic simulation and analysis of pressure regulator with relief valve's function[J]. Mechanical Engineering & Automation, 2016(3): 58-60 (in Chinese).
- [13] 滕浩, 石玉鹏, 张亮, 等. 基于 AMESim 减压阀动态特性仿真与试验研究[J]. 上海航天, 2015, 32(1): 48-53.
TENG H, SHI Y P, ZHANG L, et al. Simulative and experimental study of dynamic characteristics of pressure relief valve based on AMESim[J]. Aerospace Shanghai, 2015, 32(1): 48-53 (in Chinese).
- [14] 栾飞舟, 王启权. 组合化减压阀设计与性能分析[G] // 2016 中国航空学会流体传动与控制学术会议, 2016: 225-228.
LUAN F Z, WANG Q Q. Design and performance analysis of combined pressure reducing valve[C] // 2016 Chinese Aeronautical Society Fluid Power Transmission and Control conference, 2016: 225-228 (in Chinese).
- [15] 郭新平, 汪成文, 刘华, 等. 基于扩张状态观测器的泵控电液伺服系统滑模控制[J]. 北京航空航天大学学报, 2020, 46(6): 1159-1168.
GUO X P, WANG C W, LIU H, et al. Extended-state-observer based sliding mode control for pump-controlled electro-hydraulic servo system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(6): 1159-1168 (in Chinese).
- [16] 刘华, 汪成文, 郭新平, 等. 电液负载敏感位置伺服系统自抗扰控制方法[J]. 北京航空航天大学学报, 2020, 46(11): 2131-2139.
LIU H, WANG C W, GUO X P, et al. Active disturbance rejection control method for position servo system based on electrohydraulic load sensing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(11): 2131-2139 (in Chinese).
- [17] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 减压阀 一般要求, GB/T 12244—2006 [S]. 北京: 中国标准出版社, 2007.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. General requirements for pressure reducing valves: GB/T 12244—2006 [S]. Beijing: Standards Press of China, 2007 (in Chinese).
- [18] 张正春, 卢安乐, 郑张利, 等. 一种气体用核电减压阀: CN212004443U [P]. 2020-11-24.
ZHANG Z C, LU A L, ZHENG Z L, et al. A kind of gas pressure reducer in nuclear power plant: CN212004443U [P]. 2020-11-24 (in Chinese).
- [19] 闫耀保, 赵艳培, HONGSON P. 锥型阀芯的高压气动减压阀设计分析[J]. 流体传动与控制, 2011(2): 1-5.
YAN Y B, ZHAO Y P, HONGSON P. Analysis of a pneumatic high pressure reducing valve with throttle spool[J]. Fluid Power Transmission and Control, 2011(2): 1-5 (in Chinese).
- [20] 赵大磊, 于志远, 姚晓先. 气动减压阀的静动态特性计算[J]. 战术导弹控制技术, 2005(2): 84-87.
ZHAO D L, YU Z Y, YAO X X. Computation of static and dynamic characteristics for pneumatic pressure regulator[J]. Tactical Missile Control Technology, 2005(2): 84-87 (in Chinese).
- [21] 李荣泉, 谢建蔚, 刘义. 精密气体减压阀的静态特性分析[J]. 流体传动与控制, 2014(4): 38-41.
LI R Q, XIE J W, LIU Y. Static characteristic analysis of a precision pneumatic regulator[J]. Fluid Power Transmission and Control, 2014(4): 38-41 (in Chinese).
- [22] 尤裕荣, 曾维亮. 逆向卸荷式气体减压阀的静态特性分析[J]. 火箭推进, 2005, 31(4): 15-20.
YOU Y R, ZENG W L. Analysis on static characteristics of reverse balanced pressure reducing valve[J]. Journal of Rocket Propulsion, 2005, 31(4): 15-20 (in Chinese).
- [23] 尤裕荣, 曾维亮. 逆向卸荷式气体减压阀的动态特性仿真[J]. 火箭推进, 2006, 32(3): 24-30.
YOU Y R, ZENG W L. Simulation on reverse-balanced pneumatic pressure reducing valve dynamic characteristic[J]. Journal of Rocket Propulsion, 2006, 32(3): 24-30 (in Chinese).
- [24] 罗强, 刘聪聪, 樊蕾, 等. 反向式气体减压阀静态特性仿真分析[J]. 阀门, 2015(3): 8-11.
LUO Q, LIU C C, FAN L, et al. Simulation analyses on static characteristics of reverse pressure reducing valve[J]. Valve, 2015(3): 8-11 (in Chinese).
- [25] 罗强. 反向式气体减压阀静态和动态特性研究[D]. 哈尔滨: 哈尔滨工业大学, 2010.
LUO Q. Study on static and dynamic characteristics of reverse pressure reducing valve[D]. Harbin: Harbin Institute of Technology, 2010 (in Chinese).
- [26] 顾存行, 毛虎平, 王强, 等. 基于 AMESim 的直动式减压阀动态特性仿真分析[J]. 机械设计与制造, 2017(5): 234-237.
GU C X, MAO H P, WANG Q, et al. Simulation analysis of direct-acting pressure reducing valves dynamic characteristics based on the AMESim[J]. Machinery Design & Manufacture, 2017(5): 234-237 (in Chinese).
- [27] 刘丽娇, 路波, 马云龙, 等. 国内气动减压阀产品检测结果分析[J]. 液压气动与密封, 2016, 36(2): 66-70.
LIU L J, LU B, MA Y L, et al. Test result analysis of the domestic compressed air pressure regulator products[J]. Hydraulics Pneumatics & Seals, 2016, 36(2): 66-70 (in Chinese).
- [28] 邓欣爱. 一种低成本气动减压阀的设计与验证[D]. 北京: 北京工业大学, 2017.
DENG X A. Designation and verification of a low-cost pressure reducing valve[D]. Beijing: Beijing University of Technology, 2017 (in Chinese).

Design and characteristics of reverse direct-acting high-pressure reducing valve for pneumatic actuator

LIU Yanbin¹, WANG Xuesheng^{1,*}, QIN Xinya¹, WANG Hao¹, CHEN Qinzhu¹, ZHAO Sai²

(1. Key Laboratory of Pressure Systems and Safety, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China; 2. Shanghai Spaceflight Precision Machinery Institute, Shanghai 201600, China)

Abstract: Pneumatic actuator is the key component of the flight attitude control system for air-defense missile, directly influences flying stability and attitude control ability. Missile-borne high-pressure gas provides power for pneumatic actuator through pressure reduction, which can reduce the space occupied by the pneumatic system and increase missile range by carrying more gas. For a particular type of missile, a reverse non-balanced direct-acting high-pressure reducing valve with conical clack was designed. Mathematical models of thermodynamic and static analysis were built for the valve and the design & check software were developed. The simulation models with steady and unsteady inlet pressure were built based on AMESim, and the characteristics of pressure, flow rate and spool displacement were analyzed. Results show that the pressure reducing valve has good pressure and flow characteristics under design parameters, and the theoretical calculation and simulation results agree well.

Keywords: pneumatic actuator; pressure reducing valve; design method; AMESim; characteristic analysis

Received: 2021-06-02; **Accepted:** 2021-08-29; **Published online:** 2021-09-29 09:18

URL: kns.cnki.net/kcms/detail/11.2625.V.20210927.2043.001.html

Foundation item: Industry-University-Research Cooperation Fund of the Eighth Research Institute of China Aerospace Science and Technology Corporation (SAST2020-111)

* **Corresponding author.** E-mail: wangxs@ecust.edu.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0005

高速干摩擦机械密封的端面变形及摩擦磨损

马润梅, 赵祥, 陈潇竹, 李双喜*, 杨海超

(北京化工大学 流体密封技术研究中心, 北京 100029)



摘 要: 针对机械密封在高速干摩擦状态下, 因设计不当产生端面过度变形和磨损而引起的密封失效问题, 建立了热-结构耦合数值计算模型, 分析了密封的温度场和端面变形。试验测试了静环温升, 分析了动静环端面特征, 探讨了高速干摩擦状态下的磨损机制。研究表明: 建立的有限元模型能准确地预测密封的温度和端面变形, 计算值和试验值相差小于 11%; 密封端面峰值温度对转速更敏感, 随着运转时间的延长, 温度先迅速增加后逐渐变缓; 静环易产生锥度变形, 造成端面接触压力和磨损不均匀, 静环座的“匡正”作用能够改善这类变形; 摩擦转移膜的存在状态对密封的温升、端面粗糙度起关键作用, 动环表面喷涂 Cr_2O_3 等金属氧化物, 能较好地保持致密的石墨转移膜, 减轻密封的磨损。研究结果为机械密封的设计、优化和应用提供了基础。

关键词: 干摩擦; 机械密封; 温度场; 端面变形; 摩擦磨损

中图分类号: V219; TB42

文献标志码: A

文章编号: 1001-5965(2022)07-1174-09

机械密封由于其结构灵活、性能可靠等特点, 在航空航天、化工等领域中的应用十分广泛。特殊工况下, 机械密封无润滑液体, 处于干摩擦状态, 高转速下往往面临着因温度高、端面变形大所引起的过度磨损的问题, 很大程度上限制了机械密封及整体设备的可靠性和使用寿命^[1]。

机械密封由相对旋转的动静环在补偿元件和介质压力的作用下, 相互贴合以形成端面密封。运行时端面的贴合状态和摩擦磨损情况是决定机械密封性能的关键因素。高速下机械密封易发生严重的磨损, 造成密封的失效。通过运用有限元分析软件和数值分析技术可以对机械密封的摩擦热和端面变形进行求解, 以此来反映机械密封的摩擦磨损情况^[2-3]。高斌超等^[4]建立了热力耦合计算模型, 发现在温度和压力的共同作用下端面形成收敛型泄漏间隙, 研究表明, 实际的密封面由

大量微凸体组成, 这些微凸体可以采用分形模型进行表征和模拟^[5], 并在此基础上对泄漏^[6]和磨损^[7]进行估算, 一定程度上解释了端面的磨损机理。魏龙等^[8]对端面摩擦系数采用分形理论进行表征, 结合试验分析了弹簧比压、密封压力和转速对摩擦系数的影响, 能够预测一定工况条件下的摩擦系数。王计辉等^[9]基于 ABAQUS 分析平台, 依据 W-M 分形模型将粗糙面实体化, 发现接触端面上的温度场和应力分布不均匀, 为密封环的开裂、磨损提供了研究基础。

目前关于机械密封磨损的研究比较深入, 各种数值计算、有限元分析等技术已经相对成熟, 但对于高速机械密封的温升和端面变形的理论研究和试验测量结果之间还存在较大的偏差, 干摩擦机械密封的实际磨损情况和端面状态还未得到充分的解释。本文基于热-结构耦合有限元分析方法, 建

收稿日期: 2021-01-06; 录用日期: 2021-01-29; 网络出版时间: 2021-02-19 14:38

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210219.1351.003.html

基金项目: 国家重点研发计划 (2018YFB2000800)

* 通信作者。E-mail: buclsx@126.com

引用格式: 马润梅, 赵祥, 陈潇竹, 等. 高速干摩擦机械密封的端面变形及摩擦磨损[J]. 北京航空航天大学学报, 2022, 48(7): 1174-1182. MA R M, ZHAO X, CHEN X Z, et al. End face deformation and friction and wear of high-speed dry friction mechanical seal[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1174-1182 (in Chinese).

立了机械密封数值计算模型,分析了高速干摩擦机械密封的端面温度和变形;试验测量了静环温升和端面参数,分析了干摩擦状态下的磨损机制。

1 密封结构

机械密封示意图如图 1 所示。动环与旋转轴固定并随之运转,静环安装在静环座内,在波形弹簧的作用下贴紧动环,以对磨损起到补偿作用。因密封介质为空气,动静环端面之间不存在流体膜,密封处于干摩擦状态。高速下密封的摩擦磨损是决定密封性能的关键。

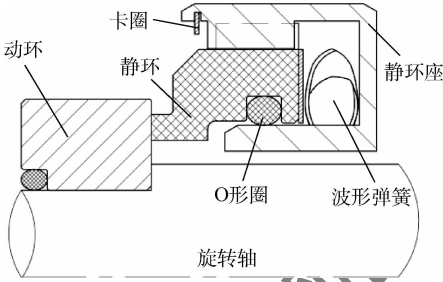


图 1 密封结构示意图

Fig.1 Schematic diagram of sealing structure

2 数值模拟

考虑到静环座对静环的变形和径向倾覆的限制,针对机械密封动静环及静环座建立分析模型,采用热-结构多物理场耦合有限元分析方法,得到动静环温度场、端面变形等结果,分析其对密封磨损的影响。模拟分析流程如图 2 所示。

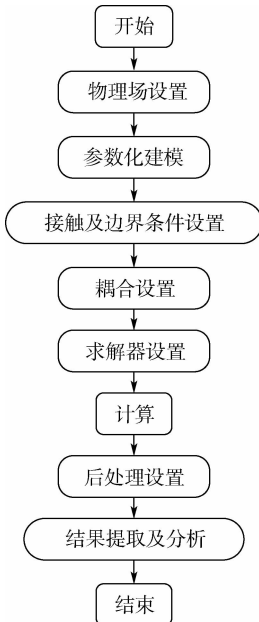


图 2 模拟分析流程

Fig.2 Flow chart of simulation analysis

2.1 几何模型

机械密封动静环为轴对称结构,为减少计算

量,采用二维轴对称模型。动静环分别建立,之间设置接触,形成装配体,如图 3 所示。各模型材料属性参数如表 1~表 3 所示。

对密封动环背侧和内径、静环座背侧和外径设置固定约束;静环内侧和动环边界 DE 为密封介质压力 P,静环背侧为弹簧作用力 F。

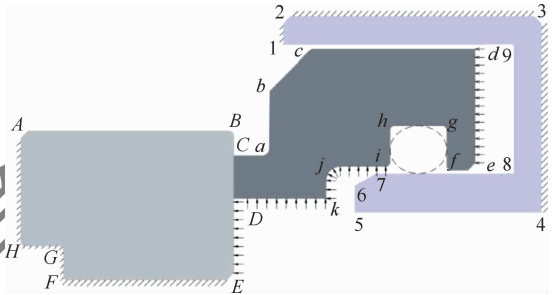


图 3 几何模型及力边界

Fig.3 Geometric model and force boundary

表 1 静环材料属性

Table 1 Stationary ring material properties

参数	数值
材料	M106D
导热系数/(W·(m·K) ⁻¹)	140
热膨胀系数/K ⁻¹	5 × 10 ⁻⁵
弹性模量/GPa	15
泊松比	0.3

表 2 动环材料属性

Table 2 Material properties of moving ring

参数	数值
材料	38CrMoAlA
导热系数/(W·(m·K) ⁻¹)	80
热膨胀系数/K ⁻¹	5.1 × 10 ⁻⁵
弹性模量/GPa	600
泊松比	0.24

表 3 静环座材料属性

Table 3 Material properties of stationary ring seat

参数	数值
材料	0Cr18Ni9
导热系数/(W·(m·K) ⁻¹)	13.8
热膨胀系数/K ⁻¹	1.654
弹性模量/GPa	204
泊松比	0.25

2.2 热边界

动环与介质接触部分设置外部强制对流,静环及静环座与介质接触部分设置弱对流。接触端面作为摩擦热源,不计流体搅拌热。默认模型具有各向同性,温度场、应力应变场符合轴对称分布。在此基础上进行热力学和结构力学之间的耦合作用。

1) 端面热源。忽略流体搅拌热后,密封端面摩擦热流密度^[10]为

$$q_m = \frac{\pi D_m n f P_c}{60} \quad (1)$$

式中： q_m 为热流密度， W/m^2 ； D_m 为接触面平均直径， m ； f 为端面摩擦系数； n 为转速， r/min ； P_c 为端面接触压力， Pa ，其按式(2)计算^[11]：

$$P_c = P_s + \Delta P(k - \lambda) \quad (2)$$

式中： P_s 为弹簧比压， Pa ， $P_s = F_s/A$ ， A 为弹簧作用面积， m^2 ； ΔP 为密封压差， Pa ； k 为载荷系数； λ 为气体反压系数，此处取 0.7。

2) 对流换热系数。动静环热边界所采用的对流换热系数的计算式^[12]为外部强制对流换热系数：

$$\alpha_1 = \frac{\beta}{D_0} 0.135 [(0.5 Re_c^2 + Re_t^2) Pr]^{0.33} \quad (3)$$

弱对流换热系数：

$$\alpha_2 = \frac{0.023\beta}{2S_s} (\varepsilon \cdot Re^{0.8} \cdot Pr^{0.4}) \quad (4)$$

式中： β 为流体的导热系数， $W/(m \cdot K)$ ； D_0 为动环外径； Re_c 反映旋转搅拌效应， Re_t 反映横向绕流； Pr 为普朗特常数； Re 为雷诺数； S_s 为静环端面外周与腔体的距离， m ； ε 为修正系数。

其他固定边界及与 O 形圈、弹簧相接触的边界简化为绝热边界。设置动静环密封端面的接触连续性。

2.3 温度场

温度场峰值出现在动静环接触面处，并向四周扩散，如图 4 所示，远离热源的温度值更低。石墨材料导热效果好，温度场梯度小，温度分布较为均匀。动环的对流换热较强，平均温度比静环低。静环座与静环之间的间隙隔绝了大部分热量，热辐射和部分弱对流传递的热量引起的静环座温升很小。温度场分布符合实际工程案例^[6]。

静环端面峰值温度 T 对转速 n 变化更加敏感，随着转速的增加，峰值温度近似线性增加，如图 5 所示。其原因在于：密封摩擦热和转速成正比(见式(1))，转速增加，对流传热带走的热量低于摩擦生热，引起温度的升高。

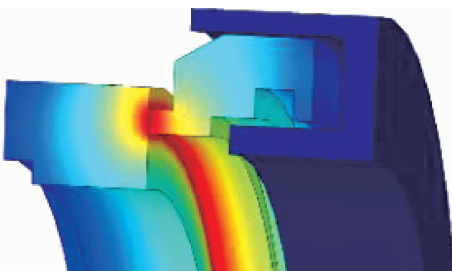


图 4 密封环温度场云图

Fig. 4 Temperature field contour of sealing ring

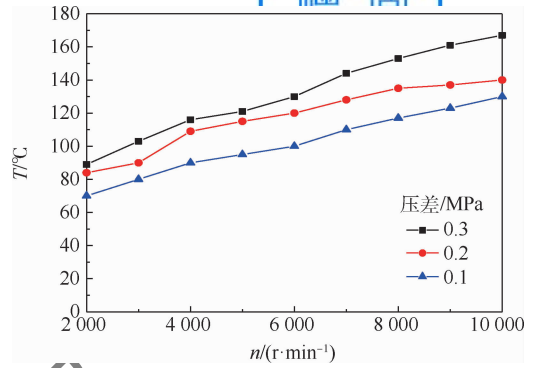


图 5 不同转速下的峰值温度

Fig. 5 Peak temperature at different rotating speeds

2.4 端面变形

图 6 为静环变形云图，端面外径变形最大，端面内径变形较小。静环外径产生一定的挤压变形，说明静环座对静环的变形有一定的矫正作用。

动静环端面的贴合状态影响着密封的应力应变和磨损。提取的静环端面变形如图 7 所示。动静环端面并未完全贴合，呈现部分脱离的状态。内径处端面脱离程度最大，沿径向向外逐渐减小，在这一点处完全贴合，并在外力的作用下导致动环产生微小的变形，形成正的位移。其原因在于：弹簧力的作用中心更靠近静环外径，静环受到以端面外径为中心的倾覆力矩，产生了一定的锥度。将静环座的影响纳入考虑因素后，静环端面整体变形幅度减小，未出现端面脱离的情况。

图 8 和图 9 为动静环端面接触压力 P_c 的分布情况。端面外径处应力最大，靠近内径压力逐渐减小，并出现压力为 0 的区域，说明此处的动静

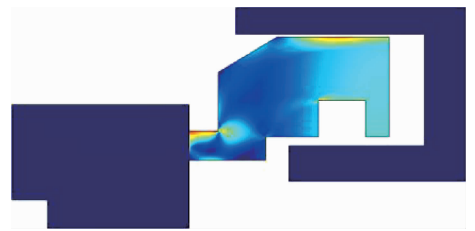


图 6 静环变形云图

Fig. 6 Stationary ring deformation contour

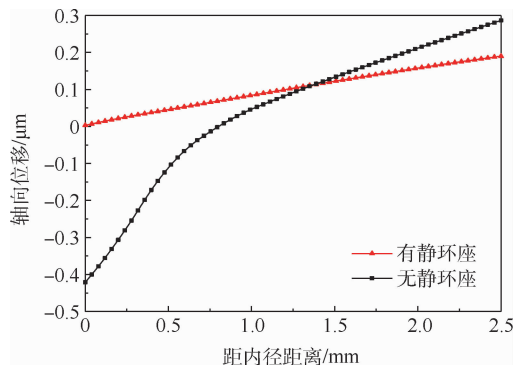


图 7 静环端面变形

Fig. 7 End face deformation of stationary ring

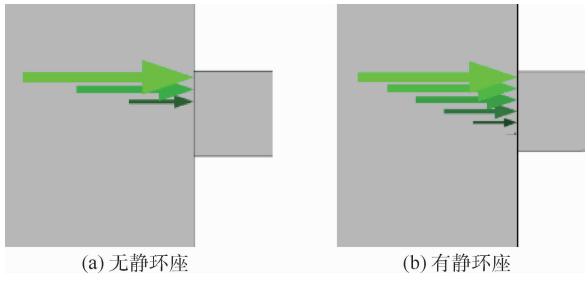


图 8 动静环端面接触压力分布

Fig. 8 Contact pressure distribution of moving and stationary rings end face

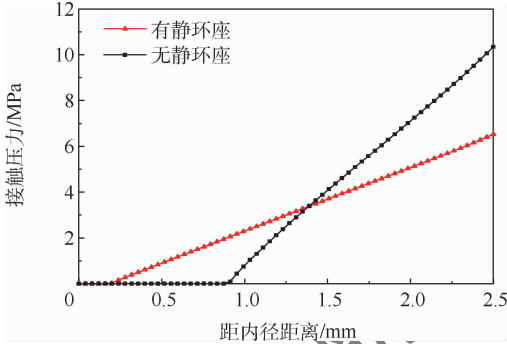


图 9 动静环端面接触压力

Fig. 9 Contact pressure of moving and stationary ring end face
环端面已经脱离。在静环座的矫正下,端面接触压力分布较为均匀,压力峰值降低。

在外力的影响下,静环端面形成内径脱离、外径贴合的“收敛型”间隙,易在外径处形成应力峰值,造成外径磨损大、内径磨损小的“畸形”端面。在进行静环结构设计时,应充分考虑外力的作用

中心,避免静环倾覆力矩的形成。静环座能够对静环的变形起到“匡正”作用,减少静环的锥度变形,提高端面的贴合程度。同时,随着接触压力峰值的降低,外径处表面微凸体的弹塑性变形减小,微凸体之间的碰撞、剪切得到改善,对减少静环磨损、提高服役寿命有一定的帮助。

3 试验

针对模拟分析的密封结构,设计了试验装置并搭建了试验台。试验获取了静环温度,测得端面的磨损情况,对比了试验前后密封环端面粗糙度、平面度和表面微观形貌的变化。

3.1 试验装置

试验系统主要由主体装置、气源系统和测量采集系统 3 大部分组成,如图 10 所示。采用 2 套机械密封背对背安装,在中部产生高压区。该系统可进行 0 ~ 15 000 r/min 转速下的试验,密封试验台如图 11 所示。

试验采用 TR200 粗糙度仪、BY-150 平面光带检测仪测量动静环密封面参数,采用 3D 光学形貌仪观测密封的摩擦磨损行为。

3.2 试件及试验内容

试件采用软硬配副,软环材料为 M106D 石墨,硬环采用 38CrMoAlA、38CrMoAlA 表面喷涂 Cr₂O₃,如图 12 所示。试验所用密封介质为室温空气,最高转速达 8 000 r/min。

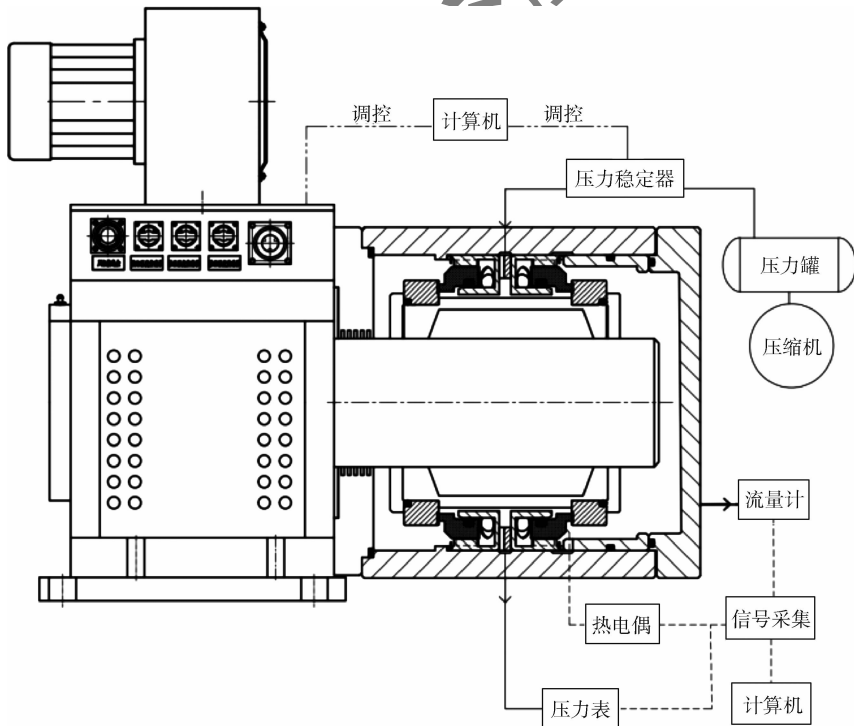


图 10 试验系统

Fig. 10 Test system

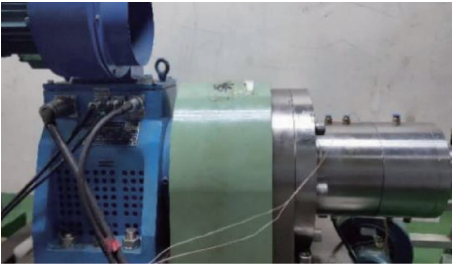


图 11 密封试验台

Fig. 11 Sealing test bench



图 13 测温点布置示意图

Fig. 13 Layout of temperature measurement points



图 12 动静环实物图

Fig. 12 Physical picture of moving and stationary ring

在不同的介质压力和转速下,采用 K 型热电偶贴在静环端面附近,测定密封静环温度并形成温升曲线,测温点布置如图 13 所示。分析试验前后动静环端面参数变化趋势,观察工况和材料对密封温度场、端面磨损状态的影响。

K 型热电偶采集的信号通过信号隔离变送器进行信号降噪,并通过 NI 数据采集卡采集转换降噪后的信号,通过编写的 Labview 程序处理,将输入信号处理转换为电压值与温度并进行实时记录,最终将 3 个测温点的数值取平均值进行输出。

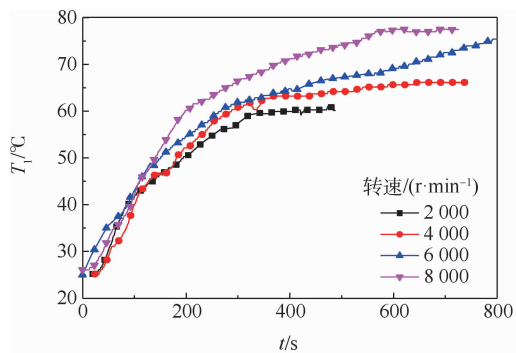
3.3 温度分析

不同压差 ΔP 下,试验温度 T_1 随时间 t 的变化如图 14 所示。可以看出,随着运转时间的延长,温度先迅速增加,后逐渐变缓,高转速下密封温度达到平衡用时更长。温度升高引起密封环和介质温度之间的差异逐渐增大,促使对流传热效果增加而达到产热和散热的逐渐平衡,升温也逐渐变缓。转速对温升的前期影响较小,不同转速下温升速率差异不大。后期低转速下温度逐渐变缓达到平衡,高转速的产热大,温度继续增加,达到更高的平衡温度。

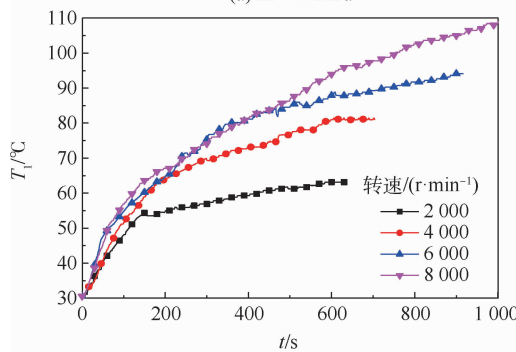
介质压差 ΔP 增加,密封达到的平衡温度 T_1 更高,如图 15 所示。研究表明,动静环的接触实际上是表面微凸体之间的接触。在介质压力和弹簧力的作用下,动静环端面产生一定的接触载荷,表面微凸体产生弹塑性变形,运转时将引起大量的挤压、破裂等。这种挤压变形将随着介质压力的增大而加剧,从而将更多的机械能转化成摩擦热,引起密封温度的升高。

试验测量了在相同条件下,2 种动环材料对静环端面温度 T_1 的影响。如图 16 所示,可以看到,相比于表面未喷涂 Cr_2O_3 的动环,表面喷涂可以达到较低的平衡温度。并且后者在试验过程中,温度变化趋势比较平缓,未出现大的波动。表示此时动静环之间的摩擦磨损既比较轻微,又较为稳定。相反,前者的温度波动则较剧烈,这与动静环表面第三体的形成和破裂有关,将在下文进行解释。

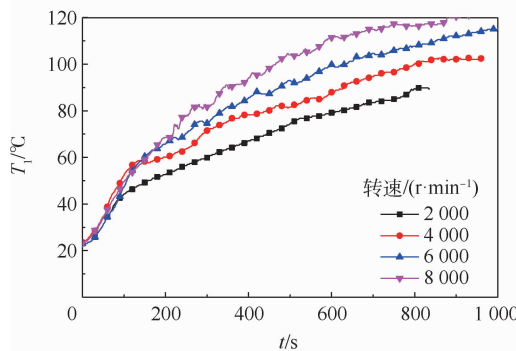
将模拟结果和试验结果对比分析可见(见图 17),两者之间的差异小于 11%,证明了有限元分析模型能够比较准确地反映实际温度。出现误差的主要原因有:①试验时热电偶测量的并非端



(a) $\Delta P=0$ MPa



(b) $\Delta P=0.1$ MPa



(c) $\Delta P=0.2$ MPa

图 14 转速对温度的影响

Fig. 14 Effect of rotating speed on temperature

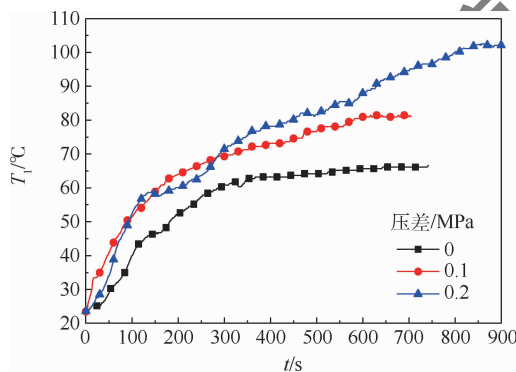


图 15 压差对温度的影响

Fig. 15 Effect of differential pressure on temperature

看出,试验前的静环端面为平行光带,表示此时的平面度很好,在 $1 \mu\text{m}$ 以内。图 18(b)为试验后的静环端面,图中的光带呈指向同一方向的“箭头”状,表示平面度变差且端面内外侧不一致。采用千分表测得静环端面内外径的高度差,如图 19 所示,外径较内径偏低 $3 \mu\text{m}$ 左右,说明外径处磨损较大,反映出外径处受到更大的接触比压,这与计算得到的端面接触压力的分布相一致,侧面证明了模拟分析的合理性。

2) 端面磨损。试验采用 TR200 粗糙度仪测得 2 种动环试验前后的表面粗糙度 R_a ,如图 20 所

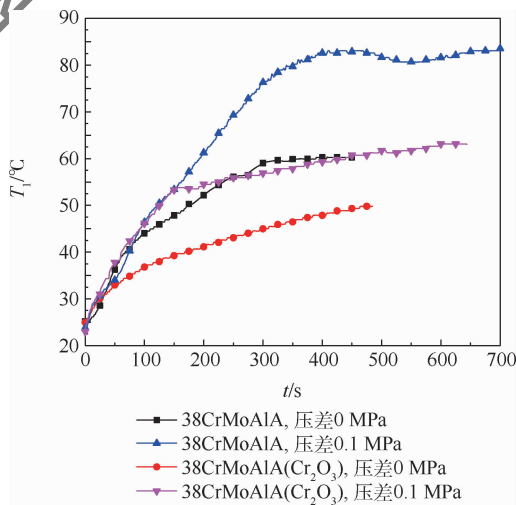


图 16 动环材料对温度的影响

Fig. 16 Effect of moving ring material on temperature

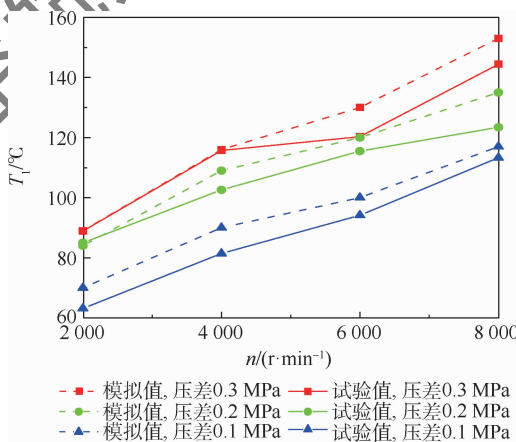


图 17 模拟与试验结果对比

Fig. 17 Comparisons of simulation and test results

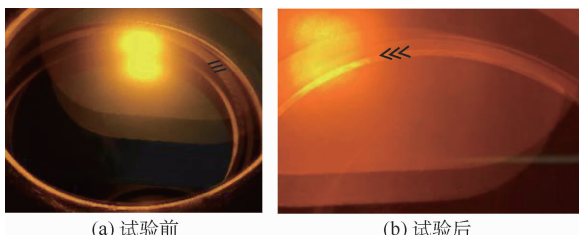


图 18 静环平面度对比

Fig. 18 Flatness comparison of stationary ring

面峰值温度,测量值偏低;②试验未达到最终的平衡温度。

3.4 端面磨损分析

1) 端面变形。试验对运转前后的静环端面进行平面度检测,如图 18 所示。从图 18(a)可以



图 19 端面内外径高度差测量

Fig. 19 Height difference between inner and outer diameter of end face

示。可以看出,试验后 2 种动环表面的轮廓曲线波动较大,波峰波谷差值增加,表面变得更为粗糙。喷涂 Cr_2O_3 的动环表面粗糙度值由 0.020 增加到 0.114,未喷涂的动环表面粗糙度值由 0.029 增加到 0.357。可见前者在运转过程中能够保持较低的粗糙度。在干摩擦条件下,低的粗糙度对应的摩擦系数较小,因此摩擦产热少,达到的平衡温度也较低。

试验前后动静环端面的微观形貌如图 21 所示。静环材料为浸金属钨的石墨,经过抛光后较硬的金属磨损少,在软基体石墨的“支撑作用”下呈凸起状,起到保护基体的“阴影效应”。石墨材料为层状结构,具有良好的自润滑性能。磨损之后,大量石墨粉堆积在静环表面,金属凸起逐渐消失,并且剥落的大金属颗粒在石墨表面形成切削和犁沟,具有典型的磨粒磨损特征。

从试验后动环表面微观形貌可以看出,2 种材料的动环表面均残留一定量的石墨薄层,形成一定厚度的摩擦转移膜。外力作用下,动静环微凸体相互接触,较软的石墨胶合在金属表面,滑移时石墨被撕脱,形成剥落的块状磨屑,具有黏着磨损的特征,并在高速摩擦、挤压的作用下形成致密的第三体^[14],起到一定的润滑减摩、速度协调的作用^[15]。

相较于未喷涂 Cr_2O_3 的动环,喷涂材料可以在基体表面形成硬度更高、含有一定比例气孔的硬质相。这种硬质相,一方面在运行过程中产生的变形小,致密的转移膜贴合在表面不易片状剥

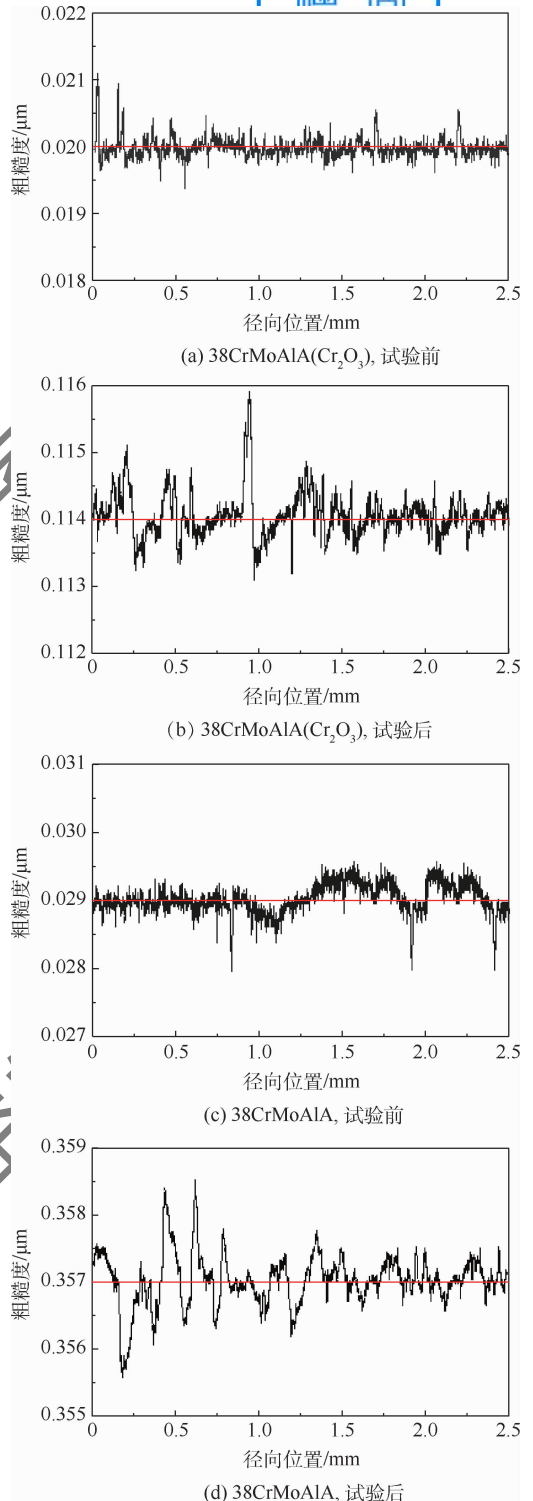


图 20 动环表面粗糙度

Fig. 20 Surface roughness of moving ring

落;另一方面,气孔的存在为摩擦转移膜提供了良好的“着床”,石墨粉渗入到气孔中可以明显提高转移膜在动环表面的“附着力”,在运行过程中保持性较好。转移膜的存在状态对静环的温升起到关键作用,运行过程中未喷涂的动环表面摩擦转移膜不连续,存在“形成”、“破碎”的交替过程,造成摩擦系数的波动,从而反映在温升上。喷涂 Cr_2O_3 的动环表面摩擦转移膜连续而又平整,运行

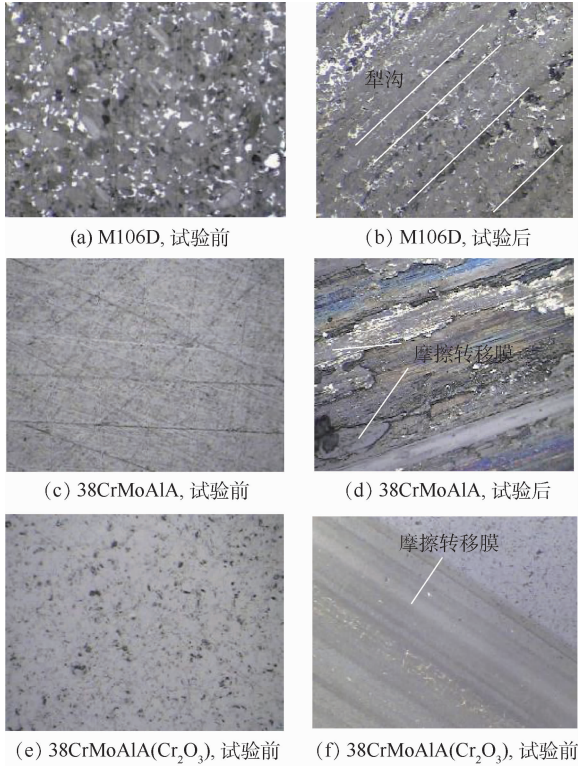


图21 试验前后端面微观形貌

Fig.21 Micro morphology of end face before and after test

过程中能保持低而稳定的摩擦系数,温升也比较平稳。

4 结论

通过有限元计算和试验,分析了干摩擦机械密封的温升、端面变形及磨损,得到以下结论:

1) 采用的热-结构耦合有限元数值计算模型能较好地预测密封的温升和变形,可为密封环结构的优化设计提供技术手段。

2) 静环在外力作用下产生锥度变形,形成“收敛型”间隙,造成端面接触压力分布不均匀,内外径磨损不一致。采用静环在内、静环座在外的结构,并控制静环外径和静环座之间的间隙可以发挥静环座的“匡正”作用,改善此类变形。

3) 机械密封在干摩擦条件下具有磨粒磨损和黏着磨损的特征。摩擦转移膜的存在状态对运行中密封的温升、端面粗糙度等的变化起关键作用。高硬度材料的变形小,转移膜不易片状剥落,动环材料表面强化有利于保持完整连续的摩擦转移膜。

4) 动环表面喷涂 Cr_2O_3 等金属氧化物可为摩擦转移膜提供良好的“着床”,能较好地保持致密的第三体膜,有助于减轻密封的摩擦磨损,延长使用寿命。

参考文献 (References)

- [1] OJILE J O, TEIXEIRA J A, CAR-MODY C. Mechanical seal failure analysis[J]. Tribology Transactions, 2010, 53(4): 630-635.
- [2] 赵伟刚,张鹏鹏,任姗姗,等. 液体火箭发动机涡轮泵机械密封磨损机理研究[J]. 火箭推进, 2017, 43(3): 10-16. ZHAO W G, ZHANG P P, REN S S, et al. Research on wear mechanism of mechanical seal for turbopump in liquid rocket engine[J]. Journal of Rocket Propulsion, 2017, 43(3): 10-16 (in Chinese).
- [3] 刘进祥,穆塔里夫·阿赫迈德. 机械密封摩擦副界面温升分析[J]. 机床与液压, 2020, 48(14): 138-141. LIU J X, MUTELLIP A. Analysis of interface temperature rise of friction pairs of mechanical seals[J]. Machine Tool & Hydraulics, 2020, 48(14): 138-141 (in Chinese).
- [4] 高斌超,孟祥铠,李纪云,等. 机械密封热力耦合有限元模型与密封性能分析[J]. 摩擦学学报, 2015, 35(5): 550-556. GAO B C, MENG X K, LI J Y, et al. Thermal-mechanical coupled finite element model and seal performance analysis of mechanical seals[J]. Tribology, 2015, 35(5): 550-556 (in Chinese).
- [5] 李小彭,杨泽敏,潘五九,等. 接触式机械密封端面的分形磨损模型[J]. 振动、测试与诊断, 2020, 40(5): 841-846. LI X P, YANG Z M, PAN W J, et al. Fractal wear model of contact mechanical seal[J]. Journal of Vibration, Measurement & Diagnosis, 2020, 40(5): 841-846 (in Chinese).
- [6] 孙见君,顾伯勤,魏龙. 基于分形理论的接触式机械密封泄漏模型[J]. 化工学报, 2006, 57(7): 1626-1631. SUN J J, GU B Q, WEI L. Leakage model of contacting mechanical seal based on fractal geometry theory[J]. CIESC Journal, 2006, 57(7): 1626-1631 (in Chinese).
- [7] 魏龙,张鹏高,房桂芳. 机械密封端面混合摩擦热计算分形模型[J]. 液压与气动, 2020(7): 112-117. WEI L, ZHANG P G, FANG G F. Calculation fractal model of mixed friction heat between the end faces for mechanical seals[J]. Chinese Hydraulics & Pneumatics, 2020(7): 112-117 (in Chinese).
- [8] 魏龙,张鹏高,刘其和,等. 接触式机械密封端面摩擦系数影响因素分析与试验[J]. 摩擦学学报, 2016, 36(3): 354-361. WEI L, ZHANG P G, LIU Q H, et al. Influencing factors analysis and experiments of friction coefficient between the end faces for contact mechanical seals[J]. Tribology, 2016, 36(3): 354-361 (in Chinese).
- [9] 王计辉,陈志,顾灿鸿,等. 机械密封在干摩擦状态下的摩擦界面热力耦合分析[J]. 摩擦学学报, 2019, 39(6): 737-745. WANG J H, CHEN Z, GU C H, et al. Thermo-mechanical coupling analysis of friction interface of mechanical seals under dry friction[J]. Tribology, 2019, 39(6): 737-745 (in Chinese).
- [10] 何永明,穆塔里夫·阿赫迈德,刘毅龙. 油泵用机械密封摩擦副界面热-结构耦合分析[J]. 流体机械, 2014, 42(6): 21-25. HE Y M, MUTELLIP A, LIU Y L. Thermal-structure coupling analysis of the friction pair interface for heat pump mechanical

- seal[J]. Fluid Machinery, 2014, 42(6):21-25 (in Chinese).
- [11] 蔡仁良, 顾伯勤, 宋鹏云. 过程装备密封技术[M]. 2版. 北京: 化学工业出版社, 2006:149.
CAI R L, GU B Q, SONG P Y. Sealing technology of process equipment[M]. 2nd ed. Beijing: Chemical Industry Press, 2006:149 (in Chinese).
- [12] 彭旭东, 顾永泉. 不同相态机械密封的性能计算[J]. 流体机械, 1994, 22(8):20-24.
PENG X D, GU Y Q. Performance calculation of mechanical seals with different phases[J]. Fluid Machinery, 1994, 22(8):20-24 (in Chinese).
- [13] 王永乐, 刘杰, 李凤成, 等. 基于 COMSOL 的高温双端面机械密封热力耦合分析[J]. 流体机械, 2019, 47(3):26-30.
WANG Y L, LIU J, LI F C, et al. Analysis of thermal-mechanical coupling of double mechanical seals under high temperature based on COMSOL software[J]. Fluid Machinery, 2019, 47(3):26-30 (in Chinese).
- [14] 韩晓明, 高飞, 宋宝韞, 等. 摩擦速度对铜基摩擦材料摩擦磨损性能影响[J]. 摩擦学学报, 2009, 29(1):89-96.
HAN X M, GAO F, SONG B Y, et al. Effect of friction speed on friction and wear performance of Cu-matrix friction materials[J]. Tribology, 2009, 29(1):89-96 (in Chinese).
- [15] 韩晓明, 高飞, 符蓉, 等. 三体摩擦体系中材料摩擦特性的研究进展[J]. 中国材料进展, 2009, 28(2):8-13.
HAN X M, GAO F, FU R, et al. Recent progress in tribological theories involving the behavior of the third body[J]. Materials China, 2009, 28(2):8-13 (in Chinese).

End face deformation and friction and wear of high-speed dry friction mechanical seal

MA Runmei, ZHAO Xiang, CHEN Xiaozhu, LI Shuangxi*, YANG Haichao

(Research Center of Fluid Sealing Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: In order to solve the problem of seal failure caused by excessive deformation and wear of mechanical seal end face due to improper design under high-speed dry friction conditions, a thermal structure coupled numerical calculation model was established to analyze the temperature field and end face deformation of mechanical seal. The temperature rise of the stationary ring was tested, the characteristics of the end face of the stationary ring were analyzed, and the wear mechanism under high speed dry friction was discussed. The results show that: the established finite element model can accurately predict the temperature and end face deformation of the seal, and the difference between the calculated value and the experimental value is less than 11%; the peak temperature of the seal end face is more sensitive to the rotating speed, and with the extension of the operating time, the temperature first increases rapidly and then gradually slows down; the static ring is prone to taper deformation, resulting in the uneven contact pressure and wear of the end face, and the "correction" effect of the stationary ring seat can improve this kind of deformation; the existence of friction transfer film plays a key role in the temperature rise and surface roughness of the seal. The moving ring surface is sprayed with Cr₂O₃ metal oxides, which can better maintain the dense graphite transfer film and reduce the wear of the seal. The research results provide a basis for the design, optimization and application of mechanical seals.

Keywords: dry friction; mechanical seal; temperature field; end face deformation; friction and wear

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

DOI: 10.13700/j.bh.1001-5965.2021.0026

抽吸气流量对催化惰化系统性能影响

王晨臣^{1,*}, 潘俊¹, 王洋洋¹, 段伟杰²

(1. 航空工业南京机电液压工程研究中心 航空机电系统综合航空科技重点实验室, 南京 211106;

2. 航空工业直升机设计研究所, 天津 333001)

摘 要: 基于耗氧型惰化系统惰化原理, 建立了绿色低温催化惰化系统(3CGIS)的AMESim 仿真模型, 研究了绿色低温催化惰化系统抽吸气流量对惰化时间的影响, 以及飞行包线内燃油箱气相空间氧气体积分数变化。将计算结果与试验数据进行对比, 结果表明, 飞行包线内燃油箱气相空间氧气体积分数计算结果与试验结果基本一致, 验证了仿真模型的正确性。在此基础上, 得到抽吸气流量与惰化时间近似呈反比关系; 当惰化时间一定时, 抽吸气流量随载油率的降低而增加; 针对下降阶段燃油箱气相空间氧气体积分数可能超过 12%, 提出一种双流量惰化模式设计方法, 可保证氧气体积分数在整个飞行包线内低于 12%。仿真结果为绿色低温催化惰化系统的设计与优化提供了依据。

关键词: 耗氧惰化; 低温催化; 抽吸气流量; 惰化时间; 氧气体积分数; AMESim 仿真

中图分类号: V37; V228

文献标志码: A

文章编号: 1001-5965(2022)07-1183-07

飞机燃油系统的起火爆炸是导致各类飞机失事的主要原因之一^[1-2]。1998年, 美国航空规章咨询委员会(ARAC)在回顾近40年来全世界运输类飞机营运记录后发现, 发生过16起飞机燃油箱爆炸事故, 造成了多达530余人的死亡^[3]。

采用中空纤维膜制取富氮气体的机载制氮惰化系统, 由于其经济高效等优势, 成为了目前应用最为广泛的飞机燃油箱惰化系统^[4-5]。不过从近年的应用现状来看, 该惰化系统仍存在较多问题, 如需要从发动机引气导致飞机代偿损失大, 系统入口需求压力较高导致使用场合受到限制, 膜丝易堵塞及对臭氧敏感导致膜性能严重衰减, 吹扫到外界的燃油蒸气污染环境等^[6-8]。

鉴于此, 研究者们开始进行新一代机载惰化技术的研究工作, 其中基于低温催化反应的机载绿色惰化系统(green on-board inert gas generation system, GOBIGGS)由于无需从发动机引气、质量

轻、无污染等优势得到了广泛关注^[9-10]。其基本原理为: 通过抽取燃油箱气相空间气体, 在催化反应器内进行低温无焰的催化反应, 消耗其中的氧气并生成大量二氧化碳和水蒸气, 将反应后的混合气体通入燃油箱以降低燃油箱气相空间氧气体积分数, 从而实现燃油箱惰化^[11-12]。

国外自2006年起就开始了新型燃油箱耗氧型惰化系统的相关研究工作, 如美国怀特-帕特森空军基地、Honeywell和Phyre公司, 其中Phyre公司已完成样机并进行试验测试。研究表明, GOBIGGS系统可迅速将燃油箱气相空间氧气体积分数降低至9%以下, 且有望进一步将催化反应换热器和反应后气体换热器的质量尺寸缩小至现有样机的1/40^[13]。相较之下, 中国对耗氧型惰化系统的研究还处于理论研究系统与系统架构权衡阶段。例如, 冯诗愚、邵垒等^[14-16]对含有外界补气和回热器的耗氧型惰化系统进行了惰化性能计算

收稿日期: 2021-01-16; 录用日期: 2021-04-24; 网络出版时间: 2021-05-07 08:59

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210506.1821.005.html

*通信作者。E-mail: 583812642@qq.com

引用格式: 王晨臣, 潘俊, 王洋洋, 等. 抽吸气流量对催化惰化系统性能影响[J]. 北京航空航天大学学报, 2022, 48(7): 1183-1189. WANG C C, PAN J, WANG Y Y, et al. Effect of suction flow rate on performance of catalytic inerting system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1183-1189 (in Chinese).

研究。彭孝天等^[17]对不包含补气和回热器的耗氧型惰化系统进行了研究,分析了环境温度对系统中“水”性能的影响。谢辉辉等^[18]对耗氧型惰化系统核心部件催化反应器工作性能进行了理论研究;冯诗愚等^[19]还通过加入传质源项的 CFD 仿真试验分析了加热和压力降低对燃油溶解氧逸出使得燃油箱气相空间氧气体积分数升高的影响。王苏明等^[20]通过 CFD 仿真对比了耗氧型惰化系统和中空纤维膜惰化系统惰化性能差异。

耗氧型惰化系统设计中一个关键参数在于系统抽吸气流量的确定,不仅决定了系统的惰化能力和地面初始惰化时间,还关系到系统各部件之间参数性能匹配,因此,了解抽吸气流量对耗氧型惰化系统惰化性能的影响,对耗氧型惰化系统抽吸气流量的设计至关重要。鉴于此,本文基于耗氧型惰化系统工作原理建立耗氧型惰化系统 AMESim 仿真模型,计算分析抽吸气流量对惰化性能的影响,

研究抽吸气流量对地面初始惰化时间、飞行包线下燃油箱上部氧气体积分数的影响,据此得到耗氧型惰化系统抽吸气流量的设计方法。

1 惰化系统原理及假设

本文依据耗氧型惰化系统工作原理提出了一种绿色低温催化惰化系统(low temperature controllable oxygen consumed catalytic green inerting system, 3CGIS),惰化系统流程如图 1 所示。该系统通过抽气风机抽取燃油箱内气相空间混合气体,送入催化反应器内预热后进行低温催化反应(催化反应温度低于 200℃),消耗其中的燃油蒸气和氧气并生成大量二氧化碳和水蒸气,催化反应后的混合气体经换热器冷却降温及水分离器除水净化后,经过合理的充填方式通入燃油箱以降低燃油箱气相空间氧气体积分数,从而达到燃油箱惰化的目的。

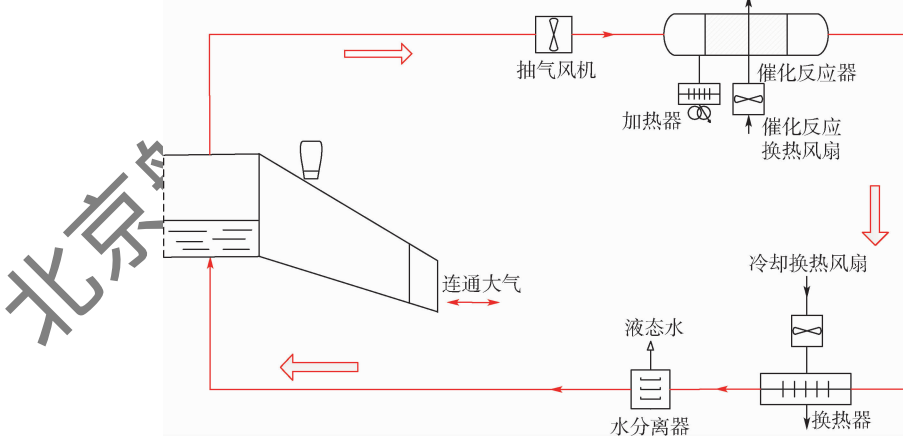


图 1 3CGIS 惰化系统流程

Fig. 1 Flowchart of 3CGIS inerting system

做以下 7 个基本假设:

1) 燃油蒸气由复杂的碳氢化合物组成,本文认为其分子式统一表示为 C_aH_b ,对于国产 RP-3 燃油,其分子式为 $C_{10.05}H_{20.42}$,不考虑燃油蒸气中存在的氧、硫和氮等微量元素。

2) 外界环境气体由氮气、氧气组成,忽略大气中及燃油中溶解的水蒸气;燃油箱抽吸气的燃油蒸气浓度在整个惰化过程保持不变。

3) 忽略燃油箱结构影响,认为燃油箱内气体温度、压力、组分等均均匀分布。

4) 不考虑燃油箱内气体的溶解逸出,燃油箱气相空间初始氧气体积分数为 21%。

5) 催化反应器反应放热由催化反应换热风扇带走,催化反应器在整个惰化过程中不会飞温。

6) 催化反应器出口换热器的冷却能力足够

将反应器出口混合气体冷却至 30℃ 以下。

7) 当进入催化反应器的混合气中氧气充足时,催化反应效率认为达到 100%;当氧气不充足时,按催化效率计算。

2 3CGIS 惰化系统仿真建模

2.1 3CGIS 惰化系统模型

AMESim 平台实现了多学科领域的系统过程建模和仿真,基于图形化的系统仿真建模环境,可使用户从繁琐的数学建模中解放,更专注于物理系统本身的设计。因此,本文通过 AMESim 仿真平台与第 1 节所建立的系统流程和假设搭建 3CGIS 惰化系统仿真模型,如图 2 所示。该仿真模型左侧可对油料和气体组分、飞机姿态信息、外界大气环境等进行定义;燃油箱有 4 个接口,分别为发动机燃油消耗、连通外接大气、风机抽吸气和

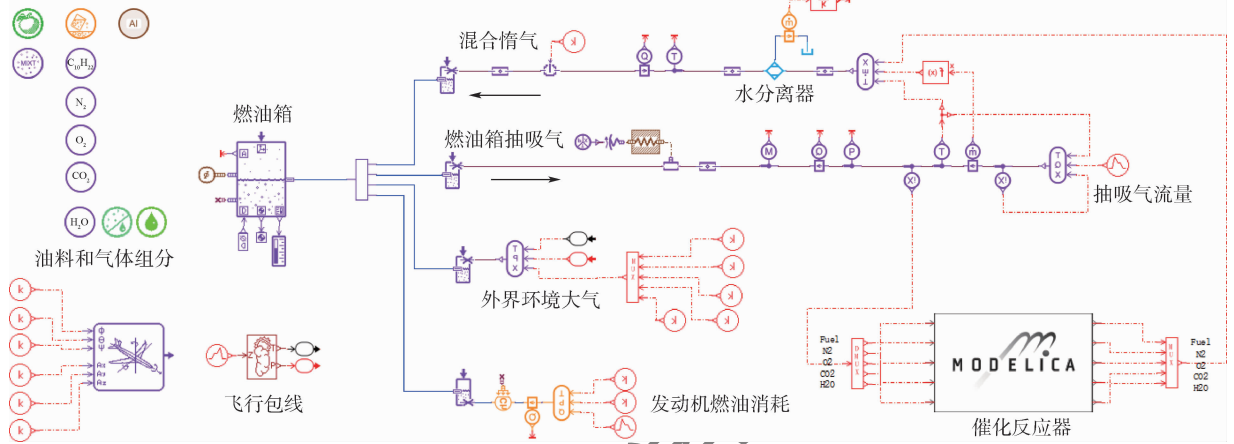


图 2 3CGIS 惰化系统 AMESim 仿真模型

Fig. 2 AMESim simulation model of 3CGIS inerting system

混合惰气回填; 催化反应器使用 Modelica 模块实现, 仿真模型中加入多个传感器以监测各个位置的状态信息。

2.2 催化反应器模型

根据模型假设, 催化反应器内催化反应方程可近似认为



当抽吸气内氧气体积摩尔浓度充足, 即抽吸气内 $(a + b/4)M_{O_2} \geq M_{C_aH_b}$ 时, 认为催化反应器内催化反应完全进行, 抽吸气内包含的燃油蒸气完全转换为二氧化碳和水; 当抽吸气内氧气体积摩尔浓度不充足, 即抽吸气内 $(a + b/4)M_{O_2} < M_{C_aH_b}$ 时, 认为催化反应器内催化反应不完全进行, 催化反应按催化效率进行计算。

催化反应器催化效率 η_{cat} 为

$$\eta_{cat} = \frac{M_{C_aH_b,r}}{M_{C_aH_b}} \quad (2)$$

式中: M_{O_2} 为进入催化反应器的氧气体积摩尔浓度; $M_{C_aH_b}$ 为进入催化反应器的燃油蒸气摩尔浓度; $M_{C_aH_b,r}$ 为催化反应器内催化反应消耗的燃油蒸气体积摩尔浓度。

3 仿真计算及验证

3.1 仿真模型验证

为验证所搭建的 3CGIS 惰化系统 AMESim 仿真模型的正确性, 采用该模型对 Phyer 公司在 FAA 大西洋城市测试中心进行技术验证的测试条件进行计算, FAA 模拟环境测试选用了短程的 FL-350 飞行包线, 其总飞行时间约为 105 min, 所惰化的燃油箱体积为 0.481 m^3 , 装载有 64 L 的 JP-8 燃油, 在爬升、降落和海平面滑行阶段抽吸流量为 56.6 L/min , 在巡航阶段吸气流量为 28.3 L/min ,

将计算结果与试验数据进行对比, 如图 3 所示^[21]。结果显示, 本文仿真计算结果与试验数据变化趋势基本一致。其中, 初始阶段惰化系统开始工作后, 燃油箱气相空间氧气体积分数迅速减小, 在爬升和巡航阶段始终保持极低的氧气体积分数, 开始下降后由于外界环境气体大量涌入导致氧气体积分数迅速上升, 随后稳定在某一范围, 且由于外界环境压力变化导致氧气体积分数不断波动, 全飞行包线内氧气体积分数最大差别不超过 1.5%, 说明仿真模型具有较高的准确性。

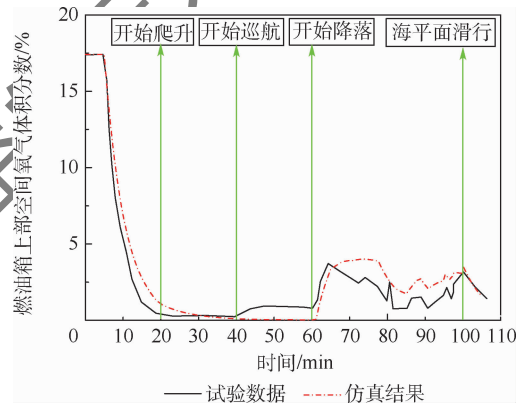


图 3 仿真结果与试验数据对比

Fig. 3 Comparison of simulation results and test data

3.2 抽吸气流量对地面初始惰化时间影响

选择某体积为 46 m^3 的中央翼燃油箱, 基于第 2 节所搭建的 3CGIS 惰化系统 AMESim 仿真模型, 在初始载油率为 0%、50% 和 97% 时, 计算惰化系统抽吸气流量对地面初始惰化时间的影响。

对于民机而言, 当燃油箱气相空间氧气体积分数低于 12% 时, 即认为燃油箱是惰化的, 因此, 本文将惰化系统启动至燃油箱气相空间氧气体积分数降低至 12% 所需时间视为惰化时间^[22]。在不同抽吸气流量下计算得到相应的惰化时间, 对计算结果进行线性拟合, 得到惰化时间随抽吸气流量变

化的关系,如图4~图6所示,其中惰化时间为5、10、15和20 min所需要的抽吸气流速如表1所示。

根据计算结果可知,当载油率一定时,惰化系统抽吸气流速与惰化时间近似呈反比关系,随着抽吸气流速的增加,惰化时间迅速减少,然后变化趋势趋于平缓;相同惰化时间下,燃油箱载油率越低,上部气相空间体积也就越大,所需要的抽吸气流速也越大。

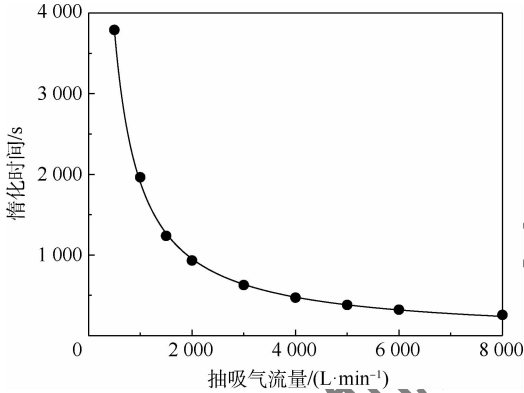


图4 载油率0%下惰化时间随抽吸气流速变化
Fig.4 Inerting time varies with suction flow rate at 0% fuel load

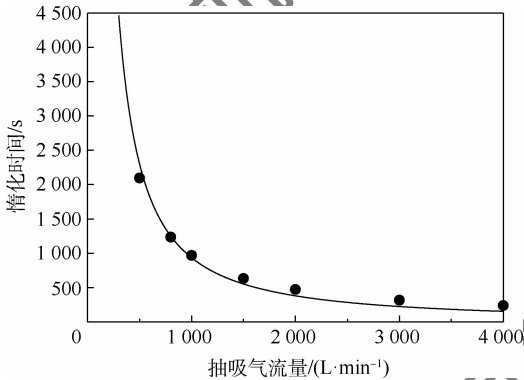


图5 载油率50%下惰化时间随抽吸气流速变化
Fig.5 Inerting time varies with suction flow rate at 50% fuel load

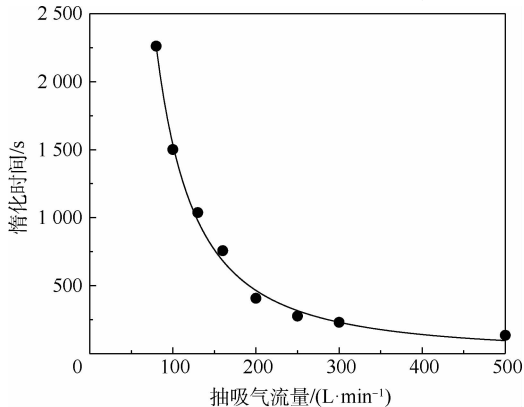


图6 载油率97%下惰化时间随抽吸气流速变化
Fig.6 Inerting time varies with suction flow rate at 97% fuel load

表1 不同惰化时间所需抽吸气流速

Table 1 Required suction flow rate under different inerting time

惰化时间/min	抽吸气流速/(L·min ⁻¹)		
	载油率0%	载油率50%	载油率97%
5	6 160	2 400	256
10	3 100	1 400	172
15	2 070	1 030	137
20	1 550	826	115

3.3 飞行包线下燃油箱气相空间氧气体积分数计算

对体积为46 m³的中央翼燃油箱选择惰化时间10 min和20 min所对应的抽吸气流速,对某飞行包线下燃油箱气相空间氧气体积分数进行计算,飞行包线包括滑出、起飞、爬升、巡航、下降、进场和滑入7个阶段,总耗时825 min,如表2所示,在该飞行包线下燃油箱气相空间氧气体积分数随时间变化如图7和图8所示。

根据计算结果可知,在下降阶段,由于燃油箱外界环境压力的迅速升高,外界空气会迅速且大量涌入燃油箱,造成燃油箱气相空间氧气体积分数升高,当抽吸气流速越低时,燃油箱气相空间氧

表2 飞行包线信息

Table 2 Flight envelope information

状态	轮挡时间/min	飞行高度/m	轮挡耗油/kg
滑出	7	0	189
起飞	2	0~457	630
爬升	29	457~12 000	4 992
巡航	755	12 000	67 987
下降	21	12 000~457	373
进场	6	457	240
滑入	5	0	135

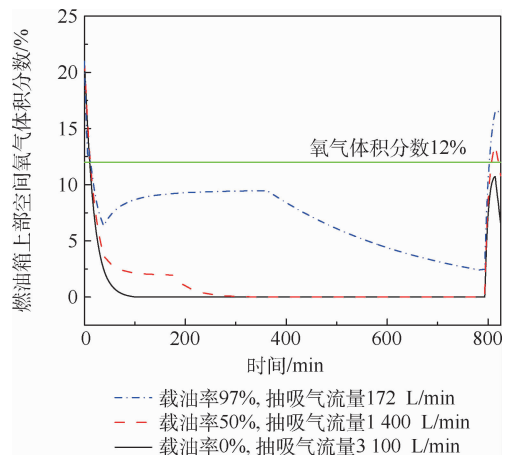


图7 惰化时间10 min对应飞行包线下氧气体积分数变化
Fig.7 Variation of oxygen volume fraction in flight envelope at 10 min inerting time

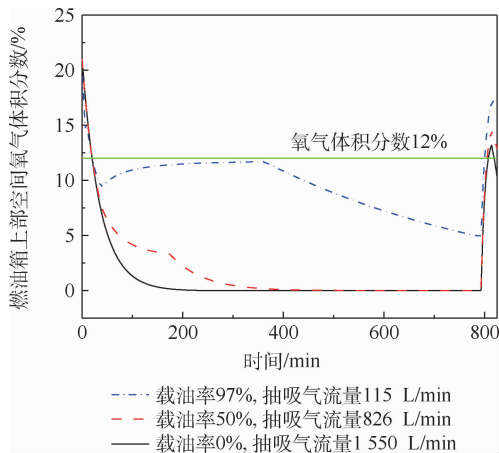


图 8 惰化时间 20 min 对应飞行包线下氧气体积分数变化

Fig. 8 Variation of oxygen volume fraction in flight envelope at 20 min inerting time

气体积分数上升程度越高。除抽吸气流量 3 100 L/min 外, 其余抽吸气流量在下降阶段燃油箱气相空间氧气体积分数均会超过 12%, 因此, 有必要在下降阶段根据惰化需要采取另一个抽吸气流量, 以保证燃油箱气相空间氧气体积分数在整个飞行包线内始终低于 12%。

为此, 本文提出一种适用于 3CGIS 惰化系统抽吸气流量确定的设计方法, 即双流量惰化模式, 该模式在滑出、起飞、爬升、巡航阶段采用初始状态抽吸气流量, 在下降、进场、滑入阶段采用下降状态抽吸气流量。初始状态的抽吸气流量根据载油率、惰化时间来确定; 下降状态的抽吸气流量与下降阶段开始时刻燃油箱气相空间体积大小、氧气体积分数和下降速率等因素有关。先根据经验取一个抽吸气流量估值, 代入 AMESim 仿真模型计算, 若下降阶段氧气体积分数超过 12%, 则适当增加下降阶段抽吸气流量, 再次计算直至氧气体积分数维持在 12%, 得到下降阶段抽吸气流量。

根据该设计方法, 得到不同惰化时间和载油率下双流量惰化模式对应的抽吸气流量, 如表 3 所示。在该抽吸气流量下计算得到飞行包线内燃油箱气相空间氧气体积分数变化关系如图 9 和图 10 所示。

表 3 双流量惰化模式下抽吸气流量

Table 3 Suction flow rate in dual-flow inerting mode

惰化时间/ min	载油率/ %	抽吸气流量/(L·min ⁻¹)	
		滑出、起飞、爬升、 巡航阶段	下降、进场、 滑入阶段
10	0	3 100	2 280
10	50	1 400	2 280
10	97	172	2 450
20	0	1 550	2 280
20	50	826	2 280
20	97	115	2 550

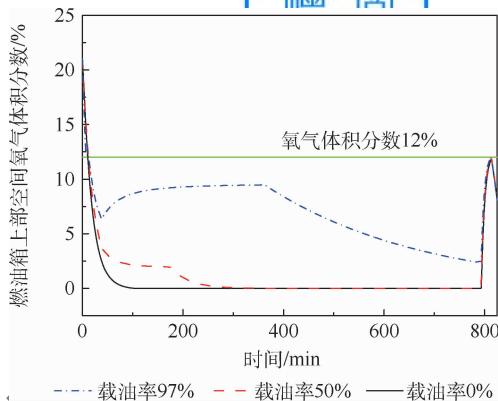


图 9 惰化时间 10 min 双流量模式下氧气体积分数变化

Fig. 9 Variation of oxygen volume fraction in dual-flow inerting mode at 10 min inerting time

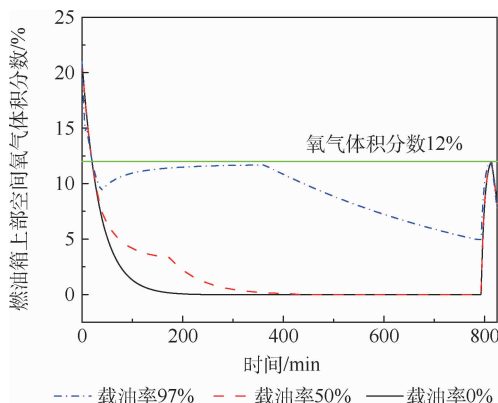


图 10 惰化时间 20 min 双流量模式下氧气体积分数变化

Fig. 10 Variation of oxygen volume fraction in dual-flow inerting mode at 20 min inerting time

根据计算结果可知, 采用双流量惰化模式对应的抽吸气流量状态可有效将燃油箱气相空间氧气体积分数在整个飞行包线内控制在 12% 以下, 从而满足燃油箱惰化要求。

4 结 论

在建立耗氧型惰化系统 AMESim 仿真模型的基础上, 分析了惰化系统抽吸气流量对地面初始惰化时间、飞行包线内燃油箱气相空间氧气体积分数的影响, 结论如下:

1) 惰化系统抽吸气流量与惰化时间近似呈反比关系, 随着抽吸气流量的增加, 地面初始惰化时间迅速减少, 然后变化趋势趋于平缓; 相同惰化时间下, 燃油箱载油率越低, 所需要的抽吸气流量越大。

2) 下降阶段由于外界空气大量涌入燃油箱, 会使燃油箱内氧气体积分数迅速升高, 当抽吸气流量较低时, 燃油箱气相空间氧气体积分数会超过 12%。

3) 提出一种适用于 3CGIS 惰化系统抽吸气

流量确定的双流量惰化模式设计方法,该惰化模式下可有效将燃油箱气相空间氧气体积分数在整个飞行包线内控制在 12% 以下,从而满足燃油箱惰化要求。

参考文献 (References)

- [1] MANATT S A. Fuel tank inerting system; US, 4556180 [P]. 1985-12-03.
- [2] 刘小芳,刘卫华. 飞机供氧和燃油箱惰化技术概况[J]. 北华航天工业学院学报,2008,18(3):4-7.
LIU X F, LIU W H. Outline of airborne oxygen supplied and fuel tanks inerted[J]. Journal of North China Institute of Aerospace Engineering, 2008, 18(3):4-7 (in Chinese).
- [3] Fuel Tank Harmonization Working Group Task Group 1. Service history/fuel tank safety level assessment[R]. [S. l.]: Aviation Rulemaking Advisory Committee (ARAC), 1998.
- [4] 冯诗愚,卢吉,刘卫华,等. 机载制氮系统中空纤维膜分离特性[J]. 航空动力学报,2012,27(6):1332-1339.
FENG S Y, LU J, LIU W H, et al. Separation performance of hollow fiber membrane for on-board inerting gas generating system[J]. Journal of Aerospace Power, 2012, 27(6):1332-1339 (in Chinese).
- [5] LANGTON R, CLARK C, HEWITT M, et al. Aircraft fuel systems[M]. New York: John Wiley & Sons, 2010:225-237.
- [6] ABRAMOWITZ A, BORIS P. Characterization of an oxygen/nitrogen permeable membrane system; DOT/FAA/AR-95/91 [R]. [S. l.]: FAA Report, 1996.
- [7] CAVAGE W M. The cost of implementing ground-based fuel tank inerting in the commercial fleet; DOT/FAA/AR-00/19 [R]. [S. l.]: FAA Report, 2000.
- [8] SUMMER S M. Cold ambient temperature effects on heated fuel tank vapor concentrations; DOT/FAA/AR-TN99/93 [R]. [S. l.]: FAA Report, 2000.
- [9] 王明波,邵垒,潘俊,等. 耗氧型燃油箱惰化技术历史和现状[J]. 航空科学技术, 2016, 27(7):1-7.
WANG M B, SHAO L, PAN J, et al. History and current status of oxygen consumption based fuel tank inerting technology[J]. Aeronautical Science & Technology, 2016, 27(7):1-7 (in Chinese).
- [10] 刘凤春,邱献双. 一种新型的飞机油箱催化惰化系统[J]. 航空科学技术, 2011(4):27-29.
LIU S C, QIU X S. A new fuel tank catalytically inerting system [J]. Aeronautical Science & Technology, 2011(4):27-29 (in Chinese).
- [11] LIMAYE S, KOENIG D. Catalytic reactive component reduction system and methods for the use thereof; US, 11/994801 [P]. 2008-08-21.
- [12] MORRIS R, MILLER J, LIMAYE S. Fuel deoxygenation and aircraft thermal management; AIAA 2006-4027 [R]. Reston: AIAA, 2006.
- [13] WALKER S, JUNG W, ROBERTSON S. Demonstration of a novel catalyst based green on board inert gas generation system (GOBIGGS) for fuel tank inerting[C]. The American Helicopter Society 69th Annual Forum, 2013:1-10.
- [14] 冯诗愚,李超越,邵垒,等. 一种燃油箱绿色惰化系统地面惰化性能分析[J]. 航空动力学报, 2017, 32(2):268-274.
FENG S Y, LI C Y, SHAO L, et al. Analysis on ground-based inerting performance of a fuel tank green on-board inert gas generation system [J]. Journal of Aerospace Power, 2017, 32(2):268-274 (in Chinese).
- [15] FENG S Y, PENG X T, CHEN C, et al. Effect of air supplementation on the performance of an onboard catalytic inerting system[J]. Aerospace Science and Technology, 2020, 97(2):1-8.
- [16] 邵垒. 飞机燃油箱耗氧型惰化技术理论和实验研究[D]. 南京:南京航空航天大学, 2018.
SHAO L. Theoretical and experimental study of oxygen consumed inerting technology for aircraft fuel tank [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2018 (in Chinese).
- [17] 彭孝天,冯诗愚,周利彪,等. 温度对耗氧型惰化系统产水性能影响[J]. 航空动力学报, 2020, 35(8):1622-1627.
PENG X T, FENG S Y, ZHOU L B, et al. Effect of temperature on water production performance of oxygen-consuming inerting system [J]. Journal of Aerospace Power, 2020, 35(8):1622-1627 (in Chinese).
- [18] 谢辉辉,冯诗愚,彭孝天,等. 耗氧型惰化系统反应器性能理论研究[J]. 北京航空航天大学学报, 2019, 45(2):2312-2319.
XIE H H, FENG S Y, PENG X T, et al. Theoretical of reactor performance in oxygen consumption based inerting system [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(2):2312-2319 (in Chinese).
- [19] FENG S Y, LI C Y, PENG X T, et al. Oxygen concentration variation in ullage of an inert aircraft fuel tank determined by the dissolved oxygen evolution [J]. Chinese Journal of Aeronautics, 2020, 33(7):1919-1928.
- [20] 王苏明,冯诗愚,李宗祺,等. 燃油箱耗氧惰化与中空膜惰化的数值模拟及比较[J]. 北京航空航天大学学报, 2020, 46(5):1032-1038.
WANG S M, FENG S Y, LI Z Q, et al. Numerical simulation and comparison of oxygen consumption inerting and hollow membrane inerting in fuel tank [J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(5):1032-1038 (in Chinese).
- [21] LIMAYE S, ROBERTSON S, KOENIG D, et al. Development of a "green" on-board inert gas generation system [C] // Proceedings of the 15th Triennial International Fire & Cabin Safety Research Conference, 2007.
- [22] 中国民用航空局. 中国民用航空规章第 25 部: 运输类飞机适航标准; CCAR-25-R4 [S]. 北京: 中国民用航空局, 2011: 245-246.
Civil Aviation Administration of China. Chinese civil aviation regulations. Part 25. Airworthiness standards for transport category aircraft; CCAR-25-R4 [S]. Beijing: Civil Aviation Administration of China, 2011:245-246 (in Chinese).

Effect of suction flow rate on performance of catalytic inerting system

WANG Chenchen^{1,*}, PAN Jun¹, WANG Yangyang¹, DUAN Weijie²

(1. Aviation Key Laboratory of Science and Technology on Electromechanical System Integration,

AVIC Nanjing Engineering Institute of Aircraft System, Nanjing 211106, China;

2. AVIC Helicopter Design and Research Institute, Tianjin 333001, China)

Abstract: Based on the principle of oxygen-consuming inerting system, the AMESim simulation model of low temperature controllable oxygen consumed catalytic green inerting system (3CGIS) was established. The effect of the suction flow rate of 3CGIS on the inerting time, and the change of the oxygen volume fraction in the fuel tank ullage under the flight envelope were researched. The prediction of oxygen volume fraction under the entire flight envelope was verified against the experimental data, showing a satisfactory agreement, and its validity was validated with the comparison of results. On the basis of modeling, the suction flow rate is approximately inversely proportional to the initial inerting time were obtained; under a certain inerting time, the required suction flow rate increases with the decrease of fuel load. Aiming at the possibility that the oxygen volume fraction in the fuel tank ullage during the descent stage may exceed 12%, a dual-flow inerting mode design method is proposed to ensure that the oxygen volume fraction is less than 12% under the entire flight envelope. The results can be provided as a reference for design and optimization of low temperature controllable oxygen consumed catalytic green inerting system.

Keywords: oxygen-consuming inerting; low temperature catalysis; suction flow rate; inerting time; oxygen volume fraction; AMESim simulation

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

DOI: 10.13700/j.bh.1001-5965.2021.0030

二维有限长度柔性壁面上 T-S 波演化的数值研究

全下
文载

洪正, 叶正寅*

(西北工业大学 航空学院, 西安 710072)

摘 要: 受自然界鸟类羽毛的柔性特征启发, 利用数值模拟的方式开展了柔性壁面对亚声速边界层中 T-S 波演化的影响研究。刚性壁面上的数值结果与线性理论吻合得很好, 验证了数值方法的可靠性。在此基础上, 将部分刚性壁面替换为柔性壁面, 结果表明, 柔性壁面可以抑制 T-S 波在空间上的增长, 从而推迟边界层流动转换。壁面的变形不只跟随 T-S 波的波形, 还因为柔性段与刚性段相接的前缘和后缘引起与扰动源频率相同的更大尺度的壁面波动, 壁面的实际变形由这几种波叠加而成。开展的参数研究结果表明, 增大表面的质量密度对于柔性壁面衰减扰动的效果几乎没有影响; 增大表面张力和增加底部支撑的弹性系数可以增加壁面的刚性, 减小壁面变形的幅度; 增加阻尼可以抑制柔性段前后缘产生的大尺度壁面波动的传播, 而对跟随 T-S 波的变形影响不大。总体上, 柔性壁面的变形程度越大, 其扰动的抑制效果越强。

关键词: 柔性壁面; 有限长度; T-S 波; 边界层转换; 减阻

中图分类号: V211.3

文献标志码: A

文章编号: 1001-5965(2022)07-1190-10

在人类飞行器的发展历史中, 从鸟类飞行中得到过多种启示^[1]。对鸟类的研究初期大多从生物学角度出发, 随着观察和分析的深入, 逐渐认识到空气动力学的奥秘^[2]。鸟类经过千万年的进化, 形成了带弯度的厚前缘和薄的后缘形状的翅膀。飞机机翼的形状正是从鸟类仿生而来, 厚的前缘有利于提高最大升力系数和失速迎角。观察鸽子可以发现, 厚的翅膀前缘是由柔软而致密的覆羽层叠而成, 具有弹性的特征, 能够在压力作用下产生局部变形, 这与机翼刚性的表面明显不同。北美金鸫鸟飞越超过 4 000 km 的大洋每小时仅损失其体重的 0.06%^[3]; 信天翁几乎不用拍动翅膀就可以长时间飞翔于海面之上^[4]。如此惊人的飞行效率必然对应着极低的飞行阻力, 而在同样条件下, 层流边界层的摩擦阻力要比湍流

边界层小得多。因此, 可以推测出鸟翼表面应是低阻力的层流边界层流动占据主导地位。因此, 羽毛的柔性特征是否对延迟或者抑制层流到湍流的转换起作用是本研究的出发点。

柔软的表面并非只存在于鸟类身上, 最早引起科学家们注意的生物是海豚。Gray 悖论指出海豚的肌肉力量不足以维持游动所需的能量, 因此推测海豚的皮肤具有减阻的效果^[5]。Kramer^[6-7]最早使用自制的海豚皮开展了柔性壁面对减阻的实验, 研究表明, 海豚皮最高达到了 50% 的减阻效果。然而, 后续不同研究者的实验结果并不一致, 柔性壁面的减阻效果存在争议^[8]。直到 Carpenter 和 Garrad^[9-10]通过稳定性理论证明了柔性壁面能抑制转捩线性阶段小扰动的增长, 从而具有推迟转捩的效果。这一理论结果在 Lee 和 Kim^[11]

收稿日期: 2021-01-20; 录用日期: 2021-03-07; 网络出版时间: 2021-03-25 20:45

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210324.1625.004.html

基金项目: 国家自然科学基金(12072281)

*通信作者。E-mail: yezy@nwpu.edu.cn

引用格式: 洪正, 叶正寅. 二维有限长度柔性壁面上 T-S 波演化的数值研究[J]. 北京航空航天大学学报, 2022, 48(7): 1190-1199.
HONG Z, YE Z Y. Numerical investigation on evolution of T-S wave on a two-dimensional compliant wall with finite length[J].
Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1190-1199 (in Chinese).

的实验、Davies 和 Carpenter^[12]、Wang 等^[13] 的数值模拟中都得到了验证。Duncan^[14] 和 Kireiko^[15] 开展了柔性变形对完全发展的湍流边界层减阻的理论研究,使用同样的 DNS(direct numerical simulation)模型,却得到了阻力增加或统计上对流场无影响的结果^[16-18],这与 Duncan^[14] 和 Kireiko^[15] 的理论预测相反,也与 Choi 等^[19] 实验中获得的 7% 阻力减小矛盾。Xia^[18] 和 Zhang^[20] 等通过数值模拟的方式开展了被动壁面变形与湍流边界层之间相互作用的机理研究。Jozsa 等^[21-22] 系统地研究了壁面主动或被动产生切向变形对湍流边界层减阻的效果,指出柔性壁面的各向异性对于取得实际的减阻十分重要。

自然转捩通常在低湍流条件下(小于 1%)发生,常见于亚声速飞行器的外部流动中。自然转捩过程可划分为 5 个特征阶段:感受阶段、线性阶段、非线性饱和阶段、二次失稳阶段和层流到湍流的破碎阶段^[23]。鸟类柔软的翅膀前缘是边界层开始出现的位置,就转捩而言,很可能对应着边界层内 T-S 波线性增长的初期阶段。已有的研究主要以海豚生存环境中的水为流动介质,对鸟类而言的空气介质下的研究很少。本文基于描述空气流动的可压缩 Navier-Stokes 方程,采用高精度的数值方法,开展柔性壁面对 T-S 波空间演化影响的数值研究。首先,通过与线性理论对比及网格无关性验证证明了数值方法的可靠性;然后,不同壁面条件下的结果对比验证了柔性壁面对扰动的衰减效果;最后,改变柔性壁面参数研究了不同参数的影响。

1 研究模型及数值方法

1.1 计算模型

对于二维 T-S 波在柔性壁面上演化特征的数值研究,本文采用的计算域模型如图 1 所示。截取平板上某一段长度的边界层流动为基本流,T-S 波通过在刚性壁面上游位置人为施加扰动来激发。

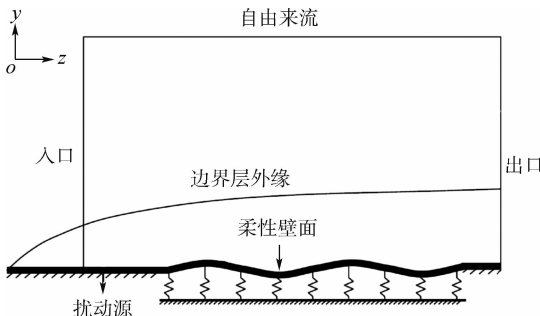


图 1 计算域模型示意图

Fig. 1 Schematic of computational domain

T-S 波先经过刚性壁面,再经过一段有限长度的柔性壁面,最终回到刚性壁面上到达出口。

入口处的流向位置记为 x_{in} , 扰动源的位置记为 x_r , 柔性壁面开始和结束的位置分别记为 x_{es} 和 x_{ee} , 出口处的位置记为 x_{out} 。在本文的计算条件下,以参考长度无量纲化后的各参数如表 1 所示。计算域的无量纲高度取为 50。

表 1 计算域流向参数

Table 1 Streamwise parameters of computational domain

参数	数值
x_{in}	55.85
x_r	135.08
x_{es}	158.88
x_{ee}	777.10
x_{out}	900

1.2 控制方程及边界条件

对于空气的运动,采用二维的可压缩 Navier-Stokes 方程来描述,无量纲化后的守恒形式如下:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{E}_c}{\partial x} + \frac{\partial \mathbf{F}_c}{\partial y} = \frac{1}{Re} \left(\frac{\partial \mathbf{E}_v}{\partial x} + \frac{\partial \mathbf{F}_v}{\partial y} \right) \quad (1)$$

式中: \mathbf{U} 为守恒变量; \mathbf{E}_c 和 \mathbf{E}_v 分别为 x 方向上的对流量和黏性通量; \mathbf{F}_c 和 \mathbf{F}_v 分别为 y 方向上的对流量和黏性通量。

$$\left\{ \begin{aligned} \mathbf{U} &= \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{pmatrix} \\ \mathbf{E}_c &= \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (\rho e + p)u \end{pmatrix} \\ \mathbf{F}_c &= \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (\rho e + p)v \end{pmatrix} \\ \mathbf{E}_v &= \begin{pmatrix} 0 \\ \tau_{xx} \\ \tau_{xy} \\ u\tau_{xx} + v\tau_{xy} - q_x \end{pmatrix} \\ \mathbf{F}_v &= \begin{pmatrix} 0 \\ \tau_{yx} \\ \tau_{yy} \\ u\tau_{yx} + v\tau_{yy} - q_y \end{pmatrix} \end{aligned} \right. \quad (2)$$

式中: ρ 、 u 、 v 、 e 、 p 分别为气体的密度、 x 方向速度、 y 方向速度、总能、压力; τ_{xx} 、 τ_{xy} 、 τ_{yx} 、 τ_{yy} 为黏性应力; q_x 、 q_y 为热流量。由于方程数少于未知变量

数,需补充完全气体方程以封闭方程组。无量纲化方程所用的参考速度为自由来流速度 U_∞ ,参考长度为扰动源施加位置处的边界层位移厚度 δ_r^* ,其余参考量均取自由来流值。在本文的参数下, $Ma = U_\infty/c_\infty = 0.2$, $Re = \rho_\infty U_\infty \delta_r^*/\mu_\infty = 400$,其中, c_∞ 为声速, ρ_∞ 为密度, μ_∞ 为黏性系数。

本文所考虑的柔性壁面是由弹簧均匀支撑且张紧的薄膜结构,采用与流体方程相同的参考量,无量纲化后的表面位移控制方程为

$$m \frac{\partial^2 \eta}{\partial t^2} + d \frac{\partial \eta}{\partial t} - T \frac{\partial^2 \eta}{\partial x^2} + k\eta = -p'_w \quad (3)$$

式中: η 为结构表面的法向位移; m 和 T 分别为膜表面的质量密度和表面张力; d 和 k 分别为支撑弹簧的阻尼系数和弹性系数; p'_w 为壁面上的脉动压力,即壁面变形的驱动力。

计算域入口对应于平板前缘之后位置 x_{in} 处的边界层剖面,可通过 Blasius 相似解或数值计算完整的平板边界层来得到该处的变量分布,再给定到入口边界。出口变量全部采用外插方式,简单且可以较好地避免出口反射的问题。上边界采用无反射的远场边界条件,下边界中刚性壁面部分采用无滑移条件,柔性壁面上的速度与壁面法向位移相关联,即

$$(u_w, v_w) = \begin{cases} (0, 0) & x < x_{cs} \text{ 或 } x > x_{ce} \\ (-\eta \frac{\partial \bar{u}}{\partial y}, \frac{\partial \eta}{\partial t}) & x_{cs} < x < x_{ce} \end{cases} \quad (4)$$

式中: \bar{u} 为平均流向速度。

在扰动源位置 x_r 处添加随时间以正弦函数形式变化的周期法向吹吸速度,用来激发边界层中相同频率的 T-S 波:

$$v_w = A \sin(\omega t) \quad (5)$$

式中: ω 为扰动的频率; A 为扰动的幅值。

1.3 数值方法

Navier-Stokes 方程的数值求解采用半离散的形式,即先求空间导数项,再进行时间上的推进。采用基于结构网格的格心有限差分法进行空间离散^[24]。以 x 方向上的通量导数求解为例,具体步骤如下:

步骤 1 基于格心处已知的流场原始变量值,利用五阶迎风线性格式重构出面心处的原始变量值的左迎风值和右迎风值 Q^L 、 Q^R ,其中, $Q = \{\rho, u, v, p\}$ 。

步骤 2 在面心处利用 Roe 方法求解当地的黎曼问题,从而获得面心处的对流通量 E_c (Q^L , Q^R),黏性通量没有迎风特性,可直接由中心值得到,即 $E_v((Q^L + Q^R)/2)$ 。

步骤 3 通过二阶中心差分得到格心处的通量导数 $\partial E_c/\partial x$ 和 $\partial E_v/\partial x$ 。

至此,空间离散完成,得到了控制体上的净通量。时间推进采用隐式求解中常用的 LU-SGS 迭代方法。

结构方程同样采用隐式迭代方法求解。时间项采用二阶向后差分,即

$$\left(\frac{\partial^2 \eta}{\partial t^2}\right)_i = \frac{\eta_i^{m+1} - 2\eta_i^n + \eta_i^{n-1}}{\Delta t^2} \quad (6)$$

$$\left(\frac{\partial \eta}{\partial t}\right)_i = \frac{3\eta_i^{m+1} - 4\eta_i^n + \eta_i^{n-1}}{2\Delta t} \quad (7)$$

空间项采用二阶中心差分,即

$$\left(\frac{\partial^2 \eta}{\partial x^2}\right)_i = \frac{\eta_{i+1}^{m+1} - 2\eta_i^{m+1} + \eta_{i-1}^{m+1}}{\Delta x^2} \quad (8)$$

理论上,当 $m \rightarrow \infty$ 时, $\eta^{m+1} \rightarrow \eta^{n+1}$ 。流体和固体的求解采用紧耦合的形式,每个实时间步内,交替求解流体和固体方程直至两者都满足收敛条件。本文计算中,残差下降 4 个量级以上即认为收敛。

1.4 数值方法验证

准确地产生 T-S 波并捕捉其空间演化是开展研究的必要基础,因此,先通过对比刚性壁面上的数值结果与线性稳定性理论来验证程序的正确性。

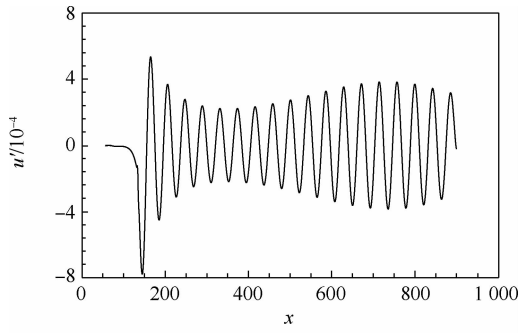
本节计算了 2 个不同频率的 T-S 波,对应扰动源的无量纲频率参数为 $F_1 = \frac{\omega_1}{Re} \times 10^6 = 140$ 和

$$F_2 = \frac{\omega_2}{Re} \times 10^6 = 156$$

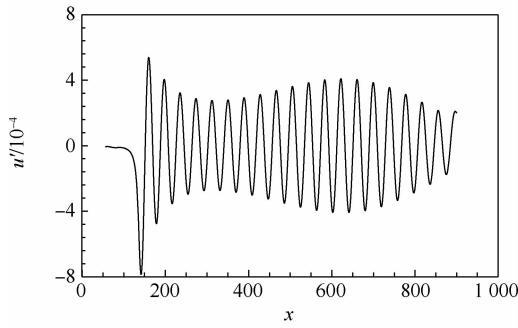
扰动的幅值均为自由来流速度的 1/1 000,即 $A_1 = A_2 = 0.001$ 。采用的计算网格在流向方向上均匀,法向上对壁面进行加密,使得距离壁面最近的网格满足 $y^+ \approx 1$ 的要求,网格单元数为 800×80 。时间步长取为 T-S 波周期的 1/1 000,总计算时长为 20 个周期。

沿 $y = 0.65$ 位置上的流向脉动速度的空间分布如图 2 所示(u' 为流量速度的脉动量)。可以看到,在扰动源的下游,经过短暂的过渡之后,脉动速度就已经呈现出典型的 T-S 波的特征,以变幅度的正弦波形式向下游传播。扰动产生后,扰动的幅值先下降再增长最后又下降。幅值变化趋势改变的位置称为中性点,图 3 对比了本文计算得到的中性点位置与线性理论及文献[13,25]的计算结果(F 为无量纲频率)。可以看到,本文数值方法得到的中性点位置与线性稳定性理论预测的结果符合得很好,说明数值方法对于捕捉 T-S 波的空间演化是有效的。

网格的差异有可能对数值结果产生显著的影响,需要验证数值结果对于网格已经没有明显的



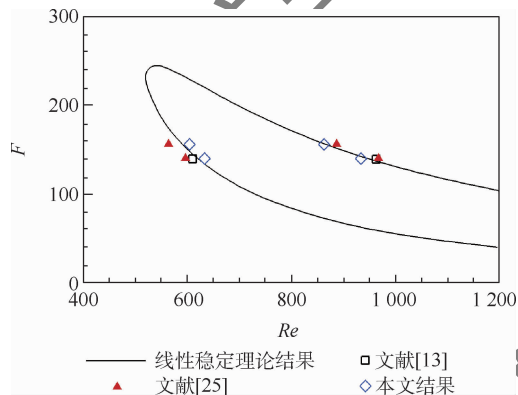
(a) $F_1 = 140$



(b) $F_2 = 156$

图 2 $y = 0.65$ 位置处的流向脉动速度沿空间的分布

Fig. 2 Spatial distribution of streamwise velocity fluctuation along $y = 0.65$

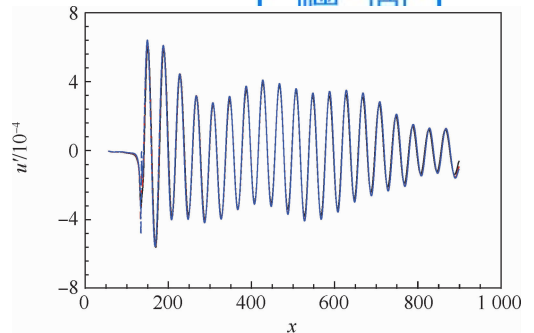


— 线性稳定理论结果 □ 文献[13]
▲ 文献[25] ◇ 本文结果

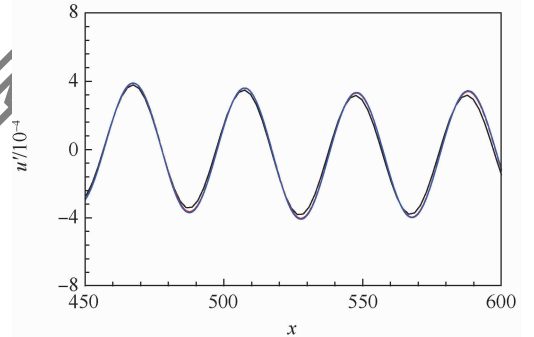
图 3 本文计算与线性稳定性理论得到的中性点位置对比

Fig. 3 Comparison of neutral points obtained from computations and linear stability theory

依赖性。考虑到流动量的变化主要体现在流向方向上,因此简单起见,仅在流向方向上均匀加密网格,法向网格数保持不变。因为最终研究的是流体和固体耦合的问题,所以网格无关性的验证将柔性壁面的变形也考虑在内。算例中扰动源的频率和幅值为 F_2 和 A_2 。图 4 给出了流向网格单元数分别为 400、800、1 600 时在 $y = 0.65$ 位置处的流向脉动速度分布。3 套网格下的结果总体特征是一致的,而从局部放大图来看,相比于流向网格数为 400 时,800 的网格可以得到更高的峰值和更低的谷值,即对波的耗散更小。流向网格数目继续加密到 1 600 后,与 800 网格相比已没有明显的差异,说明流向网格数目取为 800 时网格的影响已经很小,因此,本文计算中流向网格数均取 800。



(a) 全局视图



(b) 局部视图

— 网格单元数400 - - - 网格单元数800
· · · 网格单元数1 600

图 4 不同网格下 $y = 0.65$ 位置处流向脉动速度沿空间的分布

Fig. 4 Spatial distribution of streamwise velocity fluctuation along $y = 0.65$ with different grids

2 柔性壁面上 T-S 波的演化

如式(3)所示,控制柔性结构自身性质的参数有 4 个,即质量密度 m 、阻尼系数 d 、表面张力 T 和弹性系数 k 。为了研究刚性壁面和柔性壁面上 T-S 波演化的差异,以及不同结构参数的影响,考虑了 5 组不同的参数组合。无量纲化后的各组参数如表 2 所示。

扰动源的频率和幅值取第 1 节算例中的 F_2 和 A_2 。所有参数下的计算均采用相同的网格和计算设置,时间步长取为 T-S 波周期的 1/1 000,共计算 20 个周期。

表 2 不同柔性壁面的参数

Table 2 Parameters of different compliant walls

序号	m	d	T	k
1	1.45	0.1	14.5	0.069
2	7.25	0.1	14.5	0.069
3	1.45	0.5	14.5	0.069
4	1.45	0.1	29	0.069
5	1.45	0.1	14.5	0.69

2.1 刚性壁面与柔性壁面的对比

首先,通过对比第 1 组参数下的柔性壁面和刚性壁面上的数值结果,研究柔性壁面对 T-S 波

空间演化的影响。

图5给出了刚性壁面和柔性壁面下整个空间上的流向脉动速度分布。可以看到,2种壁面条件下流向脉动速度沿壁面法向的演化规律是相似的,即脉动从壁面处开始增长,在壁面附近达到最大/小值,随后下降/增长到反方向的极值,最后随着距离的增加脉动速度逐渐趋近于零。不同的是,刚性壁面处脉动速度为零,柔性壁面因为壁面运动在壁面处的脉动速度不为零,如图5中加粗的黑线所示。从图5中可以看出,流向脉动速度在刚进入柔性壁面部分时相对于刚性壁面幅值得到了增强,而离开柔性壁面后,幅值相比于刚性壁面上要更小。

定量地来看,在流向脉动强度最大值附近 $y = 0.65$ 位置处进行截取,图6给出了该位置处的流

向脉动速度沿流向距离的分布曲线。图中虚线标示了柔性壁面开始和结束的位置。可以清楚地看到,从刚性壁面进入柔性段后,尽管脉动幅度先是有显著的增加,但因为柔性壁面具有比刚性壁面更小的不稳定区域,在柔性段结束位置及出口处的脉动幅值相对于刚性壁面已有明显的衰减。这说明有限长的柔性壁面可以抑制 T-S 波扰动沿空间的发展。此外,注意到进入柔性壁面部分后, T-S 波的波长相比于刚性壁面变长了,这与 Wang 等^[1]的数值结果是一致的。

柔性壁面是在 T-S 波带来的压力脉动驱动下产生变形的,但是壁面变形的形式却不只是跟随 T-S 波的波形,这其中柔性壁面的前缘和后缘起到了重要的作用。图7给出了壁面刚开始运动, T-S 波部分通过柔性壁面时刻的壁面变形。从图7(a)中可以看到,上游的扰动到达柔性段前缘后,激发了柔性结构本身的振动,振动的波长显然比 T-S 波的波长要长得多。图7(b)中,柔性段前缘产生的结构振动到达后缘之后,产生了向前传播的反射波。因此,在 T-S 波完全通过之后,图8(a)中最终的壁面变形实际上由3个波叠加而来。图8(b)中对壁面变形进行了傅里叶变换,频域结果中3个峰值清晰地反映了这3个波成分(α 为流向波数)。波数最小的峰值对应于波长最长,由柔性段前缘产生的壁面波动;波数最大的峰值则对应于波长最小的 T-S 波;中间的峰值波

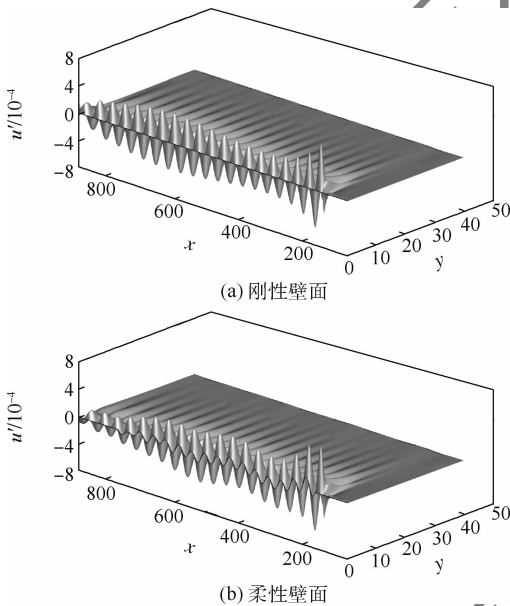


图5 刚性壁面和柔性壁面条件下的流向脉动速度空间分布
Fig.5 Spatial distribution of streamwise velocity fluctuation with rigid wall and compliant wall

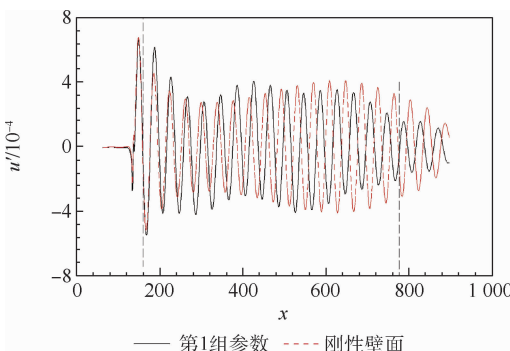
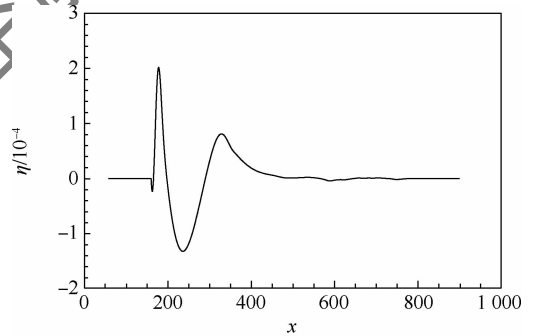
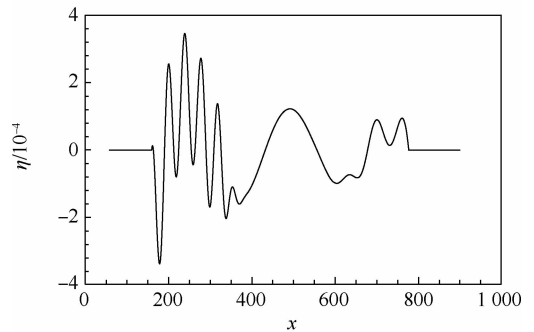


图6 柔性壁面和刚性壁上 $y = 0.65$ 处流向脉动速度的空间分布
Fig.6 Spatial distribution of streamwise velocity fluctuation along $y = 0.65$ on compliant wall and rigid wall



(a) $t = 95$



(b) $t = 450$

图7 不同时刻的壁面变形
Fig.7 Wall deformation at different times

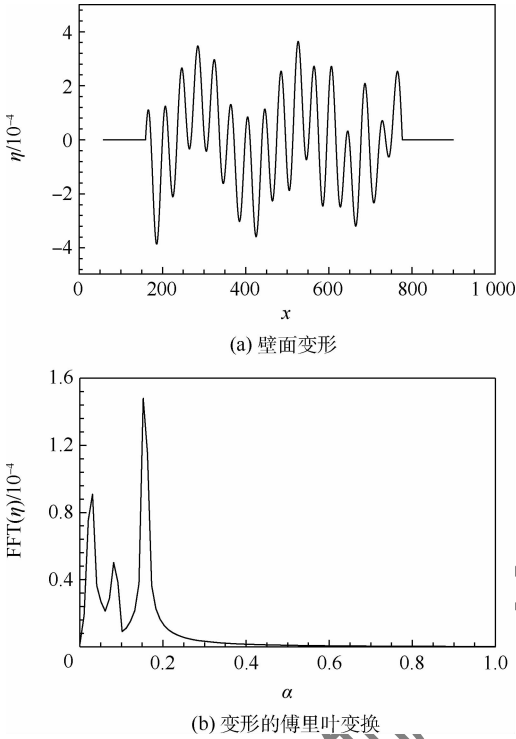


图 8 最终的壁面变形及其傅里叶变换结果

Fig. 8 Final wall deformation and its Fourier transform results

数对应的则是前面所述的由柔性段后缘反射产生的壁面波动。尽管 2 种波的波长不同,但由同一个扰动源所激发,因此其频率都与 T-S 波的频率相同。

图 8(a) 所示的壁面变形中,柔性段前缘产生的大尺度壁面波动对壁面的变形产生了明显的影响,壁面的这种大尺度运动反过来必然也会对经过其表面的流动产生影响。但 Wang 等^[13]指出,这种形式的运动对流场的影响随着法向距离的增长而快速衰减。因此,在图 6 ($y = 0.65$ 处)的结果中,壁面大尺度运动带来的影响相比壁面处已经弱了很多。柔性段前后缘除了对壁面变形的影响外,还会对脉动动能的产生或者消耗起作用。Davies 和 Carpenter^[12]对槽道流动研究的结果表明,刚性壁面到柔性壁面的突然转换使得前缘处的雷诺应力生成项突增、脉动增强,后缘则恰好相反。这与图 6 中过柔性段前缘之后脉动幅值有明显增加是一致的,到达后缘前脉动已经衰减了很多,因此,在图 6 中经过后缘之后脉动的变化不是很明显。

2.2 结构质量密度的影响

为了研究柔性表面质量密度参数对壁面变形及 T-S 波演化的影响,相对于第 1 组参数,第 2 组参数中其余参数保持不变,将质量密度增加至 5 倍。

图 9 给出了壁面的变形及其傅里叶变换结果。与第 1 组参数的结果对比中可以看到,增大

的表面质量密度明显增强了柔性段前缘带来的大尺度壁面波动,相比之下对 T-S 波带来的变形影响较小。而对于柔性段后缘反射的壁面波动,增大的质量密度则是削弱了其强度,并且波数有所增加,对应的波长减小。图 10 给出了 $y = 0.65$ 处的流向脉动速度的空间分布。可以看到,由于壁面变形形式的变化,柔性段部分上的扰动发展与第 1 组参数的结果在峰值和谷值处表现出了一些差异,但波长几乎是相同的。柔性段结束后 2 组结果基本相同,因此就对扰动的衰减效果来看,增大表面质量密度没有明显的作用。

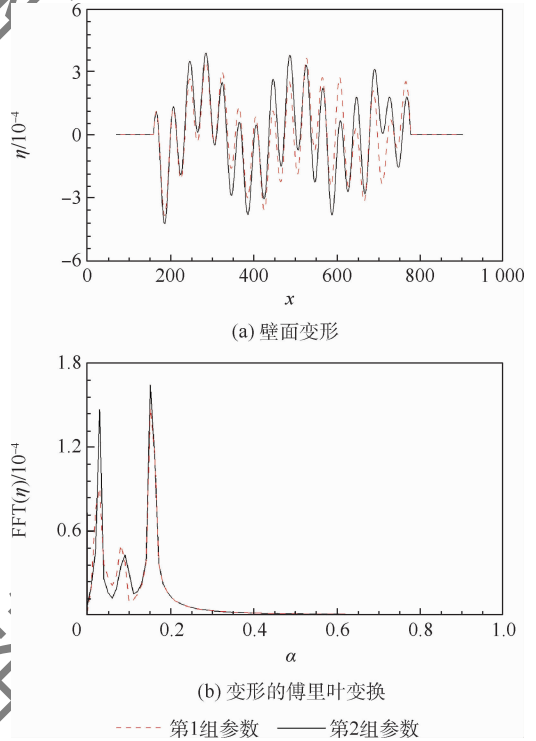


图 9 表面的质量密度对壁面变形的影响

Fig. 9 Influence of surface mass density on wall deformation

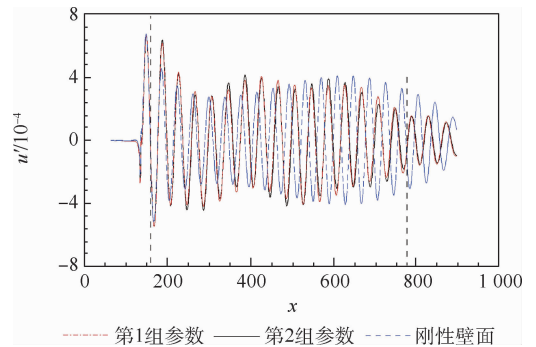


图 10 表面的质量密度对 T-S 波空间演化的影响

Fig. 10 Influence of surface mass density on spatial evolution of T-S wave

2.3 结构阻尼的影响

相比于第 1 组参数,第 3 组参数中只增大了弹簧支撑的阻尼系数,以此来研究阻尼系数对壁面变形及柔性壁面之上扰动发展的影响。

图 11 给出了壁面处的变形及其傅里叶变换的结果。可以看到,增大的阻尼系数对于柔性段前缘和后缘产生的壁面波动起到明显的抑制作用,但波数保持不变。而对于 T-S 波带来的变形,增大的阻尼系数并不会起到类似的抑制效果。从壁面开始变形到扰动完全通过柔性段的过程中观察到,增大阻尼系数直观的效果是柔性段前后缘产生的壁面波动在向后/向前传播过程中衰减得更快。因此在图 12 中,柔性段中后部的变形相比于第 1 组参数结果,T-S 波之外的波动显得更弱了。增大阻尼系数后,柔性壁面抑制 T-S 波扰动的效果削弱了,正对应了图 11 中 T-S 波带来的变形相比于第 1 组参数的结果不仅没有衰减,反而得到了增强。

2.4 结构表面张力的影响

第 4 组参数相比于第 1 组参数,保持其余参数不变,将柔性表面的张力提高 1 倍,以此来研究表面张力的影响。

从图 13(a)中给出的壁面变形结果可以看出,相比于第 1 组参数的结果,增大表面张力后壁面的变形量明显降低。直观地,张力越大,则表面的刚性越强,壁面变形也就越困难。与之相对应的,图 13(b)所示的傅里叶变换结果中,T-S 波对应的变形幅度明显下降。柔性段前缘带来的壁面波动在增加表面张力后也有一定的衰减,而由柔性段后缘反射而来的壁面波动则几乎被完全抑制了。图 14 中给出了 $y=0.65$ 处流向脉动速度的空间分布。可以看到,壁面变形被削弱之后,柔性壁面对 T-S 波发展的抑制作用也减弱了。另外,柔性表面刚性的增强也使得其上的 T-S 波的波长相较于第一组参数结果变短了,这在图 13(b)和图 14 中都有体现。

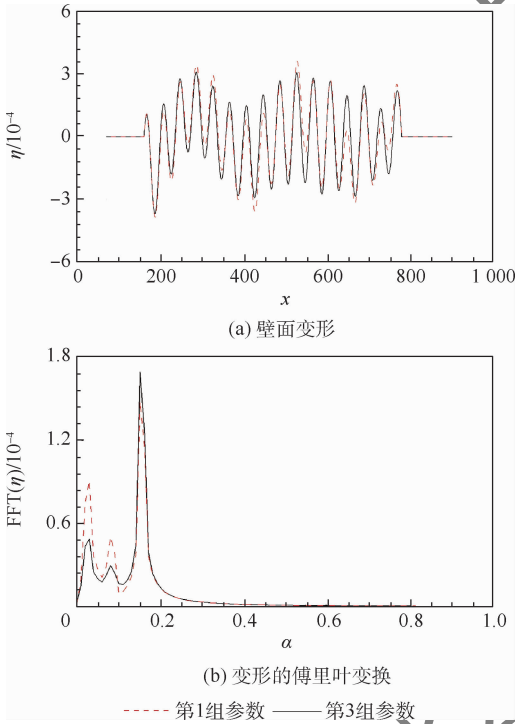


图 11 阻尼对壁面变形的影响

Fig. 11 Influence of damping on wall deformation

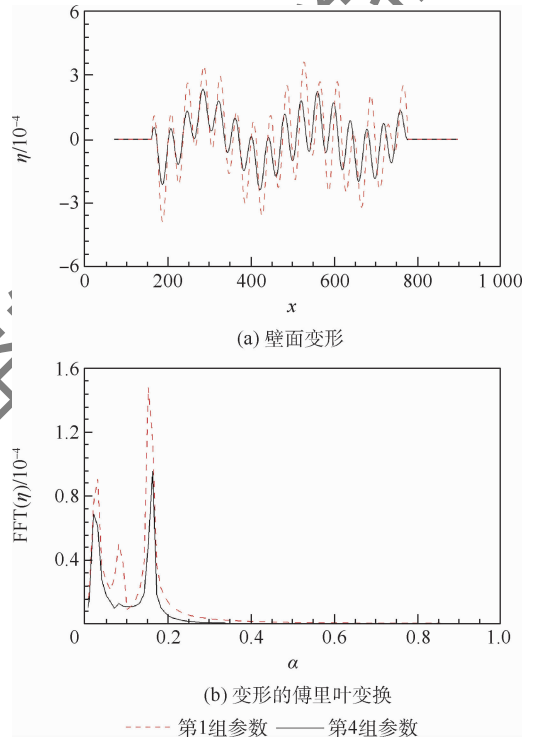


图 13 表面张力对壁面变形的影响

Fig. 13 Influence of surface tension on wall deformation

2.5 结构弹性系数的影响

为了研究支撑的弹性系数的影响,与第 1 组参数相比,保持其余参数不变,第 5 组参数仅将弹性系数增加至原来的 10 倍。

显然,底部支撑的弹性系数增大,柔性壁面的刚性必然会增强,其变形幅度应会变小。图 15(a)给出的壁面位移结果正对应了这样的情形。除了

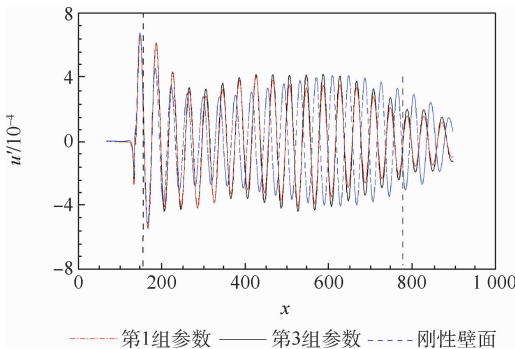


图 12 阻尼对 T-S 波空间演化的影响

Fig. 12 Influence of damping on spatial evolution of T-S wave

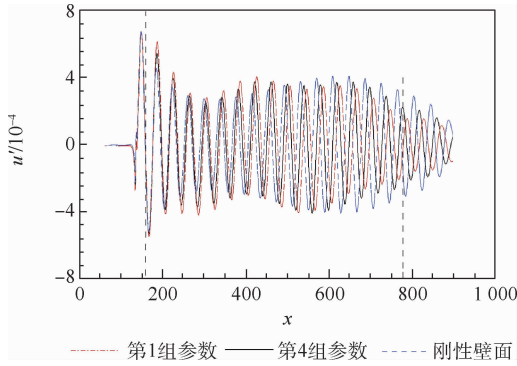


图 14 表面张力对 T-S 波空间演化的影响
Fig. 14 Influence of surface tension on spatial evolution of T-S wave

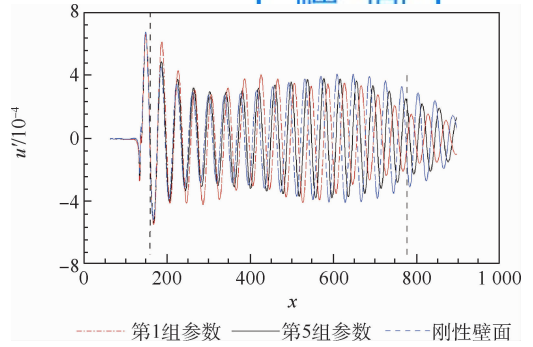
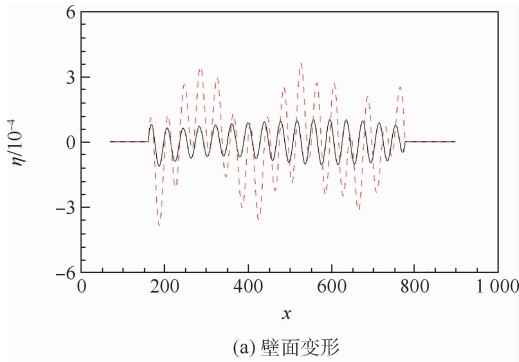
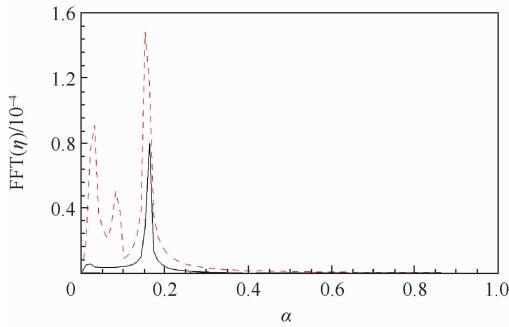


图 16 弹性系数对 T-S 波空间演化的影响
Fig. 16 Influence of elastic coefficient on spatial evolution of T-S wave



(a) 壁面变形



(b) 变形的傅里叶变换

--- 第 1 组参数 — 第 5 组参数

图 15 弹性系数对壁面变形的影响

Fig. 15 Influence of elastic coefficient on wall deformation 变形的幅值减小了之外,图 15(a)中的壁面变形中明显可以看出 T-S 波对应的变形占据了绝对的主导,柔性段前后缘产生的大尺度波动的效果几乎没有了。图 15(b)中对壁面变形的傅里叶变换结果更直接地表明了这样的结论,T-S 波对应的变形幅值约减小了一半,而柔性段前后缘引起的壁面波动幅值已经接近于零。尽管增大表面的张力和增加支撑的弹性系数都可以增大表面的刚性,但产生的结果却有明显的差异。增大支撑的弹性系数对于柔性段前后缘引起的壁面波动的抑制作用相比增大表面张力要强烈得多。图 16 给出了 $y = 0.65$ 处流向脉动速度随空间的分布。与第 1 组参数结果相比,柔性表面刚性的增强带来的是对 T-S 波抑制效果的削弱。

3 结 论

本文针对二维有限长度柔性壁面上的 T-S 波演化规律开展了数值研究,在选取的参数范围内,得到了以下主要结论:

1) 与刚性壁面相比,柔性壁面对 T-S 波扰动的增长具有抑制的效果,因此具备推迟边界层转捩的潜力。

2) 柔性壁面的变形除了跟随 T-S 波的波形外,还会由有限长柔性壁面的前缘和后缘引起向后/向前传播的更大尺度的壁面波动。

3) 增大阻尼系数对柔性段前后缘引起的壁面波动有明显的抑制作用,而 T-S 波对应的变形则受影响较小;增大表面张力和增大底部支撑的弹性系数都会提高柔性壁面的刚性,导致壁面变形幅度降低,其中增大弹性系数可以完全抑制柔性段前后缘带来的壁面波动,此时壁面变形只跟随 T-S 波的波形。

4) 对于壁面变形波动具有抑制作用的因素,如增大阻尼、增大表面张力和底部支撑的弹性系数,都会使得柔性壁面相对于刚性壁面的扰动衰减效果降低。

参考文献 (References)

[1] 屈秋林,王晋军. 鸟类飞行空气动力学对人类飞行的启示 [J]. 物理,2016,45(10):640-644.
QU Q L, WANG J J. Human flight inspired by the aerodynamics of bird flight [J]. Physics, 2016, 45 (10) : 640-644 (in Chinese).

[2] VIDELER J J. Avian flight [M]. Oxford: Oxford University Press, 2006.

[3] LINCOLN F C, PETERSON S R. Migration of birds [M]. [S. l.] : US Fish and Wildlife Service, US Department of the Interior, 1979.

[4] CROXALL J P, SILK J R, PHILLIPS R A, et al. Global circum-navigations: Tracking year-around ranges of nonbreeding alba-

- trosses[J]. *Science*, 2005, 307(5707): 249-250.
- [5] GRAY J. Studies in animal locomotion VI-The propulsive powers of the dolphin[J]. *Journal of Experimental Biology*, 1936, 13(2): 192-199.
- [6] KRAMER M O. Boundary layer stabilization by distributed damping[J]. *Journal of the Aerospace Sciences*, 1960, 27(1): 69.
- [7] KRAMER M O. Boundary layer stabilization by distributed damping[J]. *Naval Engineers Journal*, 1962, 74(2): 341-348.
- [8] GAD-EL-HAK M. Compliant coating for drag reduction[J]. *Progress in Aerospace Sciences*, 2002, 38(1): 77-99.
- [9] CARPENTER P W, GARRAD A D. The hydrodynamic stability of flow over Kramer-type compliant surfaces. Part 1. Tollmien-Schlichting instabilities[J]. *Journal of Fluid Mechanics*, 1985, 155: 465-510.
- [10] CARPENTER P W, GARRAD A D. The hydrodynamic stability of flow over Kramer-type compliant surfaces. Part 2. Flow-induced surface instability [J]. *Journal of Fluid Mechanics*, 1986, 170: 199-232.
- [11] LEE C, KIM J. Control of the viscous sublayer for drag reduction[J]. *Physics of Fluids*, 2002, 14(7): 2523-2529.
- [12] DAVIES C, CARPENTER P W. Numerical simulation of the evolution of Tollmien-Schlichting waves over finite compliant panels[J]. *Journal of Fluid Mechanics*, 1997, 335: 361-392.
- [13] WANG Z, YEO K S, KHOO B C. On two-dimensional linear waves in Blasius boundary layer over viscoelastic layers[J]. *European Journal of Mechanics-B/Fluids*, 2006, 25(1): 33-58.
- [14] DUNCAN J H. The response of an incompressible, viscoelastic coating to pressure fluctuations in a turbulent boundary layer [J]. *Journal of Fluid Mechanics*, 1986, 171: 339-363.
- [15] KIREIKO G V. Interaction of wall turbulence with a compliant surface[J]. *Fluid Dynamics*, 1990, 25(4): 550-554.
- [16] XU S, REMPFER D, LUMLEY J. Turbulence over a compliant surface: Numerical simulation and analysis[J]. *Journal of Fluid Mechanics*, 2003, 478: 11-34.
- [17] KIM E, CHOI H. Space-time characteristics of a compliant wall in a turbulent channel flow [J]. *Journal of Fluid Mechanics*, 2014, 756: 30-53.
- [18] XIA Q J, HUANG W X, XU C X. Direct numerical simulation of turbulent boundary layer over a compliant wall [J]. *Journal of Fluids and Structures*, 2017, 71: 126-142.
- [19] CHOI K S, YANG X, CLAYTON B R, et al. Turbulent drag reduction using compliant surfaces [J]. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 1997, 453(1965): 2229-2240.
- [20] ZHANG C, WANG J, BLAKE W, et al. Deformation of a compliant wall in a turbulent channel flow [J]. *Journal of Fluid Mechanics*, 2017, 823: 345-390.
- [21] JOZSA T I. Analytical solutions of incompressible laminar channel and pipe flow driven by in-plane wall oscillations [J]. *Physics of Fluids*, 2019, 31(8): 083605.
- [22] JOZSA T I, BALARAS E, KASHTALYAN M, et al. Active and passive in-plane wall fluctuations in turbulent channel flows [J]. *Journal of Fluid Mechanics*, 2019, 866: 689-720.
- [23] LEE C, JIANG X. Flow structures in transitional and turbulent boundary layers [J]. *Physics of Fluids*, 2019, 31(11): 111301.
- [24] LIAO F, YE Z Y, ZHANG L X. Extending geometric conservation law to cell-centered finite difference method on stationary grids [J]. *Journal of Computational Physics*, 2015, 284: 419-433.
- [25] FASEL H, KONZELMANN U. Nonparallel stability of a flat-plate boundary layer using the complete Navier-Stokes equations [J]. *Journal of Fluid Mechanics*, 1990, 221: 311-347.

Numerical investigation on evolution of T-S wave on a two-dimensional compliant wall with finite length

HONG Zheng, YE Zhengyin *

(School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Inspired by the flexible characteristics of bird feathers, numerical simulations are used to study the influence of the compliant wall on the evolution of T-S wave in the subsonic boundary layer flow. First, the numerical results on the rigid wall are in good agreement with the linear stability theory, which verifies the reliability of the adopted numerical methods. On this basis, part of the rigid wall is replaced with a compliant wall, and the results show that the compliant wall can suppress the spatial growth of T-S wave, thus delaying the flow transition. Furthermore, the deformation of the compliant wall not only follows the waveform of T-S wave, but also includes larger-scale vibrations with the same frequency as the disturbance source, which are caused by the leading edge and trailing edge of the compliant section. The actual deformation of the compliant wall is a superposition of these waves. Later parameter study shows that increasing the surface mass density has almost no effect on the compliant wall in terms of attenuating disturbance. Increasing the surface tension or increasing the elastic coefficient of the foundation can increase the stiffness of the compliant wall and thus reduce the amplitude of wall deformation. Increasing the damping can suppress the propagation of large-scale wall vibrations generated at the leading edge and trailing edge of the compliant section, while having little effect on the deformation directly corresponding to T-S wave. The overall trend is that when the amplitude of wall deformation decreases, the attenuation effect of the compliant wall on T-S wave decreases.

Keywords: compliant wall; finite length; T-S wave; boundary-layer transition; drag reduction

Received: 2021-01-20; **Accepted:** 2021-03-07; **Published online:** 2021-03-25 20:45

URL: kns.cnki.net/kcms/detail/11.2625.V.20210324.1625.004.html

Foundation item: National Natural Science Foundation of China (12072281)

* **Corresponding author.** E-mail: yezy@nwpu.edu.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0023

基于状态空间涡格法的阵风减缓分析

杨澜, 安朝*, 谢长川, 杨超

(北京航空航天大学 航空科学与工程学院, 北京 100083)



摘 要: 阵风响应计算及阵风减缓控制系统设计是飞行器气动弹性分析中的重要内容。基于状态空间涡格法(VLM)建立飞行器阵风气动模型,给出有限元结构模态及控制面模态广义自由度与涡格法控制点边界条件的插值关系,建立适用于复杂模型的阵风响应分析方法,弥补了传统阵风响应分析方法需要进行有理函数拟合或迭代计算资源消耗大等不足。在此基础上,基于经典PID控制方法设计阵风减缓控制系统,仿真得到离散阵风及 von Karman 连续阵风激励下的系统开/闭环时域响应情况,对比响应幅值计算减缓率。仿真计算结果表明:根据所提方法建立的阵风响应分析方法准确,阵风减缓控制系统能有效降低原气动弹性系统的阵风响应。

关键词: 气动弹性; 气动伺服弹性; 阵风响应; 阵风减缓控制系统; 涡格法(VLM)

中图分类号: V215.3

文献标志码: A **文章编号:** 1001-5965(2022)07-1200-10

飞行器在飞行中遭遇阵风扰动后机体结构会产生颠簸,干扰驾驶员操作,影响乘坐舒适性,增加气动弹性载荷,缩短结构疲劳寿命,严重威胁飞行安全^[1-2]。随着现代飞行器设计要求的逐渐提高,复合材料使用率及结构柔性的增大使得飞行器对阵风扰动的敏感性进一步提高,阵风扰动对于飞行器的影响更加复杂^[3-4]。2004年,美国“Helios 太阳神”高空长航时无人机在夏威夷上空试飞时遭遇突风,机翼发生持续的上反变形,出现发散的俯仰运动,结构破坏解体,NASA在其研究报告中将事故原因直指缺乏对该类型飞机的结构、气动等多学科耦合分析方法,并着重强调了气动伺服弹性分析和主动控制阵风载荷减缓研究的重要性^[5]。由此可见,阵风响应及阵风减缓分析研究对于提高现代飞行器的飞行性能、保障飞行安全具有重要意义。

利用势流理论面元法进行气动力计算是飞行器气动载荷分析的重要方法^[6]。面元法在求解

精度上相较二维气动力建模方法提高很多且计算效率高,满足理论研究及工程应用需求。偶极子格网法(double lattice method, DLM)和涡格法(vortex lattice method, VLM)是2类常用的气动力分析方法,能够分别在频域及时域上给出气动力计算结果,在气动弹性分析特别是阵风响应分析上应用广泛。

基于偶极子格网法进行弹性飞行器阵风响应的分析主要包括频域方法、状态空间方法及混合建模方法^[7-8]。频域方法建立偶极子格网法气动力模型,求解包括与阵风相关的广义非定常气动力影响系数(aerodynamic influence coefficients, AIC)矩阵,建立频域气动弹性响应方程来计算系统阵风响应。状态空间方法是最便于系统仿真及控制器设计的建模方法^[9],为得到气动弹性系统的状态空间方程形式,需要在广义非定常气动力影响系数矩阵的基础上进行有理函数拟合,得到气动力时域表达形式,该方法需要先建立频域气

收稿日期: 2021-01-15; 录用日期: 2021-03-19; 网络出版时间: 2021-04-27 11:28

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210426.1847.004.html

*通信作者: E-mail: ac@buaa.edu.cn

引用格式: 杨澜, 安朝, 谢长川, 等. 基于状态空间涡格法的阵风减缓分析[J]. 北京航空航天大学学报, 2022, 48(7): 1200-1209.
YANG L, AN C, XIE C C, et al. Gust load alleviation analysis based on vortex lattice method in state-space form[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1200-1209 (in Chinese).

动力模型,再进行频域时域拟合,计算较繁琐,同时,有理函数拟合的准确性取决于气动力滞后根的选取,容易出现较大误差^[8]。混合建模方法则结合了前两者的建模特点,仅对结构模态及控制面模态自由度引起的气动力做有理函数拟合,而与阵风相关的气动力仍采用频域方法求解,该方法一般仅用于阵风响应分析。

涡格法相较于偶极子格网法具有突出的优势。偶极子格网法作为频域气动力方法,使用了小扰动的简谐振动假设,且无法考虑机翼的面内运动。在进行时域仿真或与其他时域模型结合时,有理函数拟合等方法对气动力计算准确性存在影响,而涡格法是时域三维气动力求解方法,不需要简谐振动假设,可直接计算任意运动下的非定常气动力。基于涡格法进行弹性飞行器阵风响应的分析主要基于松耦合迭代算法进行结构与气动模型的耦合求解^[6]。在该分析框架下针对飞行器整机或部件进行阵风响应分析,其计算结果精度令人满意^[10-13]。特别地,在涡格法计算过程中,给定精确几何边界条件后,阵风响应分析方法同样适用于机翼大变形情况。随着分析模型复杂性的增加,基于涡格法的松耦合迭代求解方法同样会使问题分析效率大幅降低,同时松耦合迭代算法也使得分析方法难以应用于控制器设计,无法进行阵风减缓研究。将涡格法表达为状态空间的形式并结合结构动力学方程进行动力学响应分析是解决该问题的有效方法^[14-15]。然而,目前相关研究往往集中在与简单梁模型的结合上,缺少有限元模型与状态空间涡格法模型的耦合关系,同时,耦合控制系统的阵风减缓研究较少。

在阵风减缓控制器设计方面,国内外学者提出了一系列基于经典控制理论和现代控制理论的阵风减缓主动控制方法,包括工程实用性较好的经典PID控制^[16]、线性二次高斯(linear quadratic Gaussian, LQG)控制^[17-18]、鲁棒控制^[19]及广义预测控制(generalized predictive control, GPC)^[20]等。在阵风减缓控制方面,利用相关控制方案驱动控制面偏转,均能够取得较好的减缓效果。

本文在状态空间形式的涡格法气动力模型基础上,结合结构有限元模型推导涡格法控制点边界条件与结构模态及控制面模态广义自由度插值关系,建立适用于复杂机翼模型的阵风响应状态空间方程并进行阵风响应分析。与ZAERO软件计算结果相比,阵风响应幅值计算准确度满足要求。在该状态空间模型基础上,结合经典PID控制方法,建立以翼尖加速度为减缓目标的阵风减

缓控制系统,在施加离散阵风及连续阵风激励情况下,通过Simulink仿真验证本文所设计阵风减缓控制系统的有效性。

1 状态空间涡格法

1.1 势流理论及涡格法

当流体的流动问题满足高雷诺数、无分离流及不考虑跨声速效应的假设时,流场满足Euler方程。当引入正压气体假设及不可压无旋流假设后,可得到全速度势方程,流场的Laplace方程为

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0 \quad (1)$$

式中: Φ 为速度势函数; x, y, z 为空间物理坐标。

式(1)为偏微分方程,假定速度势函数后,结合具体的边界条件即可求解出流场的速度分布情况。

速度势可分为扰动速度势 Φ^* 和远场速度势 Φ_∞ 。无穷远处的边界条件要求机翼对无穷远处扰动为零,而机翼壁面要求满足的边界条件为控制点处“法向不穿透”,即控制点处壁面法向速度为零。

$$\nabla(\Phi^* + \Phi_\infty) \cdot \mathbf{n} = 0 \quad (2)$$

式中: \mathbf{n} 为控制点处的法向量列阵。此条件也被称为Neumann边界条件。

引入来流速度矢量 \mathbf{V}_∞ 、来流扰动速度矢量 \mathbf{V}_g 及壁面运动速度矢量 \mathbf{V}_b , Neumann边界条件可改写为

$$(\nabla\Phi^* + \mathbf{V}_\infty + \mathbf{V}_g - \mathbf{V}_b) \cdot \mathbf{n} = 0 \quad (3)$$

式中: $\nabla\Phi^*$ 为机翼壁面处的诱导速度。壁面运动速度矢量包括刚性运动速度及弹性运动速度。

基于势流理论的涡格法要求在机翼中弧面沿弦向和展向划分为若干四边形气动力网格。在每个1/4弦线处布置涡线段,4段等强度涡线首尾相连构成一个涡环。选取3/4弦线中点为控制点,在控制点处需要满足Neumann边界条件。尾涡部分同样用涡环表示,在每一时间步中,气动面沿后缘不断脱出一排尾涡涡环。给定强度涡线段对空间任意一点的诱导速度可以通过Biot-Savart定理计算,Neumann边界条件可写为

$$\mathbf{K}\Gamma = \mathbf{RHS} \quad (4)$$

式中: \mathbf{K} 为诱导系数矩阵,矩阵中的元素 K_{ij} 表示第*i*个涡环对第*j*个控制点的诱导速度局部法向分量; Γ 为涡环强度; \mathbf{RHS} 包括壁面运动速度、来流速度在控制点局部法向的投影:

$$\mathbf{RHS} = -(\mathbf{V}_\infty + \mathbf{V}_g - \mathbf{V}_b) \cdot \mathbf{n} \quad (5)$$

基于Neumann边界条件求解涡环强度即可进一步计算气动力分布情况^[6]。图1为涡格法涡环示意图。

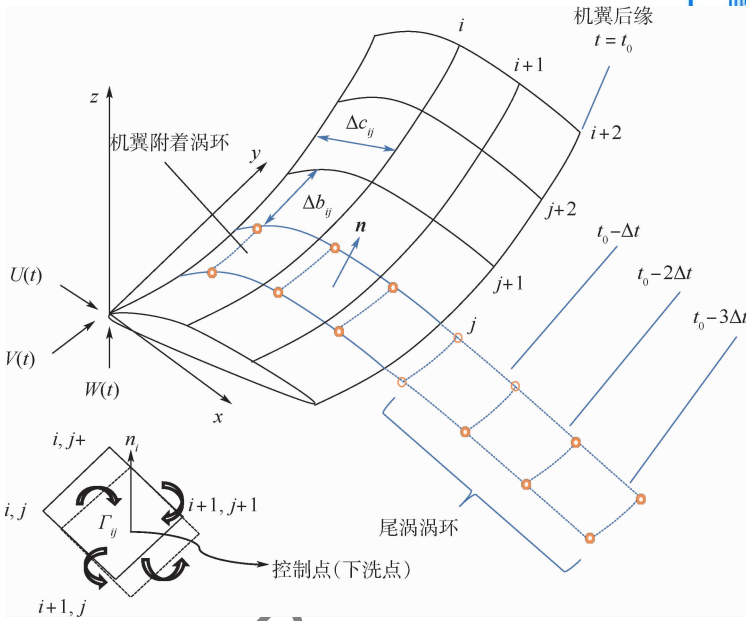


图1 涡格法涡环示意图

Fig. 1 Illustration of vortex ring in VLM

1.2 涡格法的状态空间方程

设机翼表面附着涡强度列向量为 Γ_b , 最靠近机翼后缘的第一排尾涡强度列向量为 Γ_{w0} , 其他尾涡强度列向量为 Γ_{wl} , N_b 、 N_{w0} 、 N_{wl} 分别为附着涡、第一排尾涡和其他尾涡数量。式(4)可写为^[6]

$$K_b \Gamma_b + K_{w0} \Gamma_{w0} + K_{wl} \Gamma_{wl} = -w \tag{6}$$

式中: $w = (V_{\infty} + V_g - V_b) \cdot n$; K_b 、 K_{w0} 和 K_{wl} 分别为附着涡、第一排尾涡和其他尾涡的诱导系数矩阵。

根据后缘 Kutta 条件, 脱尾涡的过程保持涡强守恒:

$$\dot{\Gamma}_{w0} = \frac{1}{\Delta t} (C_1 \Gamma_b - \Gamma_{w0}) \tag{7}$$

根据 Kelvin 定理进一步规定尾涡脱出后保持涡强不变^[6]:

$$\dot{\Gamma}_{wl} = \frac{1}{\Delta t} [(C_2 - I) \Gamma_{wl} + C_3 \Gamma_{w0}] \tag{8}$$

式中: C_1 、 C_2 、 C_3 为表征涡对应关系的常数提取矩阵, 只包含 0 和 1。

综合式(6)~式(8)可得涡格法状态空间方程形式为

$$\dot{\Gamma}_w = A_a \Gamma_w + B_a w \tag{9}$$

式中: $\Gamma_w = [\Gamma_{w0} \ \Gamma_{wl}]^T$; A_a 、 B_a 为状态空间系数矩阵, 仅与气动面几何形状及涡的划分有关。

$$A_a = \begin{bmatrix} -\frac{1}{\Delta t} (I - C_1 K_b^{-1} K_{w0}) & -\frac{1}{\Delta t} C_1 K_b^{-1} K_{wl} \\ \frac{1}{\Delta t} C_3 & \frac{1}{\Delta t} (C_2 - I) \end{bmatrix} \tag{10}$$

$$B_a = \begin{bmatrix} -\frac{1}{\Delta t} C_1 K_b^{-1} & 0 \end{bmatrix}^T \tag{11}$$

根据非定常 Bernoulli 原理, 气动网格上的压强差可表示为

$$\Delta p_{ij} = \rho_{\infty} \left[V_{l,ij} \tau_1 \frac{\partial \Gamma_{b,ij}}{\partial \tau_1} + V_{l,ij} \tau_2 \frac{\partial \Gamma_{b,ij}}{\partial \tau_2} + \frac{\partial \Gamma_{b,ij}}{\partial t} \right] \tag{12}$$

式中: 下标 ij 表示网格单元分量; ρ_{∞} 为来流大气密度; V_l 为气动力作用点当地速度矢量; τ_1 和 τ_2 分别为网格气动力作用点当地展向和弦向方向的切向量。假设气动面扰动速度远小于来流速度, 可认为 $V_l = V_{\infty}$ 。

式(12)中涡强的变化量可由如下差分方程求解:

$$\begin{cases} \frac{\partial \Gamma_{b,ij}}{\partial \tau_1} \approx \frac{\Gamma_{b,ij} - \Gamma_{b,i-1,j}}{\Delta c_{ij}} \\ \frac{\partial \Gamma_{b,ij}}{\partial \tau_2} \approx \frac{\Gamma_{b,ij} - \Gamma_{b,i,j-1}}{\Delta b_{ij}} \end{cases} \tag{13}$$

式中: Δc_{ij} 和 Δb_{ij} 分别为网格单元展向和弦向长度。

作用在气动网格上的气动力可表示为

$$F_{ij} = \Delta p_{ij} S_{ij} n \tag{14}$$

式中: F_{ij} 为气动网格上的气动力向量; S_{ij} 为气动网格面积。

综合式(6)、式(12)、式(13)和式(14)可得气动力向量 F 为

$$F = B_1 w + B_2 \dot{w} + B_3 \Gamma_w \tag{15}$$

式中: B_1 、 B_2 、 B_3 为系数矩阵。

涡格法状态空间方程可完整表达为

$$\begin{cases} \dot{x}_{\text{aero}} = A_{\text{aero}} x_{\text{aero}} + B_{\text{aero}} u_{\text{aero}} \\ \dot{y}_{\text{aero}} = C_{\text{aero}} x_{\text{aero}} + D_{\text{aero}} u_{\text{aero}} \end{cases} \quad (16)$$

式中:状态变量 $x_{\text{aero}} = \Gamma_w$; 输入量 $u_{\text{aero}} = [w \quad \dot{w}]^T$; 输出量 $y_{\text{aero}} = F$; $A_{\text{aero}} = A_a$, $B_{\text{aero}} = [B_a \quad O]$; $C_{\text{aero}} = B_3$; $D_{\text{aero}} = [B_1 \quad B_2]$ 。

需要说明的是,涡格法状态空间形式在推导时采用的速度势方程及边界条件与经典曲面涡格法相同,因此两者的适用范围和限制条件一致。在低马赫数范围(一般小于 0.4)且忽略翼型及翼型厚度影响的情况下,该方法能够取得准确的计算结果。同时,该方法对机翼气动面上反角及变形大小等没有限制。

2 阵风响应与减缓分析

2.1 结构动力学模型

考虑模型适用性要求,采用有限元模态方法建立结构动力学模型,模态坐标下的结构动力学方程为

$$M_s \ddot{q}_s + B_s \dot{q}_s + K_s q_s = F_s \quad (17)$$

式中: M_s 、 B_s 和 K_s 分别为广义质量阵、阻尼阵和刚度阵; q_s 为广义模态坐标; F_s 为广义力。

模态坐标与物理坐标的关系为

$$u_s = \Phi q_s \quad (18)$$

式中: u_s 为物理空间坐标; Φ 为模态矩阵。

2.2 结构气动插值关系

基于三维曲面样条插值理论^[21],可给出有限元模态广义自由度与气动面控制点法向量改变量的关系。设变形空间为 W ,初始构型空间为 X ,变形后构型空间 $Y = W + X$ 。建立构型空间到变形空间的可微映射 $P(X)$:

$$W = P(X) \quad (19)$$

则存在切映射 DP 使得如下关系成立:

$$\chi = DP \cdot v \quad (20)$$

$$\kappa = (I + DP)v \quad (21)$$

式中: v 、 χ 、 κ 分别为 X 、 W 、 Y 空间中切向量。

气动面位移 W_A 插值可由结构位移 W_s 与映射关系求解:

$$W_A = P^{-1}(X_A, X_S)P(X_S)W_s = EW_s \quad (22)$$

式中: E 为位移插值矩阵; P^{-1} 表示对 P 求逆。

同样的,气动面速度 V_A 插值可由结构速度 V_s 与映射关系求解:

$$V_A = EV_s \quad (23)$$

由某一结构位移 u_s 导致的气动面上任意点切向量 τ 的变化 $\Delta\tau$ 同样可由映射关系求解:

$$\Delta\tau = E_{\tau 0} u_s \quad (24)$$

式中: $E_{\tau 0}$ 为物理位移到切向量变化的插值矩阵。

考虑物理位移与模态空间广义自由度关系式(18),则有

$$\Delta\tau = E_{\tau} q_s \quad (25)$$

式中: $E_{\tau} = E_{\tau 0} \Phi u_s$ 为模态空间广义自由度到切向量变化的插值矩阵。

考虑结构变形后空间法向量与 2 个切向量的叉乘关系:

$$n = (\tau_1 + \Delta\tau_1) \times (\tau_2 + \Delta\tau_2) = n_0 + \tau_1 \times \Delta\tau_2 + \Delta\tau_1 \times \tau_2 + \Delta\tau_1 \times \Delta\tau_2 \quad (26)$$

式中: n_0 和 τ_1 、 τ_2 分别为未变形状态下翼面某点处的法向量和 2 个方向切向量。

忽略二阶小量 $\Delta\tau_1 \times \Delta\tau_2$,结合方程(24)可得变形后法向量与模态空间广义自由度关系为

$$n = n_0 + \Delta n = n_0 + E_N q_s \quad (27)$$

式中: E_N 为模态空间广义自由度即结构模态坐标到法向量变化量的插值矩阵。

耦合涡格法状态空间模型与结构动力学方程的关键是将法向下洗速度 w 中与弹性振动相关的部分表达为结构模态坐标 q_s 的函数。考虑弹性变形带来的扰动速度变化及对气动面控制点法向量的影响:

$$w = (V_{\infty} + V_s - V_b) \cdot (n_0 + \Delta n) = V_{\infty} n_0 + V_g n_0 - V_b n_0 + V_{\infty} \Delta n + V_g \Delta n - V_b \Delta n \quad (28)$$

忽略扰动速度与法向量增量相乘形成的高阶项,下洗速度可表达为

$$w = V_{\infty} n_0 + V_g n_0 - V_b n_0 + V_{\infty} \Delta n \quad (29)$$

不考虑刚体运动自由度,壁面运动速度 V_b 完全由弹性振动导致:

$$V_b n_0 = E u_s = E \Phi \dot{q}_s \quad (30)$$

综合式(27)、式(29)和式(30),则有

$$w = V_{\infty} n_0 + V_g n_0 - E \Phi \dot{q}_s + V_{\infty} E_N q_s \quad (31)$$

即可得到法向下洗速度 w 与结构模态坐标 q_s 的耦合关系。

2.3 气动弹性系统状态空间方程

改写式(31)形式为

$$w = R_0 + R_g v_g + R_{1s} q_s + R_{2s} \dot{q}_s \quad (32)$$

式中: $R_0 = V_{\infty} n_0$; $R_g = V_g n_0 v_g^{-1}$; $R_{1s} = E \Phi$; $R_{2s} = V_{\infty} E_N$; v_g 为空间流场中的阵风速度向量。

为便于引入控制面自由度,引入舵面偏转模态 Φ_C ,舵面偏转模态与弹性振动模态处理方式一致,即将舵面弹性偏转带来的气动力变化通过对下洗速度处法向量变化 Δn 及运动速度 V_b 的影响体现,类似有

$$\begin{cases} \mathbf{R}_{1s}^c = \mathbf{E}\Phi_C \\ \mathbf{R}_{2s}^c = \mathbf{V}_\infty \mathbf{E}_N^c \end{cases} \quad (33)$$

则考虑舵面偏转的下洗速度为

$$\mathbf{w} = \mathbf{R}_0 + \mathbf{R}_g \mathbf{v}_g + \mathbf{R}_{1s} \mathbf{q}_s + \mathbf{R}_{2s} \dot{\mathbf{q}}_s + \mathbf{R}_{1s}^c \delta + \mathbf{R}_{2s}^c \dot{\delta} \quad (34)$$

$$\dot{\mathbf{w}} = \mathbf{R}_g \dot{\mathbf{v}}_g + \mathbf{R}_{1s} \dot{\mathbf{q}}_s + \mathbf{R}_{2s} \ddot{\mathbf{q}}_s + \mathbf{R}_{1s}^c \dot{\delta} + \mathbf{R}_{2s}^c \ddot{\delta} \quad (35)$$

式中: δ 为舵面偏转。

结合式(15), 结构广义外力为

$$\mathbf{F}_s = \Phi \mathbf{E}_F (\mathbf{B}_1 \mathbf{w} + \mathbf{B}_2 \dot{\mathbf{w}} + \mathbf{B}_3 \Gamma_w) \quad (36)$$

式中: \mathbf{E}_F 为力插值矩阵。

综合式(34)~式(36)可得

$$\mathbf{F}_s = \mathbf{K}_0 + \mathbf{K}_1 \mathbf{q}_s + \mathbf{K}_2 \dot{\mathbf{q}}_s + \mathbf{K}_3 \ddot{\mathbf{q}}_s + \mathbf{K}_4 \mathbf{v}_g + \mathbf{K}_5 \dot{\mathbf{v}}_g + \mathbf{K}_6 \delta + \mathbf{K}_7 \dot{\delta} + \mathbf{K}_8 \ddot{\delta} + \mathbf{K}_9 \Gamma_w \quad (37)$$

式中: \mathbf{K}_i 为系数矩阵。

结合气动模型(9), 气动方程为

$$\begin{aligned} \dot{\Gamma}_w &= \mathbf{A}_a \Gamma_w + \mathbf{B}_a \mathbf{w} = \mathbf{A}_a \Gamma_w + \mathbf{A}_0 + \mathbf{A}_1 \mathbf{q}_s + \\ &\mathbf{A}_2 \dot{\mathbf{q}}_s + \mathbf{A}_3 \mathbf{v}_g + \mathbf{A}_4 \delta + \mathbf{A}_5 \dot{\delta} \end{aligned} \quad (38)$$

式中: \mathbf{A}_i 为系数矩阵。

综合式(17)、式(37)和式(38)可得气动弹性

状态空间模型为

$$\dot{\mathbf{x}} = \mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{u} \quad (39)$$

$$\mathbf{x} = [\dot{\mathbf{q}}_s \ \mathbf{q}_s \ \Gamma_w]^T \quad (40)$$

$$\mathbf{u} = [\delta \ \dot{\delta} \ \ddot{\delta} \ \mathbf{v}_g \ \dot{\mathbf{v}}_g] \quad (41)$$

考虑结构信号输出, 常用为某传感器安置点加速度信号, 则

$$\mathbf{u}_s = \Phi_r \ddot{\mathbf{q}}_s \quad (42)$$

式中: Φ_r 为传感器位置的模态矩阵。

值得说明的是, 将状态空间模型(39)中的输入项 \mathbf{v}_g 、 $\dot{\mathbf{v}}_g$ 置零, 即可给出指定风速下无阵风激励的气动弹性系统动响应计算结果。在初始小风速条件下, 不断增大来流风速, 计算不同来流风速下的气动弹性系统响应, 直至系统动响应发散, 即可计算得到气动弹性系统颤振速度, 利用该特点可进行气动力建模及气动弹性系统建模准确性的验证。

2.4 阵风模型

离散阵风模型是经典常用的工程阵风扰动模型, 其表现为确定性的风速变化, 可用来表征离散风切变、大气紊流中的峰值及地形诱导气流等。离散阵风风场剖面示意图如图2所示。研究中使用“sin”型离散阵风模型, 该离散阵风模型与“1-cos”型离散阵风模型性质相近, 且易于在风洞试验中进行模拟, 其数学表达式为

$$w_g(l) = \begin{cases} 0 & l < 0 \\ w_{g0} \sin\left(\frac{2\pi l}{L_g}\right) & 0 \leq l \leq L_g \\ 0 & l > L_g \end{cases} \quad (43)$$

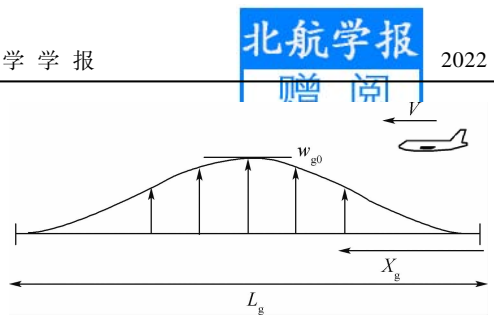


图2 离散阵风风场剖面示意图

Fig.2 Schematic diagram of discrete gust field

$$l = t - \frac{x_g - x_0}{V} \quad (44)$$

式中: w_g 为阵风速度; w_{g0} 为阵风速度幅值; L_g 为阵风尺度; t 为仿真时间; x_g 为飞行器位置; x_0 为飞行器初始位置, 一般可置零; V 为来流速度。

离散阵风指定为一维阵风, 在展向保持均匀不变, 阵风速度垂直于来流方向, 考虑前述状态空间形式表达, 阵风加载形式为

$$\mathbf{v}_g = [0 \ 0 \ w_g(l)]^T \quad (45)$$

连续阵风模型一般指大气紊流, 是通过对大气紊流现象进行理论和试验研究得到的功率谱模型。本文使用更为真实反映大气紊流特性的 von Karman 连续阵风模型, 同样只考虑垂直突风, 其功率谱密度函数为

$$\Phi_w(\omega) = \sigma_w^2 \frac{\tau_w}{\pi} \cdot \frac{1 + \frac{8}{3}(1.339\omega\tau_w)^2}{[1 + (1.339\omega\tau_w)^2]^{11/6}} \quad (46)$$

式中: τ_w 为阵风尺度与来流风速的比值; σ_w 为阵风均方根速度; ω 为圆频率。

将白噪声信号通过符合功率密度函数的成形滤波器即可给出时域内的阵风速度 w_g , 阵风加载形式与式(45)一致。

2.5 控制系统设计

PID 系统具有结构简单、易于工程实现的特点, 因此, 控制减缓分析基于经典 PID 控制律进行。由于真实气动弹性系统不可避免地存在环境干扰, 噪声信号通常比系统响应变化更为剧烈, 微分环节降低了系统的信噪比, 利于稳定控制。因此, 一般情况下仅使用 PI 控制, 并在反馈回路中加入低通滤波器, 进一步过滤高频信号。由于系统的输出为加速度信号, 经过一次积分为速度信号, 形成速度反馈, 这对气动弹性系统具有气动阻尼作用, 可以加快系统响应衰减。另外, 纯积分环节具有实部为零的极点, 控制系统临界稳定, 微小的非零均值即可导致控制系统失稳, 因此积分环节由惯性环节来实现。

针对离散与连续阵风, 选取翼尖垂向加速度作为反馈信号, 通过控制增益并转化为控制信号传给舵机, 驱动控制面, 控制框图如图3所示。舵

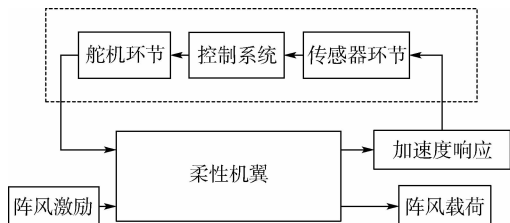


图 3 阵风减缓控制方案

Fig. 3 Gust alleviation control scheme

机环节及传感器环节均假设为理想环节。

阵风减缓控制系统的效果评估用阵风减缓率 η 描述,形式如下:

$$\eta = \frac{a_{open} - a_{close}}{a_{open}} \times 100\% \quad (47)$$

式中: a_{open} 为开环状态最大输出响应; a_{close} 为闭环最大输出响应。

3 算例分析

3.1 柔性机翼模型

采用的柔性机翼模型为以铝制十字梁为主梁的大展弦比机翼模型。设计模型及有限元模型如图 4 和图 5 所示。利用第 1 节状态空间形式涡格法建立气动模型,气动面如图 6 所示。

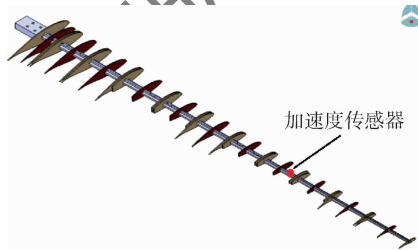


图 4 柔性机翼 CAD 模型

Fig. 4 CAD model of flexible wing

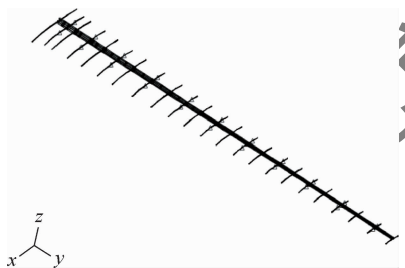


图 5 柔性机翼有限元模型

Fig. 5 Finite element model of flexible wing

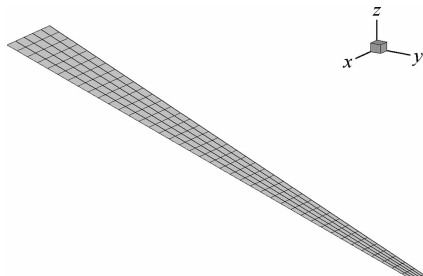


图 6 柔性机翼气动面模型

Fig. 6 Aerodynamics model of flexible wing

机翼主梁为翼尖到翼根线性增大的十字梁,主梁位置位于弦长的 40%。主梁在 MSC. Nastran 中以 CBEAM 梁单元模拟,翼肋以大刚度梁单元模拟,质量特性以集中质量点模拟。加速度传感器布置于距翼尖 1/3 翼展处。柔性机翼模型设计参数如表 1 所示。

柔性翼控制面面积为 $1.68 \times 10^{-5} \text{ m}^2$,平均弦长为 0.04 m,据翼根距离 0.804 5 m,机翼模型主要模态信息如表 2 所示。

表 1 柔性机翼设计参数

Table 1 Design parameters of the wing model

参数	数值
质量/kg	2.819 9
半展长/m	1.542
翼根弦长/m	0.261
翼梢弦长/m	0.069
扭转角/(°)	-2.0
展弦比	9.3
翼型	超临界翼型

表 2 柔性机翼模型主要模态信息

Table 2 Modes analysis results of the wing model

模态阶数	模态频率/Hz	模态特征
1	3.35	垂直一弯
2	5.18	水平一弯
3	10.05	垂直二弯
4	17.17	水平二弯
5	25.68	垂直三弯
6	33.05	一阶扭转

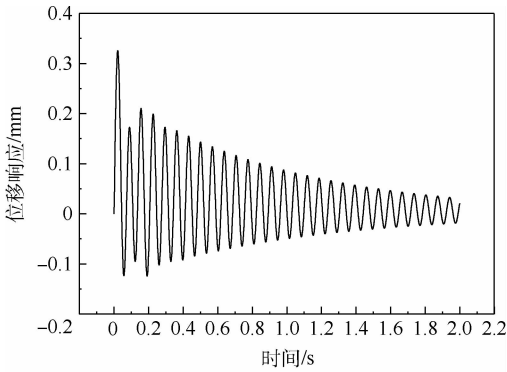
3.2 颤振分析

以海平面密度为基准,将阵风输入项置零,自 10 m/s 来流速度起始不断增大来流速度输入计算气动弹性系统动响应,观察时域响应曲线发散情况。来流风速 51.5 m/s、51.7 m/s、51.8 m/s 及 52.0 m/s 四种情况下的机翼时域响应曲线如图 7 所示。在 51.5 m/s 来流风速下,气动弹性系统阻尼较大,风速为 51.7 m/s 时,系统阻尼减小。当来流速度达到 51.8 m/s 时,机翼出现响应发散现象,风速为 52.0 m/s 时,发散响应明显。仿真计算结果显示,该机翼颤振速度为 51.8 m/s。

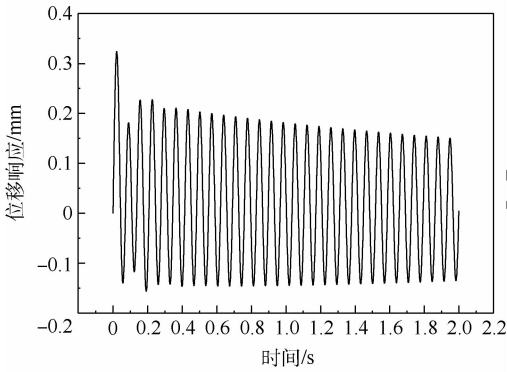
图 8 给出该机翼模型在 ZAERO (ZONA6 亚声速气动模型) 中颤振分析得到的 V - g 曲线及 V - f 曲线。可见,机翼模型颤振速度为 52.0 m/s 时为标准的弯扭耦合颤振。ZAERO 计算得到的颤振速度与基于状态空间涡格法建立的气动弹性系统响应方程计算结果一致性很好,基于状态空间涡格法建立的气动弹性系统响应模型准确,可以用于气动弹性响应分析求解。

3.3 阵风响应分析

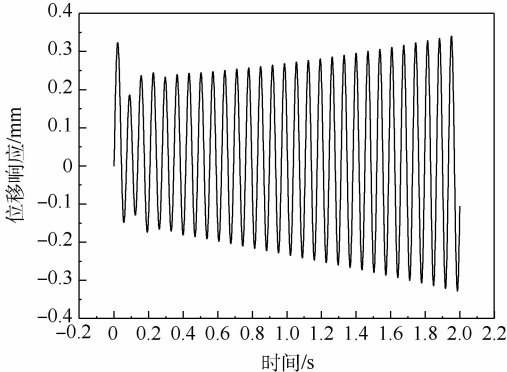
以海平面密度为基准,给定前述离散阵风



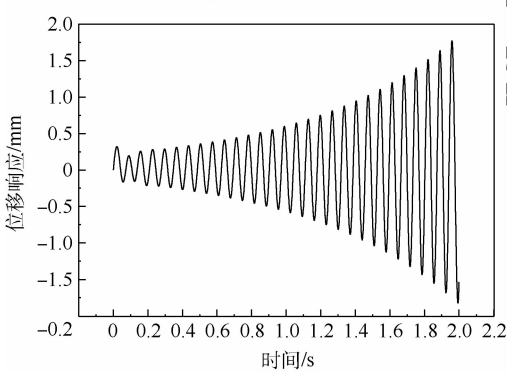
(a) 来流速度51.5 m/s



(b) 来流速度51.7 m/s



(c) 来流速度51.8 m/s

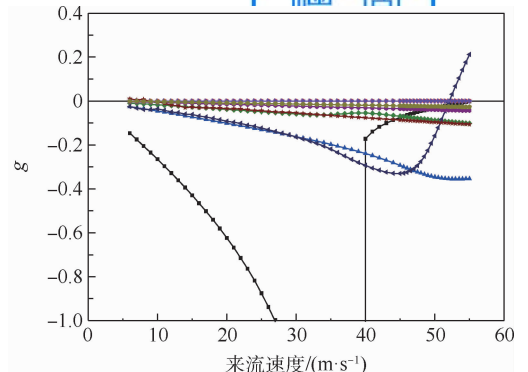


(d) 来流速度52.0 m/s

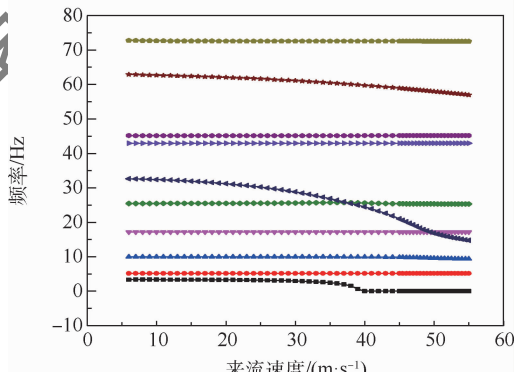
图7 不同风速下位移时域响应

Fig. 7 Displacement response with different velocities

模型作为系统激励,阵风响应输出点为加速度传感器位置处。以 32 m/s 风速及 28 m/s 为例,频率 3.6 Hz 阵风激励下,加速度及位移响应计算情况如图 9 ~ 图 12 所示。阵风速度幅值给定 505 mm/s。



(a) V-g 曲线



(b) V-f 曲线

- mode 1
- mode 2
- mode 3
- mode 4
- mode 5
- mode 6
- mode 7
- mode 8
- mode 9
- mode 10

图8 ZAERO 颤振分析结果

Fig. 8 Flutter analysis results from ZAERO

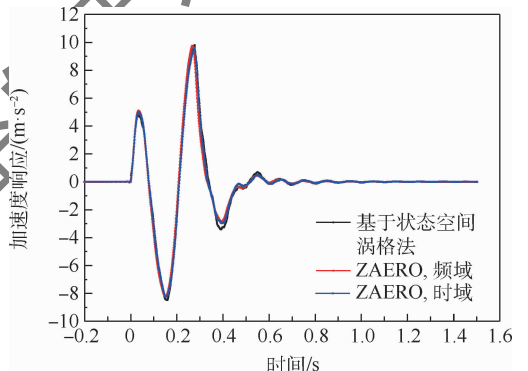


图9 风速 32 m/s 下加速度时域响应

Fig. 9 Acceleration response in time domain ($V = 32$ m/s)

由时域仿真结果可见,基于状态空间涡格法的阵风建模方法与 ZAERO 频域及时域计算方法的对比,结果表明准确度很高。表 3 及表 4 给出了 2 种工况下最大响应幅值对比及误差情况,其中 ZAERO 计算结果为时域方法计算结果。基于状态空间涡格法的计算值略大于 ZAERO 计算值,加速度响应幅值最大误差不超过 3.2%,位移响应幅值最大误差不超过 5.6%。ZAERO 频域计算方法得到的结果相比时域计算方法要更接近于基于状态空间涡格法得到的计算结果。

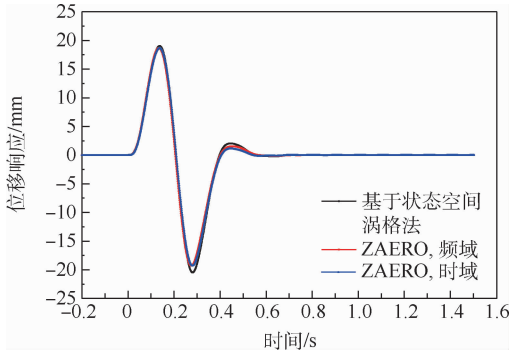


图 10 风速 32 m/s 下位移时域响应

Fig. 10 Displacement response in time domain ($V = 32 \text{ m/s}$)

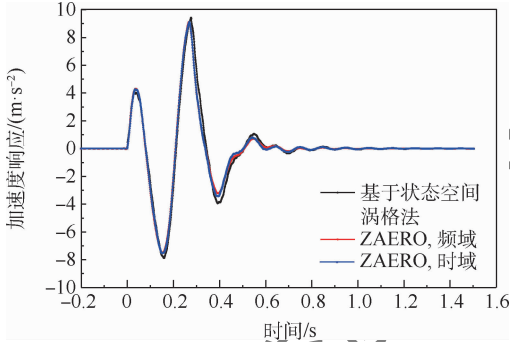


图 11 风速 28 m/s 下加速度时域响应

Fig. 11 Acceleration response in time domain ($V = 28 \text{ m/s}$)

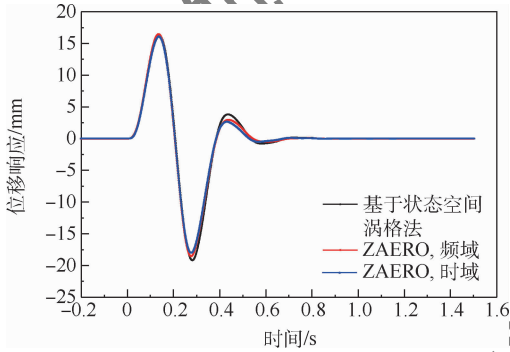


图 12 风速 28 m/s 下位移时域响应

Fig. 12 Displacement response in time domain ($V = 28 \text{ m/s}$)

表 3 风速 32 m/s 下计算结果对比

Table 3 Results contrast with 32 m/s velocity

方法	位移响应/mm	加速度响应/($\text{m} \cdot \text{s}^{-2}$)
状态空间涡格法	20.45	9.80
ZAERO	19.37	9.59
偏差/%	5.6	2.2

表 4 风速 28 m/s 下计算结果对比

Table 4 Results contrast with 28 m/s velocity

方法	位移响应/mm	加速度响应/($\text{m} \cdot \text{s}^{-2}$)
状态空间涡格法	19.18	9.43
ZAERO	18.32	9.14
偏差/%	4.7	3.2

3.4 阵风减缓分析

采用经典 PID 控制方法构建阵风减缓控制系统,其控制参数如图 13 所示。其中, K_p 、 K_i 、 K_d 分别为比例、积分、微分环节的增益系数。在控制环

节前增加低通滤波器,滤掉高频信号。

在离散阵风激励情况下,仍以风速 32 m/s 及 28 m/s、阵风频率 3.6 Hz 为例,给出开/闭环状态下机翼加速度及位移时域响应仿真结果。仿真计算结果如图 14 ~ 图 17 所示。

风速为 32 m/s 时,开环状态下,机翼加速度

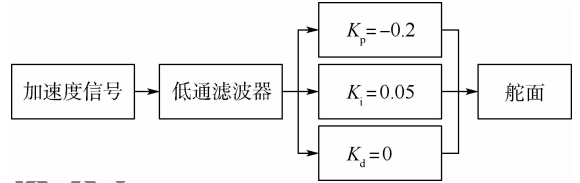


图 13 PID 控制参数

Fig. 13 PID controller parameters

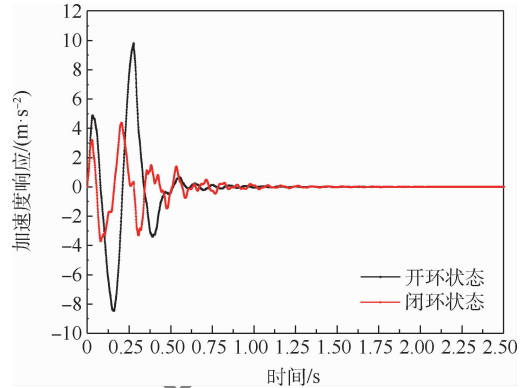


图 14 风速 32 m/s 下开/闭环时域加速度响应

Fig. 14 Open/close loop acceleration response ($V = 32 \text{ m/s}$)

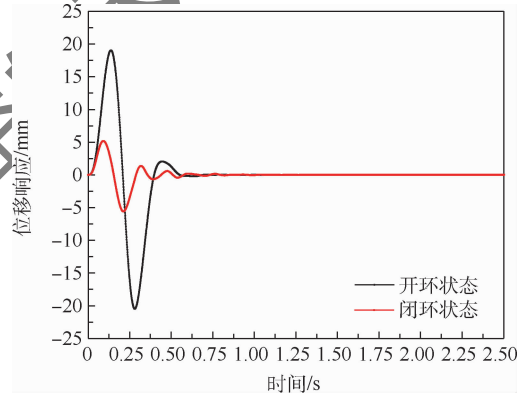


图 15 风速 32 m/s 下开/闭环时域位移响应

Fig. 15 Open/close loop displacement response ($V = 32 \text{ m/s}$)

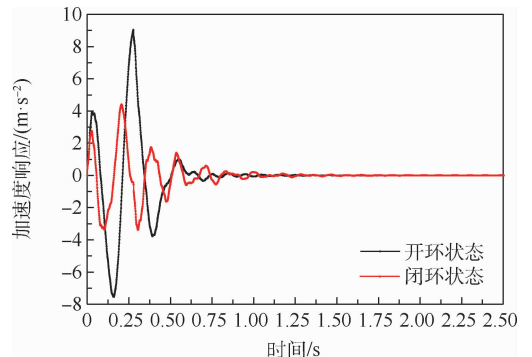


图 16 风速 28 m/s 下开/闭环时域加速度响应

Fig. 16 Open/close loop acceleration response ($V = 28 \text{ m/s}$)

最大响应幅值接近 9.80 m/s^2 , 打开阵风减缓控制系统后, 机翼加速度最大响应幅值为 4.37 m/s^2 , 减缓率接近 55.4%。位移减缓率接近 70.7%。风速为 28 m/s 时, 开环状态下, 机翼加速度最大响应幅值接近 9.43 m/s^2 , 打开阵风减缓控制系统后, 机翼加速度最大响应幅值为 4.41 m/s^2 , 减缓率接近 53.2%, 位移减缓率接近 72.7%。结果表明, 控制系统可以有效减缓加速度传感器处的结构响应。

在连续阵风激励下, 以来流速度 28 m/s 为例, 给定 von Karman 连续阵风速度均方根为 2.0 m/s , 阵风尺度为 760.0 m 。计算开/闭环状态下机翼加速度及位移时域响应, 图 18 和图 19 为

仿真计算结果。结果显示, 对于连续阵风激励, 控制系统同样可以减缓加速度传感器处的结构响应, 对于位移响应信号, 减缓效果显著; 对于加速度响应信号, 均方根值减缓率接近 21%。

4 结论

1) 本文基于涡格法状态空间形式建立了一套新的阵风响应分析框架, 弥补了传统响应分析方法需要进行有理函数拟合或时域迭代求解效率低等缺陷。该方法能够应用于复杂飞行器模型的气动弹性响应分析计算, 具有重要的参考价值。

2) 将本文方法的计算结果与经典气动弹性分析软件 ZAERO 颤振及离散阵风响应分析结果进行了对比, 结果表明计算误差很小, 结果准确可靠。

3) 本文基于涡格法状态空间形式建立了一套阵风减缓分析方法。基于经典 PID 控制方法设计阵风减缓控制律, 仿真计算结果显示, 阵风减缓控制系统能够有效减缓结构响应。

参考文献 (References)

- [1] 杨超. 飞行器气动弹性原理[M]. 2版. 北京: 北京航空航天大学出版社, 2016.
YANG C. Aerelastic theory of aircraft[M]. 2nd ed. Beijing: Beihang University Press, 2016 (in Chinese).
- [2] 金长江, 肖业伦. 大气扰动中的飞行原理[M]. 北京: 国防工业出版社, 1992.
JIN C J, XIAO Y L. Flight principle with atmosphere turbulence[M]. Beijing: National Defence Industry Press, 1992 (in Chinese).
- [3] SU W H, CESNIK C E S. Dynamic response of highly flexible flying wing[J]. AIAA Journal, 2011, 49(2): 324-339.
- [4] AN C, YANG C, XIE C C, et al. Flutter and gust response analysis of a wing model including geometric nonlinearities based on a modified structural ROM[J]. Chinese Journal of Aeronautics, 2020, 33(1): 48-63.
- [5] NOTT T E, BROWN J M, PEREZ-DAVIS M E, et al. Investigation of the helios prototype aircraft mishap: HQ 04-283[R]. Hampton: NASA Langley Research Center, 2004.
- [6] KARA J, PLOTKIN A. Low-speed aerodynamics: From wing theory to panel method[M]. Cambridge: Cambridge University Press, 2001.
- [7] AZOULAY D, KARPEL M. Characterization of method for computation of aeroservoelastic response to gust excitation: AIAA-2006-1938[R]. Reston: AIAA, 2006.
- [8] KARPEL M, MOULIN B, CHEN P C. Dynamic response of aeroservoelastic systems to gust excitation[J]. Journal of Aircraft, 2005, 42(5): 1264-1272.
- [9] 陈磊, 吴志刚, 杨超, 等. 多控制面机翼阵风减缓主动控制与风洞试验验证[J]. 航空学报, 2009, 30(12): 2250-2256.
CHEN L, WU Z G, YANG C, et al. Active control and wind

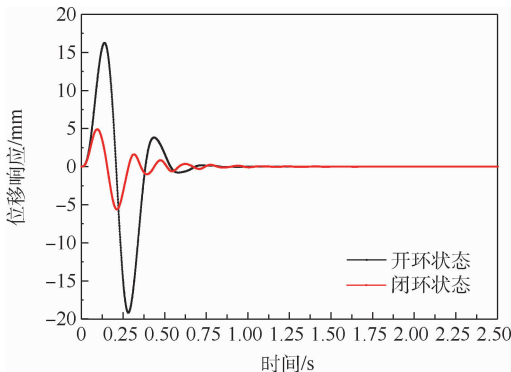


图 17 风速 28 m/s 下开/闭环时域位移响应

Fig. 17 Open/close loop displacement response ($V = 28 \text{ m/s}$)

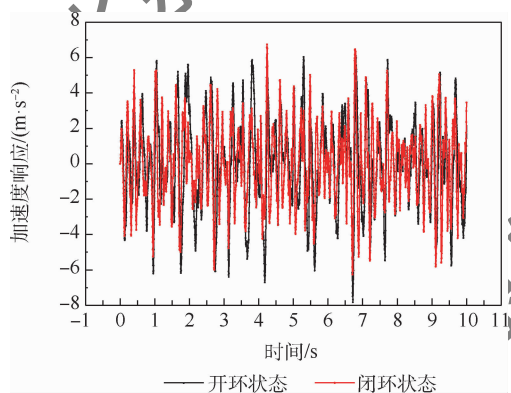


图 18 连续阵风时域加速度响应

Fig. 18 Acceleration response with continuum gust

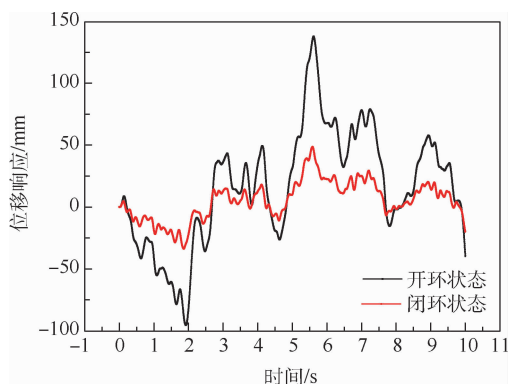


图 19 连续阵风时域位移响应

Fig. 19 Displacement response with continuum gust

- tunnel test verification of multi-control surfaces wing for gust alleviation[J]. *Acta Aeronautica et Astronautica Sinica*, 2009, 30(12):2250-2256 (in Chinese).
- [10] WANG Z, CHEN P C, LIU D D, et al. Nonlinear-aerodynamics/nonlinear-structure interaction methodology for a high-altitude long-endurance wing [J]. *Journal of Aircraft*, 2010, 47(2): 556-566.
- [11] SU W H, CESNIK C E S. Dynamic response of highly flexible flying wings[C]//Proceeding of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. Reston: AIAA, 2006:412-435.
- [12] LIU Y, XIE C C, YANG C, et al. Gust response analysis and wind tunnel test for a high-aspect ratio wing[J]. *Chinese Journal of Aeronautics*, 2016, 29(1):91-103.
- [13] GUO D, XU M, CHEN S L. Nonlinear gust response analysis of free flexible aircraft[J]. *International Journal of Intelligent System Application (IJISA)*, 2013, 5(2):1-15.
- [14] WERTER N P M, DE BREUKER R. A novel dynamic aeroelastic framework for aeroelastic tailoring and structural optimisation [J]. *Composite Structures*, 2016, 158:369-386.
- [15] WERTER N P M, DE BREUKER R, ABDALLA M M. Continuous-time state-space unsteady aerodynamic modeling for efficient loads analysis[J]. *AIAA Journal*, 2018, 56(3):905-916.
- [16] WU Z G, CHEN L, YANG C, et al. Gust response modeling and alleviation scheme design for an elastic aircraft [J]. *Science China Technological Sciences*, 2010, 53(11):3110-3118.
- [17] 聂雪媛, 杨国伟. 基于 CFD 降阶模型的阵风减缓主动控制研究[J]. *航空学报*, 2015, 36(4):1103-1111.
- NIE X Y, YANG G W. Gust alleviation active control based on CFD reduced-order models[J]. *Acta Aeronautica et Astronautica Sinica*, 2015, 36(4):1103-1111 (in Chinese)
- [18] LIU X, SUN Q, COOPER J E. LQG based model predictive control for gust load alleviation[J]. *Aerospace Science and Technology*, 2017, 71:499-509.
- [19] YAGHI L, RAVEH D E, IDAN M. Deformation control of highly flexible aircraft in trimmed flight and gust encounter [J]. *Journal of Aircraft*, 2018, 55(2):829-840.
- [20] DAY Y T, YANG C, WANG C L. Strategy for robust gust response alleviation of an aircraft mode [J]. *Control Engineering Practice*, 2017, 60(3):211-217.
- [21] 谢长川. 飞行器气动弹性稳定性静/动耦合理论与试验研究[D]. 北京:北京航空航天大学, 2011.
- XIE C C. Static/dynamics coupling theory and test study of aircraft aeroelastic stability [D]. Beijing:Beihang University, 2011 (in Chinese).

Gust load alleviation analysis based on vortex lattice method in state-space form

YANG Lan, AN Chao*, XIE Changchuan, YANG Chao

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: Gust response and gust load alleviation control system design is an important issue in aeroelasticity. This paper presents a gust response model based on the vortex lattice method (VLM) in state-space formulation and gives the couple relationships between finite element method modes/control surface modes and boundary conditions of VLM, which can be applied to complicated aircraft model. This method can avoid the disadvantages of the traditional gust response analysis method with no requirement of rational function assessment and iteration calculation with lots of resources. Introducing the traditional PID control algorithm, a gust load alleviation system is given, and gust time response of open loop/closed loop under a discrete gust and von Karman continuum gust excitation are presented. The alleviation effect can be solved by contrasting the response amplitude. The simulation results show that gust response analysis results based on this method are accurate and the gust load alleviation control system can alleviate the load response of the original aeroelastic system effectively.

Keywords: aeroelasticity; aeroservoelasticity; gust response; gust load alleviation control system; vortex lattice method (VLM)

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

DOI: 10.13700/j.bh.1001-5965.2021.0021

磁悬浮转子状态反馈解耦自抗扰控制方法

尹增愿¹, 蔡远文², 任元^{2,*}, 王卫杰², 陈晓岑³, 于春森¹

(1. 航天工程大学 研究生院, 北京 101416; 2. 航天工程大学 宇航科学与技术系, 北京 101416;
3. 陆军研究院, 北京 100190)



摘 要: 针对磁悬浮控制敏感陀螺(MSCSG)转子偏转通道强耦合及航天器姿态测量过程中受扰失稳问题,提出了一种磁悬浮转子偏转解耦抗干扰控制方法。分析了转子两自由度偏转耦合现象,设计了基于状态反馈的解耦控制器;建立了 MSCSG 在姿态测量过程中航天器的姿态运动对磁悬浮转子产生的干扰力矩模型,采用自抗扰控制器(ADRC)抑制磁悬浮转子的外部干扰;对所建立的扩展状态观测器(ESO)跟踪性和系统稳定性进行了分析,通过调节 ADRC 中非线性状态误差反馈控制律系数,实现了系统有界输入条件下的稳定。仿真结果表明:状态反馈解耦能够实现偏转自由度的完全解耦,ESO 具有良好的跟踪性能,ADRC 较传统 PID 控制方法具有更好的抗干扰性能。

关键词: 磁悬浮控制敏感陀螺(MSCSG); 磁悬浮转子; 姿态测量; 状态反馈解耦; 自抗扰控制(ADRC)

中图分类号: V448.2

文献标志码: A

文章编号: 1001-5965(2022)07-1210-12

随着新一代高精度对地观测和遥感卫星的快速发展,对航天器的超稳、超精、超敏姿态控制系统提出了更高的要求^[1-2]。一方面,传统机械结构的执行器由于振动、摩擦等干扰很难实现航天器高精度姿态控制^[3-4];另一方面,传统的控制系统采用执行机构和测量单元串联大闭环结构,导致控制力矩的输出滞后于姿态的测量^[5]。为解决传统机械结构和控制系统带来的问题,磁悬浮控制敏感陀螺(magnetically suspended control and sensitive gyroscope, MSCSG)是现有解决航天器姿态控制系统缺陷的理想装置。

MSCSG 是一种航天器姿态控制与姿态测量于一体的装置^[6-7]。MSCSG 的三自由度平动采用主动磁轴承控制,两自由度偏转采用洛伦兹力磁轴承控制。由于洛伦兹力磁轴承(Lorentz force magnetic bearing, LFMB)力/力矩和电流的线性

关系, MSCSG 在输出两自由度控制力矩的同时还能够实现两自由度姿态角速率的测量。由于陀螺效应的存在,磁悬浮转子的两自由度偏转存在耦合现象,导致转子很难实现高精度控制。此外,当航天器高动态运动,需要 MSCSG 转子偏转快速响应,否则会使转子的相对运动过大,转子出现碰撞保护轴承的现象,导致转子失稳。但是,当实现转子偏转快速响应时,其抗扰性能很难得到保证^[8]。

为有效抑制陀螺效应的影响,任元等^[9-10]提出陀螺效应控制的稳定性判据。以磁悬浮控制力矩陀螺为研究对象,北京航空航天大学房建成团队提出采用逆系统控制、模态分离等一系列高稳定、快响应的磁悬浮转子控制方法^[11-12],但是该方法需要大量运算,增加了控制系统的复杂性。基于陀螺效应的解耦方法,传统解耦方法有滤波交叉 PID 控制解耦,该方法在分散 PID 控制基础

收稿日期: 2021-01-14; 录用日期: 2021-03-28; 网络出版时间: 2021-04-19 08:29

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210415.1540.002.html

基金项目: 国家自然科学基金(52075545, 11802342); 北京市“高创计划”青年拔尖人才项目(2017000026833ZK23)

* 通信作者: E-mail: renyuan_823@aliyun.com

引用格式: 尹增愿, 蔡远文, 任元, 等. 磁悬浮转子状态反馈解耦自抗扰控制方法[J]. 北京航空航天大学学报, 2022, 48(7): 1210-1221. YIN Z Y, CAI Y W, REN Y, et al. Decoupled active disturbance rejection control method for magnetically suspended rotor based on state feedback[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1210-1221 (in Chinese).

上引入交叉控制项对耦合系统解耦,但是其解耦性能和抗干扰性能有待提高^[13-14]。文献[15-16]提出一种模糊逻辑控制方法,但是其最佳应用场合是大滞后非精确模型控制。此外,为了实现磁悬浮转子两自由度偏转解耦控制,还有 LQR 控制方法^[17]、Cholesky 分解降秩方法^[18]、最小二乘支持向量积解耦方法^[19]、微几何方法^[20]、预测模型控制方法^[21]等。

MSCSG 偏转控制是一个复杂的多效应扰动系统,除需对两自由度解耦外,所设计控制器必须要具有强鲁棒性、快响应、抗扰动性^[22-23]。PID 方法原理简单,便于实现,但机动性差,对指令信号跟踪相位滞后;滑模变结构控制具有快响应、参数变化不敏感优点,但在变结构切换过程中容易引起系统抖振;自适应控制的抗扰动性能有一定的提高,但是通过对误差信号简单处理将导致系统的控制精度降低和动态品质变差,无法实现高精度控制^[24]。因此,上述控制方法无法同时对转子偏转的快速响应和扰动抑制。

基于磁悬浮转子系统的偏转自由度耦合及受扰问题,本文结合磁悬浮转子的实际工作状况,提出了一种状态反馈解耦自抗扰控制(active disturbance rejection control, ADRC)方法。该方法通

过状态反馈实现了磁悬浮转子偏转两自由度的完全解耦,结合 ADRC 实现了磁悬浮转子的抗扰动控制^[25]。ADRC 采取了 PID 误差反馈控制的核心理念。本文所提状态反馈解耦方法能够实现磁悬浮转子偏转自由度的完全解耦,采用 ADRC 实现了转子的抗干扰和快响应控制。

1 MSCSG 工作原理

15 N·m·s 的 MSCSG 结构如图 1 所示。其主要由转子、径向磁轴承、轴向磁轴承、LFMB、高精度电涡流位移传感器、陀螺房、电机等组成。其中,径向磁轴承控制转子的径向运动,轴向磁轴承控制转子的轴向运动,LFMB 控制转子两自由度偏转运动,电涡流位移传感器检测转子的位移和角位移信号,陀螺房安装固定定子。磁悬浮转子采用双球面结构,轴向磁轴承和径向磁轴承磁极均采用球面结构。转子能够实现 ±0.3° 偏转,通过转子的偏转能够实现输出两自由度控制力矩,从而控制航天器姿态。

本文定义了如下坐标系:

惯性坐标系 $o_i-x_iy_iz_i$;选用地心赤道坐标系代替惯性坐标系,原点 o_i 位于地心, x_i 和 z_i 轴分别指向春分点和北极,并与 y_i 轴构成右手坐标系。

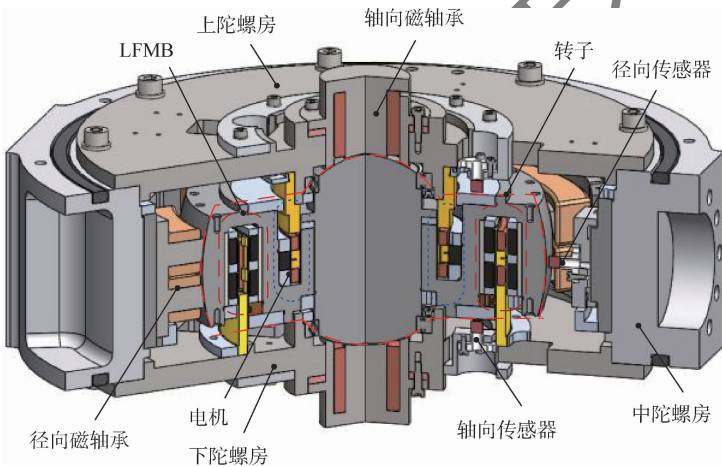


图 1 15 N·m·s 磁悬浮控制敏感陀螺截面示意图

Fig. 1 Section diagram of 15 N·m·s MSCSG

转子坐标系 $o_r-x_r y_r z_r$; x_r 和 y_r 轴位于转子赤道面内且相互垂直, z_r 轴和转子转轴重合,符合右手螺旋定则。

定子坐标系 $o_s-x_s y_s z_s$; x_s 和 y_s 轴位于径向磁轴承定子赤道面内且垂直指向径向磁轴承定子磁极面, z_s 轴垂直于径向磁轴承定子赤道面。

假设在惯性坐标系下磁悬浮转子的姿态角速度和角加速度为 $(\dot{\alpha}, \dot{\beta}, \dot{\Omega})$ 和 $(\ddot{\alpha}, \ddot{\beta}, \ddot{\Omega})$ 。根据转子动力学可得转子偏转动力学方程如下:

$$\begin{bmatrix} J_x \ddot{\beta} + J_z \Omega \dot{\alpha} \\ J_y \ddot{\alpha} - J_z \Omega \dot{\beta} \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \end{bmatrix} \quad (1)$$

式中: T_x 和 T_y 为转子运动产生的两自由度控制力矩; $\dot{\alpha}$ 指向 x_i 轴, $\dot{\beta}$ 指向 y_i 轴, $\dot{\Omega}$ 指向 z_i 轴; J_x 、 J_y 和 J_z 为转子绕各轴的转动惯量, $J_x = J_y$ 。

假设在惯性坐标系下航天器的姿态角速度和姿态角加速度为

$$\begin{cases} \omega = (\omega_x \ \omega_y \ \omega_z) \\ \dot{\omega} = (\dot{\omega}_x \ \dot{\omega}_y \ \dot{\omega}_z) \end{cases} \quad (2)$$

当 MSCSG 安装在航天器上,此时航天器和 MSCSG 陀螺房固联,则磁悬浮转子在惯性坐标系下的姿态角速度由 2 部分组成,即转子相对于陀螺房的姿态角速度和陀螺房在惯性坐标系下的姿态角速度。因此,转子相对于航天器的姿态角速度和角加速度可表示为

$$\begin{cases} \boldsymbol{\omega}_r = (\boldsymbol{\omega}_x + \dot{\alpha} & \boldsymbol{\omega}_y + \dot{\beta} & \boldsymbol{\omega}_z + \Omega) \\ \dot{\boldsymbol{\omega}}_r = (\dot{\boldsymbol{\omega}}_x + \ddot{\alpha} & \dot{\boldsymbol{\omega}}_y + \ddot{\beta} & \dot{\boldsymbol{\omega}}_z + \dot{\Omega}) \end{cases} \quad (3)$$

由于 $\Omega \gg \omega_z$,在忽略航天器 Z 轴运动时,联立式(1)和式(3)可得陀螺技术方程为

$$\begin{bmatrix} J_x(\dot{\omega}_y + \ddot{\beta}) + J_z\Omega(\omega_x + \dot{\alpha}) \\ J_y(\dot{\omega}_x + \ddot{\alpha}) - J_z\Omega(\omega_y + \dot{\beta}) \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \end{bmatrix} \quad (4)$$

令

$$\begin{cases} T_{rx} = J_x\dot{\omega}_y + J_z\Omega\omega_x \\ T_{ry} = J_y\dot{\omega}_x - J_z\Omega\omega_y \end{cases} \quad (5)$$

式中: T_{rx} 和 T_{ry} 为航天器运动对磁悬浮转子施加的力矩。

MSCSG 在航天器姿态测量过程中,其控制转子偏转的装置是 LFMB,结构示意图如图 2 所示^[26]。LFMB 根据通电导体在磁场中受电磁力的原理制成。当相对位置的绕组线圈电流大小相等、方向相反时,其产生的电磁力大小相等,方向相反,组成一对力偶,此力偶控制转子的偏转。LFMB 产生的力偶可表示为

$$\begin{cases} M_x = (2nBi_xL)2l_m \\ M_y = (2nBi_yL)2l_m \end{cases} \quad (6)$$

式中: M_x 、 M_y 为 LFMB 产生的控制力矩; i_x 、 i_y 分

别为线圈中驱动转子绕 x_r 、 y_r 轴偏转的激励电流; n 为线圈的匝数; B 为磁场强度; L 为单匝线圈周向长度; l_m 为线圈所在圆周的半径。

当航天器机动,高速旋转的磁悬浮转子定轴性遭到破坏,转子相对于定子会产生运动,此时,LFMB 产生的力矩能够抵抗转子的相对运动,有

$$\begin{cases} T_x = M_x \\ T_y = M_y \end{cases} \quad (7)$$

联立式(4)~式(6)可得 MSCSG 转子平衡方程为

$$\begin{bmatrix} J_x\ddot{\beta} + J_z\Omega\dot{\alpha} + T_{rx} \\ J_y\ddot{\alpha} - J_z\Omega\dot{\beta} + T_{ry} \end{bmatrix} = \begin{bmatrix} k_T i_x \\ k_T i_y \end{bmatrix} \quad (8)$$

式中: $k_T = 4nBLl_m$ 为 LFMB 的扭转刚度。

分析式(5)和式(8),可以得到如下结论:转子两自由度偏转之间存在相互耦合,严重影响磁悬浮转子控制精度;当 MSCSG 姿态测量时,航天器的姿态运动可以看成对 MSCSG 转子施加干扰力矩 T_{rx} 和 T_{ry} 。通过对干扰力矩的分析可知,随着航天器姿态角或者姿态角速度的增大,干扰力矩逐渐增大,干扰力矩导致转子轴心轨迹跳动量增大,严重情况下导致转子碰撞保护轴承,影响转子的稳定性。

基于磁悬浮转子姿态测量系统的偏转自由度耦合及干扰力矩扰动问题,必须要实现磁悬浮转子的解耦抗干扰控制。本文提出了一种状态反馈解耦自抗扰控制方法。通过状态反馈的方法实现磁悬浮转子偏转两自由度的完全解耦,结合 ADRC 实现磁悬浮转子的抗扰动控制。

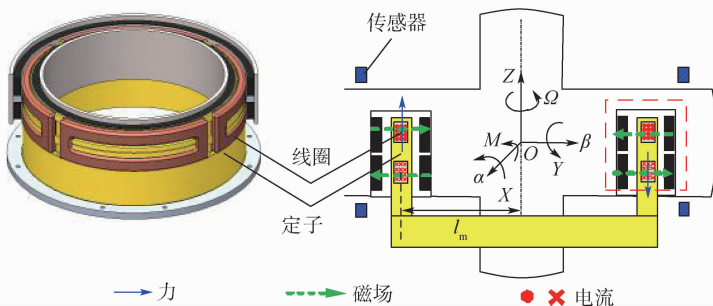


图 2 洛伦兹力磁轴承结构示意图

Fig. 2 Structure diagram of LFMB

2 状态反馈解耦自抗扰控制器设计

2.1 状态反馈解耦设计

根据式(8)可知,当两自由度偏转输入信号为 0 时,在不考虑干扰力矩扰动的情况下,对式(8)进行拉普拉斯变换可得

$$\begin{cases} J_x s^2 \beta(s) + J_z \Omega s \alpha(s) = k_T i_x(s) e^{-\tau s} \\ J_y s^2 \alpha(s) - J_z \Omega s \beta(s) = k_T i_y(s) e^{-\tau s} \end{cases} \quad (9)$$

式中: $e^{-\tau s}$ 为 LFMB 控制系统的延时现象对电流作用的效果。

根据式(9)可求得磁悬浮转子偏转角度和控制电流之间的关系为

$$\begin{bmatrix} \alpha(s) \\ \beta(s) \end{bmatrix} = e^{-\tau s} \begin{bmatrix} \frac{k_T H}{J_x J_y s^3 + H^2 s} & \frac{k_T J_x}{J_x J_y s^2 + H^2} \\ \frac{k_T J_y}{J_x J_y s^2 + H^2} & -\frac{k_T H}{J_x J_y s^3 + H^2 s} \end{bmatrix} \begin{bmatrix} i_x(s) \\ i_y(s) \end{bmatrix} \quad (10)$$

式中: $H = J_z \Omega$ 为转子的角动量。

根据式(10),可以求得偏转自由度传递函数矩阵为

$$\begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} = \begin{bmatrix} \frac{k_T H}{J_x J_y s^3 + H^2 s} & \frac{k_T J_x}{J_x J_y s^2 + H^2} \\ \frac{k_T J_y}{J_x J_y s^2 + H^2} & -\frac{k_T H}{J_x J_y s^3 + H^2 s} \end{bmatrix} e^{-\tau s} \quad (11)$$

根据转子两自由度偏转动力学方程,可以得到两自由度偏转通道的状态空间表达式为

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \quad (12)$$

式中:状态量 $x = [\alpha \quad \dot{\alpha} \quad \beta \quad \dot{\beta}]^T$;控制量 $u = [i_x \quad i_y]^T$;控制矩阵 $B = \begin{bmatrix} 0 & 0 & 0 & k_T/J_x \\ 0 & k_T/J_y & 0 & 0 \end{bmatrix}^T$;

输出矩阵 $C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$;状态矩阵

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & J_z \Omega / J_y \\ 0 & 0 & 0 & 1 \\ 0 & -J_z \Omega / J_x & 0 & 0 \end{bmatrix}。$$

为实现磁悬浮转子两自由度偏转解耦控制,本文通过状态反馈实现偏转自由度的解耦设计。状态反馈解耦是通过状态量的反馈实现耦合系统的完全解耦。反馈状态量选取直接影响解耦性能。

状态反馈解耦自抗扰控制流程如图 3 所示。图中: $G(s)$ 为转子的动力学的传递函数, $u(s)$ 为反馈解耦控制率传递函数,解耦器 $u(s)$ 与被控对象 $G(s)$ 构成广义被控对象 $N(s)$ 。ADRC 包括 $C_1(s)$ 和 $C_2(s)$, $C_1(s)$ 和 $C_2(s)$ 主要用来控制两自

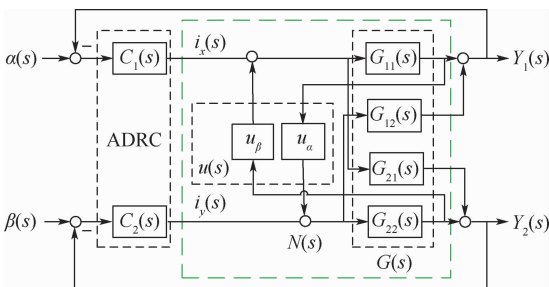


图 3 磁悬浮转子控制系统

Fig. 3 Magnetically suspended rotor control system

由度磁悬浮转子偏转。采用 ADRC 的主要优点是能够提高整个系统的抗干扰性能。此外,由于转子的固有参数波动,使状态反馈解耦带来扰动力矩,ADRC 能够抑制转子参数波动带来的干扰力矩影响,提高系统鲁棒性能。

根据 MSCSG 转子偏转自由度状态空间表达式,可以求得状态空间结构性指数 d_i ,其满足:

$$C_i A^l B \neq 0 \quad l = 0, 1, 2, \dots, m - 1 \quad (13)$$

且状态空间结构性指数 d_i 是介于 0 到 $m - 1$ 之间的一个最小整数 l 。式中: C_i 为输出矩阵第 i 行向量。

根据状态方程,求得结构性指数 $d_1 = 1, d_2 = 1$ 。定义解耦性判据矩阵 E ,解耦性判据矩阵 E 可表示为

$$E = \begin{bmatrix} C_1 A^{d_1} B \\ C_2 A^{d_2} B \end{bmatrix} \quad (14)$$

系统能够实现状态反馈解耦的充要条件是解耦性判据矩阵满秩,由于 $\det E = \begin{vmatrix} 0 & k_T/J_y \\ k_T/J_x & 0 \end{vmatrix} \neq 0$,此系统能够通过状态反馈实现解耦。为了实现解耦,状态反馈解耦率可以描述为

$$u = Lv - Kx \quad (15)$$

式中: v 为转子偏转两自由度的控制电流量。

根据求得的状态空间结构性指数,得到解耦后的第 1 个子系统阶次和第 2 个子系统的阶次表示为

$$\begin{cases} H_\alpha = d_1 + 1 = 2 \\ H_\beta = d_2 + 1 = 2 \end{cases} \quad (16)$$

定义 $K = -E^{-1}F, L = E^{-1}$,矩阵 F 可通过式(17)求得

$$F = \begin{bmatrix} C_1 A^{d_1+1} \\ C_2 A^{d_2+1} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{H}{k_T} & 0 & 0 \\ 0 & 0 & 0 & \frac{H}{k_T} \end{bmatrix} \quad (17)$$

解耦之后的状态矩阵为 $A' = A + BK =$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{控制矩阵为 } B' = BL = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T。$$

因此,解耦之后系统的传递函数矩阵为

$$G_s = C(sI - A')^{-1}B' = \begin{bmatrix} \frac{1}{s^2} & 0 \\ 0 & \frac{1}{s^2} \end{bmatrix} \quad (18)$$

状态反馈解耦之后的系统是二重纯积分系统。系统极点在虚轴上,系统处于临界稳定状态。为了实现系统稳定,仍需要经过极点的配置,使系统极点落在复平面的左半平面上。设配置之后的极点为 (p_1, p_2, p_3, p_4) 。设极点配置状态反馈增益矩阵为 $F_s = \begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{23} & p_{24} \end{bmatrix}$ 。

因此,极点配置之后的特征方程为

$$\prod_{i=1}^4 (s - p_i) = C(sI - A' - B'F_s)B' \quad (19)$$

根据式(19)可求得极点配置状态反馈增益矩阵。

在考虑系统稳定性、响应时间、超调等因素后,经过比较,选取极点配置状态反馈增益矩阵 $F_s = \begin{bmatrix} -1 & -10 & 0 & 0 \\ 0 & 0 & -1 & -10 \end{bmatrix}$ 时,系统具有较好的动态响应。解耦和极点配置后的控制框图如图4所示。

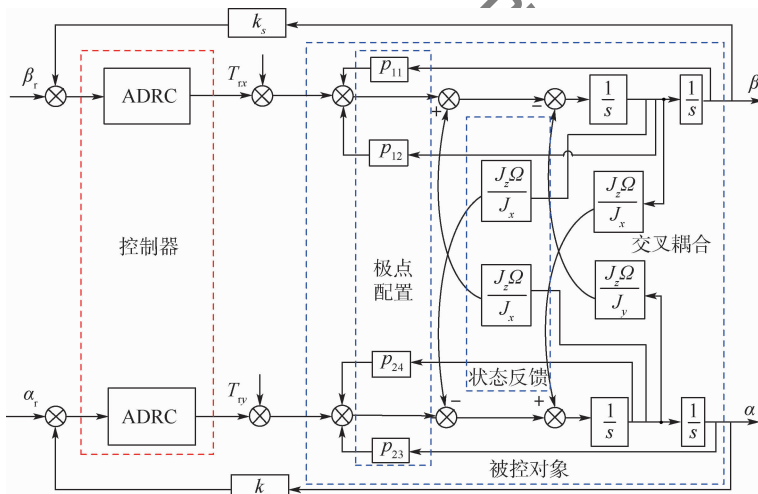


图4 解耦之后的磁悬浮转子控制系统

Fig. 4 Decoupled magnetically suspended rotor control system

2.2 自抗扰控制器设计

根据 MSCSG 转子平衡方程(8)可知,外界角速度的输入等效于对 MSCSG 施加一个干扰力矩,且外界输入角速度越大,等效干扰力矩越大。当干扰力矩过大时,会出现转子碰撞保护轴承的现象,严重影响磁悬浮转子磁悬浮的稳定性。此外,控制系统会产生一定的控制延时,加剧了扰动力矩对磁悬浮转子的干扰作用。

采用传统的 PID 控制方法、滑模控制方法难以兼顾系统的响应速度及抗干扰性能。因此,本文将能够同时调节系统响应速度和抗干扰性能的 ADRC 应用于转子偏转抗干扰快响应控制。通过设计 ADRC 实现实时消除干扰力矩对磁悬浮转子的干扰,以此提高 MSCSG 的控制精度。除此之外,ADRC 还能消除转子固有属性波动使解耦不充分带来的干扰,提高系统鲁棒性能。

ADRC 是一种非线性控制器,主要由 3 部分组成,即非线性微分跟踪器(TD)、扩张状态观测器(ESO)、非线性状态误差反馈控制律(NLSEF)。ADRC 的控制原理如图 5 所示。图中:参考输入为 r_α ,外界扰动力矩为 T_d ,本文中 T_d 主要包括 T_{rx} 和 T_{ry} , b 为 ESO 对控制器 $u(t)$ 观测的观测系数。TD 为参数输入安排过渡过程,使相应的输出在有

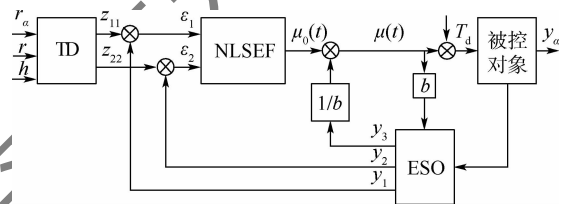


图5 自抗扰控制器控制框图

Fig. 5 Control block diagram of ADRC

限时间内平稳地跟踪输入信号,得到光滑的输入信号,并提取其微分信号。ESO 是 ADRC 的核心,采用双通道补偿的方法改造对象模型,将非线性、不确定的系统进行观测和补偿系统的扰动。NLSEF 可以实现 ESO 观测状态误差与被控对象控制变量的非线性组合,实现对 TD 输出与 ESO 给出的状态变量估计取误差,形成状态变量误差。

由于采用一般的近似微分处理方法,得到的结果受噪声信号干扰严重,本文采用 TD 的数学模型,数学表达式表示为

$$\begin{cases} z_{11} = z_{22} \\ z_{22} = fhan(z_{11} - r_\alpha, z_{22}, r, h) \end{cases} \quad (20)$$

式中: r_α 为磁悬浮转子偏转单自由度参考输入角位移; z_{11} 为对 r_α 安排的过渡过程; z_{22} 为过渡过程 z_{11} 的微分信号; r 和 h 为 TD 增益系数; $fhan$ 函数可表示为

$$\begin{cases} d = rh \\ d_0 = hd \\ \vartheta = z_{11} + hz_{22} \\ a_0 = \sqrt{d^2 + 8r|\vartheta|} \\ a = \begin{cases} z_{22} + \frac{a_0 - d}{2}\text{sign}(\vartheta) & |\vartheta| > d_0 \\ z_{22} + \frac{\vartheta}{h}\text{sign}(\vartheta) & |\vartheta| < d_0 \end{cases} \\ \text{fhan} = - \begin{cases} r\text{sign}(a) & |a| > d \\ r\frac{a}{d} & |a| < d \end{cases} \end{cases} \quad (21)$$

从控制的角度看,控制被控对象时,只需要考虑外界干扰对系统的影响即可。显而易见,这些影响系统输出的扰动会反映在控制系统的输出中。因此,只需要采取一些方法对扰动产生的影响进行估计就可以提取出扰动信息,从而通过补偿的方法进行干扰抑制。利用 ESO 的思想,将影响输出的扰动信息转换成新状态变量,通过反馈控制就可以实现扰动的控制。对于 MSCSG 而言,经过状态反馈解耦后,磁悬浮转子的偏转控制可以简化为二阶线性系统,如下:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(x_1, x_2, T_{ra}(t)) + (b - b_0)u(t) + b_0u(t) \\ y = x_1 \end{cases} \quad (22)$$

式中: $T_{ra}(t)$ 为系统的外部扰动; f 代表转子系统的状态空间; b_0 为系统的已知变量。

如果将作用于开环系统的所有总扰动的实时作用量扩充成新的状态变量 x_3 , 记作

$$x_3 = f(x_1, x_2, T_{ra}(t)) + (b - b_0)u(t) \quad (23)$$

记 $\dot{x}_3 = \alpha(t)$ 。那么转子偏转系统可以扩张成新的线性系统:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 + bu(t) \\ \dot{x}_3 = \alpha(t) \\ y = x_1 \end{cases} \quad (24)$$

构造其扰动观测器如下:

$$\begin{cases} \dot{z}_1 = z_2 - g_1(z_1 - x_1) \\ \dot{z}_2 = z_3 - g_2(z_1 - x_1) + b_0u \\ \dot{z}_3 = -g_3(z_1 - x_1) \end{cases} \quad (25)$$

根据式(25)可知,状态估计误差可以表示为

$$\begin{cases} e_1 = z_1 - x_1 \\ e_2 = z_2 - x_2 \\ e_3 = z_3 - x_3 \end{cases} \quad (26)$$

对式(26)求导可得误差状态方程为

$$\begin{cases} \dot{e}_1 = e_2 - g_1(e_1) \\ \dot{e}_2 = e_3 - g_2(e_1) \\ \dot{e}_3 = -\alpha(t) - g_3(e_1) \end{cases} \quad (27)$$

只要保证 $\alpha(t)$ 有界,就可以选取适当的 $g_1(e_1)$ 、 $g_2(e_1)$ 、 $g_3(e_1)$, 保证式(27)对系统原点稳定。

在 ESO 构造过程中, g 函数最常用的选择函数可表示为

$$\text{fal}(\nu, \tau, \gamma) \begin{cases} \nu\gamma^{1-\tau} & \nu \leq 0.5 \\ |\nu|^\tau \text{sign}(\nu) & \nu > 0.5 \end{cases} \quad (28)$$

式中: ν 为 fal 函数的输入; τ 主要影响跟踪效果; γ 为 fal 函数的参数, 主要影响 fal 函数的滤波效果。

为了实现状态观测更好的跟踪效果和滤波效果, 以及便于进行观测器的性能分析, fal 常取线性函数。因此, 构造的 ESO 可表示为

$$\begin{cases} \dot{e} = z_1 - x_1 \\ \dot{z}_1 = z_2 - \beta_1 e_1 \\ \dot{z}_2 = z_3 - \beta_2 e_1 + b_0 u \\ \dot{z}_3 = -\beta_3 e_1 \end{cases} \quad (29)$$

式中: β_1 、 β_2 、 β_3 为相应的状态观测量的增益系数。

根据图 5 可知, ADRC 的 NLSEF 可以表示为

$$u_0(t) = \sigma_1 \text{fnl}(\varepsilon_1) + \sigma_2 \text{fnl}(\varepsilon_2) \quad (30)$$

式中: σ_1 、 σ_2 为 NLSEF 的增益; fnl 函数可以表述为

$$\text{fnl}(\kappa) = \begin{cases} \text{sign}(\kappa)\sqrt{|\kappa|} & |\kappa| > 0.1 \\ \kappa & |\kappa| \leq 0.1 \end{cases} \quad (31)$$

本文在保证控制系统性能优越的条件下, 为了提高系统运行效率, 减小系统响应时间, 选用 $\text{fnl}(\kappa) = \kappa$ 。因此, 本文实现了非线性误差反馈控制率的线性化。

转子偏转自抗扰控制的控制率可以表示为

$$u(t) = u_0(t) - \frac{1}{b}z_3 \quad (32)$$

2.3 控制器性能分析

根据磁悬浮转子动力学模型和 ESO 模型, 磁悬浮转子 ADRC 的状态空间可表示为

$$\dot{X} = A_d X + B_1 \mu + B_r T_r \quad (33)$$

式中: μ 为状态反馈后系统的输入; T_r 为系统的扰动; $A_d = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$; $B_1 = [0 \ b \ 0]^T$; $B_r = [0 \ 0 \ 1]^T$ 。

经过 ESO 观测后的变量可表示为

$$\dot{Z} = A_d Z + B_1 \mu + B_r \tilde{T}_r + L_0 (X - Z) \quad (34)$$

式中: \tilde{T}_r 为 ESO 观测到的系统的扰动; $L_0 = \begin{bmatrix} \beta_1 & 0 & 0 \\ \beta_2 & 0 & 0 \\ \beta_3 & 0 & 0 \end{bmatrix}^T$ 。

将式(33)和式(34)联立可得

$$\dot{X} - \dot{Z} = (A_d - L_0)(X - Z) + B_r(T_r - \tilde{T}_r) \quad (35)$$

根据式(35)可以计算出

$$|\lambda I - (A_d - L_0)| = \lambda^{n+1} + l_1\lambda^n + l_n\lambda + l_{n+1} \quad (36)$$

为使观测误差趋近于0,矩阵 $A_d - L_0$ 特征值必须全部有负实部。为便于分析计算,按如下方式选取,使得

$$\lambda^{n+1} + l_1\lambda^n + l_n\lambda + l_{n+1} = (\lambda + p)^{n+1} \quad p > 0 \quad (37)$$

根据式(37)可以得出

$$l_i = \frac{(n+1)!p^i}{i!(n+1-i)!} \quad i = 1, 2, \dots, n+1 \quad (38)$$

将式(38)代入式(35)可知

$$\begin{cases} \dot{x}_i - \dot{z}_i = \frac{(n+1)!p^i}{i!(n+1-i)!}(x_i - z_i) & i = 1, 2, \dots, n \\ \dot{x}_{n+1} - \dot{z}_{n+1} = -p^{n+1} + T_r - \tilde{T}_r \end{cases} \quad (39)$$

式(39)分别除以 p^{i-1}, p^n ,可以得到

$$\begin{cases} \frac{\dot{x}_i - \dot{z}_i}{p^{i-1}} = -\frac{(n+1)!}{i!(n+1-i)!}p(x_i - z_i) & i = 1, 2, \dots, n \\ \frac{\dot{x}_{n+1} - \dot{z}_{n+1}}{p^n} = -p + \frac{T_r - \tilde{T}_r}{p^n} \end{cases} \quad (40)$$

构造状态观测误差为

$$\psi_i = \frac{\dot{x}_i - \dot{z}_i}{p^{i-1}} \quad (41)$$

记 $L_i = \frac{(n+1)!}{i!(n+1-i)!}$,式(41)可以写为

$$\begin{cases} \dot{\psi}_i = -l_p\psi_i \\ \dot{\psi}_{n+1} = -p\psi_{n+1} + \frac{T_r - \tilde{T}_r}{p^n} \end{cases} \quad i = 1, 2, \dots, n \quad (42)$$

式(42)中能够保证观测误差趋近于零。因此,构造的状态变量能够准确地跟踪系统的状态变量。

根据 ADRC 控制原理可知,其反馈控制误差可表示为

$$\dot{\varepsilon} = A_3\varepsilon + A_4\psi \quad (43)$$

$$\text{式中: } A_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \sigma_1 & \sigma_2 & -1 \end{bmatrix}; A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \sigma_1 & \sigma_2 & -1 \end{bmatrix}.$$

分析可知,反馈控制误差包括 2 部分:式(43)右边第 1 项是反馈控制误差项,第 2 项是

状态观测估计误差项。通过对 ESO 观测误差分析可知,状态观测估计误差项为 0。

式(43)等效于

$$\dot{\varepsilon} = A_3\varepsilon \quad (44)$$

因此,存在 σ_1 和 σ_2 使得矩阵 A_3 特征根 λ_1, λ_2 和 λ_3 都具有负实部,有

$$\varepsilon = \lim_{t \rightarrow 0} \sum_{i=1}^3 h_i \exp(\lambda_i t) = 0 \quad (45)$$

通过上述分析可知,系统能够实现有界输入下的有界稳定。

图 6 和图 7 为输入为阶跃信号和正弦信号时 ESO 对状态量的观测效果。

根据图 6 和图 7 可知,设计的状态变量 z_1, z_2

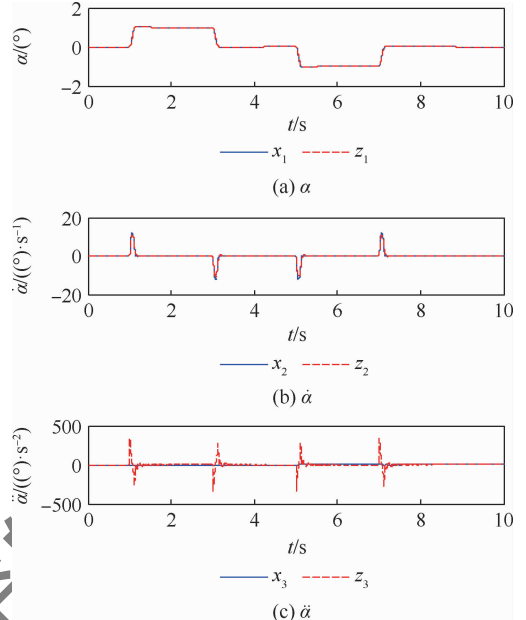


图 6 ESO 对阶跃输入的观测效果

Fig. 6 Observation of step input by ESO

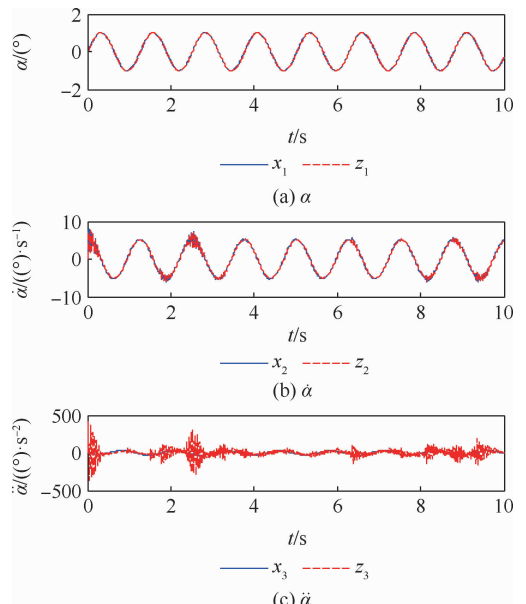


图 7 ESO 对正弦输入的观测效果

Fig. 7 Observation of sinusoidal input by ESO

对系统变量 x_1, x_2 具有很好的观测性。由于系统是纯积分系统,只有 $\alpha(t)$ 在跳变时跟踪效果有误差,在其他时刻能较好地反映系统变量。因此,构造的 ESO 具有较好的跟踪性能。

3 仿真分析

为了验证本文所提的状态反馈解耦自抗扰控制的有效性和优越性,将本文方法和传统交叉反馈解耦 PID 控制对比仿真。MSCSG 系统参数和控制系统仿真参数如表 1 和表 2 所示。

表 1 MSCSG 系统参数

Table 1 MSCSG system parameters

参数	数值
$J_x, J_y / (\text{kg} \cdot \text{m}^2)$	0.009 7
$J_z / (\text{kg} \cdot \text{m}^2)$	0.028 7
$\Omega / (\text{r} \cdot \text{min}^{-1})$	5 000
$k_T / (\text{N} \cdot \text{m} \cdot \text{A}^{-1})$	2.8

表 2 控制系统参数

Table 2 parameters of the control system

参数	数值
(σ_1, σ_2)	(490, 122)
b_0	1
h	0.01
λ	1
$(\beta_1, \beta_2, \beta_3)$	(14, 8 000, 100)
γ	0.05
γ	0.1

3.1 解耦性能分析

图 8 为阶跃输入仿真结果。 $t = 0.5 \text{ s}$ 时,转子 α 施加一个终值为 1° 的阶跃信号, $t = 2.5 \text{ s}$ 时,转子 β 施加一个终值为 -1° 的阶跃信号。分别采用交叉 PID 控制、PID + 状态反馈解耦和 ADRC + 状态反馈解耦。根据仿真结果可知,采用交叉 PID 解耦,在 $t = 0.5 \text{ s}$ 时, α 偏转 1° 时, β 会产生 0.12° 的耦合。当 $t = 2.5 \text{ s}$ 时, β 偏转 -1° 时, β 会产生 -0.12° 的耦合。因此,采用交叉 PID 解耦,会使 x, y 轴产生约 12% 的耦合跳动。当采用 PID 中状态反馈解耦控制和 ADRC + 状态反馈解耦控制, α 的运动不会对 β 的运动产生影响; β 的运动不会对 α 的运动产生影响。通过对比仿真可知,采用状态反馈解耦的效果明显优于传统交叉 PID 反馈解耦的效果,且状态反馈解耦能够实现磁悬浮转子两自由度的完全解耦。

为了进一步验证解耦性能,改变输入角位移。图 9 为正弦输入解耦效果对比仿真。当对 α 输入 $\alpha = \sin(2\pi t)$, β 输入值是 0。采用交叉 PID 解耦, β 会产生明显的耦合,耦合角位移为 $\beta = 0.318 \times$

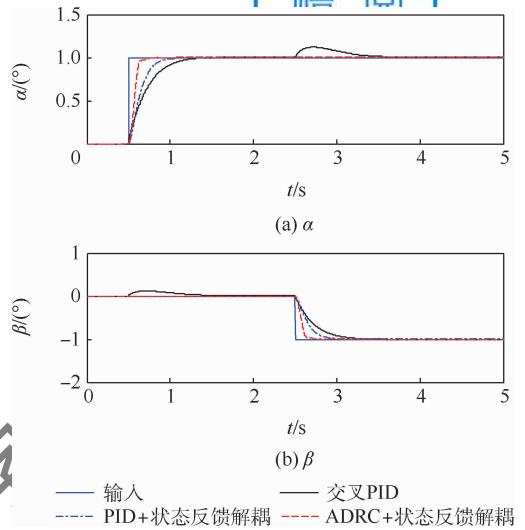


图 8 阶跃信号输入解耦性能对比

Fig. 8 Comparison of decoupling performance of step signal input

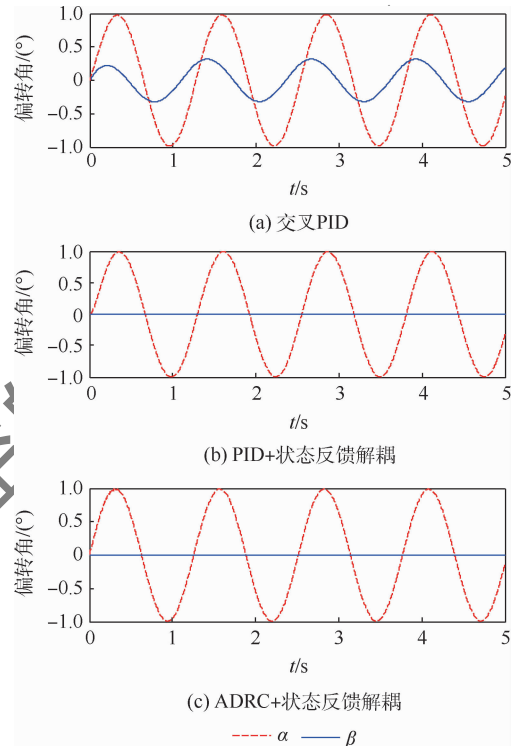


图 9 正弦信号输入解耦性能对比

Fig. 9 Comparison of decoupling performance of sinusoidal signal input

$\sin(2\pi t)$ 。但是,采用 PID + 状态反馈解耦和 ADRC + 状态反馈解耦控制,偏转两自由度不会产生相互扰动。

通过对比可知,采用状态反馈解耦可以实现磁悬浮转子两自由度偏转的完全解耦。

3.2 灵敏度分析

为了验证 ADRC 的响应速度优于传统交叉 PID 控制的响应速度,设置如下对比仿真。仿真实验结果如图 10 所示。 α 的偏转指令如黑线所

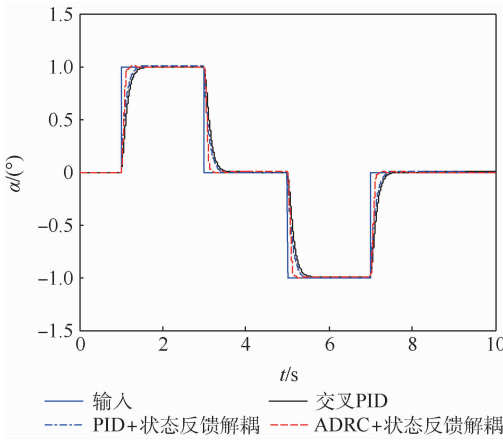


图10 响应速度对比仿真结果

Fig. 10 Simulation results of response speed

示,1 s时,令转子正向偏转 1° ,3 s时令转子偏转恢复为 0° ;5 s时令转子反方向偏转 1° ,7 s时刻再令转子偏转恢复为 0° 。采用交叉PID控制,转子偏转跟踪响应时间为0.53 s,采用PID+状态反馈解耦控制,转子偏转跟踪响应时间为0.46 s,采用ADRC+状态反馈解耦控制,转子偏转跟踪响应时间为0.25 s。仿真结果表明,与传统交叉PID方法相比,ADRC+状态反馈解耦控制方法响应速度明显提高,偏转响应时间较传统交叉PID方法降低了52.8%。

3.3 抗干扰性能分析

为了验证控制器的抗干扰性能,图11为抗干扰仿真结果。图11(a)在 $t=1$ s时,对磁悬浮转子单自由度角位移施加一个正弦往复扰动,扰动力矩可用 $T_\alpha = \sin(2\pi t)$ 表示。对比交叉PID控制方法、PID+状态反馈解耦控制方法和ADRC+状态反馈解耦控制方法的仿真结果得到,采用交叉PID控制,施加的扰动会使磁悬浮转子产生幅值为 0.127° 的同频偏转。采用PID+状态反馈解耦控制,施加的扰动会使磁悬浮转子产生幅值为 0.026° 的同频往复偏转;采用ADRC+状态反馈解耦控制,磁悬浮转子偏转的幅值为 0.0029° 。对比发现,在低频扰动下,采用ADRC+状态反馈解耦控制方法能够很好地抑制干扰力矩的扰动。

在仿真参数不变的情况下,提高扰动力矩的频率,图11(b)为扰动力矩 $T_\alpha = \sin(4\pi t)$ 。采用交叉PID控制,施加的扰动会使磁悬浮转子产生幅值为 0.106° 的同频偏转。采用PID+状态反馈解耦控制,施加的扰动会使磁悬浮转子产生幅值为 0.052° 的同频往复偏转;采用ADRC+状态反馈解耦控制,磁悬浮转子偏转的幅值为 0.008° 。进一步提高干扰力矩的扰动频率,扰动力矩 $T_\alpha = \sin(10\pi t)$ 和 $T_\alpha = \sin(60\pi t)$,根据仿真结果可知,

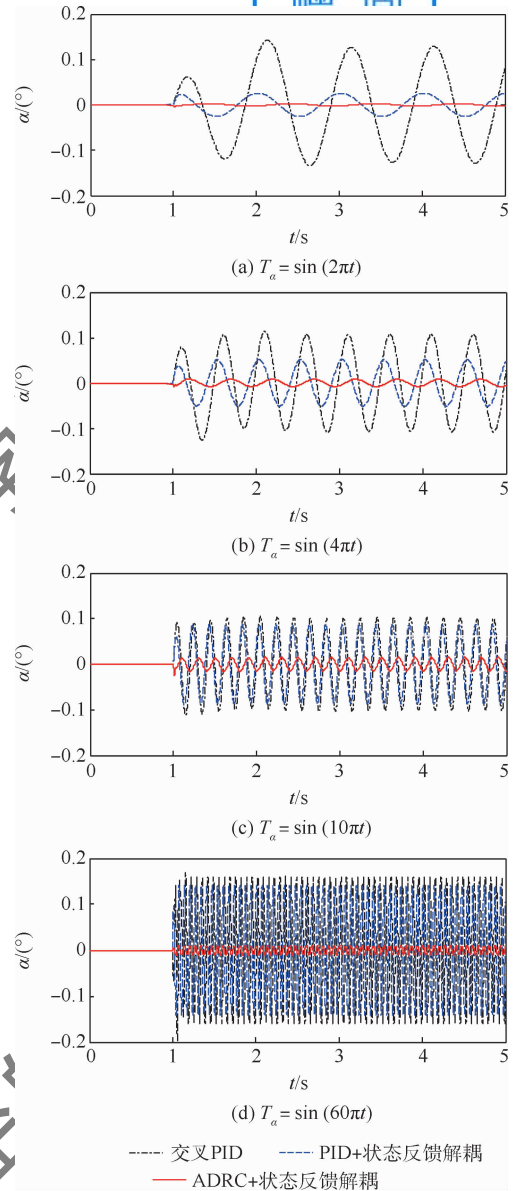


图11 抗干扰性能对比仿真结果

Fig. 11 Comparison of simulation results of anti-interference performance

采用ADRC能够有效地抑制干扰力矩对磁悬浮转子的影响。

通过对不同频率成分的扰动仿真实验可知,采用状态反馈解耦自抗扰控制方法,其抗扰动性能不会随着扰动频率的改变而降低,且具备独立调节跟踪性和抗扰性的特点,具有较强的鲁棒性。因此,ADRC抗扰性能优于PID控制器。

此外,采用ADRC能提高状态反馈解耦的容错性能。当转子的固有参数波动,如电阻阻值的跳变、线圈电流随温度的提高或者转速波动而发生改变。此时,状态反馈解耦不能够实现完全解耦,会带来扰动力矩。采用ADRC还能抑制转子参数波动带来的干扰力矩的影响,提高系统鲁棒性能。如图12(a)所示,当转子转速发生波动,转

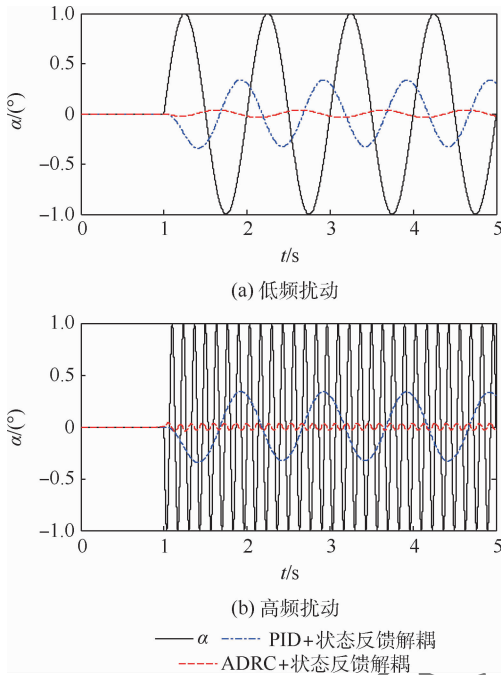


图12 解耦不充分情况下的抗干扰性能
Fig.12 Anti-interference performance under the condition of insufficient decoupling

速为 $(5\ 000 \pm 4)$ r/min, 此时, 状态反馈解耦不能实现转子完全解耦。当状态反馈解耦不充分时, 黑线代表 $t=1$ s, α 位移输入幅值 1° 的正弦往复偏转, 通过 PID+状态反馈解耦控制, 蓝线表示 β 产生 0.34° 的同频往复偏转。红线通过 ADRC+状态反馈解耦控制, β 会产生 0.038° 的同频往复偏转。采用 ADRC, 抵抗扰动的能力提高了一个数量级。图 12(b) 为仿真条件一定的情况下, 继续增加输入信号的频率, 采用本文方法和传统方法的抗干扰效果。

通过对比可知, ADRC 具有显著降低耦合的效果, 能提升系统容错能力和鲁棒性能。由于磁悬浮转子系统本来就是一个多耦合效应的系统, ADRC 在磁悬浮转子偏转控制方面具有较高的工程实践价值。

4 结论

基于 MSCSG 转子两自由度偏转控制, 本文提出了一种状态反馈解耦自抗扰控制方法, 在保证响应时间的条件下, 不仅能够实现转子的两自由度偏转完全解耦, 而且能够提高系统的响应时间。此外, 所提出的自抗扰控制方法, 能够主动抑制磁悬浮转子所受干扰力矩的扰动, 提高系统的鲁棒性。所提的状态反馈解耦自抗扰控制方法能提高控制系统的容错能力, 保证系统在承受外界干扰能正常工作。本文方法在磁悬浮转子偏转控制领

域具有较高的工程实践价值。主要结论如下:

1) 转子分别沿 x 、 y 轴偏转过程中, 相比于采用传统交叉 PID 方法会导致转子沿 x 、 y 轴产生约 12% 的耦合跳动, 基于状态反馈解耦自抗扰控制方法可以完全消除转子径向偏转两自由度间的耦合。

2) 基于状态反馈解耦自抗扰控制方法可兼顾系统的跟随性和抗扰性。偏转响应时间较传统交叉 PID 方法降低了 52.8%; 正弦扰动力矩产生的偏转波动幅值较交叉 PID 方法明显减小。

3) 当转子固有属性发生波动, 导致状态反馈不完全解耦时, 基于状态反馈解耦自抗扰控制方法能够弥补不完全解耦带来的扰动力矩的影响, 对于提高控制系统的鲁棒性具有重要意义。

参考文献 (References)

- [1] SUN L, ZHENG Z. Disturbance observer-based robust backstepping attitude stabilization of spacecraft under input saturation and measurement uncertainty [J]. IEEE Transactions on Industrial Electronics, 2017, 64(10): 7994-8002.
- [2] SUN L, HUO W, JIAO W. Adaptive backstepping control of spacecraft rendezvous and proximity operations with input saturation and full-state constraint [J]. IEEE Transactions on Industrial Electronics, 2017, 64(1): 480-492.
- [3] XU S, CUI N, FAN Y, et al. Active vibration suppression of flexible spacecraft during attitude maneuver with actuator dynamics [J]. IEEE Access, 2018, 6: 35327-35337.
- [4] LIN Z, LIN S, WU S, et al. Vibration control of a flexible spacecraft system with input backlash [J]. IEEE Access, 2019, 7: 87017-87026.
- [5] SI H, SHAO X, ZHANG W. MLP-based neural guaranteed performance control for MEMS gyroscope with logarithmic quantizer [J]. IEEE Access, 2020, 8: 38596-38605.
- [6] REN Y, CHEN X, CAI Y, et al. Attitude-rate measurement and control integration using magnetically suspended control and sensitive gyroscopes [J]. IEEE Transactions on Industrial Electronics, 2018, 65(6): 4921-4932.
- [7] XU G F, CAI Y W, REN Y, et al. Application of a new Lorentz force-type tilting control magnetic bearing in a magnetically suspended control sensitive gyroscope with cross-sliding mode control [J]. Transactions of the Japan Society for Aeronautical and Space Sciences, 2018, 61(1): 40-47.
- [8] 夏长峰, 蔡远文, 任元, 等. MSCSG 转子系统的扩展双频 Bode 图稳定性分析方法 [J]. 宇航学报, 2018, 39(2): 168-176.
XIA C F, CAI Y W, REN Y, et al. Extended dual-frequency Bode diagram stability analysis method for MSCSG rotor system [J]. Journal of Astronautics, 2018, 39(2): 168-176 (in Chinese).
- [9] FANG J, REN Y, FAN Y. Nutation and precession stability criterion of magnetically suspended rigid rotors with gyroscopic effects based on positive and negative frequency characteristics

- [J]. IEEE Transactions on Industrial Electronics, 2014, 61(4):2003-2014.
- [10] REN Y, SU D, FANG J. Whirling modes stability criterion for a magnetically suspended flywheel rotor with significant gyroscopic effects and bending modes[J]. IEEE Transactions on Power Electronics, 2013, 28(12):5890-5901.
- [11] FANG J C, REN Y. Decoupling control of magnetically suspended rotor system in control moment gyros based on an inverse system method[J]. IEEE/ASME Transactions on Mechatronics, 2012, 17(6):1133-1144.
- [12] REN Y, FANG J C. High-stability and fast-response twisting motion control for the magnetically suspended rotor system in a control moment gyro[J]. IEEE/ASME Transactions on Mechatronics, 2013, 18(5):1625-1634.
- [13] AHRENS M, KUCERA L, LARSONNEUR R. Performance of a magnetically suspended flywheel energy storage device [J]. IEEE Transactions on Control Systems Technology, 1996, 4(5):494-502.
- [14] FAN Y, FANG J. Experimental research on the nutational stability of magnetically suspended momentum flywheel in control moment gyroscope (CMG) [C] // Proceedings of 9th International Symposium on Magnetic Bearings, 2004:116-121.
- [15] HUNG J Y. Magnetic bearing control using fuzzy logic [J]. IEEE Transactions on Industry Applications, 1995, 31(6):1492-1497.
- [16] SOBHAN P V S, KUMAR G V N, AMARNATH J. Rotor levitation by active magnetic bearings using fuzzy logic controller [C] // 2010 International Conference on Industrial Electronics, Control and Robotics. Piscataway: IEEE Press, 2010:27-29.
- [17] BENOMAIR A M, BASHIR F A, TOKHI M O. Optimal control based LQR-feedback linearisation for magnetic levitation using improved spiral dynamic algorithm [C] // 2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR). Piscataway: IEEE Press, 2015:24-27.
- [18] ZHANG Y C, SUN G J, ZHANG Y J. Experimental verification for zero power control of 0.5 kWh class flywheel system using magnetic bearing with gyroscopic effect [C] // Proceedings of International Conference on Machine Learning and Cybernetics. Piscataway: IEEE Press, 2002:4-5.
- [19] REN Y, FANG J. High-precision and strong-robustness control for an MSCMG based on modal separation and rotation motion decoupling strategy [J]. IEEE Transactions on Industrial Electronics, 2014, 61(3):1539-1551.
- [20] XIE J J, LIU G, WEN T. Composite compensation for load torque of active magnetic bearing in DGMSCMG [J]. Optical and Precision Engineering, 2015, 23(8):2211-2219.
- [21] KANG M S, LYOU J, LEE J K. Sliding mode control for an active magnetic bearing system subject to base motion [J]. Mechatronics, 2010, 20(1):171-178.
- [22] 刘强, 赵勇, 代峰燕, 等. 磁悬浮陀螺飞轮用隐式洛伦兹力磁轴承 [J]. 光学精密工程, 2018, 26(2):399-409.
- LIU Q, ZHAO Y, DAI F Y, et al. Implicit Lorentz force magnetic bearing for magnetically suspended gyro flywheel [J]. Optical and Precision Engineering, 2018, 26(2):399-409 (in Chinese).
- [23] 夏长峰, 蔡远文, 任元, 等. 磁悬浮控制敏感陀螺转子前馈解耦内模控制 [J]. 北京航空航天大学学报, 2018, 44(3):480-488.
- XIA C F, CAI Y W, REN Y, et al. Feedforward decoupling and internal model control for rotor of magnetically suspended control and sensing gyroscope [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(3):480-488 (in Chinese).
- [24] BASARAN S, SIVRIOGLU S. Novel repulsive magnetic bearing flywheel system with composite adaptive control [J]. IET Electric Power Applications, 2019, 13(5):676-685.
- [25] HAN J. From PID to active disturbance rejection control [J]. IEEE Transactions on Industrial Electronics, 2009, 56(3):900-906.
- [26] 尹增惠, 蔡远文, 王卫杰, 等. 一种组合磁钢叠加磁场洛伦兹力磁轴承设计方法 [J]. 宇航学报, 2018, 39(7):56-64.
- YIN Z Y, CAI Y W, WANG W J, et al. A Lorentz force magnetic bearing design method with composite magnetic steel and superimposed magnetic field [J]. Journal of Aerospace, 2018, 39(7):56-64 (in Chinese).

Decoupled active disturbance rejection control method for magnetically suspended rotor based on state feedback

YIN Zengyuan¹, CAI Yuanwen², REN Yuan^{2,*}, WANG Weijie², CHEN Xiaocen³, YU Chunmiao¹

(1. School of Postgraduate Management, Space Engineering University, Beijing 101416, China;

2. Department of Space Science and Technology, Space Engineering University, Beijing 101416, China;

3. Army Academy, Beijing 100190, China)

Abstract: A decoupled anti-interference control method for rotor tilting is proposed to address the strong coupling of rotor deflection channels of magnetically suspended control and sensitive gyroscope (MSCSG), as well as disturbance instability during spacecraft attitude measurement. The coupling of rotors tilting with two degrees of freedom is analyzed, and a decoupling controller is designed based on state feedback. The MSCSG interference torque to the magnetically suspended rotor produced by attitude motions during spacecraft attitude measurement is deduced. The anti-interference control of the rotor is realized by the active disturbance rejection control (ADRC). The tracking performance and stability of the constructed extended state observer (ESO) are analyzed. The stability of the system with bounded inputs is achieved by adjusting the nonlinear state error feedback coefficients. Simulation results show that the state feedback decoupling can realize the complete decoupling of tilting freedom degrees, that ESO has good tracking performance, and that ADRC has better anti-interference performance than the traditional PID control method.

Keywords: magnetically suspended control and sensitive gyroscope (MSCSG); magnetically suspended rotor; attitude measurement; state feedback decoupling; active disturbance rejection control (ADRC)

Received: 2021-01-14; **Accepted:** 2021-03-28; **Published online:** 2021-04-19 08:29

URL: kns.cnki.net/kcms/detail/11.2625.V.20210415.1540.002.html

Foundation items: National Natural Science Foundation of China (52075545, 11802342); Beijing Youth Top-Notch Talent Support Program (2017000026833ZK23)

* **Corresponding author.** E-mail: renyuan_823@aliyun.com

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

DOI: 10.13700/j.bh.1001-5965.2021.0013

基于高效注意力和上下文感知的目标跟踪算法

柏罗, 张宏立*, 王聪

(新疆大学 电气工程学院, 乌鲁木齐 830047)

摘 要: 基于匹配思想的孪生网络算法缺乏对目标的整体性感知, 容易出现对目标状态估计不够精准和在复杂环境中跟丢的现象。为此, 在孪生网络的基础上设计了2个轻量级的模块来实现更精准、更鲁棒的目标跟踪。在提取特征的主干网络之后, 嵌入一个高效通道注意力模块, 实现高效提取目标特征并增强差异化表示, 使网络更侧重于目标信息; 模板匹配之后的特征通过一个局部上下文感知模块, 增强网络对目标的整体感知, 以应对跟踪过程中复杂多变的环境; 采用 Anchor-free 的状态估计策略实现对目标的精准估计。实验结果表明: 所提算法 SiamCC 在数据集 OTB100、VOT2016 和 VOT2018 上的测试结果均好于 DaSiamRPN、ATOM 等算法, 并且跟踪速度达到了 85 帧/s。

关键词: 机器视觉; 目标跟踪; 孪生网络; 通道注意力; 上下文感知

中图分类号: TP391.4

文献标志码: A **文章编号:** 1001-5965(2022)07-1222-11

目标跟踪在计算机视觉领域一直是前前沿问题, 跟踪的对象是无类别限制的任一物体, 一般表现为在视频中精准无误地估计出跟踪物体的位置和状态。目标跟踪在视频监控、自动驾驶、行为分析等方面有着广阔的应用前景^[1]。

为了跟踪无类别限制的任意目标, 跟踪器一般只能在初始帧知道待跟踪目标的信息, 再逐帧估计出跟踪目标的位置和状态, 其通常以矩形框的形式表示。近年来, 跟踪领域涌现出了非常多优秀的算法。平方误差最小滤波器(MOSSE)^[2]首次把相关滤波运用到目标跟踪, 相关滤波的思想后续发展了核相关滤波器(KCF)^[3]和高效卷积算子(ECO)^[4]。随着深度学习的进一步发展, 特别是卷积神经网络(CNN)在计算机视觉的运用, 目标跟踪算法开始尝试用卷积神经网络来提取目标特征^[5]。Wang 和 Yeung^[6]第一次把深度模型运用到目标跟踪领域。Ma 等^[7]提出的基于多层卷积特征的跟踪器(HCFT)考虑了不同层目

标特征的差异。

全卷积孪生网络(SiamFC)^[8]采取离线训练、在线匹配跟踪且不更新模板的机制, 较好地平衡了跟踪性能与速度。He 等^[9]运用了双重孪生网络(SASiam), 分别学习语义特征和外观特征, 更好地刻画了目标特征。Li 等^[10]采取区域提议网络(RPN)进行目标分类和定位(SiamRPN), 替代 SiamFC 的多尺度估计, 使网络分别输出目标分类和边框回归。Fan 和 Ling^[11]采用级联多层学习策略逐步提高了 RPN 的准确性(C-RPN)。Shen 等^[12]将跟踪问题看作关键部分搜索, 并以关键部分引导模板更新。Zhang 和 Peng^[13]探讨了基于孪生网络的跟踪器的主干网络设计问题。

本文采用残差卷积神经网络(ResNet)^[14]作为提取特征的主干网络, 考虑到跟踪器需要得到丰富的语义信息才能更加稳健地跟踪目标, 而以往 SiamFC、SiamRPN 等算法只采用了 5 层的主干网络 AlexNet^[15], 显然对目标特征的语义信息表

收稿日期: 2021-01-11; 录用日期: 2021-01-22; 网络出版时间: 2021-03-10 09:09

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210309.1340.008.html

基金项目: 国家自然科学基金(51767022, 51967019)

* 通信作者。E-mail: 1606829274@qq.com

引用格式: 柏罗, 张宏立, 王聪. 基于高效注意力和上下文感知的目标跟踪算法[J]. 北京航空航天大学学报, 2022, 48(7): 1222-1232. BAI L, ZHANG H L, WANG C. Target tracking algorithm based on efficient attention and context awareness[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1222-1232 (in Chinese).

达不够丰富,容易导致跟踪失败。由于跟踪问题对速度有着较高的要求,采用网络层数和计算开销适中的主干网络——ResNet18。

为使网络提取的特征更加丰富,本文基于残差注意力的孪生网络算法(RASNet)^[16]采用了计算量较大的残差注意力网络^[17]。设计了一个计算量很小的高效通道注意力模块(efficient channel attention module, ECAM),通过赋予特征通道不同的权重,判别特征通道表示的差异,加强对目标特征的感知,进而提升跟踪的性能。

对目标的整体感知有助于增强跟踪算法的鲁棒性,但现有算法很少考虑到这一点。本文设计了一个高效的局部上下文感知模块(local context awareness module, LCAM),通过获取的局部上下文信息有助于应对目标的形变、复杂背景的干扰等情况,有效减少了目标估计不准确和跟丢现象。

在跟踪器对目标状态估计部分,本文采用分类与回归并行的策略。不同于 SiamRPN、C-RPN 基于锚点(Anchor)的状态估计,采用无锚点(Anchor-free)机制,以更加合理的方式实现对目标状态的估计,避免了人为设计 Anchor 的不足及超参的设置。

1 本文算法

1.1 孪生网络跟踪算法

孪生网络跟踪算法不仅在跟踪性能上有着不错的表现,且在速度上有着明显的优势,在跟踪领域有着巨大的影响。其主要思想是:通过大量数据对学习相似性表示,采用卷积方式匹配 2 帧之间最相似的区域,从响应值较大的区域估计目标的状态。

由于孪生网络跟踪算法通常仅以视频序列的

第 1 帧为模板,在后续帧中匹配出最为相似的目标,并且在跟踪过程中不更新模板,而当目标发生形变、旋转甚至模糊时容易造成目标跟丢现象。当跟踪环境出现显著干扰时,如环境的光照变化、复杂背景等,也易发生跟踪漂移现象。正因如此,本文采用提取目标特征信息更强的残差卷积神经网络,并设计了 2 个轻量级模块来丰富目标的特征信息,增强目标的整体感知,以应对上述跟踪难点。

本文算法延续了 SiamFC 这种简洁、高效的最相似匹配思想,不同之处在于:采取了多层特征匹配策略,分别对最大响应区的目标进行状态估计,再把 3 层结果加权融合。算法的整体框架如图 1 所示。首先,采用特征提取网络 ResNet18 的 Conv3、Conv4、Conv5 的特征,为保证 Conv3、Conv4、Conv5 输出特征大小一致,改变原始网络 Conv4 和 Conv5 的卷积步长。提取的特征经过高效通道注意力(ECA)模块(注意力机制在加强目标特征的同时可在一定程度上削弱其他干扰特征),再将模板特征和搜索特征逐通道地进行特征匹配。匹配后的特征经过局部上下文感知(LCA)模块获取目标的上下文信息,加强目标整体感知,有利于对后续的目标状态估计。然后,对目标特征分别进行分类及边框估计。最后,加权融合得到最终结果。

1.2 高效通道注意力模块

Hu 等^[18]提出了首个通道注意力网络 SENet(squeeze-and-excitation networks),在增加一定的计算开销下更好地提取特征。SENet 将每个通道独立看待,并赋予其不同的权重,进而增强或削弱某些特征;CBAM^[19]在 SENet 的基础上考虑到特征的空间信息,通过结合通道注意力和空间注意

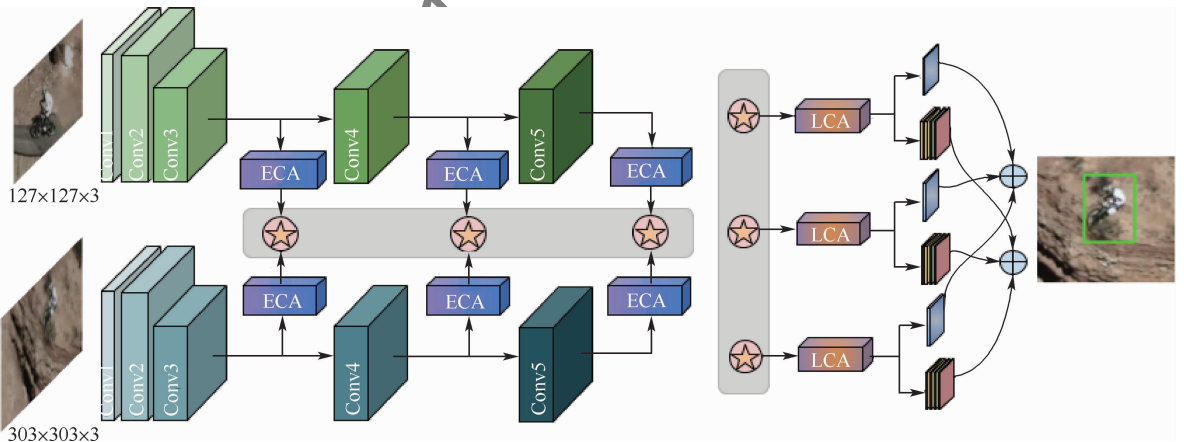


图 1 基于高效注意力和上下文感知的孪生网络目标跟踪算法框架

力在分类任务上取得了巨大的成功;Wang 等^[20]则充分分析了 SENet 的不足之处,提出了不降维度的、更高效的通道注意力网络 ECANet。

本文设计的高效通道注意力模块 ECAM 如图 2 所示。将输入的特征先进行逐通道全局均值池化(global average pooling, GAP)和全局最大值池化(global average pooling, GAP)和全局最大值

池化(global max pooling, GMP),在不减少通道数的情况下进行跨通道交互学习,以增加很少的计算量的代价来有效提高对特征表示。跨通道数 $k=3$ 。值得注意的是,设计的高效通道注意力模块在提升跟踪性能的同时几乎不会影响跟踪速度。

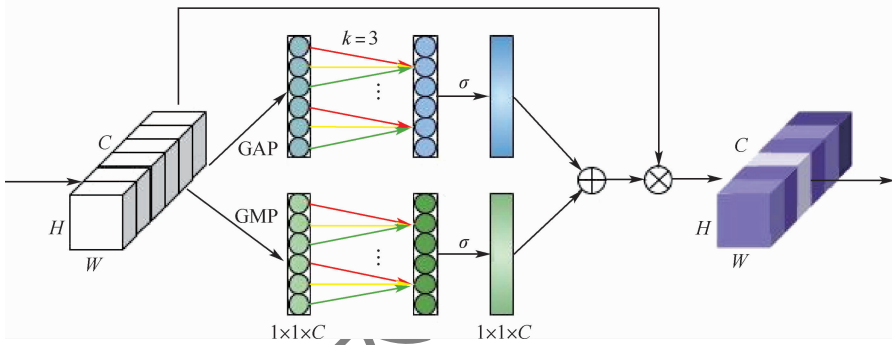


图 2 高效通道注意力模块

Fig. 2 Efficient channel attention module

先分别计算输入特征每个通道 c_i 的 GAP 和 GMP:

$$g_i = G(c_i) \tag{1}$$

式中: g_i 分别为每个通道计算的均值或最大值; G 为均值或最大值计算。

再计算相邻通道之间交互学到的 g'_i :

$$g'_i = \sum_{j=1}^k \alpha^j g_i^j, g_i^j \in \Omega_i^k \tag{2}$$

式中: Ω_i^k 为 g_i 的 k 个相邻通道的集合; α^j 为学习到的权重。

经过 Sigmoid 函数,就能得到 c_i 的权重:

$$w_i = \sigma(\lambda_A g'_{A_i} + \lambda_M g'_{M_i}) \tag{3}$$

式中: σ 为 Sigmoid 函数; λ 为求和权重,由网络训练得到;下标 A 和 M 分别表示均值池化和最大值池化。

最终将模块的输入特征与权重 w 逐通道相乘,达到增强原有特征的效果。

1.3 局部上下文感知模块

Wang 等^[21]设计了 NLNet (non-local neural networks)——一种非局部运算,将某个位置的响

应计算为所有位置特征的加权总和,使感受野增大,以捕获有效的全局信息,而不是只关注局部区域信息。大量实验证明,NLNet 在分类、检测和分割等任务上都有不错的效果,然而很少有研究把非局部信息嵌入到孪生网络中来实现对目标的整体感知。

受 CCNet^[22]的启发,本文设计了一种高效的局部上下文感知模块 LCAM。局部上下文感知模块可以把远程信息简化为每个像素点的十字型区域,通过十字交叉感知就可以快速地捕获非局部信息,网络能有效感知目标的上下文信息,进而增强目标状态估计。设计的局部上下文感知模块如图 3 所示。

将输入的特征图分别经过 3 个 1×1 的卷积进行降维,通过减少特征通道数来减少计算量,生成对应的特征 Q, K, V 。

先计算特征 Q, K 的亲密度(Affinity):

$$a_{i,u} = Q_u \cdot \Omega_{i,u}^T \tag{4}$$

式中: $a_{i,u}$ 为 Q_u 和 $\Omega_{i,u}^T$ 亲和程度; Q_u 为 Q 中位置 u

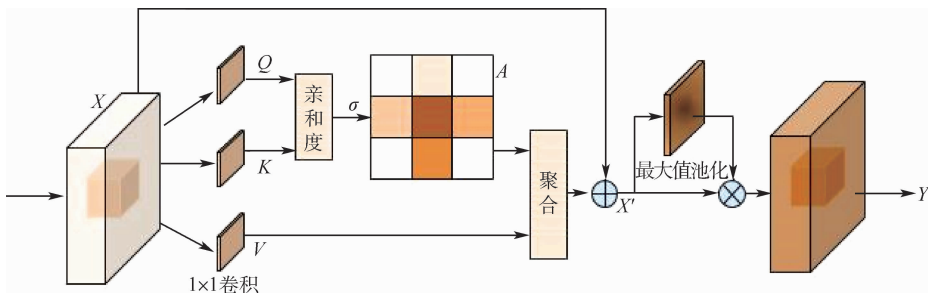


图 3 局部上下文感知模块

Fig. 3 Local context awareness module

的向量; Ω_u 为 K 中对应位置 u 的同一行或列的特征向量的集合。经过 Softmax 函数,得到亲和度特征 A 。

再通过聚合 (Aggregation) 操作把 V 中对应位置 u 的同一行或列的特征值 $\Phi_{i,u}$ 进行点乘;加上原始输入特征 X ,就得到了聚合特征 X' 。

$$X'_u = \sum_{i \in |\Phi_u|} A_{i,u} \Phi_{i,u} + X_u \quad (5)$$

最终在聚合特征的空间维度上进行最大值池化 (MaxPool),赋予空间上每个位置不同的权重。

1.4 Anchor-free 目标状态估计器

人为设计的 Anchor 带有先验知识的同时也可能存在认知偏见,对于不同的任务,存在不够灵活通用的问题,往往需要重新设计。而对于目标跟踪的对象往往是任意的、不确定的物体,使用人为设定的 Anchor 是不够准确的。由于跟踪目标在跟踪过程中往往会发生巨大的形变,Anchor 的存在对于目标的状态估计会造成影响。另外,在设计 Anchor 中往往含有大量参数,而且跟踪结果往往对这些超参数很敏感,需要耗时费力地调参。

本文采取的 Anchor-free 策略受 FoveaBox^[23] 的启发,不需要人为设计 Anchor,从根源上避免了上述问题,对于任一物体都可更准确地进行估计。例如,在数据集 OTB100 的 Gym 视频中 (见图 4),运动员肢体发生了非常大的形变,基于 Anchor 的 SiamRPN++^[24] (蓝色框) 和 DaSiamRPN^[25] (青色框) 算法都不能准确地回归出运动员的真实边框,而本文算法回归出的框 (红色) 能很恰当地包含运动员的手脚等细节。注意,由人为标注的框 (绿色) 也存在不恰当的地方。

具体来说,本文将样本标注 (GroundTruth) 内的不同区域差别看待,仅选择目标中心区域为正样本,分配方式如下:

$$\begin{cases} |x - c_x| < 0.5w\lambda_1 \\ |x - c_y| < 0.5h\lambda_1 \end{cases} \quad (6)$$



图 4 不同算法在 Gym 视频上的测试结果

Fig. 4 Test results of different algorithms on Gym video

$$\begin{cases} |x - c_x| > 0.5w\lambda_2 \\ |x - c_y| > 0.5h\lambda_2 \end{cases} \quad (7)$$

式中: c_x, c_y 和 w, h 分别为标注框的中心点坐标和框的长、宽; λ_1, λ_2 为超参数,用以限定正负样本的区域,分别设置为 0.6、0.9。当采样点 (x, y) 满足式 (6) 时,将其标为正样本;反之,满足式 (7) 为负样本;介于正负样本之间的区域选择忽略。

如图 5 所示,绿色框为目标的 GroundTruth,红色框内为正样本区域 (positive area),蓝色框与绿色框之间为负样本区域 (negative area),而介于红蓝之间的是算法采样时忽略的区域。

对于正样本区域的每个样本 (x, y) ,都可以得到一个映射关系: $f(x, y) = (x_1, y_1, x_2, y_2)$,其中, $(x_1, y_1), (x_2, y_2)$ 分别表示预测框的左上角、右下角坐标,使用 Smooth L1 函数计算预测值与 gt 之间的损失。也就是说,算法期望正样本区域内的点回归出与真实边框尽可能一致的边框,进而精准估计出目标的状态,详细内容可参见文献 [23]。

分类损失函数选择 Focal loss^[26],其对于难易样本不均衡有显著的优化效果。最终的损失 L 为 $L = L_{reg} + \lambda L_{cls}$ (8)

为在训练时较好地平衡回归损失 L_{reg} 和分类损失 L_{cls} 之间的差距,将 λ 设为 10。

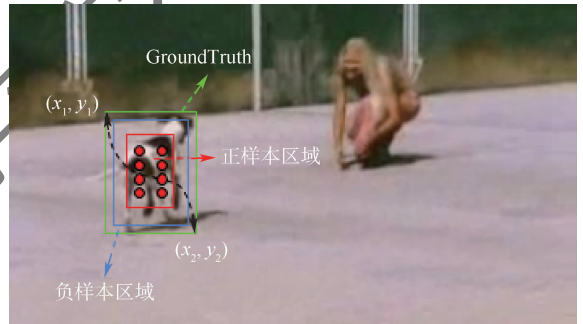


图 5 正负样本区域的分配方式示例

Fig. 5 Example of allocation of positive and negative sample areas

2 实验分析

2.1 实验平台

实验平台的 CPU 为 Intel i7 8700, GPU 为 NVIDIA RTX2070 (8 GB 显存),内存为 32 GB,操作系统为 64 位的 Ubuntu18.04,开发环境为基于 Python3.6 的 PyTorch1.2。

2.2 实验细节

训练集选择 GOT-10k^[27] 和 LaSOT^[28] 两大公开标准数据集。GOT-10k 数据集包含了超过

10 000个真实运动对象的视频片段和超过150万个手动标记的边界框;LaSOT数据集包含了70个类别,共1 400个视频序列,超过300万帧,每个视频序列平均长度超过2 000帧。共训练30轮,每轮有20 000张图片对。其中,前5轮选择冻结预训练网络ResNet18的权重,后25轮将主干网络部分加入训练,进行微调。使用随机梯度下降法进行优化,学习率从0.01逐步降到0.000 01。

2.3 实验结果对比

在OTB、VOT这2类标准数据集上进行测试。结果显示,本文算法均有较好的表现,特别是在精确率和速度上有着很大的优势。

2.3.1 在OTB上的实验结果分析

OTB2013是Wu等^[29]在2013年公开的一个目标跟踪的测试基准,共有51个视频序列;2015年,在OTB2013的基础上扩充得到了OTB100(OTB2015),共有100个视频序列,包含了OTB2013,因此直接对比在OTB100上的表现。

选择近年的算法和经典代表性算法进行比较,包括SiamRPN++, ATOM^[30]、GradNet^[31]、GCT^[32]、SiamRPN、SiamFC等。从图6(图例中数字表示成功率)可以看到,本文算法在数据集OTB100上,无论是成功率(Success)还是精确率(Precision)都优于其他算法。特别是在精确率上,本文算法不仅超过了SiamRPN++,并且大幅领先于ATOM、GradNet等其他算法。在成功率上也仅次于SiamRPN++,比DaSiamRPN、SASiam高了近2%。

数据集OTB100中各类难点属性的结果如图7所示,特别是对于物体的遮挡(occlusion)、形变(deformation)、平面外旋转(out-of-plane rota-

tion),本文算法都有很好的表现。如图7(b)、(h)、(i)所示,在精确率方面分别取得了0.921、0.890、0.934,分别高于SiamRPN++2.9%、1.6%、2.1%,远超其他算法。以上结果得益于设计的局部上下文感知模块可以对目标物体进行整体感知,以及采取的Anchor-free的状态估计方法。图7中,图例括号中的数字表示精确率。

2.3.2 在VOT上的实验结果分析

视觉目标跟踪(visual object tracking, VOT)是一个专门针对单目标跟踪的挑战赛。本文主要测试了在VOT2016、VOT2018上的表现,并与近年的算法进行了比较。VOT2016和VOT2018各有60个视频序列,包括环境光照变化、目标尺度变化及运动遮挡等各种挑战。

从表1可以看出,本文算法SiamCC在数据集VOT2016上的结果好于SPM^[33]、DaSiamRPN等算法。虽然在准确性上略逊SPM,但在鲁棒性上均胜其他算法,取得了最好的结果。相比于SPM算法,SiamCC算法降低了23.8%的错误率。从图8可以看到,在期望平均重叠率(expected average overlap, EAO)方面,SiamCC好于SPM、DaSiamRPN算法,增幅分别为3.2%、9.0%。图8中,图例括号中的数字表示期望平均重叠率。

在数据集VOT2018上,SiamCC也取得了不错的结果。如表2所示,SiamCC的鲁棒性达到了0.19,好于SiamRPN++算法,降低了17.4%的错误率。从图9可以看出,本文算法在期望平均重叠率上为0.405,仅次于SiamRPN++。相比于SiamMask^[34]算法,SiamCC增幅达6.2%,并降低了47.4%的错误率。图9中,图例括号中的数字表示期望平均重叠率。

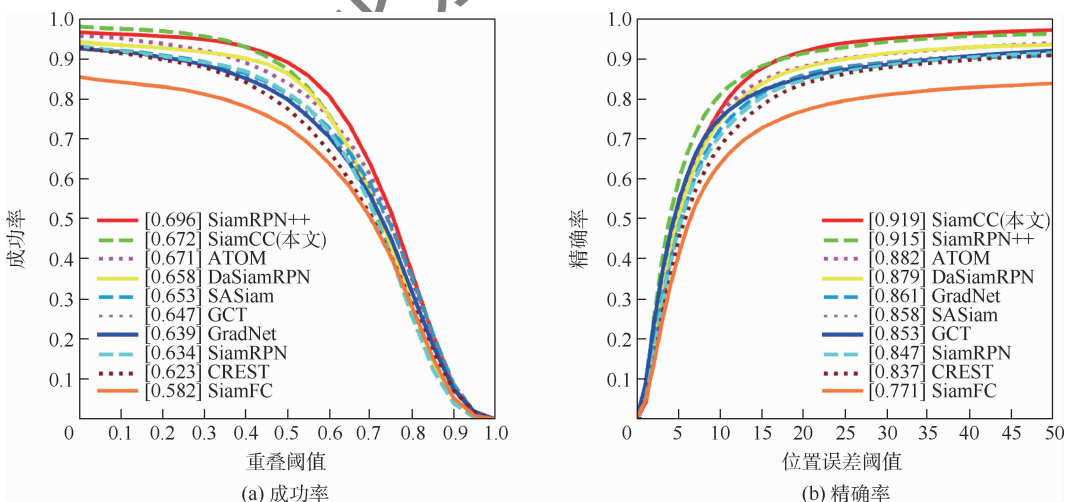
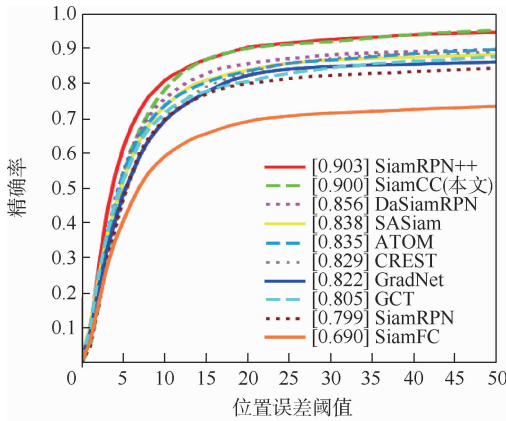
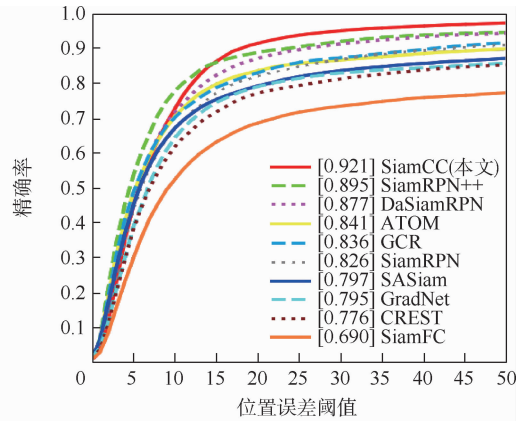


图6 不同算法在数据集OTB100上的精确率与成功率对比

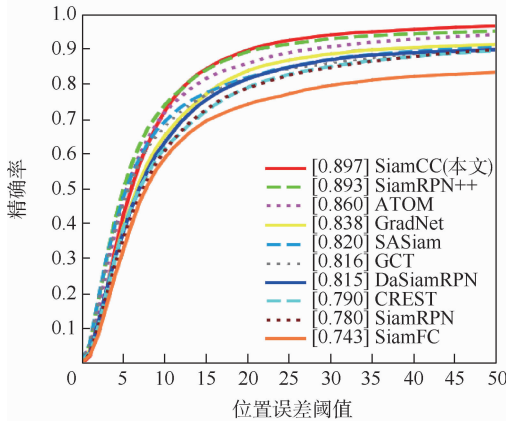
Fig. 6 Comparison of accuracy and success rates of different algorithms on the OTB100 datasets



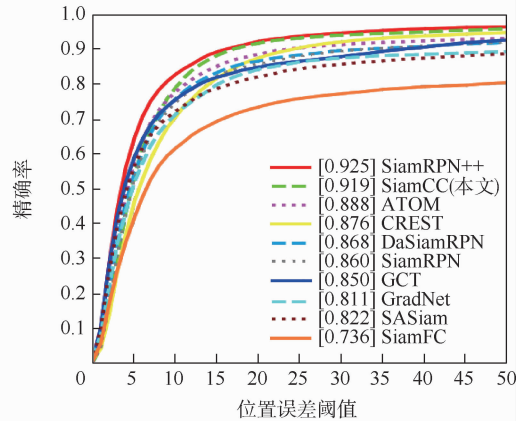
(a) 背景干扰



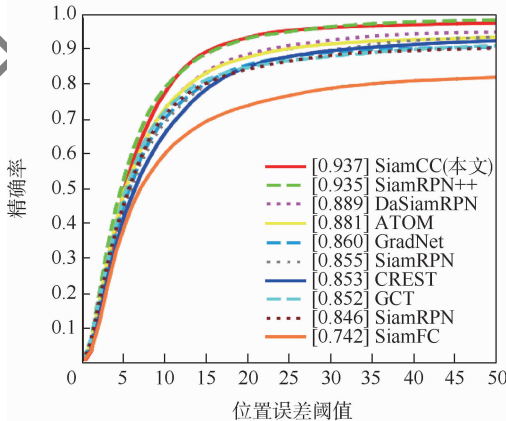
(b) 形变



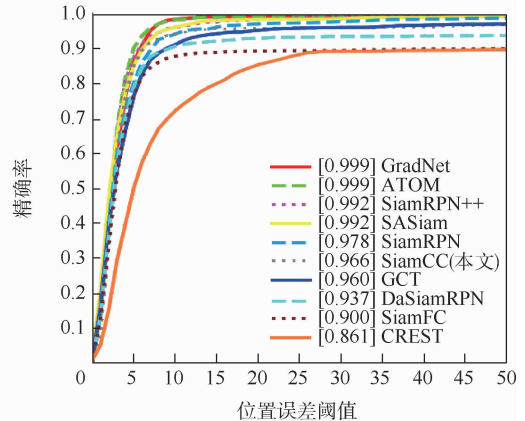
(c) 快速运动



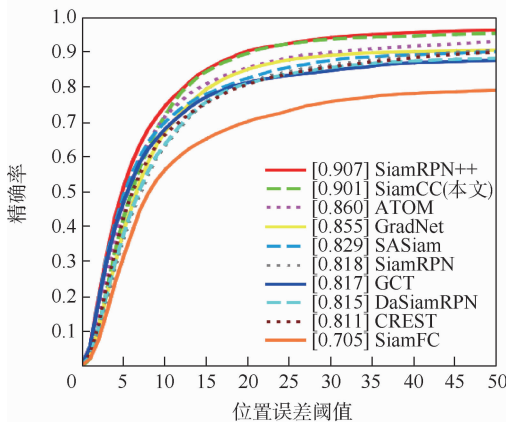
(d) 光照变化



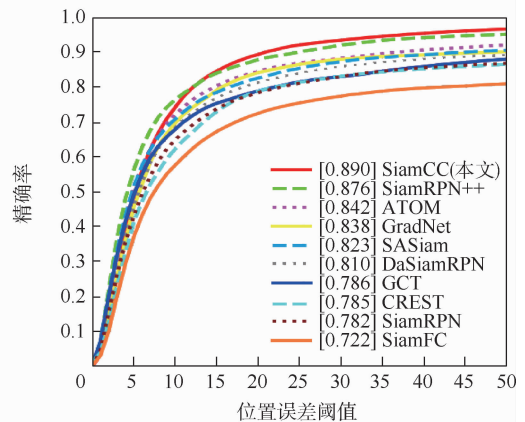
(e) 平面内旋转



(f) 低分辨率



(g) 运动模糊



(h) 遮挡

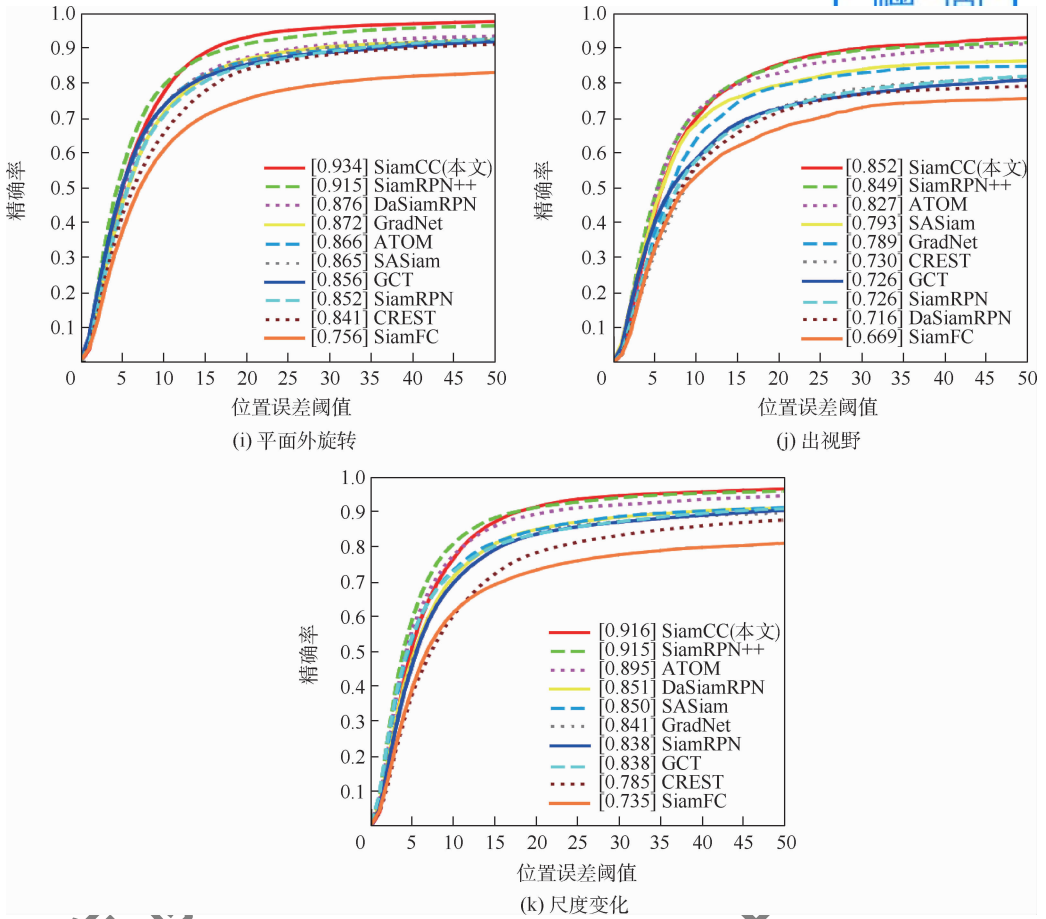


图7 不同算法在11类挑战下的精确率对比

Fig.7 Comparison of accuracy rates of different algorithms under 11 types of challenges

表1 不同算法在数据集 VOT2016 上的测试结果对比

Table 1 Comparison of test results of different algorithms on the VOT2016 datasets

算法	准确性	鲁棒性	EAO
SiamCC(本文)	0.61	0.16	0.448
SPM	0.62	0.21	0.434
DaSiamRPN	0.61	0.22	0.411
ECO	0.55	0.20	0.375
C-RPN	0.59	0.27	0.363
SiamRPN	0.56	0.26	0.344
STRCF	0.55		0.313
SASiam	0.54	0.34	0.291
MDNet	0.51	0.25	0.283
SiamFC	0.53	0.46	0.235

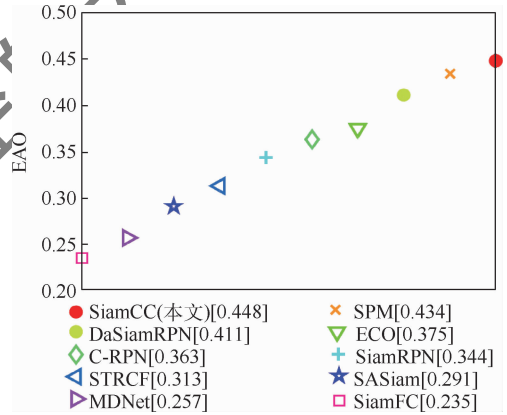


图8 不同算法在数据集 VOT2016 上期望平均重叠率对比

Fig.8 Comparison of EAO of different algorithms on the VOT2016 datasets

2.3.3 跟踪性能与速度之间的权衡分析

由于采用了计算量适中的主干网络,并设计了2个轻量级模块 ECAM 和 LCAM,在提升跟踪性能的同时几乎不会影响速度,本文算法 SiamCC 在跟踪速度和性能上更加均衡。

如图 10 所示,本文算法在跟踪性能上略好于 ATOM,但在速度上快了近 3 倍,达到了 85 帧/s,远超实时需求。虽然在速度上 SPM、DaSiamRPN 略快于 SiamCC,但在跟踪性能上均弱于 SiamCC。

与 SiamRPN ++ 算法的对比中,为了更加公平,分别对比了基于 ResNet50 和 ResNet18 两个版本的 SiamRPN ++。可以从图 10 中看出,同是以 ResNet18 为主干网络的 SiamCC,在速度和性能上均好于 SiamRPN ++ ResNet18;虽然在跟踪性能上略输于 SiamRPN ++ ResNet50,但速度上 SiamCC 快了 2.5 倍。由此得出,SiamCC 在跟踪性能和速度方面更加均衡。

表 2 不同算法在数据集 VOT2018 上的测试结果对比

2.4 消融实验

Table 2 Comparison of test results of different algorithms on the VOT2018 datasets

算法	准确性	鲁棒性	EAO
SiamCC(本文)	0.58	0.19	0.405
SiamRPN++	0.60	0.23	0.414
ATOM	0.59	0.20	0.401
SiamMask	0.61	0.28	0.380
SPM	0.58	0.30	0.338
DaSiamRPN	0.56	0.34	0.326
ECO	0.48	0.27	0.280
GradNet	0.51	0.38	0.247
SiamRPN	0.49	0.46	0.244
SASiam	0.50	0.46	0.236

为了说明本文设计的 2 个轻量级模块的有效性,进行了消融实验。使用同样的训练数据和方式,以没有 ECAM 和 LCAM 模块为基准,训练了只加 ECAM 和只加 LCAM;同时也训练了未做改进的 ECANet 和 CCANet 模块。在数据集 OTB2013 和 VOT2016 上进行测试。从表 3 可以对比分析出:

1) 对比第 1、2、4 行数据,在原有基准网络上添加 ECANet 模块后,测得 OTB2013 成功率和 VOT2016 期望平均重叠率分别仅仅提升 0.6% 和 0.7%;而设计的 ECAM 模块能分别提升 1.2% 和 2.1%。

2) 对比第 1、3、5 行,同样在原有基准网络上添加 CCNet 模块后,测得 OTB2013 成功率和 VOT2016 期望平均重叠率分别仅仅提升 0.9% 和 1.2%;而设计的 LCAM 模块能分别提升 2.3% 和 3.4%。

3) 对比第 1、6、7 行,在原有基准网络上同时添加 ECANet、CCNet 模块后,测得 OTB2013 成功率和 VOT2016 期望平均重叠率分别仅仅提升 1.7% 和 2.9%;而设计的 ECAM、LCAM 模块能分别提升 3.5% 和 7.4%。

从速度对比上可以看出,增加的 2 个轻量级模块影响很小。进而证明,本文设计的 2 个轻量级模块 ECAM 和 LCAM 确实在提升性能的同时几乎不会带来太多的计算开销。

2.5 实验结果定性分析

为了对比 SiamCC 与其他算法的差异,选择 OTB100 的测试结果进行定性分析。测试结果如图 11 所示。6 个视频序列包括了遮挡、形变、复杂背景、旋转、快速移动等挑战。绿色为标注框,红色为本文算法 SiamCC,蓝色、黄色、青色、橘色分别为 SiamRPN++、ATOM、DaSiamRPN、GradNet 算法。

在 Bird1 测试序列中,SiamCC 能有效应对出现的遮挡情况,并在后续帧中鲁棒地克服物体的巨大形变。在 Board 和 Soccer 这 2 个序列中,由

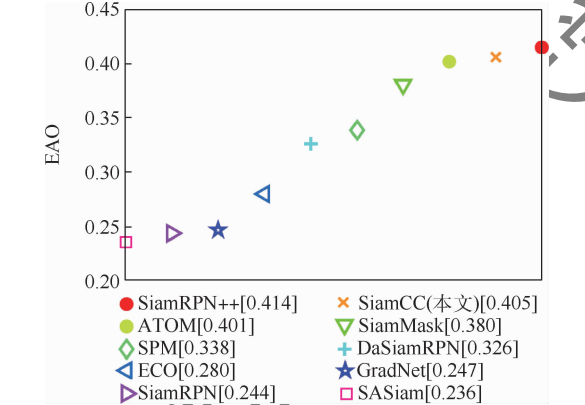


图 9 不同算法在数据集 VOT2018 上期望平均重叠率对比
Fig. 9 Comparison of EAO of different algorithms on the VOT2018 datasets

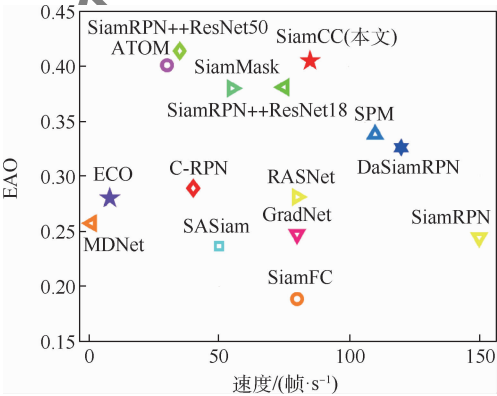


图 10 不同算法在数据集 VOT2018 上的性能与速度对比
Fig. 10 Comparison of performance and speed of different algorithms on VOT2018 dataset

表 3 本文算法在数据集 OTB2013 和 VOT2016 上的消融实验

Table 3 Ablation experiment of the proposed algorithm on OTB2013 and VOT2016 datasets

模块	模块	模块	模块	OTB2013 成功率	VOT2016 EAO	速度/(帧·s ⁻¹)
				0.661	0.417	93
ECANet				0.665	0.420	91
	CCNet			0.667	0.422	88
		ECAM		0.669	0.426	91
			LCAM	0.676	0.431	88
ECANet	CCNet			0.672	0.429	85
		ECAM	LCAM	0.684	0.448	85



图 11 各类算法在不同视频下的跟踪结果

Fig. 11 Display of tracking results of various algorithms with different videos

于复杂背景的存在,很容易干扰到跟踪器,如 Board 第 352 帧和 Soccer 第 129 帧,其他算法都相继出现跟丢的现象,而 SiamCC 却能有效地跟上目标;当物体发生旋转时,Board 第 498 帧中,其他算法并没有很好地估计出目标的大小。在 Diving 和 Jump 视频里,由于人体姿态发生了巨大的形变,对跟踪算法有着极高的挑战,SiamCC 能准确地估计人体的尺寸大小且没有发生丢失现象,如 Diving 第 106、213 帧,Jump 第 34、52、100 帧。而在 Ironman 视频里,由于光照变化及物体的快速移动,其他算法都出现了跟丢的现象,虽然 SiamCC 也有一点飘移,仍能比较稳健地跟踪上目标。

3 结 论

本文提出的基于高效注意力和上下文感知的目标跟踪算法,在所设计的 2 个轻量级模块作用下,可以有效地学习到目标的丰富信息和局部上下文信息,来应对物体的严重形变和复杂环境的干扰,并且不会带来计算上的开销。

1) 高效通道注意力模块 ECAM 在不减少通

道数的前提下,通过交互式学习特征通道之间的关系,丰富了目标的特征信息,有效解决了目标发生严重形变和旋转时的情况。

2) 局部上下文感知模块 LCAM 以高效的方式捕获目标的局部上下文信息,增强网络对目标的整体感知,进而应对部分遮挡时复杂环境的干扰。

3) 在标准数据集上有着不错的表现。在测试集 OTB100 上的成功率和精确率分别达到了 0.672 和 0.919;在测试集 VOT2016 和 VOT2018 上期望平均重叠率分别达到了 0.448 和 0.405;并且能以 85 帧/s 的速度超实时跟踪。

今后的工作会尝试采用更深、更高效的主干网络来提取特征,采用更大的数据集来训练网络模型,得到更好的跟踪结果。

参考文献 (References)

- [1] 孟磊,杨旭. 目标跟踪算法综述[J]. 自动化学报, 2019, 45 (7): 1244-1260.
- MENG L, YANG X. A survey of object tracking algorithms[J]. Acta Automatica Sinica, 2019, 45 (7): 1244-1260 (in Chi-

- nese).
- [2] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 2544-2550.
- [3] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (3): 583-596.
- [4] DANELLJAN M, BHAT G, SHAHBAZ K F, et al. ECO: Efficient convolution operators for tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6638-6646.
- [5] 李玺, 查宇飞, 张天柱, 等. 基于深度学习的目标跟踪算法发展综述 [J]. 中国图象图形学报, 2019, 24 (12): 2057-2080.
LI X, ZHA Y F, ZHANG T Z, et al. A survey of visual object tracking algorithms based on deep learning [J]. Journal of Image and Graphics, 2019, 24 (12): 2057-2080 (in Chinese).
- [6] WANG N, YEUNG D Y. Learning a deep compact image representation for visual tracking [C] // Proceedings of the 26th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 809-817.
- [7] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 3074-3082.
- [8] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking [C] // European Conference on Computer Vision. Berlin: Springer, 2016: 850-865.
- [9] HE A, LUO C, TIAN X, et al. A twofold Siamese network for real-time object tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4834-4843.
- [10] LI B, YAN J, WU W, et al. High performance visual tracking with Siamese region proposal network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8971-8980.
- [11] FAN H, LING H. Siamese cascaded region proposal networks for realtime visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7952-7961.
- [12] SHEN J, TANG X, DONG X, et al. Visual object tracking by hierarchical attention Siamese network [J]. IEEE Transactions on Cybernetics, 2019, 50 (7): 3068-3080.
- [13] ZHANG Z, PENG H. Deeper and wider Siamese networks for real-time visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4591-4600.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
- [15] KEIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems, 2012: 1097-1105.
- [16] WANG Q, TENG Z, XING J, et al. Learning attentions: Residual attentional Siamese network for high performance online visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4854-4863.
- [17] WANG F, JIANG M, QIAN C, et al. Residual attention network for image classification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 3156-3164.
- [18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.
- [19] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C] // European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.
- [20] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11534-11542.
- [21] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7794-7803.
- [22] HUANG Z, WANG X, HUANG L, et al. CCNet: Criss-cross attention for semantic segmentation [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 603-612.
- [23] KONG T, SUN F, LIU H, et al. FoveaBox: Beyond anchor-based object detection [J]. IEEE Transactions on Image Processing, 2020, 29: 7389-7398.
- [24] LI B, WU W, WANG Q, et al. SiamRPN++: Evolution of Siamese visual tracking with very deep networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4282-4291.
- [25] ZHU Z, WANG Q, LI B, et al. Distractor-aware Siamese networks for visual object tracking [C] // European Conference on Computer Vision. Berlin: Springer, 2018: 101-117.
- [26] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.
- [27] HUANG L, ZHAO X, HUANG K. GOT-10k: A large high-diversity benchmark for generic object tracking in the wild [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43 (5): 1562-1577.
- [28] FAN H, LIN L, YANG F, et al. LaSOT: A high-quality benchmark for large-scale single object tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5374-5383.
- [29] WU Y, LIM J, YANG M H. Online object tracking: A benchmark [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 2411-2418.

- [30] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: Accurate tracking by overlap maximization [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4660-4669.
- [31] LI P, CHEN B, OUYANG W, et al. GradNet: Gradient-guided network for visual object tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6162-6171.
- [32] GAO J, ZHANG T, XU C. Graph convolutional tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4649-4659.
- [33] WANG G, LUO C, XIONG Z, et al. SPM-tracker: Series-parallel matching for real-time visual object tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3643-3652.
- [34] WANG Q, ZHANG L, BERTINETTO L, et al. Fast online object tracking and segmentation: A unifying approach [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 1328-1338.

Target tracking algorithm based on efficient attention and context awareness

BAI Luo, ZHANG Hongli*, WANG Cong

(School of Electrical Engineering, Xinjiang University, Urumqi 830047, China)

Abstract: The matching-based Siamese network algorithm often lacks the overall perception of a target, which easily leads to inaccurate target state estimation and target missing in complex environments. Therefore, this paper designs two lightweight modules on the basis of the twin network to achieve more accurate and robust target tracking. An efficient channel attention module is embedded into the backbone network after its construction for feature extraction. Efficient extraction of target features and enhanced differential representation are achieved, so that the network pays more attention to the target information. The features after template matching pass a local context awareness module, thus enhancing the network's overall perception of the target to deal with the complex and changeable environment in the tracking process. The Anchor-free state estimation strategy is used to achieve accurate estimation of the target. Experimental results show that on the datasets OTB100, VOT2016 and VOT2018, SiamCC algorithm outperforms DaSiamRPN algorithms and ATOM algorithm, with the tracking speed reaching 85 frame/s.

Keywords: machine vision; target tracking; Siamese network; channel attention; context awareness

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

DOI: 10.13700/j.bh.1001-5965.2021.0006

基于 DDQN 的片上网络混合关键性消息调度方法



李国梁, 李峭*, 徐亚军, 熊华钢

(北京航空航天大学 电子信息工程学院, 北京 100083)

摘要: 对片上网络(NoC)承载的混合关键性消息进行实时调度是其应用于航空电子系统片上多核通信的关键。为解决可满足性模理论(SMT)法求解效率低、低优先级消息等待延迟大的问题,提出了一种基于双深度Q网络(DDQN)的混合关键性消息调度方法。将虫孔交换机制下的消息调度问题建模为马尔可夫决策过程,建立包含环境、动作、状态、奖励的多层感知调度模型;随机生成多组分布不同的混合关键性消息作为训练样本,采用DDQN算法求解该调度模型;在此基础上,提出并实现了带孔隙DDQN算法,在保证时间触发(TT)消息可调度前提下为速率约束(RC)消息预留用于虫孔交换的时隙。算例研究表明:所提方法的求解时长及TT消息确定性端到端延迟的平均值均低于SMT法;带孔隙DDQN算法的RC消息延迟较不带孔隙DDQN算法和SMT法显著降低。

关键词: 片上网络(NoC); 时间触发(TT)机制; 双深度Q网络(DDQN); 混合关键性消息; 消息调度

中图分类号: Y247; TP393

文献标志码: A 文章编号: 1001-5965(2022)07-1233-09

随着分布式综合模块化航空电子(distributed integrated modular avionics, DIMA)架构的发展,未来航电系统将进一步朝着智能化、微内核和片上系统(system on chip, SoC)的方向发展^[1-2]。片上网络(network on chip, NoC)借鉴分布式计算机系统的互连概念,将运算与通信解耦合,解决了总线结构SoC通信效率低、扩展性差、功耗高三大难题,为航电系统片上多核带来一种全新的通信方式。为适应复杂多变的空中环境,机载网络需要支持不同实时性需求的应用间通信,因此能否将NoC应用到航电领域的关键在于能否对其承载的混合关键性消息进行实时调度。

NoC消息调度是综合考虑时序分配和路径规划的NP难题^[3],混合整数线性规划(mixed inte-

ger linear programming, MILP)^[3]、元启发式算法^[4-6]和可满足性模理论(satisfiability modulo theories, SMT)法^[7]是当前常用的3种方法。其中, MILP法与SMT法通过求解无冲突约束条件得到离线调度表,已有学者通过拓扑分解^[7]、简化约束条件^[8]对SMT法进行优化,但仍未突破求解时长随网络规模增大呈指数型增长的局限性,难以应用于高可靠、强实时的航电领域;遗传算法^[4]、离散粒子群优化算法^[5]、模拟退火算法^[6]等元启发式算法虽能对调度表进行优化,但通常执行效率较低且模型难以泛化。

此外,从提高网络可靠性的角度出发,文献^[9]为每条时间触发(TT)消息设置独立的更新时间点,解决了“TIME4”模型^[10]及“TWO-

收稿日期: 2021-01-06; 录用日期: 2021-04-11; 网络出版时间: 2021-04-26 10:07

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210425.1811.003.html

基金项目: 国家自然科学基金(62071023)

* 通信作者. E-mail: avionics@buaa.edu.cn

引用格式: 李国梁, 李峭, 徐亚军, 等. 基于DDQN的片上网络混合关键性消息调度方法[J]. 北京航空航天大学学报, 2022, 48(7): 1233-1241. LI G L, LI Q, XU Y J, et al. A DDQN-based mixed-criticality messages scheduling method for network-on-chip[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1233-1241 (in Chinese).

PHASE”模型^[11]重构时由于更新不一致导致的帧丢失问题;从提高网络实时性的角度出发,文献[8]对基本SMT法的约束条件进行适应性修改,针对不同应用场景分别实现了背靠背调度算法、小延迟调度算法及大孔隙调度算法,降低了速率约束(RC)消息的等待延迟。但2种方法均是基于SMT法进行优化,若消息分布改变或网络发生故障,需要重新修改约束条件输入求解器求解,不适合在线重构。

2013年,DeepMind团队利用神经网络和Q学习(Q-learning,QL)算法玩Atari游戏^[12],将深度学习强大的感知能力和强化学习的决策能力相结合,首次提出深度强化学习(deep reinforcement learning,DRL),战胜了人类最顶尖棋手的Alpha-Go即结合了DRL与蒙特卡罗树搜索^[13]。鉴于DRL在解决搜索规划类问题时的优异表现,已有使用强化学习算法求解TT消息调度表的尝试^[14],其求解效率较SMT法显著提高,但该方法调度前假设传输路径先验已知,未对TT消息进行路径规划,也未考虑时间触发调度间隔对RC消息延迟的影响。

本文根据混合关键性消息所属通信任务的优

先级顺序,结合NoC的虚通道及虫孔交换机制,将NoC消息调度抽象为马尔可夫决策过程(Markov decision process,MDP),并建立包含环境、动作、状态、奖励的调度模型。在此基础上,随机生成多组分布不同的待调度消息组,将每条消息执行某个动作后的状态转移情况及所获奖励存入经验池,待经验池存满后随机采样若干个经验序列喂入神经网络进行训练,训练完成后保存模型指导新消息调度。同时,对基本双深度Q网络(double deep Q-network,DDQN)算法^[15]进行优化,在保证TT消息可调度性的前提下,能够给RC消息预留一个时隙的带宽,降低了RC消息等待延迟。

1 时间触发片上网络

时间触发通信机制的引入使航电网络从航空电子全双工交换式以太网(AFDX)时代迈入时间触发以太网(TTE)时代^[16]。为顺应航电网络的发展趋势,本文着眼于图1所示的网络架构,其中,航电模块为执行某种特定功能的SoC,TTE用于各航电模块间的有线互联,单一模块内部多核间通信则由时间触发片上网络(TTNoC)负责。

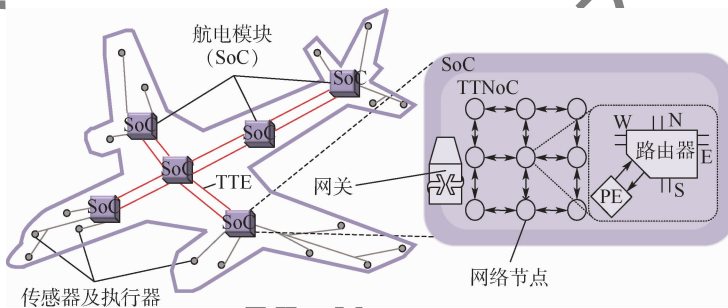


图1 航电网络系统架构

Fig. 1 System architecture of avionics network

1.1 网络结构

在航空、航天等对实时性要求较高的安全关键性领域,可以将时间触发架构引入NoC中。图1所示的TTNoC由路由器、处理单元(PE)及全双工数据链路3部分组成。其中,路由器负责接收、转发数据包;处理单元为执行计算、控制功能的嵌入式IP核,内部常集成网络接口(NI),负责路由器与处理单元间数据格式的转换。在Mesh、Torus、胖树等众多NoC拓扑结构中,Mesh结构以其良好的可扩展性成为当前应用最广泛的结构^[17],如图1所示,其节点呈网格型排列,每个路由器包含5个输入、输出端口,分别与东、西、南、北4个方向上的相邻路由器及本地处理单元相连。

1.2 消息类型

TTNoC在IEEE 802.3标准以太网数据包定义的基础上扩展时间触发通信的同步机制^[16],以使其同时支持标准以太网的事件触发(ET)消息及确定性的TT消息。

TT消息具有严格时间确定性,按照全局调度表进行转发,具有最高优先级,对应实时性要求最高的任务。

ET消息又分为RC消息和尽力传(BE)消息。其中,RC消息与AFDX兼容,优先级仅次于TT消息,占用时间触发调度孔隙传输,通常需要满足带宽分配间隔及最大帧长2个约束条件^[8];BE消息优先级最低,无服务质量保证,使用TT消息与RC消息的剩余带宽传输^[16]。鉴于BE消

息无传输确定性保证,本文在调度时只考虑 TT 消息和 RC 消息。

1.3 消息转发机制

单一芯片硬件资源有限,缓冲区深度浅,难以实现复杂通信协议,适于采用虫孔交换与虚通道(virtual channel, VC)相结合的轻量化转发机制。

虫孔交换将消息以微片(flit)为单位传输,并进一步将微片划分为头微片、体微片和尾微片。头微片携带整个分组的路由信息,体微片及尾微片按头微片建立的路由传输,尾微片传输完成后释放当前链路资源供其他消息使用^[18]。虫孔交换的优势在于:缓冲区深度较分组交换明显降低,网络阻塞时各微片就地缓存,利于负载均衡;此外,RC 消息无需等待大时隙出现,TT 消息调度间隔不小于一个微片长度即可传输 RC 消息,可降低 RC 消息等待延迟,保证其传输确定性。

虚通道技术将单一物理通道分为逻辑上互相隔离的若干个虚通道,通过合理调度虚通道实现分时复用以提高信道利用率^[18]。例如,图 2 所示的路由器每个物理通道包含 4 个虚通道,每个虚通道缓存深度为 4 个微片长度,当路由器 R1 的西输入端口同时有 A、B、C 三种不同类型的微片请求输入时,可将其存入不同虚通道的缓冲区,避免因输入竞争导致数据丢失。此外,假设路由器 R1 的南输入缓存中存满 F 微片,输出端为南向(路由器 R1 的南输出即路由器 R2 的北输入);路

由器 R2 的北输入缓存中存满 E 微片,输出端为北向。若无虚通道,路由器 R1 和路由器 R2 会一直等待对方先释放资源,造成“死锁”现象。虽然虚通道技术解决了数据丢失及“死锁”问题,但增加虚通道要相应地增加缓存及虚通道分配消耗,实际应用中,虚通道数一般不超过 4 个。

2 基于 DDQN 的调度模型

2.1 深度强化学习基础

2.1.1 马尔可夫决策过程

若某个随机过程满足式(1),即下一状态 S_{t+1} 仅由当前状态 S_t 决定,与历史状态无关,则称此过程具有马尔可夫特性^[12]。

$$P(S_{t+1} | S_t) = P(S_{t+1} | S_t, S_{t-1}, \dots, S_1, S_0) \quad (1)$$

MDP 过程是强化学习的基础^[19],可以用一个六元组 $\langle S, A, P, R, \pi, \gamma \rangle$ 表示, S 为当前过程的有限状态集合, A 为有限动作集合,策略 π 表示在特定状态下执行某个动作的概率,特别地,在状态 s 下执行动作 a 的概率为

$$\pi(a | s) = P(A_t = a | S_t = s) \quad (2)$$

P 为状态转移概率矩阵,在 s_1 采取动作 a_1 到达 s_2 的概率 $P_{s_1 \rightarrow s_2}^{a_1}$ 可表示为

$$P_{s_1 \rightarrow s_2}^{a_1} = P(S_{t+1} = s_2 | S_t = s_1, A_t = a_1) \quad (3)$$

R 为奖励矩阵,由于在状态 s_1 采取动作 a_1 可能到达若干个状态,采用所有可达状态所获奖励的期望值来衡量在 s_1 采取动作 a_1 的奖励,即

$$R_{s_1}^{a_1} = E[R_t | S_t = s_1, A_t = a_1] \quad (4)$$

单步奖励难以从整体上反映某个状态的价值,引入 G_t 表示在时刻 t 从状态 s 开始沿某条轨迹到达终点的累积奖励值,即

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \dots + \gamma^T R_{t+T} = \sum_{n=0}^T \gamma^n R_{t+n} \quad (5)$$

式中: $\gamma \in [0, 1]$ 为奖励值折扣因子,体现延时奖励对状态 s 价值的影响; T 为从状态 s 沿当前轨迹到达终点所经历的状态数。

更进一步,在状态 s 按照策略 π 可以有 multiple 到达终止状态的轨迹,所有轨迹累积奖励值的期望称为状态价值函数 $V_\pi(s)$,用来表示状态 s 的整体价值。实际操作中难以遍历所有轨迹,通常结合贝尔曼期望等式^[15]更新 $V_\pi(s)$,即

$$V_\pi(s) = E_\pi[G_t | S_t = s] = E_\pi[R_t + \gamma V_\pi(s_{t+1}) | S_t = s] \quad (6)$$

式中: s_{t+1} 表示状态 s 按策略 π 到达的下一状态。

2.1.2 Q 学习

执行某个动作的价值往往比某个状态的价值

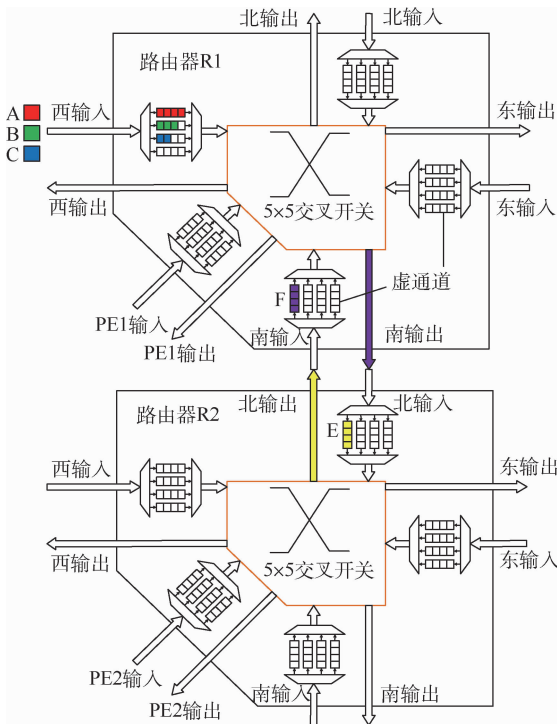


图 2 虚通道路由器结构

Fig. 2 Router architecture with virtual channel

更直观,故 Q 学习进一步定义了“状态—动作”价值函数 $Q_{\pi}(s, a)$, 用来表示仅在状态 s 强制执行动作 a , 后续状态仍按策略 π 执行所获得的奖励值。在 Q 学习中, $Q_{\pi}(s, a)$ 简称为 Q 函数, 可表示为

$$Q_{\pi}(s, a) = E[G_t | S_t = s, A_t = a] \quad (7)$$

与监督学习不同, Q 学习的状态转移概率矩阵 P 及奖励矩阵 R 是未知的, 需要与环境(本文环境为整个网络拓扑, 包括所有链路的排布方式及所有路由器的缓存占用情况)交互来更新 P 和 R 。

如图 3 所示, Q 学习用初始策略 π 与环境交互, 采用时序差分法评价策略 π , 即每执行一个动作都进行 Q 值更新。学习到策略 π 的 Q 值后, 在状态 s 的可选动作集 A 中执行最大 Q 值对应的动作即可得到策略 π^* , 即

$$\pi^*(a | s) = \begin{cases} 1 & a = \arg \max_{a \in A} Q_{\pi}(s, a) \\ 0 & \text{其他} \end{cases} \quad (8)$$

这种局部贪心的思想保证给定策略 π 总能找到更优的策略 π^* 。令 $\pi = \pi^*$ 继续训练, 直至收敛到最优策略。 Q 值的时序差分更新公式为

$$Q_{\pi}(s, a) = Q_{\pi}(s, a) + \alpha \left(R_s^a + \gamma \sum_{s'} P_{s \rightarrow s'}^a \max_{a' \in A'} Q_{\pi}(s', a') - Q_{\pi}(s, a) \right) \quad (9)$$

式中: $\alpha \in (0, 1]$ 为学习效率因子; $P_{s \rightarrow s'}^a$ 表示在 s 执行 a 转移到 s' 的概率; A' 为 s' 的动作集; S' 为 s' 的状态集。

如果应用场景中执行某个动作后状态转移是确定的, 更新公式简化为

$$Q_{\pi}(s, a) = Q_{\pi}(s, a) + \alpha \left(R_s^a + \gamma \max_{a' \in A'} Q_{\pi}(s', a') - Q_{\pi}(s, a) \right) \quad (10)$$

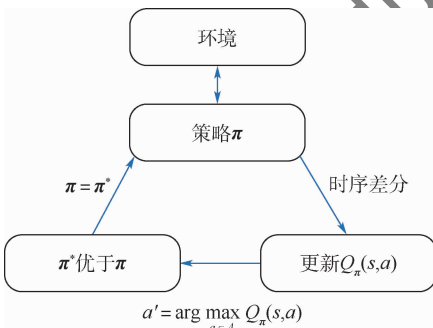


图 3 Q 学习训练过程

Fig. 3 Training process of Q-learning

2.1.3 深度 Q 网络

Q 学习用表格存储 Q 值, 对于状态数量庞大的应用场景, 维护 Q 表需占据大量存储空间且在线搜索效率较低^[19]。深度 Q 网络 (deep Q-net-

work, DQN) 结合值函数估计和神经网络技术, 通过神经网络输出 Q 值的估计值, 解决了 Q 学习的“维度爆炸”问题。

可将 DQN 的训练过程视为回归问题, 即通过反向传播更新网络参数, 以最小化预测值和标签间的偏差, 其均方误差意义下的损失函数 $L(\theta)$ 为

$$L(\theta) = E \left[\left(R_s^a + \gamma \max_{a' \in A'} Q_{\pi}(s', a', \theta) - Q_{\pi}(s, a, \theta) \right)^2 \right] \quad (11)$$

式中: $R_s^a + \gamma \max_{a' \in A'} Q_{\pi}(s', a', \theta)$ 为训练标签; θ 为神经网络参数集; $Q_{\pi}(s', a', \theta)$ 和 $Q_{\pi}(s, a, \theta)$ 为神经网络在不同“状态—动作”对下的输出预测值。

得到 $L(\theta)$ 后, 通过随机梯度下降更新 θ 使损失函数收敛即可完成训练。

上述神经网络训练时, 输出值和标签均动态变化且二者具有较强相关性, 易导致训练不稳定, 可通过引入目标网络机制解决这一问题^[20]。如图 4 所示, 将某个定长时间段设定为一个时间步, 目标网络参数在单个时间步内保持不变, 每隔 C 步将训练网络参数更新给目标网络, 单个时间步内转化为求解一个标签固定的回归问题, $L(\theta)$ 相应地变为

$$L(\theta) = E \left[\left(R_s^a + \gamma \max_{a' \in A'} Q_{\pi}(s', a', \theta^-) - Q_{\pi}(s, a, \theta) \right)^2 \right] \quad (12)$$

式中: θ 和 θ^- 分别为训练网络和目标网络参数, 且 θ^- 为前一时间步的 θ 。

此外, 较强的样本连续性也易导致训练不稳定, 可将采集到的样本存入经验池, 每隔固定时间从经验池中随机采样训练。

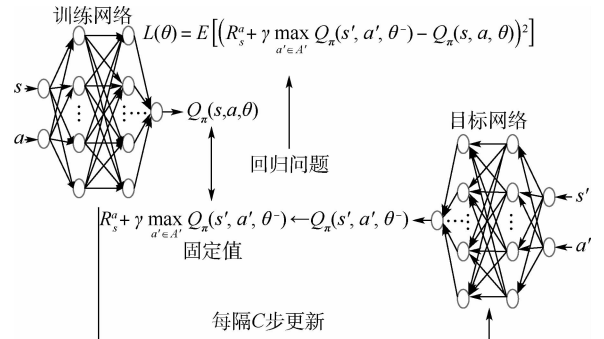


图 4 目标网络机制

Fig. 4 Target network mechanism

2.1.4 双深度 Q 网络

Q 学习与 DQN 通过执行当前策略最大 Q 值对应动作来优化策略, 容易过高估计某个动作的价值, 而 DDQN 可通过对 Q 值无偏估计来解决高估问题^[15]。DDQN 沿用图 4 所示的网络结构, 但

其更新时通过训练网络选择动作,通过目标网络来计算对应的 Q 值,其 $L(\theta)$ 为

$$L(\theta) = E \left[\left(R_s^a + \gamma Q_{\pi}(s', \arg \max_{a'} Q_{\pi}(s', a', \theta), \theta^-) - Q_{\pi}(s, a, \theta) \right)^2 \right] \quad (13)$$

DDQN 原理在于:若训练网络高估了某个动作,只要目标网络未高估此动作,仍可按目标网络计算出准确 Q 值;若目标网络高估了某个动作,只要训练网络未高估此动作,则该动作不会被选择。

2.2 基于 DDQN 的调度模型设计

2.2.1 可调度性经验感知机制

假设 M1、M2、M3 三条消息(以微片形式,下文所述消息均以微片形式传输)经过链路 1 和链路 2 到达目的节点。如图 5(a)所示,当虚通道缓存充足时,3 条消息占据不同虚通道,通过分时复用实现无碰撞传输。若其他虚通道缓存均被占用,M1、M2、M3 只能通过同一虚通道传输,而虫孔交换要求当前消息传输完成后才释放虚通道缓存,即占用同一虚通道的消息传输时间不能交叉。

如图 5(b)所示,M1 对虚通道的占用时间超过 M2 的周期,势必会导致 M2 帧丢失,但 M3 仍可利用 M1 的孔隙完成传输,由此可见,周期越长的消息越容易找到调度孔隙。此外,消息帧长及消息源端和目的端的相对距离也会影响调度难度。当消息所属应用之间无依赖关系时,可在调度前将消息按周期由短到长、源端到目的端的相对距离(以最短路径跳数衡量)由长到短、帧长由长到短进行可调度性排序。这种基于先验经验的排序可以尽量避免多条消息占用同一虚通道导致的帧丢失现象,跳过某些明显无解的状态,加速神经网络的收敛。

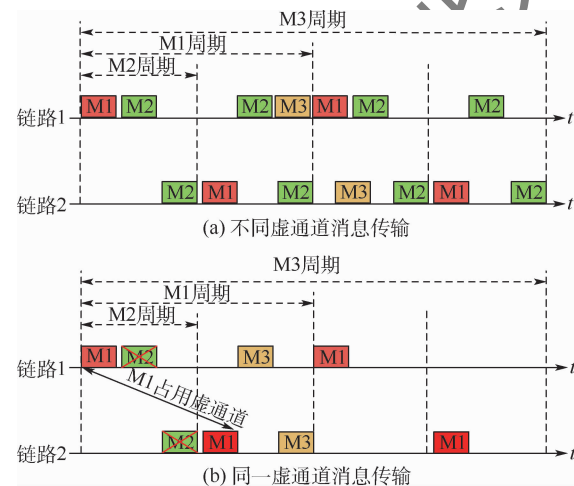


图 5 虚通道缓存对消息传输的影响

Fig. 5 Impact of virtual channel cache on message transmission

2.2.2 TTNoC 调度的马尔可夫特性

尽管 TTNoC 对不同类型消息的服务质量保证不同,同类型消息的固有属性、派发时刻也各不相同,但从增量化调度角度分析,当前待调度消息仅与此刻链路中已调度消息的排布方式及路由器虚通道缓存有关,与历史状态无关,具备式(1)所述马尔可夫特性。

借鉴时分多址的思想,将时间轴分为若干个等间隔时隙,每个时隙完成单个微片传输,将原本连续时间轴上的消息传输区间分配转化为离散的时序分配,使其符合 MDP 离散动作要求。进一步,将 TTNoC 消息调度建模为 2.1.1 节所述的 MDP 过程。调度环境为整个网络拓扑结构;DDQN 算法被视为与环境交互的智能体;状态 s 表示当前所有数据链路已调度消息的排布方式及路由器虚通道的缓存占有率;动作 a 体现路径规划下的时序分配,表示当前时刻选择某条链路的某个时隙传输当前消息;奖励值 R 的设置应使智能体倾向于选择容易求得调度表且端到端延迟较小的动作,其计算公式为

$$R = \Delta d + \frac{1}{\Delta f} + \frac{1}{c} + h \quad (14)$$

式中: Δd 为执行动作前后消息所在节点与目的节点的曼哈顿距离(两节点间的最短路径跳数)之差, $\Delta d = 1$ 表示靠近目的节点, $\Delta d = -1$ 表示远离目的节点; Δf 为单跳延迟,即执行动作前后链路派发时刻之差; c 为当前链路已调度消息数,特别地,消息数为 0 时令 $c = 0.5$; h 的设置在于使智能体尽快完成调度,若执行当前动作后消息完成调度,令 $h = 10$,否则 $h = 0$ 。

2.2.3 DDQN 模型实现

Hasselt 等^[15] 提出 DDQN 时采用多层卷积神经网络(convolutional neural network, CNN)提取图像特征,鉴于 TTNoC 消息调度不涉及图像处理,为防止模型过于复杂导致过拟合现象,简化了 DDQN 的网络结构,采取图 6 所示的多层感知(multilayer perception, MLP)神经网络模型。图中: n_{pre} 和 n_{des} 分别为当前消息所在节点和目的节点编号;其余神经元从 p_{fir} 到 p_{last} 表示网络拓扑中所有单向链路的利用率。

上述 MLP 模型中,输入层神经元个数取决于网络拓扑规模。例如, $N \times N$ 的 Mesh 结构共有 $2N(N-1)$ 条全双工链路,为使状态 s 尽可能表示当前调度表特征,使每条单向链路对应一个神经元,故其输入层共有 $4N(N-1) + 2$ 个神经元。此外,隐藏层层数过大时提升层数带来的增益不明

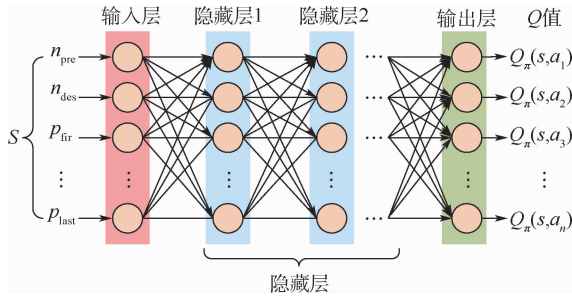


图6 多层感知神经网络模型

Fig.6 Multilayer perception model of neural network

显,故根据具体网络规模将隐藏层层数设置为2~4层,并采用Relu激活函数。输出层采用Softmax激活函数,表示智能体在状态 s 执行各个动作对应的 Q 值,输出神经元个数即状态 s 下的可选动作数。

2.2.4 带孔隙 DDQN 算法

上述基本 DDQN 算法在 TT 消息调度时,为最小化端到端延迟,在下一跳链路中应选择最早空闲时隙,此时动作数等于当前节点的度(即与当前节点有直连链路的节点数),该方法与文献[8]中的背靠背调度方法类似,此时 TT 消息首尾相接,无调度间隔,易加剧 RC 消息的等待延迟。因此,在保证 TT 消息可调度前提下(即当前 TT 消息截止期限前有若干个空闲时隙),可在执行动作时选择最早空闲时隙的后一个时隙。

带孔隙 DDQN 算法实现时,仅需修改式(14)中 Δf 即可。即 TT 消息截止期限(通常等于周期)前有若干个时隙时,若执行动作后 $\Delta f=1$,强制令 $\Delta f=3$,这样即可让智能体倾向于选择 $\Delta f=2$ 的动作,为 RC 消息预留了一个时隙。由于单个时隙时间粒度较小,空出一个时隙在提高 RC 消息传输确定性的同时并不会从根本上影响 TT 消息的端到端延迟。

采用2.1节的目标网络和经验回放机制对 MLP 模型进行训练,为保证模型 Q 值预测的准确性,随机生成多组分布不同的待调度消息作为训练样本,并采用 ϵ 贪婪搜索解决探索—利用窘境^[15];即以概率 $1-\epsilon$ 按照 Q 值选择动作,以 ϵ 的概率随机选取动作。模型训练过程如算法1所示。

2.3 基于 DDQN 的调度模型求解

算法1 DDQN 模型训练算法。

输入:目标网络更新频率 C 、单次采样规模 B 、经验池 D 、经验池规模 D_s 、消息集合 M 。

1: $\theta \rightarrow Q$ -network, $\theta^- \rightarrow$ target network, $\theta = \theta^-$

2: for each msg in M :

3: for each time step t :

4: 给定状态 s ,基于 Q 网络选动作(ϵ 贪心机制)

5: 获得奖励 R_s^a ,到达新状态 s'

6: if current buffer depth $< D_s$:

7: 将 (s, a, R_s^a, s') 存入经验池

8: else:

9: 用 (s, a, R_s^a, s') 代替最早的经验元组

10: 从经验池 D 中均匀随机采样 B 个经验元组

11: 定义 $a^{\max} = \arg \max_{a'} Q(s', a', \theta)$

12: if s' is teiminal: Target $y = R_s^a$

13: else: Target $y = R_s^a + \gamma Q(s', a^{\max},$

$\theta^-)$

14: 用SGD在损失函数 $|y - Q(s, a, \theta)|^2$ 下更新参数集 θ

15: 每 C 个时间步令 $\theta^- = \theta$

16: end for

17: end for

训练完成后,保存模型参数用来指导TTNoC消息调度,调度算法如算法2所示。

算法2 基于 DDQN 的调度算法。

输入:待调度消息集合 M_s 、MLP模型。

输出: M_s 的调度表sch。

1: for i in range(1, the number of M_s):

2: for j in range($i-1, -1, -1$):

3: 定义 temp = msg ^{j} (the j^{th} message of M_s)

4: if (msg ^{j} . period $>$ temp. period) || (msg ^{j} . period = temp. period && msg ^{j} . hop $<$ temp. hop) || (msg ^{j} . period = temp. period && msg ^{j} . hop = temp. hop && msg ^{j} . length $<$ temp. length):

5: msg ^{$j+1$} \leftarrow msg ^{j}

6: end if

7: end for

8: msg ^{$j+1$} \leftarrow temp

9: end for

10: for each msg in M_s :

11: while 当前状态未到达终止状态 do

12: 根据算法1训练的MLP模型选择动作

13: 将msg的转发时刻写入调度表

14: end while

15: end for

3 实验与分析

3.1 实验条件

3.1.1 模型训练

选取图1所示的4×4 Mesh结构验证本文算法,使用TensorFlow 2.1搭载MLP模型:输入层包含50个神经元,隐藏层层数为2,每层128个神经元。随机生成1000组包含50条消息的样本,将每条消息从源节点开始执行某个动作前后状态转移情况及所获奖励存入经验池(经验池存满后最新获得的经验值覆盖最早经验值),待经验值存满后,随机抽取64个经验序列喂入神经网络进行训练,当前消息到达目的节点后视为当前轨迹到达终止状态,所有消息到达终止状态后完成训练并保存训练模型。

3.1.2 调度实验

待调度消息的源端及目的端随机生成;TT消息的周期及RC消息的带宽分配间隔(BAG)在{1,2,4,8,16,32} ms中取值,所有消息均采用以太网帧结构,帧长范围为64~1518 Byte,加上前导符和帧起始定界符(8 Byte)及最小帧间间隔(12 Byte),消息帧总长度范围为84~1538 Byte^[16],进一步将数据帧分为4个微片,每个微片长度为21~385 Byte;链路带宽为100 Mbit/s,故单个微片传输时间为1.68~30.8 μs;将1 ms均分为20个时隙,每个时隙长50 μs,保证每个时隙可完成单个微片传输,剩余带宽为保护间隔,可供BE消息使用。

3.2 实验结果分析

3.2.1 TT消息调度分析

图7为DDQN算法与文献[7]中基于SMT的拓扑分解法(简称拓扑分解法,且拓扑分解法也采用虫孔交换机制)在4×4 Mesh结构下运行时

间对比。可以看出,随着消息数量增加,拓扑分解法求解时长近似呈指数型增长,而DDQN算法求解时长近似呈线性增长且远小于拓扑分解法。以图7中调度50条消息为例,DDQN算法求解时长仅为拓扑分解法的1.7%。

分析其原因在于:拓扑分解法需要满足一系列无冲突约束条件,随着消息数量增加,约束条件相应成倍增加,SMT求解器搜索区间也相应成倍增大,造成求解时间过长;而本文DDQN算法完成训练后,将当前状态输入到MLP模型会输出一个最优动作,故求解时长近似随消息数量呈线性增长。值得一提的是,3.1.1节所述的模型训练花费的时间未考虑在内,这是因为该部分时间消耗不属于调度过程,且对于同一个拓扑结构来讲,训练时间代价只需消耗一次。此外,训练完成后,DDQN算法不会像拓扑分解法一样由于消息分布改变造成约束条件改变,适合重构场景下的消息调度。

3.2.2 TT消息端到端延迟

消息延迟包括派发时刻与生成时刻的差值及消息在NoC中的传输延迟。假设所有待调度消息均在0时刻生成,则派发时刻可完全反映消息从处理单元产生到传入NoC的等待延迟。图8对比了同一待调度消息组分别采用DDQN算法与拓扑分解法的平均派发时刻及平均端到端延迟。由于2种算法求解时均在可调度前提下贪婪选择最早时隙,故两者端到端延迟差异不大,但DDQN算法的平均派发时刻远小于拓扑分解法,即其求得的调度表TT消息的端到端延迟小于SMT法。以图8中调度50条消息为例,DDQN算法通过缩短消息生成时刻和派发时刻的差值,使得TT消息平均端到端延迟仅为拓扑分解法的21.9%。

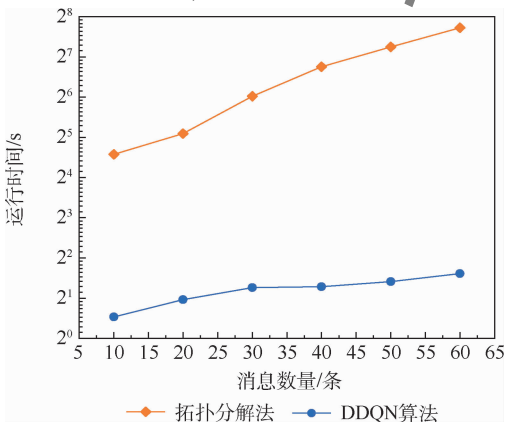


图7 DDQN算法与拓扑分解法运行时间对比

Fig.7 Execution time comparison between DDQN algorithm and TBDA-SMT method

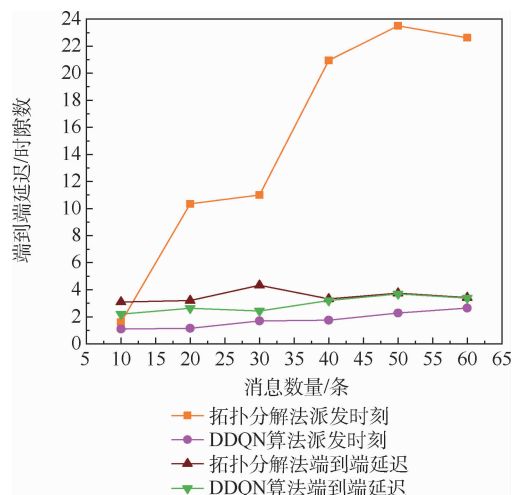


图8 DDQN算法与拓扑分解法平均派发时刻及端到端延迟对比

Fig.8 Comparison of average dispatch time and end-to-end delay between DDQN algorithm and TBDA-SMT method

3.2.3 RC 消息的传输确定性分析

虫孔交换机制下 RC 消息无需等待大孔隙出现,只要 RC 消息生成时刻所在时隙无 TT 消息占用即可完成传输。如 2.2.4 节所述,训练 MLP 模型时,在保证 TT 消息可调度前提下可预留一个时隙带宽供 RC 消息使用。

随机生成 50 条 TT 消息,分别用拓扑分解法、带孔隙 DDQN 算法及不带孔隙 DDQN 算法求解 TT 消息调度表,在此基础上对比表 1 所示 RC 消息集在 3 种算法下的等待延迟。表中: n_i 为第 i 个网络节点(按从上到下、从左到右排序)。

对比结果如图 9 所示。分析可得,带孔隙 DDQN 算法 RC 消息等待延迟最低,当链路资源充足时,该算法理论上 RC 消息最多等待一个时隙;而不带孔隙 DDQN 算法及拓扑分解法的 RC 消息等待延迟无确界保证,2 种算法在 TT 调度时贪婪搜索最早时隙会使部分 RC 消息(见图 9 中的 RC1、RC9、RC16、RC19)等待延迟过高。

表 1 RC 消息参数
Table 1 Properties of RC message

编号	源节点及目的节点	BAG/ms	生成时刻/时隙
RC1	$n_6 \rightarrow n_9$	8	3
RC2	$n_6 \rightarrow n_{10}$	32	6
RC3	$n_{15} \rightarrow n_{14}$	16	8
RC4	$n_7 \rightarrow n_{14}$	4	7
⋮			
RC18	$n_0 \rightarrow n_{11}$	2	8
RC19	$n_5 \rightarrow n_7$	16	2
RC20	$n_0 \rightarrow n_{13}$	1	8

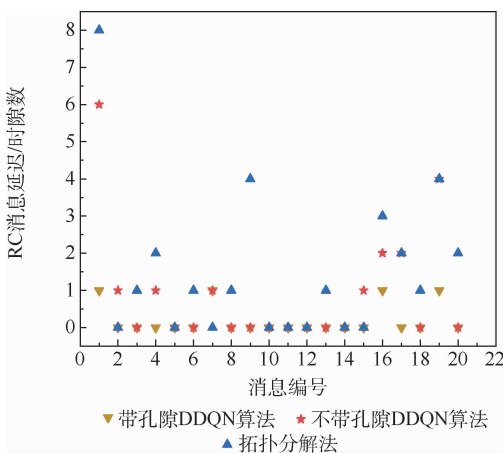


图 9 不同算法下的 RC 消息等待延迟

Fig. 9 RC message delays in different algorithms

4 结论

1) 将 NoC 消息调度问题建模为 MDP 过程,并建立包含环境、动作、状态、奖励的调度模型。

2) 采用 DDQN 算法求解调度模型,算法性能较基于 SMT 的拓扑分解法显著提升。在不考虑训练调度模型所消耗的时间前提下,以 4×4 Mesh 结构调度 50 条消息为例,DDQN 算法求解时长仅为 SMT 法的 1.7%;此外,DDQN 算法通过缩短消息生成时刻和派发时刻的差值,使得 TT 消息平均端到端延迟仅为 SMT 法的 21.9%。

3) 优化后的带孔隙 DDQN 算法理想情况下 RC 消息最多等待一个时隙,较基于 SMT 的拓扑分解法及不带孔隙 DDQN 算法大大缩短 RC 消息的等待延迟,提高其传输确定性。

本文建立的调度模型及求解方法可以对深度强化学习算法应用于 NoC 消息调度提供理论依据及实验参考。

参考文献 (References)

- [1] WOLFIG R, JAKOVLJEVIC M. Distributed IMA and DO-297: Architectural, communication and certification attributes [C] // Proceedings Digital Avionics Systems Conference. Piscataway: IEEE Press, 2008: 1. E. 4-1-1-E. 4-10.
- [2] 孔韵雯, 李峭, 熊华钢, 等. 片间综合化互连时间触发通信调度方法 [J]. 航空学报, 2018, 39(2): 321590.
KONG Y W, LI Q, XIONG H G, et al. Time-triggered communication scheduling method for off-chip integrated interconnection [J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(2): 321590 (in Chinese).
- [3] MURSHED A, OBERMAISSER H, AHMADIAN H, et al. Scheduling and allocation of time-triggered and event-triggered services for multi-core processors with networks-on-a-chip [C] // 2015 IEEE 13th International Conference on Industrial Informatics (INDIN). Piscataway: IEEE Press, 2015: 1424-1431.
- [4] 易伟, 王佳文, 潘红兵, 等. 基于蚁群混沌遗传算法的片上网络映射 [J]. 电子学报, 2011, 39(8): 1832-1836.
YI W, WANG J W, PAN H B, et al. Ant colony chaos genetic algorithm for mapping task graphs to a network on chip [J]. Acta Electronica Sinica, 2011, 39(8): 1832-1836 (in Chinese).
- [5] SAHU P K, SHAH T, MANNA K, et al. Application mapping onto mesh-based network-on-chip using discrete particle swarm optimization [J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2014, 22(2): 300-312.
- [6] CHAI S, LI Y, WANG J, et al. A list simulated annealing algorithm for task scheduling on network-on-chip [J]. Journal of Computers, 2014, 9(1): 176-182.
- [7] SHI Y F, LI Q, YANG J H, et al. A topology-based decomposition approach for time-triggered message scheduling in network-on-chip [C] // 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC). Piscataway: IEEE Press, 2019: 1-8.
- [8] 鲁俊, 何锋, 熊华钢, 等. 软件定义时间触发网络的调度算法优化 [J]. 北京航空航天大学学报, 2021, 47(5): 1004-1014.
LU J, HE F, XIONG H G, et al. Scheduling algorithms optimization in software defined time-triggered ethernet [J]. Journal of

- Beijing University of Aeronautics and Astronautics, 2021, 47 (5):1004-1014 (in Chinese).
- [9] LI Z H, WAN H, PANG Z Y, et al. An enhanced reconfiguration for deterministic transmission in time-triggered networks [J]. IEEE/ACM Transactions on Networking, 2019, 27 (3): 1124-1137.
- [10] MIZRAHI T, MOSES Y. Software defined networks: It's about time [C] // The 35th Annual IEEE International Conference on Computer Communications. Piscataway: IEEE Press, 2016: 1-9.
- [11] MARK R, NATE F, JENNIFER R, et al. Abstractions for network update [J]. SIGCOMM Computer, 2012, 42 (4): 323-334.
- [12] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing Atari with deep reinforcement learning [C] // Proceedings of Workshops at the 26th Neural Information Processing Systems 2013, 2013: 201-220.
- [13] SILVER D, HUANG A, MADDISON C, et al. Mastering the game of go with deep neural networks and tree search [J]. Nature, 2016, 529 (7587): 484-489.
- [14] 李浩若, 何锋, 郑重, 等. 基于强化学习的时间触发通信调度方法 [J]. 北京航空航天大学学报, 2019, 45 (9): 1894-1901.
LI H R, HE F, ZHENG Z, et al. Time-triggered communication scheduling method based on reinforcement learning [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45 (9): 1894-1901 (in Chinese).
- [15] HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning [C] // Association for the Advance of Artificial Intelligence, 2015: 2094-2100.
- [16] 何锋. 机载网络技术基础 [M]. 北京: 国防工业出版社, 2018: 192-196.
HE F. Fundamentals of airborne network [M]. Beijing: National Defense Industry Press, 2018: 192-196 (in Chinese).
- [17] 王炜, 乔林, 汤志忠. 片上网络互连拓扑综述 [J]. 计算机科学, 2011, 38 (10): 1-5.
- [18] 黄惠凡. 片上网络路由器设计与实现 [D]. 西安: 西安电子科技大学, 2013: 14-16.
HUANG Z F. The router design and implementation of network on chip [D]. Xi'an: Xidian University, 2013: 14-16 (in Chinese).
- [19] SU Y S, FAN R, FU X M, et al. An adaptive deep Q-network-based energy and latency-aware routing protocol design for underwater acoustic sensor networks [J]. IEEE Access, 2019, 7: 9091-9104.
- [20] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518 (7540): 529-533.

A DDQN-based mixed-criticality messages scheduling method for network-on-chip

LI Guoliang, LI Qiao*, XU Yajun, XIONG Huagang

(School of Electronics and Information Engineering, Beihang University, Beijing 100083, China)

Abstract: Real-time scheduling of mixed-criticality messages carried by network on chip (NoC) is the key to its application to on-chip multi-core communication in avionics system. A double deep Q-network (DDQN) method was proposed to solve the problem of satisfiability modulo theories (SMT) to low efficiency and high delay of low-priority messages. The message scheduling problem under wormhole switch mechanism was modeled as a Markov decision process, and a scheduling model including environment, action, state and reward was established. Then, DDQN was applied to interact with environment in different message distribution generated randomly, and the empirical sequence obtained through interaction was regarded as the training sample of the neural network. In addition, a scheduling method named pore-DDQN was implemented, that is, a time slot was reserved for rate-constrained (RC) messages on the condition that time-triggered (TT) messages can be scheduled. The case study shows that the solution time and the average end-to-end delay of TT messages of DDQN are lower than that of SMT, and the delay of RC messages with pore-DDQN is significantly lower than that of DDQN and SMT.

Keywords: network on chip (NoC); time-triggered (TT) mechanism; double deep Q-network (DDQN); mixed-criticality message; message scheduling

Received: 2021-01-06; Accepted: 2021-04-11; Published online: 2021-04-26 10:07

URL: kns.cnki.net/kcms/detail/11.2625.V.20210425.1811.003.html

Foundation item: National Natural Science Foundation of China (62071023)

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

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

DOI: 10.13700/j.bh.1001-5965.2021.0019

基于离散系数改进的 VWKNN 位置 指纹定位算法



许甜, 何泾沙, 朱娜斐*, 邓万航, 吴霜, 他永君

(北京工业大学 信息学部, 北京 100124)

摘 要: 位置指纹算法是研究室内定位技术的主要方法, 其中在线阶段的匹配算法是影响室内定位精度的主要因素之一。目前, 在线阶段的匹配算法有最近邻算法、K 近邻算法以及加权 K 近邻算法。其中, 最近邻算法和 K 近邻算法都没有考虑到不同参考点和待定位点之间的欧氏距离对定位精度的影响, 而加权 K 近邻算法虽然考虑到了欧氏距离对定位精度的影响, 对最终的定位结果采用欧氏距离归一化处理进行加权, 却没有考虑到 AP 信号的波动性对定位结果也会产生很大的影响。因此, 针对在线阶段的匹配算法作出改进, 提出了基于离散系数改进的加权 K 近邻算法。在离线阶段建立位置指纹数据库, 在在线阶段使用离散系数来反映各 AP 信号的稳定性, 进而对待定位点与参考点之间的欧氏距离进行加权, 计算出所有的加权欧氏距离后, 从中选取距离最近的 k 个参考点, 估算出待定位点的物理位置。实验结果表明: 基于离散系数改进的加权 K 近邻算法可以实现平均定位精度比 K 近邻算法提高 15% ~ 17%, 较加权 K 近邻算法提高了 11% ~ 13% 的定位效果。

关键词: 室内定位; 位置指纹; AP 信号强度; 加权 K 近邻; 离散系数

中图分类号: TN92

文献标志码: A

文章编号: 1001-5965(2022)07-1242-10

随着互联网技术及移动智能终端的不断发展与普及, 人们对基于物理位置的服务需求不断增加, 基于物理位置的各项服务技术进入了飞速发展的阶段^[1]。虽然在室外环境中 GPS 定位技术的精度已经达到米级, 但是在复杂的室内环境中, 由于信号会发生衰减及多径效应, 使得 GPS 技术并不能很好地发挥作用^[2]。据研究调查表明, 人们生活中的大部分时间都是处于室内环境中的。因此, 对于室内定位技术的研究具有极其重要的意义。

室内定位技术有很多种, 如基于 Wi-Fi 的位置指纹定位技术、蓝牙定位技术、步行者航位推算 (PDR) 定位技术、红外线定位技术及超宽带定位

技术等^[3]。而 Wi-Fi 设施在室内分布十分普遍, 给室内定位技术的实现带来了极大的便利, 利用现有的 Wi-Fi 进行室内定位具有成本低、操作方便及无需额外设备辅助等诸多优点, 已经成为室内定位技术研究的热点之一。

目前, 基于 Wi-Fi 的室内定位技术有基于信号到达时间 (TOA) 定位法、基于信号到达时间差 (TDOA) 定位法、基于信号到达角度 (AOA) 定位法及位置指纹法^[4]。TOA 定位的原理是通过终端接收到信号源节点发送信号的到达时间来获得信号源节点到待定位目标的距离^[5]。使用 TOA 算法进行定位的关键是信号源节点和移动终端必须具备时间上的同步, 否则将会出现较大的时间

收稿日期: 2021-01-14; 录用日期: 2021-05-07; 网络出版时间: 2021-06-18 09:06

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210617.1652.001.html

基金项目: 国家重点研发计划 (2019QY(Y)0601); 山东省自然科学基金 (ZR2020MF029)

* 通信作者. E-mail: zmf@bjut.edu.cn

引用格式: 许甜, 何泾沙, 朱娜斐, 等. 基于离散系数改进的 VWKNN 位置指纹定位算法[J]. 北京航空航天大学学报, 2022, 48(7): 1242-1251. XU T, HE J S, ZHU N F, et al. VWKNN location fingerprint positioning algorithm based on improved discrete coefficient [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1242-1251 (in Chinese).

误差,对定位的准确性造成很大影响。而现有的无线信号很难提供高精度的时间同步,因此这种定位方法还需要改进。TDOA 定位是一种利用时间差进行定位的方法,通过测量信号到达监测站的时间确定信号源的距离,再利用信号源到各监测站的距离确定信号的具体位置^[6]。但是,绝对时间一般较难测量,还需要额外的硬件设备监测时间。AOA 定位不需要测量时间获取距离,而是通过添加天线阵列或其他接收设备测量无线信号的发射角度,信号发射端到移动接收端的直线所形成的侧位线的延伸线的交点就是待定位点的位置^[7]。该定位方法同样需要添加额外的硬件设备,成本较高。位置指纹法与其他 3 种基于测距的定位方法相比,不需要测量 Wi-Fi 的接入点位置,也不需要使用测量时间和角度的硬件设备,具有操作便捷、成本低等优点,是室内定位方法中最广泛的研究方法^[8]。位置指纹法通过计算位置指纹的相似度确定待定位点的物理位置,其基本原理是:在离线阶段采集所有采样点的位置指纹信息,录入位置指纹数据库中;在在线阶段利用采集到的待定位点的信息和指纹数据库中的位置指纹进行匹配,得到待定位点的坐标^[9]。

因此,针对 AP 信号强度不稳定导致在线阶段的定位误差问题,本文提出了一种基于离散系数改进的加权 K 近邻 (weighted K-nearest neighbor, WKNN) 算法。离散系数可以反映单位均值上的离散程度^[10]。各 AP 信号源的离散系数可以很好地反映出每个 AP 的稳定程度。首先,计算出各 AP 的离散系数,对其进行归一化处理后计算出每一个 AP 的权值。其中,AP 信号源越稳定,其离散系数就越小,相应的权值就越大;AP 信号源越不稳定,其离散系数就越大,相应的权值就越小。这样,稳定的 AP 信号就被更多地考虑进欧氏距离的计算中,使得加权后的欧氏距离能够更好地反映待定位点与参考点之间的相似度。然后,计算待定位点与各参考点之间经过离散系数加权后的欧氏距离,对其进行排序,选出加权欧氏距离最小的 k 个参考点进行最终定位结果的计算^[11]。由于加权后的欧氏距离充分考虑到了 AP 信号的波动性对定位结果造成的影响,最终选出的 k 个最近邻的参考点更加准确,计算出的待定位点物理位置坐标也更加精确。最后,通过实验进行验证,证明了改进后算法的优越性。

1 位置指纹算法

位置指纹就是把实际场景当中的物理位置和

某种“指纹”联系起来,一个位置对应一个唯一的指纹信息。基于 Wi-Fi 的位置指纹算法实现室内定位分为 2 个阶段:离线阶段和在线阶段^[12]。在离线阶段需要设置参考点,采集各参考点的位置指纹信息,建立离线位置指纹数据库;在在线阶段,将待定位点的指纹信息与离线位置指纹数据库中的指纹信息进行匹配,得到待定位点的物理位置。图 1 为位置指纹算法定位原理。

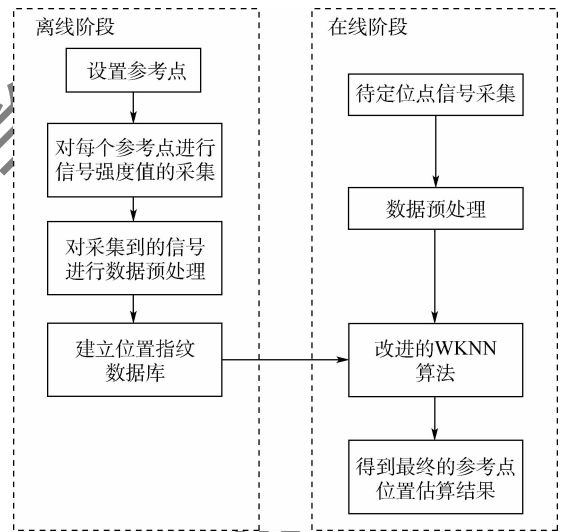


图 1 位置指纹算法定位原理

Fig. 1 Positioning schematic diagram of location fingerprint algorithm

1.1 离线阶段位置指纹数据库的建立

首先,设置合理的参考点。测量每个参考点的物理位置,选择 n 个 AP,在每一个参考点处采集这 n 个 AP 的信号强度值,录入位置指纹数据库中。位置指纹数据库结构为

$$\mathbf{LF} = \begin{bmatrix} x_1 & y_1 & \text{rssi}_1^1 & \text{rssi}_1^2 & \cdots & \text{rssi}_1^n \\ x_2 & y_2 & \text{rssi}_2^1 & \text{rssi}_2^2 & \cdots & \text{rssi}_2^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_m & y_m & \text{rssi}_m^1 & \text{rssi}_m^2 & \cdots & \text{rssi}_m^n \end{bmatrix} \quad (1)$$

式中: \mathbf{LF} 为离线位置指纹数据库,包含每一个参考点的物理位置信息及每一个 AP 在每个参考点处的信号强度值; (x_m, y_m) 为第 m 个参考点处的物理位置; rssi_m^n 为第 m 个参考点处接收到的第 n 个 AP 的信号强度值。

1.2 在线阶段匹配定位

在线阶段,需先采集待定位点处各 AP 的信号强度值,通过合适的匹配算法,用测得的信号强度值与离线位置指纹数据库中的各参考点的 AP 信号强度值进行匹配,找到与待定位点最相似的 k 个参考点;再对这最近邻的 k 个参考点的物理位置坐标进行计算,从而估算出待定位点处的物

理位置坐标。

最常用的位置指纹算法有最近邻(NN)算法^[13]、K近邻(KNN)算法^[14]及WKNN算法^[15]。

1.2.1 最近邻算法

最近邻算法的定位原理是:先采集待定位点处的各AP的信号强度值,再利用AP信号强度计算出待定位点与各参考点之间的欧氏距离,选出欧氏距离最小的参考点的位置坐标作为待定位点最终的物理位置坐标^[16]。待定位点与第*i*个参考点之间的欧氏距离计算式为

$$D_i = \sqrt{\sum_{j=1}^n (\text{RSSI}_j - \text{rssi}_i^j)^2} \quad (2)$$

式中: D_i 为待定位点与第*i*个参考点之间的欧氏距离; RSSI_j 为待定位点在第*j*个AP处测得的信号强度值; rssi_i^j 为第*i*个参考点在第*j*个AP处测得的信号强度值。

待定位点的物理位置坐标为

$$(x, y) = \min(D_i) \quad i \in (1, n) \quad (3)$$

最近邻算法相对于其他定位算法来说比较简单,花费的时间也较短,因为其只需要找出欧氏距离最小的点的位置坐标就可以。但是在实际场景当中,信号的传播是不稳定的,只把最小的欧氏距离作为判断参考点与待定位点是否最相似的指标是不准确的,会给定位的结果带来很大的误差。

1.2.2 K近邻算法

KNN算法是在最近邻算法计算出待定位点与各参考点之间的欧氏距离后,对所有的欧氏距离进行从小到大的排序,选出欧氏距离最小的*k* ($k > 1$)个参考点,把这*k*个参考点的物理位置坐标在每个维度上求平均值作为待定位点最终的物理位置坐标^[17]。

待定位点的物理位置坐标为

$$(x, y) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i) \quad k > 1 \quad (4)$$

KNN算法中最重要的2个因素就是*k*值的选取和距离的计算,*k*值中的*k*代表选出的欧氏距离最小的*k*个参考点。在*k*值的选取方面,可以通过对不同的*k*值进行实验来选取最优的*k*值。但是在距离的计算方面,由于与待定位点距离不同的参考点对定位结果的影响是不同的,欧氏距离越大的参考点,与待定位点之间的相似度越高,在计算定位结果时就应该被更多的考虑进来。如果不根据每一个参考点对定位结果的贡献程度给每个参考点赋予不同的权值进行最终定位结果的计算,就一定会对最终的定位结果产生很

大的影响。

1.2.3 加权K近邻算法

WKNN算法就是在KNN算法的基础上,考虑到与待定位点之间的欧氏距离不同的参考点对定位结果影响不同,使用欧氏距离来计算最终选出的*k*个最近邻参考点的不同的权值^[18]。使用WKNN算法计算待定位点的物理位置坐标为

$$(x, y) = \sum_{i=1}^k \omega_i (x_i, y_i) \quad k > 1 \quad (5)$$

式中: (x, y) 为待定位点的物理位置坐标; (x_i, y_i) 为最终选出的*k*个最近邻的参考点中第*i*个参考点的物理位置坐标; ω_i 为第*i*个参考点的权值, ω_i 的计算式为

$$\omega_i = \frac{1}{D_i + 1} \quad k > 1 \quad (6)$$

$$\sum_{j=1}^k \frac{1}{D_j + 1}$$

式中: D_i 为最终选出的*k*个最近邻的参考点中,每一个参考点与待定位点之间的欧氏距离。这样,欧氏距离越小,就说明该参考点与待定位点之间的相似度越高,对定位结果的影响就越大,在计算定位结果时应该被更多的考虑进计算当中,在计算时相应的权值较大。WKNN算法考虑了待定位点与不同参考点之间的相似度对定位结果产生影响的问题,通过欧氏距离对定位结果进行加权计算,可以在一定程度上提高定位的精确度^[19]。

KNN算法没有考虑欧氏距离不同的参考点对定位结果影响不同,只是对最近邻的*k*个参考点的物理坐标的各维向量取平均值进行最终待定位点坐标的计算,会对最终的定位结果造成很大的影响。WKNN算法虽然弥补了KNN算法的缺点,给欧氏距离不同的参考点赋予不同的权值进行最终定位结果的计算^[20],但是,在实际场景中,AP信号不稳定,会受到障碍物遮挡、信号反射及多径效应等影响,如果在计算中不把AP信号的稳定性考虑进来,将所有的AP信号视为一样稳定,也会在很大程度上影响到最终的定位结果。因此,在计算欧氏距离时,将AP信号的稳定性这一因素考虑进来可以在一定程度上提高室内定位的精度。

2 基于离散系数改进的VWKNN算法

由于实际场景中的环境十分复杂,不仅墙壁、障碍物的遮挡会导致移动终端对信号的接收造成影响,室内人员的大量流动也会使终端采集到的信号产生波动^[21]。对于不同的位置,每个AP信

号的稳定性是不同的。针对这种情况,本文提出了基于离散系数改进的 WKNN 算法,称为 VWKNN 算法。利用各 AP 的离散系数对欧氏距离进行加权计算,使加权后的欧氏距离能更准确地反映出待定位点与参考点之间的相似程度。

2.1 传统欧氏距离

WKNN 算法通过先计算待定位点与参考点之间的欧氏距离,再把欧氏距离从小到大进行排序,进而找出 k 个欧氏距离最小的参考点作为最近邻的 k 个参考点,之后再用这 k 个参考点的欧氏距离通过权值计算公式计算出各自的权值,计算出待定位点的物理位置坐标。

假设有 n 个 AP 信号源,则待定位点处接收到 n 个 AP 信号强度值的向量组表示为

$$\mathbf{RSSI} = [\text{RSSI}_1 \quad \text{RSSI}_2 \quad \cdots \quad \text{RSSI}_n]_{1 \times n} \quad (7)$$

位置指纹数据库中第 j 个参考点处的 n 个 AP 信号强度值的向量组表示为

$$\mathbf{rss}_j = [\text{rss}_{j1} \quad \text{rss}_{j2} \quad \cdots \quad \text{rss}_{jn}]_{1 \times n} \quad (8)$$

那么,待定位点与位置指纹数据库中第 j 个参考点处的欧氏距离计算式为

$$D_j = [(\text{RSSI}_1 - \text{rss}_{j1})^2 + (\text{RSSI}_2 - \text{rss}_{j2})^2 + \cdots + (\text{RSSI}_n - \text{rss}_{jn})^2]^{\frac{1}{2}} \quad (9)$$

使用待定位点与参考点之间信号强度值的差异来计算待定位点与参考点之间的欧氏距离虽然在一定程度上可以反映待定位点与参考点之间的相似度,然而,传统欧氏距离的计算方式并没有考虑到 AP 信号的波动对计算结果带来的影响。由于在实际的室内环境中,信号的波动会导致测量的信号强度值并不十分准确,信号越稳定,采集到的信号强度值越准确;信号越不稳定,采集到的信号强度值准确性就越低。这就使传统欧氏距离的计算方法并不能很好地反映待定位点与参考点之间的相似性。

2.2 基于离散系数改进的加权欧氏距离

离散系数可以反映单位均值上的离散程度,主要用于比较均值不同的样本之间的离散程度。使用各 AP 的离散系数来对传统欧氏距离进行加权可以把不同 AP 信号的波动性考虑进来,使计算出来的加权欧氏距离能更好地反映待定位点与参考点之间的相似程度。

假设有 n 个 AP 信号源,在待定位点处对每个 AP 进行 m 次测量,得到一个 m 行 n 列的矩阵:

$$\mathbf{RSS}_{m \times n} = \begin{bmatrix} \text{rss}_{11}^1 & \text{rss}_{11}^2 & \cdots & \text{rss}_{11}^n \\ \text{rss}_{12}^1 & \text{rss}_{12}^2 & \cdots & \text{rss}_{12}^n \\ \vdots & \vdots & \ddots & \vdots \\ \text{rss}_{m1}^1 & \text{rss}_{m1}^2 & \cdots & \text{rss}_{m1}^n \end{bmatrix} \quad (10)$$

第 i 个 AP 的离散系数 V_i 的计算式为

$$V_i = \frac{\sigma_i}{\text{avr}_i} \quad i \in (1, n) \quad (11)$$

式中: σ_i 为第 i 个 AP 信号的 m 次测量值的标准差; avr_i 为第 i 个 AP 信号的 m 次测量值的平均值。

标准差 σ_i 的计算式为

$$\sigma_i = \left\{ \frac{1}{m} [(\text{rss}_{i1}^1 - \text{avr}_i)^2 + (\text{rss}_{i2}^1 - \text{avr}_i)^2 + \cdots + (\text{rss}_{i1}^m - \text{avr}_i)^2] \right\}^{\frac{1}{2}} \quad (12)$$

平均值 avr_i 计算式为

$$\text{avr}_i = \frac{\text{rss}_{i1}^1 + \text{rss}_{i2}^1 + \cdots + \text{rss}_{i1}^m}{m} \quad (13)$$

计算出每个 AP 的离散系数后,再利用离散系数计算出每个 AP 信号的权重因子。第 i 个 AP 信号的权值计算式为

$$W_i = \frac{1/V_i}{\sum_{i=1}^n 1/V_i} \quad (14)$$

经过离散系数加权后的欧氏距离计算式为

$$\tilde{D}_j = [W_1(\text{rss}_{j1}^1 - \text{avr}_1)^2 + W_2(\text{rss}_{j2}^1 - \text{avr}_2)^2 + \cdots + W_n(\text{rss}_{jn}^1 - \text{avr}_n)^2]^{\frac{1}{2}} \quad (15)$$

式中: W_i 为第 i 个 AP 信号的权值; avr_i 为在待定位点采集到的第 i 个 AP 的 m 次信号强度值的平均值; rss_{jn}^1 为第 j 个参考点处测得的第 n 个 AP 的信号强度值。

由于各个 AP 测量 m 次信号强度的均值是不同的,使用各 AP 的离散系数可以更好地反映出每个 AP 样本数据的离散程度。离散系数大的 AP,信号强度值波动比较大,稳定性较弱;离散系数小的 AP,信号强度值波动比较小,稳定性较强。使用基于离散系数改进后的加权欧氏距离表示待定位点与参考点之间的相似度时,在权值的计算过程中,使用离散系数的倒数进行计算,可以使离散系数大的 AP 的权重因子更小,这样就可以使 AP 信号的不稳定性对最终的定位结果造成的误差有效降低。

2.3 基于离散系数改进加权欧氏距离后的 VWKNN 算法

VWKNN 算法在计算欧氏距离时,用离散系数进行加权,使通过匹配算法找到的与待定位点最近邻的 k 个参考点更精确。首先,在待定位点处对每个 AP 信号源采集 m 次信号强度值,通过离散系数加权欧氏距离公式计算待定位点与每个

参考点之间的加权欧氏距离;然后,把计算出的加权欧氏距离从小到大进行排序,选出加权欧氏距离最小的 k 个参考点,分别计算出这 k 个参考点中每个参考点的距离权重因子;最后,计算待定位点的物理位置坐标。每个参考点的距离权重因子计算式为

$$\omega_i = \frac{1}{D_i + 1} \quad k > 1 \quad (16)$$

$$\sum_{i=1}^k \frac{1}{D_i + 1}$$

式中: ω_i 为第 i 个参考点的权重因子; D_i 为最终选出的 k 个最近邻的参考点中,每个参考点与待定位点之间的基于离散系数加权后的欧氏距离。

待定位点的物理位置坐标为

$$(x, y) = \sum_{i=1}^k \omega_i (x_i, y_i) \quad k > 1 \quad (17)$$

VWKNN 算法的流程如图 2 所示。

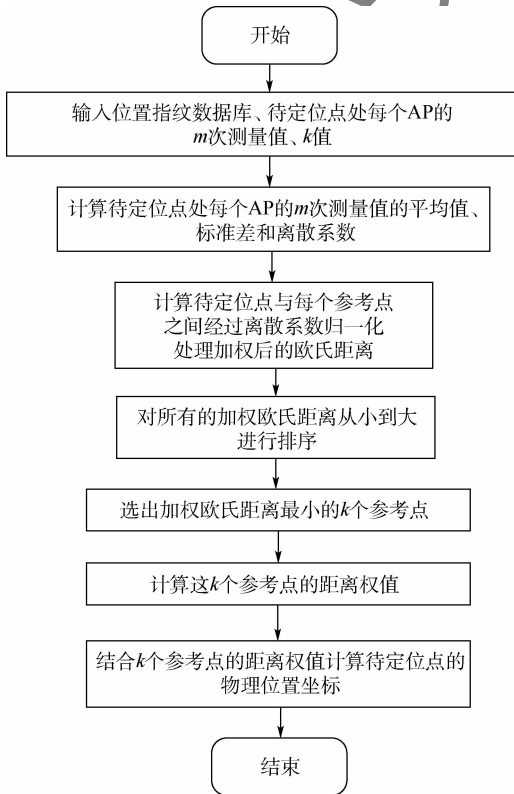


图 2 VWKNN 算法流程

Fig. 2 Flowchart of VWKNN algorithm

3 实验与分析

采用 MATLAB 软件对基于离散系数改进后的 VWKNN 算法进行仿真,并将其与 KNN 算法和 WKNN 算法进行比较,验证了 VWKNN 算法能够提高室内定位的定位精度。

3.1 实验环境的搭建

实验选择某大型购物商场的一层进行信号强

度的采集。在商场一层选择的区域用 Wi-Fi 分析仪选出 6 个比较稳定的 AP 信号源,这 6 个 AP 在实验区域内的大部分地方都是可以采集到的。选择商场一层一块长 20 m、宽 15 m 的矩形区域,该区域是商场一层一片空旷的区域,没有商店及墙体等大型遮挡物,只有小型的装饰物摆放。以 1 m 为间隔把该矩形区域划分为若干正方形网格,将网格的交叉点设置为位置指纹数据库中的参考点,共 266 个参考点。信号强度值的采集设备为 HONOR V20,型号为 PCT-AL10,系统为 Android 10。使用 Wi-Fi 分析仪 APP 采集信号强度值,每隔 1 s 采集一次信号强度值,在每个参考点处对每个 AP 采集 10 次信号强度值,用高斯滤波法对其进行数据预处理,除去噪声数据,之后将每个参考点处所有采集的非噪声信号强度值取平均值,录入到位置指纹数据库中。建立好位置指纹数据库后,在这片区域中设置 15 个待定位点,记录每个待定位点的物理位置信息,在每个待定位点处对每个 AP 采集 20 次信号强度值,对采集到的 10 次信号强度值进行数据预处理,经过数据降噪后,用于改进后 WKNN 算法定位精度的实验。

实验区域及区域内参考点与待定位点分布如图 3 所示。图中,共有 266 个参考点和 15 个待定位点。

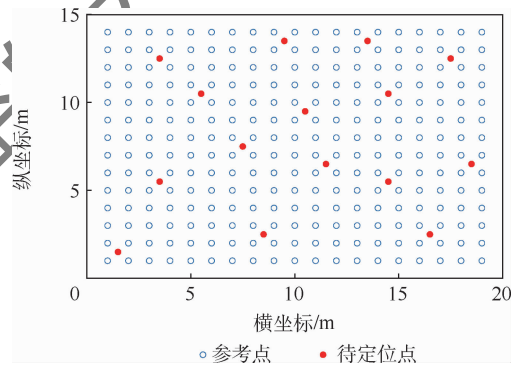


图 3 实验区域参考点及待定位点分布

Fig. 3 Reference points and distribution map of pending sites in experimental area

3.2 不同定位算法 k 值的选择

在 KNN 算法、VWKNN 算法及 WKNN 算法中, k 值的选择是十分重要的。选取不同的 k 值会直接造成定位结果的差异,为了找出各定位算法中定位误差最小的 k 值,用 15 个待定位点对 KNN 算法、WKNN 算法及 VWKNN 算法的 k 值进行实验。

首先,在建立好的离线位置指纹数据库中,使用 KNN 算法, k 值分别取 1~6 进行实验,对定位误差进行分析。每个待定位点的定位误差的计算式为

$$E_i = \sqrt{(x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2} \quad (18)$$

式中: E_i 为第 i 个待定位点的定位误差; x_{i1} 和 y_{i1} 分别为第 i 个待定位点实际位置的横坐标和纵坐标; x_{i2} 和 y_{i2} 分别为第 i 个待定位点测量得到的位置的横坐标和纵坐标。

图 4 为 KNN 算法在不同 k 值下的定位误差。因为 k 取 1 的时就是最近邻算法,所以, k 值取 1 是为了计算最近邻算法的定位误差。从图 4 中可以看出,最近邻算法的定位误差大部分高于其他算法的定位误差。当 $k=3$ 时,误差曲线基本上都低于 k 取其他值时的曲线,这就说明,KNN 算法在实验中,当 $k=3$ 时,定位误差最小。 k 取不同值时,KNN 算法的平均定位误差如表 1 所示。

从表 1 可以看出,在 KNN 算法中, $k=3$ 时平均定位误差最小。因此,在后续的实验过程中,对于 KNN 算法 k 值将一直取 3 来进行实验。

然后,对于 WKNN 算法, k 值分别取 2~6 进行实验,并对这几个 k 值的定位误差进行分析。图 5 为 WKNN 算法在不同 k 值下的定位误差。

从图 5 中可以看出,当 $k=4$ 时,WKNN 算法的定位误差曲线大部分都在 k 取其他值的定位误差之下。因此,在 WKNN 算法中, $k=4$ 时,定位效

果较好。 k 取不同值时,WKNN 算法的平均定位误差如表 2 所示。

从表 2 可以看出,实验中,WKNN 算法在 $k=4$ 时平均定位误差最小。因此,在以后的实验过程中,将会对 WKNN 算法以 k 值取 4 来进行后续的实验。

最后,对 VWKNN 算法 k 值分别取 2~6 进行实验,并对这几个 k 值的定位误差进行分析。图 6 为 VWKNN 算法在不同 k 值下的定位误差。

从图 6 中可以看出,当 $k=4$ 时,VWKNN 算法的定位误差曲线大部分都在 k 取其他值的定位

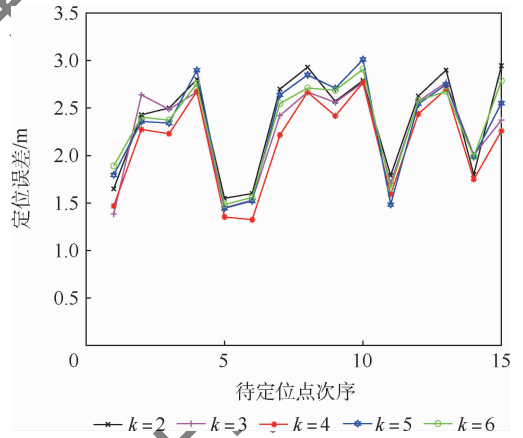


图 5 WKNN 算法在不同 k 值下的定位误差

Fig. 5 Positioning errors of WKNN algorithm under different k values

表 2 WKNN 算法的平均定位误差
Table 2. Average positioning error of WKNN algorithm

k	平均定位误差/m
2	2.37
3	2.26
4	2.14
5	2.32
6	2.33

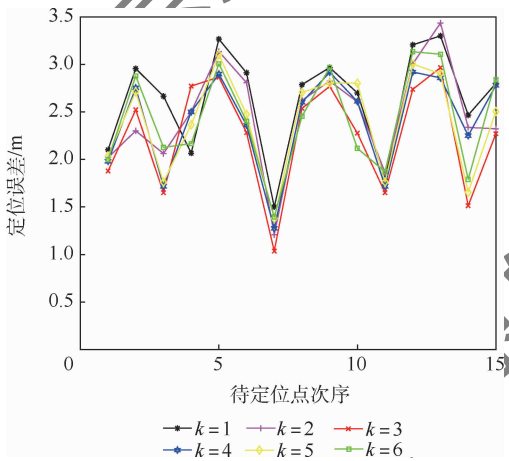


图 4 KNN 算法在不同 k 值下的定位误差

Fig. 4 Positioning errors of KNN algorithm under different k values

表 1 KNN 算法的平均定位误差

Table 1 Average positioning error of KNN algorithm

k	平均定位误差/m
1	2.63
2	2.47
3	2.25
4	2.41
5	2.40
6	2.57

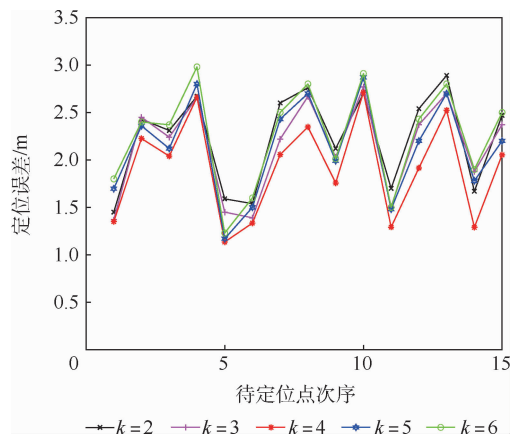


图 6 VWKNN 算法在不同 k 值下的定位误差

Fig. 6 Positioning errors of VWKNN algorithm under different k values

误差之下。因此,在 VWKNN 算法中, $k=4$ 时,定位效果较好。 k 取不同值时,VWKNN 算法的平均定位误差如表 3 所示。

从表 3 可以看出,实验中,VWKNN 算法在 $k=4$ 时平均定位误差最小。因此,在以后的实验过程中,对 VWKNN 算法的 k 值将取 4 来进行后续的实验。

表 3 VWKNN 算法的平均定位误差

Table 3 Average positioning error of VWKNN algorithm

k	平均定位误差/m
2	2.23
3	2.13
4	1.91
5	2.12
6	2.24

3.3 基于离散系数改进的 WKNN 算法

实验对基于离散系数改进的 VWKNN 算法的定位误差和 KNN 算法及 WKNN 算法的定位误差进行了对比,并对定位误差进行分析。其中,由于在 3.2 节实验中传统 WKNN 算法的最优 k 值为 4 且基于离散系数改进的 WKNN 算法只是对欧氏距离的权值计算方法进行改进,改进后的 WKNN 算法的 k 值也取 4 来进行实验。

实验依旧采用 3.2 节的 15 个待定位点进行定位坐标及定位误差的计算。在 15 个待定位点已经采集好的 AP 信号强度值的基础上,在这 15 个待定位点处手机分别朝东、南、西、北 4 个方向分别采集一次信号强度值,进行改进后的 WKNN 算法的实验。表 4 为 KNN 算法在 $k=3$ 、WKNN 算法和 VWKNN 算法在 $k=4$ 时的平均定位误差。

从表 4 可以看出,KNN 算法的定位误差是最高的,WKNN 算法的定位误差比 KNN 算法的定位误差小,而改进后的 VWKNN 算法的定位误差比传统的 WKNN 算法的定位误差小。图 7 对每个待定位点处使用 KNN 算法、WKNN 算法和 VWKNN 算法的定位误差进行了比较。

累积分布函数能够描述变量的概率分布情况,对于离散变量而言,该函数可以计算出所有小于或等于一个值出现的概率之和。从定位误差的

表 4 三种定位算法的平均定位误差

Table 4 Average positioning error of three positioning algorithms

定位算法	平均定位误差/m
KNN 算法	2.25
WKNN 算法	2.14
VWKNN 算法	1.91

累积分布函数曲线可以看出定位误差的概率分布情况。图 8 为 KNN 算法、WKNN 算法及 VWKNN 算法的定位误差累积分布概率对比。

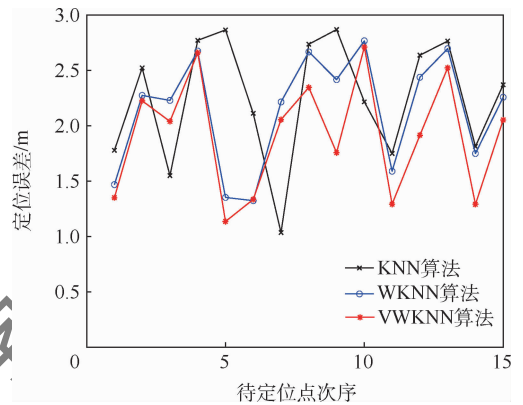


图 7 三种定位算法的定位误差比较

Fig. 7 Comparison of positioning errors of three positioning algorithms

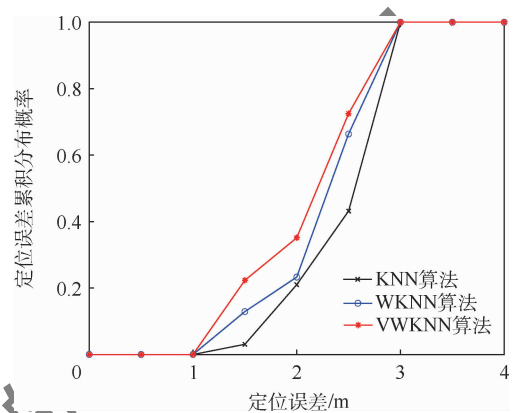


图 8 定位误差累积分布概率对比

Fig. 8 Comparison of cumulative distribution probability of positioning error

从图 7 可以更直观地看出,KNN 算法在 80% 情况下的定位误差均大于 WKNN 算法和 VWKNN 算法,并且定位误差非常不稳定。而 VWKNN 算法在每个待定位点处的定位误差在多数情况下都低于 WKNN 算法的定位误差,在少数情况下与 WKNN 算法的定位误差相近。从图 8 可以看出,KNN 算法、WKNN 算法和 VWKNN 算法的定位误差都控制在 3 m 以内,KNN 算法定位误差在 2.5 m 以内的累积分布概率为 43.1%,WKNN 算法定位误差在 2.5 m 以内的累积分布概率为 66.4%,而 VWKNN 算法定位误差在 2.5 m 以内的累积分布概率为 72.5%。可以表明,VWKNN 算法在 2.5 m 以内的定位误差累积分布概率比 KNN 算法提高了 29.4%,比 WKNN 算法提高了 6.1%。

为验证改进后的 VWKNN 算法在间隔不同的参考点部署方式下均有提高,在相同场景和相同区域下,以 2 m 和 3 m 的间隔分别对参考点进行

部署,建立离线位置指纹数据库,采取 3.2 节中设置好的 15 个待定位点,对 2 m、3 m 间隔的参考点部署形式进行定位算法精确度的实验,并将其与 1 m 间隔的参考点部署方式的实验结果进行比较。

图 9 为在 1 m、2 m、3 m 间隔的参考点部署方式下,KNN 算法在 $k=3$ 、WKNN 算法和 VWKNN 算法在 $k=4$ 时的平均定位误差比较。

表 5 为参考点分别在 1 m、2 m、3 m 间隔的部署方式下,KNN 算法在 $k=3$ 、WKNN 算法和 VWKNN 算法在 $k=4$ 时的平均定位误差。

从图 9 和表 5 中可以看出,在 1 m、2 m、3 m 间隔的参考点部署方式下,KNN 算法的定位误差都是最高的,WKNN 算法的定位误差都比 KNN 算法的定位误差小,而 VWKNN 算法的定位误差是最小的。

综上所述,基于离散系数改进后的 VWKNN 算法确实可以在一定程度上提高室内定位的精度。

强度的稳定性,并加入到各 AP 信号的权值计算当中,可以有效减少 AP 信号的波动性对最终定位结果产生的影响。

3) 实验表明,VWKNN 算法的平均定位精度比 KNN 算法提高了 15% ~ 17%,比 WKNN 算法提高了 11% ~ 13%,可以在一定程度上提高室内定位的精度,具有一定的可行性和实用价值。

参考文献 (References)

[1] YANG Q, CHEN Y Q, YIN J, et al. LEAPS: A location estimation and action prediction system in a wireless LAN environment[C] // Proceedings of network and Parallel Computing, 2004, 3222: 584-591.

[2] SKOV-PETERSEN H, BARKOW B, LUNDHEDE T, et al. How do cyclists make their way? - A GPS-based revealed preference study in Copenhagen[J]. International Journal of Geographical Information Science, 2018, 32(7): 1469-1484.

[3] MAINETTI L, PATRONO L, SERGI I. A survey on indoor positioning systems[C] // Proceedings of 2014 22nd International Conference on Software. Piscataway: IEEE Press, 2015: 111-120.

[4] 刘璐,李静.一种在非视距环境中的 TDOA/AOA 混合定位方法[J]. 通信学报, 2005, 26(5): 63-68.

LIU J, LI J. A hybrid TDOA/AOA localization method in a non-line-of-sight environment[J]. Journal of Communications, 2005, 26(5): 63-68(in Chinese).

[5] 赵强,王敬东,刘云霄,等.基于粒子群 TOA 室内定位的系统误差修正[J]. 电子测量技术, 2017, 40(8): 189-192.

ZHAO Q, WANG J D, LIU Y X, et al. Systematic error correction based on particle swarm TOA indoor positioning[J]. Electronic Measurement Technology, 2017, 40(8): 189-192(in Chinese).

[6] LUO M Z, CHEN X, CAO S, et al. Two new shrinking-circle methods for source localization based on TDoA measurements[J]. Sensors, 2018, 18(4): 1274-1279.

[7] 田增山,张千坤,周牧,等.基于 CFR 虚拟阵列天线的 AOA 室内定位[J]. 电子学报, 2018, 46(6): 1468-1474.

TIAN Z S, ZHANG Q K, ZHOU M, et al. AOA indoor positioning based on CFR virtual array antenna[C] // Acta Electronica Sinica, 2018, 46(6): 1468-1474(in Chinese).

[8] REMPEL P, BORISOV A, SIEMENS E, et al. Local system of positioning using a WiFi network[C] // MATEC Web of Conferences, 2018, 155: 01014.

[9] PRASITHSANGAREE P, KRISHNAMURTHY P, CHRYSANTHIS P K. On indoor position location with wireless LANs[C] // Proceedings of IEEE International Symposium on Personal. Piscataway: IEEE Press, 2002: 720-724.

[10] 王文森. 变异系数—一个衡量离散程度简单而有用的统计指标[J]. 中国统计, 2007(6): 41-42.

WANG W S. Coefficient of variation—A simple and useful statistical measure of dispersion[J]. China Statistics, 2007(6): 41-42(in Chinese).

[11] WANG Q, FENG Y, ZHANG X, et al. IWKNN: An effective

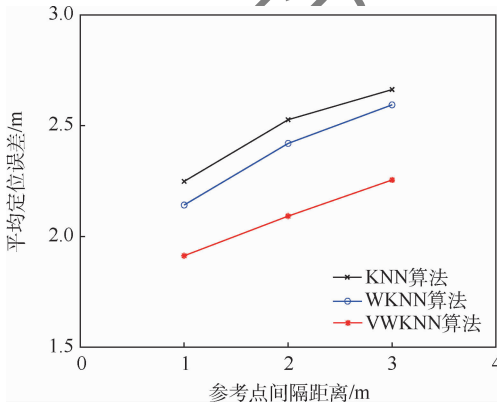


图 9 不同部署方式下 3 种定位算法的平均定位误差对比

Fig.9 Comparison of average positioning error of three positioning algorithms under different deployment modes

表 5 不同部署方式下 3 种定位算法的平均定位误差

Table 5 Average positioning error of three positioning algorithms under different deployment modes

定位算法	平均定位误差/m		
	1 m 间隔	2 m 间隔	3 m 间隔
KNN 算法	2.25	2.53	2.66
WKNN 算法	2.14	2.41	2.59
VWKNN 算法	1.91	2.09	2.25

4 结 论

1) 基于离散系数改进的 VWKNN 算法是针对传统 WKNN 算法没有考虑到 AP 信号强度不稳定的问题提出的。

2) 使用离散系数归一化处理表示 AP 信号

- bluetooth positioning method based on isomap and WKNN[J]. *Mobile Information Systems*, 2016, 2016:876585.
- [12] 薛伟,陈璟,张熠. 基于深度学习的 Wi-Fi 与 iBeacon 融合的室内定位方法[J]. *计算机工程与应用*, 2019, 55(1):29-34. XUE W, CHEN J, ZHANG Y. Indoor positioning method based on deep learning fusion of Wi-Fi and iBeacon [J]. *Computer Engineering and Applications*, 2019, 55(1):29-34 (in Chinese).
- [13] 田家英,张志华. 基于近邻法的 WIFI 室内定位改进算法研究[J]. *测绘工程*, 2018, 27(12):34-39. TIAN J Y, ZHANG Z H. Research on improved WIFI indoor localization algorithm based on nearest neighbor method[J]. *Engineering of Surveying and Mapping*, 2018, 27(12):34-39 (in Chinese).
- [14] 吴虹,王国萍,彭鸿钊,等. 一种基于 KNN 的室内位置指纹定位算法[J]. *南开大学学报(自然科学版)*, 2020, 53(6):5-9. WU H, WANG G P, PENG H Z, et al. An indoor location fingerprint algorithm based on KNN[J]. *Acta Scientiarum Naturalium Universitatis Nankaiensis*, 2020, 53(6):5-9 (in Chinese).
- [15] 田洪亮,钱志鸿,梁潇,等. 离散度 WKNN 位置指纹 Wi-Fi 定位算法[J]. *哈尔滨工业大学学报*, 2017, 49(5):94-99. TIAN H L, QIAN Z H, LIANG X, et al. Dispersion WKNN position fingerprint Wi-Fi location algorithm[J]. *Journal of Harbin Institute of Technology*, 2017, 49(5):94-99 (in Chinese).
- [16] BAI H, YAN W. Application of NN-KNN algorithm combined with cloud-based indoor positioning[J]. *Metallurgical and Mining Industry*, 2015, 7(5):383-388.
- [17] 卢选民,院文乐,邱杨,等. 一种改进的基于 KNN 的动态预测指纹定位算法[J]. *计算机应用研究*, 2017, 34(7):63-66. LU X M, YUAN W L, QIU Y, et al. An improved dynamic predictive fingerprint location algorithm based on KNN[J]. *Application Research of Computers*, 2017, 34(7):63-66 (in Chinese).
- [18] 王亚涛,王新珩,董育宁,等. 基于 Kmeans 和动态 WKNN 的两层 Wi-Fi 改进定位方法[J]. *南京邮电大学学报(自然科学版)*, 2017, 37(5):41-47. WANG Y T, WANG X H, DONG Y N, et al. Improved localization method for two-layer Wi-Fi based on Kmeans and dynamic WKNN[J]. *Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition)*, 2017, 37(5):41-47 (in Chinese).
- [19] YEN L, YAN C H, RENU S, et al. A modified WKNN indoor Wi-Fi localization method with differential coordinates [C] // *Proceedings of 2017 International Conference on Applied System Innovation (ICASI)*. Piscataway: IEEE Press, 2017:1822-1824.
- [20] WANG B, GAN X, LIU X, et al. A novel weighted KNN algorithm based on RSS similarity and position distance for Wi-Fi fingerprint positioning [J]. *IEEE Access*, 2020, 8:30591-30602.
- [21] ALIPPI C, VANINI G. A RSSI-based and calibrated centralized localization technique for wireless sensor networks [C] // *Proceedings of IEEE International Conference on Pervasive Computing & Communications Workshops*. Piscataway: IEEE Press, 2006:301-305.

VWKNN location fingerprint positioning algorithm based on improved discrete coefficient

XU Tian, HE Jingsha, ZHU Nafei*, DENG Wanhang, WU Shuang, TA Yongjun

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: The location fingerprint algorithm is the main method to study the indoor positioning technology, and the online matching algorithm is one of the main factors affecting the indoor positioning accuracy. At present, the matching algorithms in online stage include the nearest neighbor algorithm, K-nearest neighbor algorithm and weighted K-nearest neighbor algorithm. However, these three algorithms do not take into account the influence of the fluctuation of AP signal on the positioning result. In order to improve the matching algorithm in online stage, a weighted K-nearest neighbor algorithm based on the improved discrete coefficient is proposed. In offline stage the purpose is to establish a fingerprint database, in the online stage using discrete coefficient to reflect the stability of the various AP signal and treat the anchor point with weighted Euclidean distance between the reference point, calculate all the weighted Euclidean distance, choose the nearest k reference points, so as to estimate the physical location of pending sites. Finally, experiments show that the weighted K-nearest neighbor algorithm based on the improved discrete coefficient can achieve an average positioning accuracy which is 15% - 17% higher than the K-nearest neighbor algorithm and 11% - 13% higher than the weighted K-nearest neighbor algorithm.

Keywords: indoor positioning; location fingerprint; AP signal strength; weighted K-nearest neighbor; discrete coefficient

Received: 2021-01-14; **Accepted:** 2021-05-07; **Published online:** 2021-06-18 09:06

URL: kns.cnki.net/kcms/detail/11.2625.V.20210617.1652.001.html

Foundation items: National Key R & D Program of China (2019QY(Y)0601); Shandong Provincial Natural Science Foundation (ZR2020MF029)

* **Corresponding author.** E-mail: znf@bjut.edu.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0007

大行程解耦三平动微定位平台的设计与优化



林苗^{1,2}, 孟刚^{1,2}, 居勇健^{1,2}, 徐伟胜^{1,2}, 曹毅^{1,2,*}

(1. 江南大学 机械工程学院, 无锡 214122; 2. 江苏省食品先进制造装备技术重点实验室, 无锡 214122)

摘 要: 为设计具有大行程与良好解耦特性的三平动微定位平台, 提出了一种新型 2T3R 型运动副。基于 2T3R 型运动副, 设计了三平动微定位平台的结构; 采用非线性模型法建立了平台力-位移关系与丢失运动的理论模型, 并采用拉格朗日方程建立了平台固有频率的理论模型; 采用目标规划法对三平动微定位平台进行了参数优化; 通过有限元仿真验证了理论模型的正确性。理论计算与仿真研究结果表明: 平台一阶固有频率为 51.27 Hz, 在 1 mm 运动行程内, x, z 轴方向的丢失运动分别小于 0.67%、0.20%, 输入与输出完全解耦。研究结果证明了运动副、平台结构设计的有效性及其优化模型的可行性。

关键词: 柔顺机构; 结构设计; 微定位平台; 参数优化; 柔性运动副

中图分类号: V414.5; TH122

文献标志码: A **文章编号:** 1001-5965(2022)07-1252-11

柔顺机构是通过材料的弹性变形来实现运动和力传递及转换的一种机构^[1]。柔顺机构具有无间隙、免润滑、无摩擦、无磨损、运动精度高等优点, 适用于精密机械^[2]、生物医学^[3]、细胞操作^[4]等领域。

三平动微定位平台是一类典型的柔顺机构, 其多用于微装配或微操作系统^[5-7]中, 可实现系统在空间中对位置的精密定位。目前, 国内外针对三平动微定位平台的研究已取得了一定的成果。Awtar 等^[8]采用柔性薄板作为移动副, 提出了一种结构紧凑、大行程且具有高定位精度的三支链三平动微定位平台。Li 和 Hao 等^[9-10]分别基于自由度与约束空间拓扑法^[11-12]提出了 2T1R 型和 2T2R 型柔性运动副, 并设计出 3-PPPR 型与 3-PPRR 型大行程三平动微定位平台。王保兴等^[5]采用运动副反向串联与缩短支链长度的方法设计出 3-PPP 型柔性微定位平台, 抵消了动平台的输出耦合。Xu^[13]采用多种柔性运动副集成

的复合结构, 设计了一种结构紧凑的 3-RPC 型微定位平台。为了解决三支链平台非对称布局所导致的伴生转动与耦合误差等问题, 李海洋^[14]和 Zhang^[15]等分别提出了 6-PPPR 型与 6-PPP 型三平动微定位平台, 采用六支链对称布局消除了动平台的偏转与耦合位移, 但同时平台结构复杂, 不便于加工与装配。此外, Hao^[16]与曹毅^[17]等先后设计了五支链布局的三平动微定位平台, 兼顾了三支链布局结构简单与六支链布局消除误差的优点。综上所述, 国内外针对三平动微定位平台的研究均取得了大量成果, 不难发现: ①柔性运动副是平台结构的重要组成部分, 是实现平台大行程与解耦性能的关键; ②五支链布局平台兼具三支链布局平台与六支链布局平台的优点, 综合性能最佳。

基于此, 本文提出了一种新型 2T3R 型运动副, 并基于该运动副设计了一种大行程解耦五支链三平动微定位平台。分别建立了微定位平台

收稿日期: 2021-01-07; 录用日期: 2021-02-10; 网络出版时间: 2021-03-11 20:26

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210311.1605.002.html

基金项目: 江苏省“六大人才高峰”资助项目 (ZBZZ-012); 江苏省研究生科研与实践创新计划 (JSCX20_0760); 江南大学研究生科研与实践创新计划 (JNSJ19_005)

*通信作者。E-mail: caoyi@jiangnan.edu.cn

引用格式: 林苗, 孟刚, 居勇健, 等. 大行程解耦三平动微定位平台的设计与优化[J]. 北京航空航天大学学报, 2022, 48(7): 1252-1262. LIN M, MENG G, JU Y J, et al. Design and optimization of large-stroke decoupled three-translational micro-positioning platform [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1252-1262 (in Chinese).

静、动态特性的理论模型,以微定位平台的丢失运动与固有频率为目标对其进行了多目标参数优化。通过有限元仿真验证了理论模型的正确性,进一步证明了运动副、平台结构设计的有效性与优化模型的可行性。

1 结构设计

1.1 2T3R 型运动副结构设计

微定位平台多采用驱动器可固定安装的并联式结构^[14]。三平动微定位平台的每条支链都必须包含 3 个移动自由度,因此含有移动自由度的柔性运动副是微定位平台结构设计的关键。

图 1 为具有 2T3R 自由度的柔性运动副^[18]。如图 2 所示,该运动副由 2 组具有 2T2R 自由度的两杆机构反向串联而成,且 2 组两杆机构空间交错布置。基于自由度与约束空间拓扑法^[14-12],2T3R 型运动副的自由度可表示为

$$\{x, y, \theta_x, \theta_z\} \cup \{x, y, \theta_y, \theta_z\} = \{x, y, \theta_x, \theta_y, \theta_z\} \quad (1)$$

式中: $\theta_x, \theta_y, \theta_z$ 分别为绕 x, y, z 轴的转动自由度。图 3 表示 2T3R 型运动副沿 x 轴移动的运动原理。当该运动副输出沿 x 轴的位移 e_x 时,反向

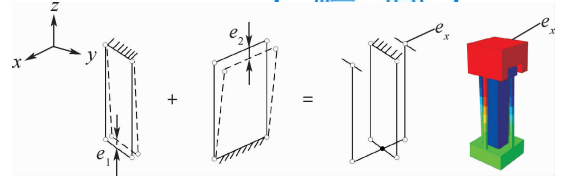


图 3 2T3R 型运动副运动原理

Fig. 3 Motion principle of 2T3R type compliant pair

两杆机构与正向两杆机构会产生沿 z 轴的寄生运动 e_1, e_2 ,二者方向相反,恰好相互抵消。同理,该运动副沿 y 轴移动时也会抵消沿 z 轴的寄生运动。基于上述可知,反向串联的连接方式使得 2T3R 型运动副可抵消运动副沿移动自由度方向运动时产生的寄生运动,从而提升平台的解耦性能。

分布柔度的柔性单元、反向串联的连接方式使 2T3R 型运动副具有较大的运动行程,且该运动副结构紧凑,尺寸参数少,可实现模块化装配。此外,该运动副变形时无应力刚化效应^[19],有利于简化平台理论建模的难度。

值得指出的是,与现有的 2T 型^[5]运动副相比,2T3R 型运动副具有无需多轴联动便可抵消寄生运动的优点。与现有的 2T1R 型^[17]运动副相比,2T3R 型运动副的结构更为简单轻巧,更易于加工与装配。此外,高自由度的运动副也增加了平台结构设计的多样性。

1.2 柔性支链设计

为实现微定位平台的 3 个移动自由度,有必要设计出含有 3 个移动自由度的柔性支链。

如图 4 所示,基于旋量理论^[20]的设计方法,分别采用柔性薄板移动副为驱动副,2T3R 型运动副为被动副,得到 PPPR 型与 PPPRR 型 2 种柔性支链,其自由度可验证为

$$\{x, y, \theta_z\} \cup \{z\} = \{x, y, z, \theta_z\} \quad (2)$$

$$\{x, z, \theta_x, \theta_y\} \cup \{y\} = \{x, y, z, \theta_x, \theta_y\} \quad (3)$$

1.3 三平动微定位平台结构设计

三平动微定位平台主要有三、五、六支链 3 种布局方式。其中,三支链平台结构简单,但动平台受力不对称,易出现伴生偏转。六支链平台结构完全对称,可消除伴生偏转,但平台整体结构复杂,不利于加工与安装。而五支链平台同时具备三支链与六支链平台的优点,综合性能最佳。

为消除动平台的伴生偏转并考虑平台的加工与安装因素,采用五支链的布局方式,将图 4 中 PPPR 型与 PPPRR 型 2 种柔性支链相结合,设计出 4-PPRR&PPPR 型三平动微定位平台,其结构

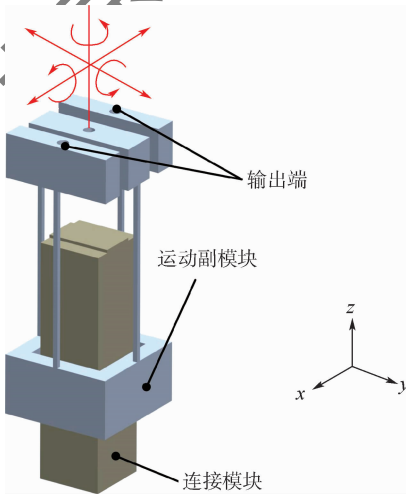


图 1 2T3R 型柔性运动副

Fig. 1 2T3R type compliant motion pair

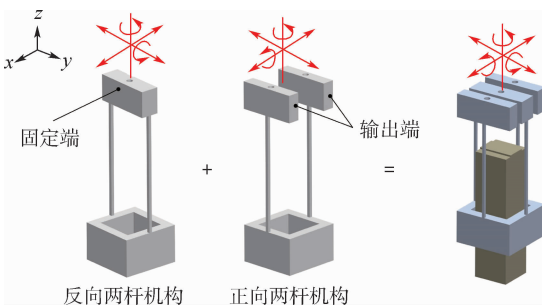


图 2 2T3R 型运动副自由度原理

Fig. 2 Freedom principle of 2T3R type motion pair

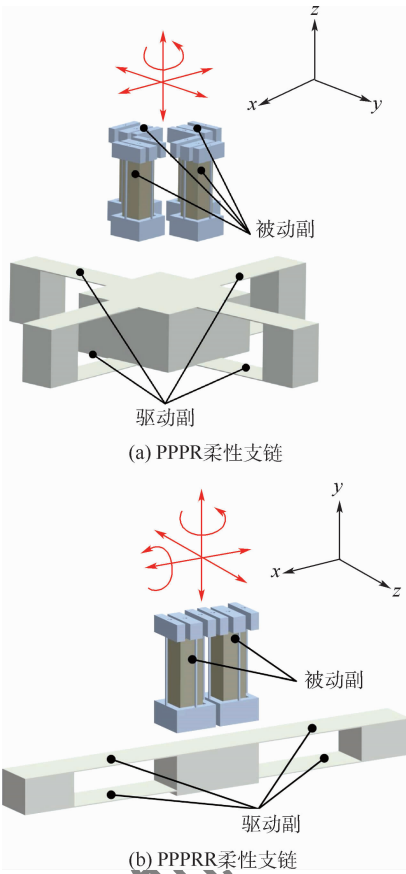


图4 含有3个移动自由度的柔性支链

Fig.4 Compliant limbs with 3 DOF of translation

如图5(a)所示。基于旋量理论^[20],该微定位平台自由度可表示为

$$4\{x, y, z, \theta_x, \theta_y\} \cap \{x, y, z, \theta_z\} = \{x, y, z\} \quad (4)$$

式中: \cap 表示取交集。

该微定位平台在 x, y 平面的支链布局如图5(b)所示,其中支链1~支链4对称布置,以消除动平台的伴生转动与输出耦合;如图5(a)~(c)所示,该平台支链1~支链4中柔性薄板、2T3R型运动副中细长杆、支链5中柔性薄板均具有相同的尺寸参数,以便于模块化加工。

类比现有的三平动微定位平台^[5-9,13-17],该平台具有以下优点:①平台中所有被动副均为同一模块的2T3R型运动副,变形时不产生寄生运动,且无应力刚化效应^[19],故平台输入输出完全解耦,理论建模难度低;②平台中柔性薄板与细长杆均为分布柔度式结构,且尺寸参数少,故该平台运动行程大,可实现模块化加工;③五支链布局消除了动平台的伴生转动,且易于加工与装配。

关于驱动方案,图5(a)所示平台具有较大的运动行程,因此拟采用可提供毫米级驱动位移的音圈电机驱动器以实现平台对位置的精密定位。

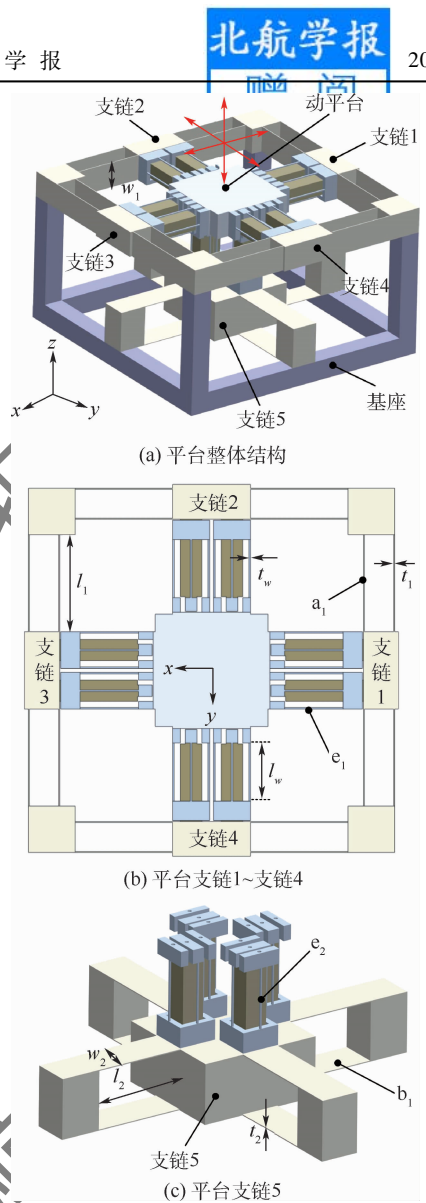


图5 4-PPRR&PPPR型微定位平台

Fig.5 4-PPRR&PPPR type micro-positioning platform

2 静、动态特性分析

2.1 静态特性分析

2.1.1 平台 x 轴力-位移关系分析

力-位移关系是平台的重要静态特性,建立平台力-位移关系的理论模型,不仅可以为后续的实验控制提供依据,还可以指导平台的优化设计。平台力-位移关系的理论分析方法主要有柔度矩阵法^[5]、伪刚体模型法^[15]、梁约束模型法^[21]及非线性模型法^[22],其中,非线性模型法在大变形、非线性条件下具有较高的理论精度。在大驱动位移作用下,图5(a)所示平台的驱动副会产生较大的非线性变形,因此选用非线性模型法对该平台的力-位移关系进行分析。

由平台结构的对称性可知,其沿 x, y 轴方向特性完全相同,因此以 x 轴方向为研究对象,建立平台在 x 轴方向的力-位移关系模型。

如图 6 所示,当沿 x 轴方向支链 1 驱动副单独施加驱动位移 δ_x 时,支链 1、支链 3 中的驱动副与支链 2、支链 4、支链 5 中的被动副产生弹性变形,通过建立驱动副中柔性薄板与被动副的力-位移关系模型,可得到平台整体沿 x 轴方向的力-位移关系。

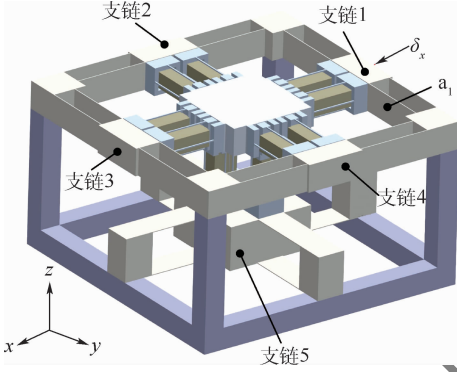


图 6 微定位平台沿 x 轴方向驱动示意图

Fig.6 Schematic diagram of micro-positioning platform driving along x axis direction

以支链 1 中柔性薄板 a_1 为例,当柔性薄板 a_1 沿 x 轴方向位移为 δ_x 时,柔性薄板 a_1 产生弯曲和拉伸变形且在 y 轴方向的投影长度保持不变,其力-位移关系模型如图 7 所示。根据柔性薄板 a_1 在 x 轴方向的受力平衡条件,其沿 x 轴方向受力与位移 δ_x 之间的关系为

$$F_{a_1x} = F_{a_1bend} + F_{a_1tens} \sin \alpha_1 \quad (5)$$

$$F_{a_1bend} = \frac{E^1 t_1^3 w_1}{l_1^3} \delta_x \quad (6)$$

$$F_{a_1tens} = \frac{l_{1tens} - l_1}{l_1} E^1 t_1 w_1 \quad (7)$$

$$l_{1tens} = \sqrt{l_1^2 + \delta_x^2} \quad (8)$$

$$\sin \alpha_1 = \frac{\delta_x}{l_{1tens}} \quad (9)$$

式中: F_{a_1bend} 、 F_{a_1tens} 分别为柔性薄板 a_1 所受弯曲、拉伸作用力; l_1 、 t_1 、 w_1 分别为柔性薄板 a_1 的长度、宽度、厚度; l_{1tens} 、 α_1 分别为柔性薄板 a_1 变形后的长度、变形前后的夹角。

必须指出的是,当柔性薄板 a_1 的宽度 t_1 与厚度 w_1 满足 $t_1 > 10w_1$ 时,弹性模量 E 应修正为平面模量 E^1 ,二者之间的关系为^[5]

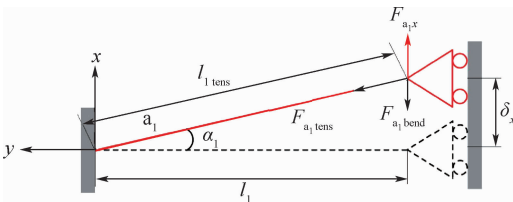


图 7 柔性薄板 a_1 的力-位移关系模型

Fig.7 Force-displacement relationship model of flexible sheet a_1

$$E^1 = \frac{E}{1 - \nu^2} \quad (10)$$

式中: ν 为材料的泊松比。

支链 1 ~ 支链 5 中被动副均为 2T3R 型运动副,该运动副为反向串联结构,变形时无应力刚化效应,其受力与变形之间为线性关系,故该运动副的刚度及力-位移关系为

$$k_{2T3R} = \frac{Et_w^4}{l_w^3} \quad (11)$$

$$F_{2T3R} = k_{2T3R} \delta_x \quad (12)$$

式中: l_w 、 t_w 分别为细长杆的长度、厚度。

根据式(5)、式(12)可得平台沿 x 轴方向施加驱动位移 δ_x 所需驱动力为

$$F_x = 8F_{a_1x} + 8F_{2T3Rx} \quad (13)$$

2.1.2 平台 z 轴力-位移关系分析

如图 8 所示,当沿 z 轴方向支链 5 驱动副单独施加驱动位移 δ_z 时,支链 1 ~ 支链 4 中的被动副与支链 5 中的驱动副产生弹性变形,同理可得平台整体在 z 轴方向的力-位移关系。

以支链 5 中柔性薄板 b_1 为例,当柔性薄板 b_1 沿 z 轴方向位移为 δ_z 时,其力-位移关系模型如图 9 所示。同理,柔性薄板 b_1 沿 z 轴方向受力与位移 δ_z 之间的关系为

$$F_{b_1z} = F_{b_1bend} + F_{b_1tens} \sin \alpha_2 \quad (14)$$

$$F_{b_1bend} = \frac{E^1 t_2^3 w_2}{l_2^3} \delta_z \quad (15)$$

$$F_{b_1tens} = \frac{l_{2tens} - l_2}{l_2} E^1 t_2 w_2 \quad (16)$$

$$l_{2tens} = \sqrt{l_2^2 + \delta_z^2} \quad (17)$$

$$\sin \alpha_2 = \frac{\delta_z}{l_{2tens}} \quad (18)$$

式中: F_{b_1bend} 、 F_{b_1tens} 分别为柔性薄板 b_1 所受弯曲、拉伸作用力; l_2 、 t_2 、 w_2 分别为柔性薄板 b_1 的长度、宽度、厚度; l_{2tens} 、 α_2 分别为柔性薄板 b_1 变形后的长度、变形前后的夹角。

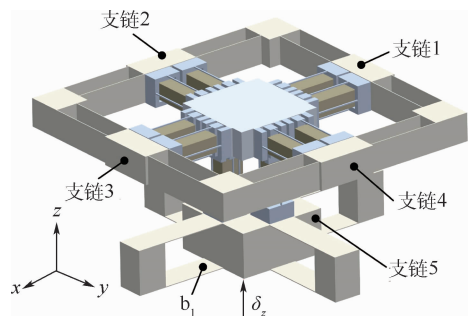
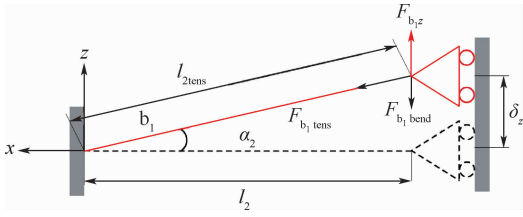


图 8 微定位平台沿 z 轴方向驱动示意图

Fig.8 Schematic diagram of micro-positioning platform driving along z axis direction

图9 柔性薄板 b_1 的力-位移关系模型Fig. 9 Force-displacement relationship model of flexible sheet b_1

同理,支链 1 ~ 支链 4 中 2T3R 型运动副的力-位移关系为

$$F_{2T3Rz} = k_{2T3R} \delta_z \quad (19)$$

因此,根据式(14)、式(19)可得平台沿 z 轴方向施加驱动位移 δ_z 所需驱动力为

$$F_z = 8F_{b1z} + 8F_{2T3Rz} \quad (20)$$

2.1.3 平台丢失运动分析

丢失运动是指平台沿驱动方向输入、输出位移之间的差值。平台中细长杆在其非自由度方向受力时会产生轴向变形,导致动平台沿驱动方向产生位移丢失,进而影响平台的定位精度,因此有必要对平台的丢失运动进行分析。

平台沿 x 轴方向支链 1 驱动副施加驱动位移 δ_x 时,支链 1 中细长杆受轴向力产生变形,基于胡克定律,平台沿 x 轴方向的丢失运动为

$$\delta_{xlost} = \frac{(F_x - 4F_{a1x})l_w}{Et_w} \quad (21)$$

平台沿 z 轴方向支链 5 驱动副施加驱动位移 δ_z 时,支链 5 中细长杆受轴向力产生轴向变形。同理,平台沿 z 轴方向的丢失运动为

$$\delta_{zlost} = \frac{(F_z - 8F_{b1z})l_w}{Et_w^2} \quad (22)$$

2.1.4 平台材料应力分析

为保证材料的变形在弹性范围内,当平台多轴联动时,有必要对平台进行材料应力分析。如图 10 所示,当沿 x, y, z 轴方向同时输入最大位移 δ_{max} 时,以支链 1 中细长杆 e_1 与支链 5 中细长杆 e_2 为例,其所承受的最大应力为

$$\sigma_{xmax} = \sigma_{xcomp} + \sigma_{xbend} \quad (23)$$

$$\sigma_{zmax} = \sigma_{zcomp} + \sigma_{zbend} \quad (24)$$

式中:

$$\sigma_{xcomp} = \frac{F_x - 4F_{a1x}}{4A} \quad (25)$$

$$\sigma_{zcomp} = \frac{F_z - 8F_{b1z}}{8A} \quad (26)$$

$$\sigma_{xbend} = \sigma_{zbend} = \frac{Et_w^4 \delta_{max}}{2Wl_w^2} \quad (27)$$

$$\delta_x = \delta_y = \delta_z = \delta_{max} \quad (28)$$

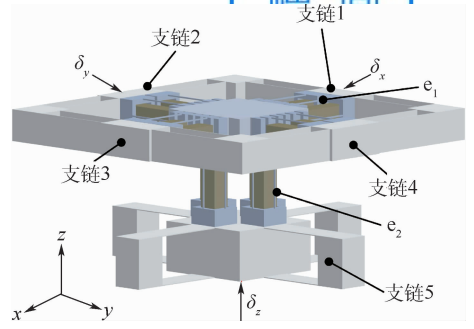


图10 微定位平台多轴联动示意图

Fig. 10 Schematic diagram of multi-axis motion for micro-positioning platform

其中: σ_{xcomp} 、 σ_{xbend} 分别为细长杆 e_1 所受压缩、弯曲应力; σ_{zcomp} 、 σ_{zbend} 分别为细长杆 e_2 所受压缩、弯曲应力; W 、 A 分别为细长杆 e_1 、 e_2 的抗弯截面系数、截面面积; δ_x 、 δ_y 、 δ_z 分别为微定位平台沿 x 、 y 、 z 轴方向输入位移。

2.1.5 平台屈曲分析

微定位平台中承受压力的细长杆有压杆失稳风险,进而影响平台的静态性能,因此需要对平台进行屈曲分析。

如图 10 所示,当微定位平台沿 x 轴方向支链 1 驱动副施加驱动位移 δ_x 时,支链 1 中被动副承受轴向压力。被动副为反向串联结构,其中正向两杆机构承受压力,故细长杆 e_1 有压杆失稳风险,其承受压力 F_{p1} 为

$$F_{p1} = \frac{1}{4}(F_x - 4F_{a1x}) \quad (29)$$

同理,当微定位平台沿 z 轴方向支链 5 驱动副施加驱动位移 δ_z 时,支链 5 中细长杆 e_2 有压杆失稳风险,其承受压力 F_{p2} 为

$$F_{p2} = \frac{1}{8}(F_z - 8F_{b1z}) \quad (30)$$

根据欧拉公式,细长杆 e_1 、 e_2 可承受的临界压力 F_{cr} 为

$$F_{cr} = \frac{\pi^2 Et_w}{\left(\frac{l_w}{2}\right)^2} \quad (31)$$

式中: I_w 为细长杆 e_1 、 e_2 的截面惯性矩。

2.2 动态特性分析

模态是微定位平台的固有振动特性,为了避免微定位平台在高频振动时出现共振现象,有必要对其进行模态分析。

如图 5(a) ~ (c) 所示,根据柔性薄板的等效刚度计算理论^[23],微定位平台支链 1、支链 3、支链 5 驱动副的等效移动刚度分别为

$$k_{drive1} = k_{drive3} = \frac{4E^1 t_1^3 w_1}{l_1^3} \quad (32)$$

$$k_{drive5} = \frac{8E^1 t_2^3 w_2}{l_2^3} \quad (33)$$

如图 11 所示,基于微定位平台各支链的并联关系,根据式(11)、式(32)、式(33)可得平台沿 x 、 z 轴方向等效输入刚度分别为

$$K_{xinput} = 8k_{2T3R} + k_{drive1} + k_{drive3} \quad (34)$$

$$K_{zinput} = 8k_{2T3R} + k_{drive5} \quad (35)$$

式中: k_{2T3R} 为 2T3R 运动副模块的移动刚度。

基于此,采用拉格朗日方程^[19]对微定位平台进行模态分析,可得平台的前三阶固有频率为

$$f_x = f_y = \sqrt{K_{xinput}/M_{ex}}/(2\pi) \quad (36)$$

$$f_z = \sqrt{K_{zinput}/M_{ez}}/(2\pi) \quad (37)$$

式中: M_{ex} 、 M_{ez} 分别为微定位平台沿 x 、 z 轴方向的等效质量,其数值为各平动部件质量之和。

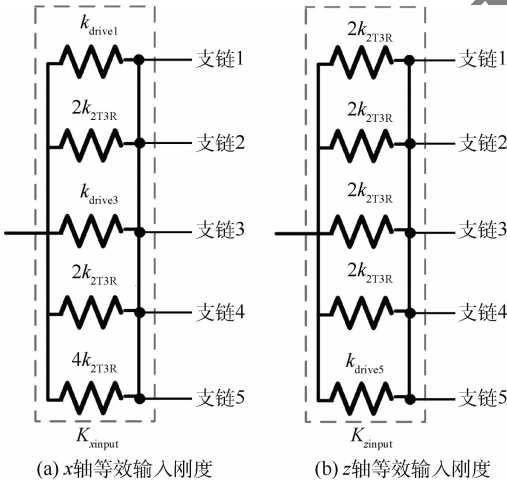


图 11 微定位平台等效输入刚度示意图

Fig. 11 Schematic diagram of equivalent input stiffness for micro-positioning platform

3 参数优化

必须指出的是,第 2 节理论分析未充分考虑微定位平台各静、动态特性间的互相影响,因此需要建立微定位平台的参数优化模型,对微定位平台的静、动态特性进行全局优化。

选择 AL-7075 作为微定位平台的材料,其弹性模量 E 为 7.17 GPa,泊松比 ν 为 0.33,屈服强度为 503 MPa,密度为 2 770 kg/m³。

基于理论分析可知,影响微定位平台静、动态特性的主要结构参数包括柔性薄板的长度 l_1 与 l_2 、宽度 t_1 与 t_2 、厚度 w_1 与 w_2 及细长杆的长度 l_w 、厚度 t_w 。因此,取优化模型的设计变量为

$$\mathbf{X} = [l_1, w_1, t_1, l_2, w_2, t_2, l_w, t_w]^T \quad (38)$$

参照文献[17]中结构的尺寸参数,并综合考虑平台的加工与安装因素,定义设计变量的范围与初始值为

$$\begin{cases} \mathbf{X}_L = [45, 18, 0.5, 35, 18, 0.5, 25, 0.5]^T \text{ mm} \\ \mathbf{X}_U = [55, 22, 1, 55, 35, 1, 35, 1]^T \text{ mm} \end{cases} \quad (39)$$

$$\mathbf{X}_0 = [50, 20, 0.5, 50, 20, 0.5, 25, 0.9]^T \text{ mm} \quad (40)$$

为方便后续实验中驱动器选型,设定音圈电机的预期最大驱动位移 δ_{max} 为 1 mm,微定位平台所需驱动力不超过 100 N。同时,微定位平台最大应力不应超过许用应力且不应发生屈曲现象。

为同时提升微定位平台的静、动态特性,参数优化时:①应减小平台沿 x 、 z 轴方向的最大丢失运动 δ_{xlost} 、 δ_{zlost} ;②应增加平台的一阶固有频率 f_z ;③应减小平台前三阶频率之间的带宽 $f_x - f_z$ 。综上,微定位平台优化模型的目标函数与约束条件可表示为

$$\begin{cases} \min f_1(\mathbf{X}) = \delta_{xlost} \\ \min f_2(\mathbf{X}) = \delta_{zlost} \\ \max f_3(\mathbf{X}) = f_z \\ \min f_4(\mathbf{X}) = f_x - f_z \\ \text{s. t.} \begin{cases} \delta_x = \delta_y = \delta_z = \delta_{max} = 1 \\ F_x \leq 100, F_z \leq 100 \\ \sigma_{xmax} \leq \frac{\sigma}{k}, \sigma_{zmax} \leq \frac{\sigma}{k} \\ F_{p1} \leq \frac{F_{cr}}{k}, F_{p2} \leq \frac{F_{cr}}{k} \end{cases} \end{cases} \quad (41)$$

式中:为保证可靠性,设定安全系数 $k=2$ 。

基于 MATLAB 中优化工具包,采用目标规划法对微定位平台进行多目标参数优化,优化后微定位平台的结构参数如表 1 所示。

分别将表 1 中结构参数与式(40)中设计变量初始值 \mathbf{X}_0 代入所推导理论模型,得到优化前后微定位平台的静、动态性能参数,如表 2 所示。

分析表 2 数据可知,优化后:①微定位平台沿 x 、 z 轴方向的丢失运动 δ_{xlost} 、 δ_{zlost} 分别减小了 25.1%、23.3%;②微定位平台的一阶固有频率 f_z 增加了 43.8%;③微定位平台前三阶频率之间的带宽 $f_x - f_z$ 减小了 97.3%。因此,微定位平台的静、动态特性均有较大提升,优化模型完全正确。

表 1 微定位平台的结构参数

Table 1 Structural parameters of micro-positioning platform

参数	数值
l_1	55.00
w_1	18.00
t_1	0.51
l_2	49.96
w_2	19.75
t_2	0.83
l_w	31.62
t_w	1.00

表2 优化前后微定位平台的静、动态性能参数

Table 2 Static and dynamic performance parameters of micro-positioning platform before and after optimization

优化前/后	$\delta_{xlost}/\mu\text{m}$	$\delta_{zlost}/\mu\text{m}$	f_z/Hz	$f_x - f_z/\text{Hz}$
优化前	9.31	2.62	34.59	23.53
优化后	6.97	2.01	49.74	0.63
优化率/%	-25.1	-23.3	43.8	-97.3

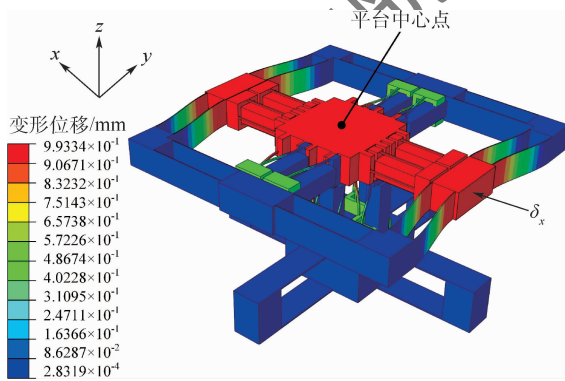
4 有限元验证

为验证所推导微定位平台静、动态特性理论模型的正确性,基于表1所确定的微定位平台结构参数,运用 Abaqus 构建有限元分析模型,分别对微定位平台的静、动态特性进行仿真验证。

4.1 静态特性验证

4.1.1 平台 x 轴静态特性验证

如图12所示,沿微定位平台 x 轴方向施加 $0 \sim 1 \text{ mm}$ 的单轴驱动位移,得到平台沿 x 轴方向各项静态性能的理论值、仿真值,如图13所示。

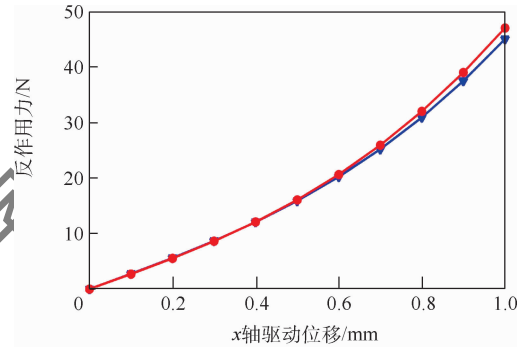
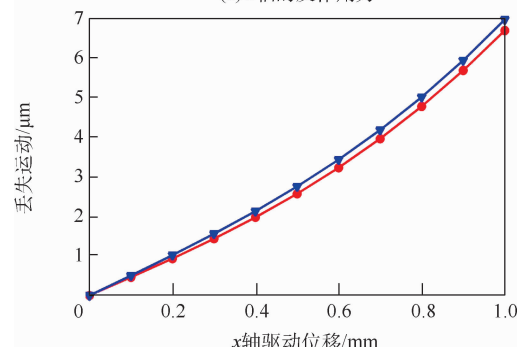
图12 微定位平台 x 轴方向静态性能有限元仿真Fig. 12 Finite element simulation of static performance for micro-positioning platform in x axis direction

如图13(a)所示,在 $0 \sim 1 \text{ mm}$ 驱动位移内,基于式(13),其反作用力理论值、仿真值的最大值分别为 45.13 N 、 47.14 N ,其仿真值与理论值之间的相对误差为 $0.2\% \sim 4.5\%$ 。

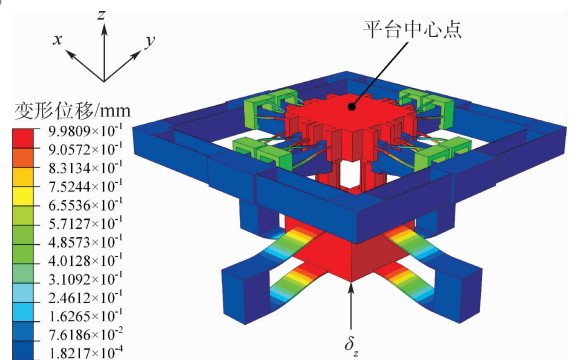
如图13(b)所示,在 $0 \sim 1 \text{ mm}$ 驱动位移内:
①基于式(21),其丢失运动理论最大值为 $6.97 \mu\text{m}$,基于有限元仿真丢失运动最大值为 $6.67 \mu\text{m}$,小于驱动位移的 0.67% ,且仿真值与理论值之间的相对误差为 $0.1\% \sim 2.9\%$;
②随着驱动位移的增加,丢失运动仿真值未发生明显突变,表明细长杆无屈曲现象。

4.1.2 平台 z 轴静态特性验证

如图14所示,沿微定位平台 z 轴方向施加 $0 \sim 1 \text{ mm}$ 的单轴驱动位移,得到平台沿 z 轴方向各项静态性能的理论值、仿真值,如图15所示。

(a) x 轴的反作用力(b) x 轴的丢失运动

—●— 仿真值 —●— 理论值

图13 微定位平台 x 轴方向静态性能验证Fig. 13 Static performance verification for micro-positioning platform in x axis direction图14 微定位平台 z 轴方向静态性能有限元仿真Fig. 14 Finite element simulation of static performance for micro-positioning platform in z axis direction

如图15(a)所示,在 $0 \sim 1 \text{ mm}$ 驱动位移内,基于式(20),其反作用力理论值、仿真值的最大值分别为 118.73 N 、 121.77 N ,其仿真值与理论值之间的相对误差为 $0.2\% \sim 2.6\%$ 。

如图15(b)所示,在 $0 \sim 1 \text{ mm}$ 驱动位移内:

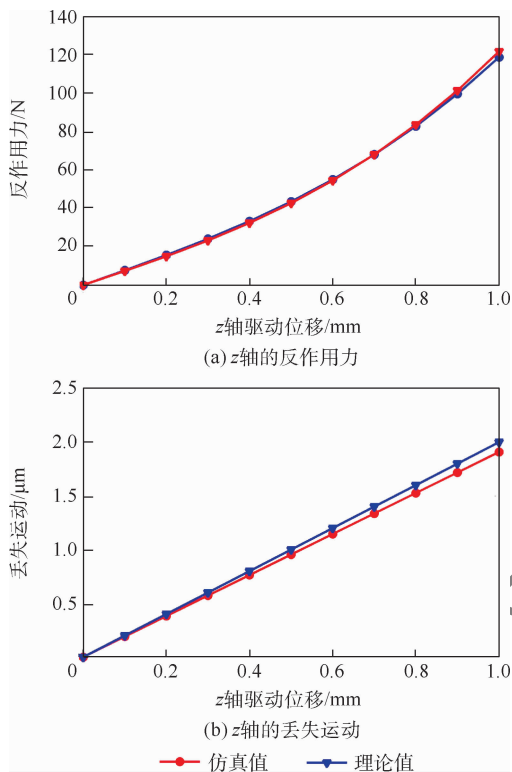


图 15 微定位平台 z 轴方向静态性能验证

Fig. 15 Static performance verification for micro-positioning platform in z axis direction

①基于式(22),其丢失运动理论最大值为 $2.01 \mu\text{m}$,基于有限元仿真丢失运动最大值分别为 $1.91 \mu\text{m}$,小于驱动位移的 0.20% ,且仿真值与理论值之间的相对误差为 $0.1\% \sim 5.0\%$;②随着驱动位移的增加,丢失运动仿真值未发生明显突变,表明细长杆无屈曲现象。

4.1.3 平台耦合位移仿真

输入耦合与输出耦合是影响平台定位精度的重要因素^[17],2T3R型运动副的反向串联结构,五支链平台的对称布置结构均可有效减小平台的输入、输出耦合位移。为验证微定位平台的解耦性能,分别对平台沿 x 、 z 轴单轴驱动时产生的输入、输出耦合位移进行仿真分析。

如图 12 所示,沿微定位平台 x 轴方向施加 $0 \sim 1 \text{ mm}$ 的单轴驱动位移,得到微定位平台输入、输出耦合位移的仿真值如图 16 所示。由图 16(a)可知,微定位平台沿 y 、 z 轴方向输入耦合位移的最大值分别为 $0.41 \mu\text{m}$ 、 $0.21 \mu\text{m}$;由图 16(b)可知,微定位平台沿 y 、 z 轴方向输出耦合位移的最大值分别为 15.7 nm 、 33.4 nm 。

如图 14 所示,沿微定位平台 z 轴方向施加 $0 \sim 1 \text{ mm}$ 的单轴驱动位移,得到微定位平台输入、输出耦合位移的仿真值如图 17 所示。可知,微定位平台沿 x 轴方向输入、输出耦合位移的最大值

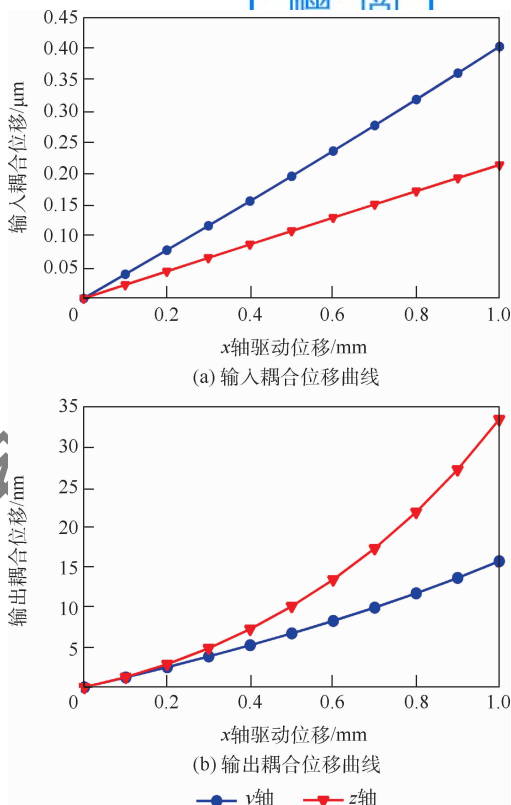


图 16 微定位平台 x 轴驱动时耦合位移仿真值

Fig. 16 Simulation value of coupling displacement for micro-positioning platform driving along x axis direction

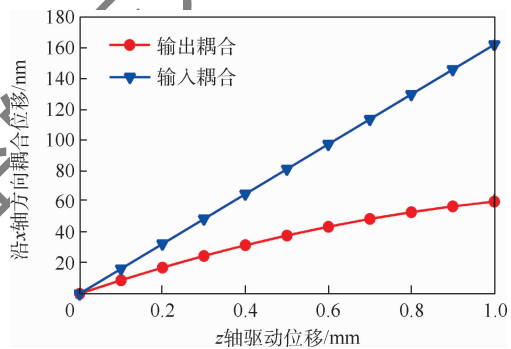


图 17 微定位平台 z 轴驱动时耦合位移仿真值

Fig. 17 Simulation value of coupling displacement for micro-positioning platform driving along z axis direction

分别为 161.9 nm 、 59.9 nm 。

基于上述仿真结果可知,微定位平台沿 x 、 z 轴方向施加单轴驱动位移产生的输入、输出耦合位移均小于驱动位移的 0.05% ,其值可忽略不计,表明平台沿 x 、 z 轴方向输入、输出完全解耦。

必须指出的是,2T3R型运动副作被动副时可有效减小平台的输入耦合位移。为进一步研究平台的输入解耦性能,将图 5(a)所示平台与采用 2T 型运动副作被动副的平台(文献[17])相比较。当分别对微定位平台施加沿 x 、 z 轴方向

1 mm 的单轴驱动位移时,2 个平台的输入耦合位移仿真值如表 3 所示。

由表 3 数据可知,与文献[17]中平台相比,图 5(a)所示平台的输入耦合位移明显更小,具备更好的输入解耦性能,进一步证明了 2T3R 型运动副与微定位平台结构设计的有效性。

表 3 微定位平台输入耦合位移仿真值

Table 3 Simulation value of input coupling displacement for micro-positioning platforms

驱动方式	耦合位移方向	输入耦合位移/ μm	
		图 5(a)平台	文献[17]平台
z 轴单轴驱动	沿 x 轴输入耦合	0.16	11.92
x 轴单轴驱动	沿 y 轴输入耦合	0.41	11.92

4.1.4 平台材料应力验证

为验证微定位平台材料的最大应力,沿 x、y、z 轴方向同时施加 1 mm 的驱动位移,则对应的平台材料应力仿真如图 18 所示。

根据仿真结果可知,平台支链 1 与支链 5 中细长杆最大应力的仿真值分别为 225.5 MPa 与 220.6 MPa。基于式(23)、式(24)可知,对应最大应力的理论值分别为 223.0 MPa 与 217.4 MPa,其仿真值与理论值之间的相对误差分别为 1.1%、1.5%。因此,微定位平台多轴联动时,各支链中细长杆最大应力值均小于材料许用应力,无应力失效风险。

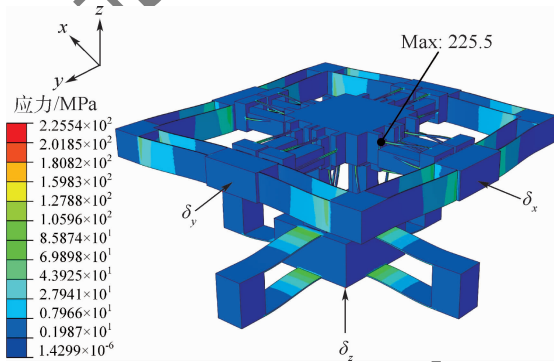


图 18 微定位平台多轴联动应力云图

Fig. 18 Stress contour of multi-axis motion for micro-positioning platform

4.2 动态性能验证

对微定位平台进行有限元模态仿真,得到平台的前 6 阶模态振型,如图 19 所示。

基于式(36)、式(37)可得微定位平台固有频率的理论值、仿真值及相对误差,如表 4 所示。可知:①平台 1~3 阶固有频率仿真值与理论值之间的相对误差小于 5.5%;②平台 1~3 阶固有频率(自由度方向)与 4~6 阶固有频率(非自由度方向)的仿真值在数值上差距明显,表明平台具有较强的动态抗干扰性能。

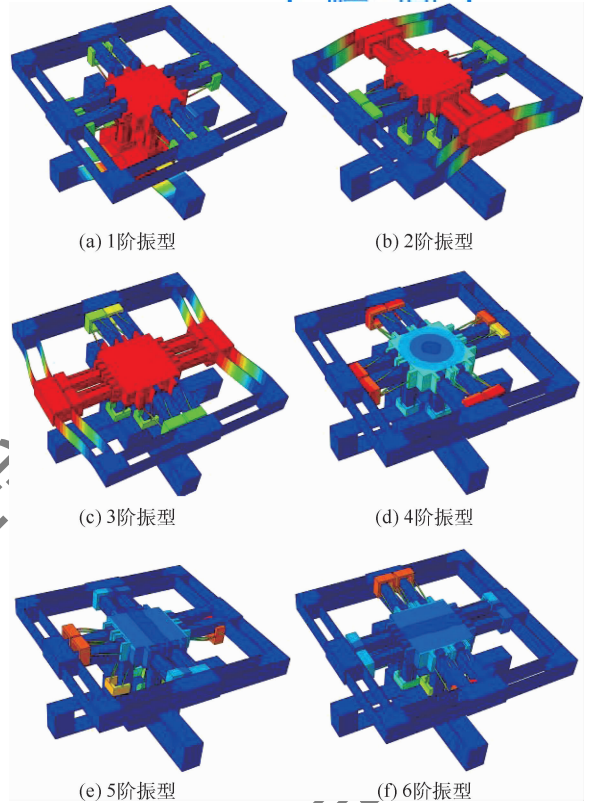


图 19 微定位平台的前 6 阶模态振型

Fig. 19 The first 6 modal shapes of micro-positioning platform

表 4 微定位平台固有频率的理论值、仿真值及相对误差
Table 4 Theoretical value, simulation value and relative error of natural frequency for micro-positioning platform

阶数	固有频率/Hz		相对误差/%
	理论值	仿真值	
1	49.74	51.27	3.1
2	50.37	52.44	4.1
3	50.37	53.11	5.4
4		208.57	
5		220.31	
6		221.65	

5 结论

1) 提出了一种新型的大行程 2T3R 柔性运动副,并基于该运动副设计了一种大行程解耦五支链三平动微定位平台。

2) 分别建立了微定位平台力-位移关系、丢失运动和固有频率的理论模型,并采用目标规划法对微定位平台进行了多目标参数优化,优化结果符合预期。

3) 通过有限元仿真验证了微定位平台理论模型的正确性,理论研究与仿真结果表明,微定位平台一阶频率为 51.27 Hz,在 1 mm 的驱动位移内,平台输入输出完全解耦且沿 x、z 轴方向丢失运动分别小于驱动位移的 0.67%、0.20%,运动

副与平台结构设计具备有效性。

参考文献 (References)

- [1] 于靖军,郝广波,陈贵敏,等. 柔性机构及其应用研究进展[J]. 机械工程学报,2015,51(13):53-68.
YU J J,HAO G B,CHEN G M, et al. State-of-art of compliant mechanisms and their applications[J]. Journal of Mechanical Engineering,2015,51(13):53-68(in Chinese).
- [2] 周睿,周辉,桂和利,等. 基于柔性铰链的二自由度微动平台分析及优化[J]. 北京航空航天大学学报,2018,44(9):199-207.
ZHOU R,ZHOU H,GUI H L, et al. Analysis and optimization of 2-DoF micro-positioning stage based on flexible hinges[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018,44(9):199-207(in Chinese)
- [3] 李杨民,汤晖,徐青松,等. 面向生物医学应用的微操作机器人技术发展态势[J]. 机械工程学报,2011,47(23):1-13.
LI Y M,TANG H,XU Q S, et al. Development status of micromanipulator technology for biomedical applications[J]. Journal of Mechanical Engineering, 2011,47(23):1-13(in Chinese).
- [4] ÖZKALE B,PARREIRA R,BEKDEMİR A, et al. Modular soft robotic micro-devices for dexterous biomanipulation[J]. Lab on a Chip,2019,19(5):778-788.
- [5] 王保兴,孟刚,林苗,等. 3-PPP型柔性并联微定位平台的设计与分析[J]. 北京航空航天大学学报,2020,46(4):798-807.
WANG B X,MENG G,LIN M, et al. Design and analysis of a 3-PPP compliant parallel micro-positioning stage[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020,46(4):798-807(in Chinese).
- [6] ZHU Z,TO S,ZHU W L, et al. Optimum design of a piezo-actuated triaxial compliant mechanism for nanocutting[J]. IEEE Transactions on Industrial Electronics, 2018,65(8):6362-6371.
- [7] WATANABE S,ANDO T. High-speed XYZ nano-positioner for scanning ion conductance microscopy[J]. Applied Physics Letters,2017,111(11):113106.
- [8] AWTAR S,USTICK J,SEN S. An XYZ parallel kinematic flexure mechanism with geometrically decoupled degrees of freedom[C]//Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,2011:119-126.
- [9] LI H Y,HAO G B,KAVANAGH R C. A new XYZ compliant parallel mechanism for micro-/nano-manipulation: Design and analysis[J]. Micromachines,2016,7(2):23.
- [10] HAO G B,LI H Y. Design of 3-legged XYZ compliant parallel manipulators with minimized parasitic rotations[J]. Robotica, 2015,33(4):787-806.
- [11] HOPKINS J B,CULPEPPER M L. Synthesis of multi-degree of freedom,parallel flexure system concepts via freedom and constraint topology (FACT). Part I:Principles[J]. Precision Engineering,2010,34(2):259-270.
- [12] HOPKINS J B,CULPEPPER M L. Synthesis of multi-degree of freedom,parallel flexure system concepts via freedom and constraint topology (FACT). Part II:Principles[J]. Precision Engineering,2010,34(2):271-278.
- [13] XU Q S. A totally decoupled piezo-driven XYZ flexure parallel micropositioning stage for micro/nanomanipulation[J]. IEEE Transactions on Automation Science and Engineering,2011,8(2):265-279.
- [14] 李海洋,郝广波,于靖军,等. 空间平动柔性并联机构的系统设计方法研究[J]. 机械工程学报,2018,54(13):57-65.
LI H Y,HAO G B,YU J J, et al. Systematic approach to the design of spatial translational compliant parallel mechanisms[J]. Journal of Mechanical Engineering, 2018,54(13):57-65(in Chinese).
- [15] ZHANG X Z,XU Q S. Design, fabrication and testing of a novel symmetrical 3-DOF large-stroke parallel micro/nano-positioning stage[J]. Robotics and Computer-Integrated Manufacturing, 2018,54:162-172.
- [16] HAO G B,KONG X W,MENG Q L. Design and modelling of spatial compliant parallel mechanisms for large range of translation[C]//Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,2010,2:329-340.
- [17] 曹毅,王保兴,孟刚,等. 大行程三平动柔性微定位平台的设计分析及优化[J]. 机械工程学报,2020,56(17):71-81.
CAO Y,WANG B X,MENG G, et al. Design analysis and optimization of large range spatial translational compliant micro-positioning stage[J]. Journal of Mechanical Engineering,2020,56(17):71-81(in Chinese).
- [18] 曹毅,孟刚,居勇健,等. 基于2T3R型柔性运动副的微定位平台:CN11785318A[P]. 2020-10-16.
CAO Y,MENG G,JU Y J, et al. Micro-positioning stage based on 2T3R compliant kinematic pair:CN11785318A[P]. 2020-10-16(in Chinese).
- [19] 杨志军,白有盾,陈新,等. 基于应力刚化效应的动态特性可调微动平台设计新方法[J]. 机械工程学报,2015,51(23):153-159.
YANG Z J,BAI Y D,CHEN X, et al. A new design method of dynamic characteristics adjustable micro motion stage based on tension stiffening[J]. Journal of Mechanical Engineering,2015,51(23):153-159(in Chinese).
- [20] YU J J,LI S Z,SU H J, et al. Screw theory based methodology for the deterministic type synthesis of flexure mechanisms[J]. Journal of Mechanisms and Robotics,2011,3(3):1194-1204.
- [21] MA F L,CHEN G M. Bi-BCM: A closed-form solution for fixed-guided beams in compliant mechanisms[J]. Journal of Mechanisms and Robotics,2017,9(1):014501.
- [22] HERPE X,WALKER R,DUNNIGAN M, et al. On a simplified nonlinear analytical model for the characterization and design optimization of a compliant XY micro-motion stage[J]. Robotics and Computer-Integrated Manufacturing,2018,49:66-76.
- [23] HOWELL L L. Compliant mechanisms[M]. New York: John Wiley and Sons,2001:302.

Design and optimization of large-stroke decoupled three-translational micro-positioning platform

LIN Miao^{1,2}, MENG Gang^{1,2}, JU Yongjian^{1,2}, XU Weisheng^{1,2}, CAO Yi^{1,2,*}

(1. School of Mechanical Engineering, Jiangnan University, Wuxi 214122, China;

2. Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment and Technology, Wuxi 214122, China)

Abstract: To design large-stroke three-translational micro-positioning platforms with excellent decoupling characteristics, a new 2T3R type motion pair was proposed. The structure of three-translational micro-positioning platform was designed based on the 2T3R type motion pair. The theoretical models of force-displacement relationship and lost motion were established by nonlinear model method, the theoretical models of platform natural frequency were established by Lagrange equation. The goal programming method was used to optimize parameters of the micro-positioning platform. The correctness of the above theoretical model was verified by finite element simulation. According to the theoretical calculation and simulation results, the first-order natural frequency of the platform is 51.27 Hz. the lost motions in x and z directions are less than 0.67% and 0.20% in 1 mm motion stroke, and the input and output motions are completely decoupled. Results show that the structure design of motion pair and platform is effective, and the optimization model is feasible.

Keywords: compliant mechanism; structural design; micro-positioning platform; parameter optimization; compliant motion pair

Received: 2021-01-07; **Accepted:** 2021-02-10; **Published online:** 2021-03-11 20:26

URL: kns.cnki.net/kcms/detail/11.2625.V.20210311.1605.002.html

Foundation items: The Six Talent Peaks Project in Jiangsu Province (ZBZZ-012); 111 Project (B18027); Postgraduate Research & Practice Innovation Program of Jiangsu Province, China (JSCX20_0760); Postgraduate Research & Practice Innovation Program of Jiangnan University, China (JNSJ19_005)

* **Corresponding author.** E-mail: caoyi@jiangnan.edu.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0020

基于 PM 算法的涡旋电磁波引信超分辨测向方法

孙丽华¹, 闫晓鹏^{1,*}, 刘强², 郝新红¹, 章鸿运¹

(1. 北京理工大学 机电动态控制重点实验室, 北京 100081;

2. 中国兵器科学研究院, 北京 100089)



摘 要: 为提高防空导弹无线电引信在低信噪比(SNR)背景下的测向精度,提出了基于传播算子算法(PM)的涡旋电磁波引信超分辨测向方法。通过相控均匀圆阵依次发射不同轨道角动量模态涡旋电磁波照射目标,基于多发单收(MISO)模型推导目标探测回波模型;通过对回波方程的近似变形构造接收信号矩阵;采用 PM 算法构造空间谱函数,进行谱峰搜索,实现对目标水平方位角的估计。对所提方法的有效性进行仿真验证,结果表明,所提方法可在低信噪比、少阵元情况下实现对目标的精确测向。

关键词: 涡旋电磁波; 轨道角动量; 引信测向; 多发单收(MISO); 传播算子算法(PM)

中图分类号: TJ43⁺ 4.1

文献标志码: A **文章编号:** 1001-5965(2022)07-1263-06

现代战争中,防空导弹武器是防空作战中能够有效应对空中威胁的主要手段。新一代防空导弹大多采用定向战斗部^[1],不仅要求引信能够获得弹目之间的距离信息,还需要具备精确的目标方位估计能力。为了提高引信测向能力,多项技术(如阵列天线技术及多通道信号处理技术)在引信系统中得到应用。目前,无线电引信通常采用多象限分区识别技术进行目标方位识别,方法主要分为振幅法和相位法。振幅法易于检测,而相位法测向精度高于振幅法,但易受噪声干扰,具有较大的随机性^[2-4]。因此,如何在低信噪比下提高测向精度成为引信领域的重要问题之一。

涡旋电磁波与传统的平面波不同,其波前是特有的螺旋结构,具有角度分集特性^[5-6],而连续变化的波前使其本身具有分辨方位的能力,从而得到广泛的重视和研究^[7-11]。文献[12]建立了涡旋雷达模型,以均匀圆阵作为收发天线推导了目

标回波的数学表达式;文献[13]推导了多发多收(multiple-input multiple-output, MIMO)和多发单收(multiple-input single-output, MISO)的涡旋雷达理论模型及其近似模型,提出了轨道角动量(orbital angular momentum, OAM)的阶数和目标的方位角坐标量存在傅里叶变换的关系,通过离散傅里叶变换可以得到目标的方位角坐标量;文献[14]借鉴空间来波信号波达角(direction of arrival, DOA)估计的思路,将涡旋雷达模型进行适当的变形,使其可以利用子空间分解的算法实现雷达目标的超分辨探测,为涡旋电磁波用于引信目标探测奠定了基础。

本文针对无线电引信对空中目标的高精度测角需求,基于 MISO 模型建立了涡旋电磁波引信目标探测回波模型;采用传播算子算法(propagator method, PM),用简单的矩阵变换代替协方差矩阵特征分解构造谱函数^[15-16],从而实现目标水平方位角的精确估计,并且通过数学仿真软件

收稿日期: 2021-01-14; 录用日期: 2021-02-22; 网络出版时间: 2021-03-09 18:22

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210308.1713.001.html

基金项目: 国家自然科学基金(61973037, 61871414)

* 通信作者. E-mail: yanxiaopeng@bit.edu.cn

引用格式: 孙丽华, 闫晓鹏, 刘强, 等. 基于 PM 算法的涡旋电磁波引信超分辨测向方法[J]. 北京航空航天大学学报, 2022, 48(7): 1263-1268. SUN L H, YAN X P, LIU Q, et al. PM based super-resolution method of azimuth detection for electromagnetic vortex wave fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1263-1268 (in Chinese).

MATLAB 仿真验证了该方法在 -10 dB 信噪比条件下测角精度高于 10° 。

1 基于 MISO 的涡旋电磁波引信目标探测回波模型

基于 MISO 的涡旋电磁波引信探测几何模型如图 1 所示。 N 个相同的各向同性阵元构成圆阵作为发射天线,各阵元均匀分布在半径为 a 的圆周上,第 n 个阵元的方位角 $\phi_n = 2\pi n/N, n = 1, 2, \dots, N$, 圆周位于 xoy 面。接收天线位于圆心处且独立于均匀圆阵,以接收天线为原点建立球坐标系,距离 r 为原点和目标连线的长度,俯仰角 $\theta \in [0, \pi/2]$ 为原点和目标的连线与 z 轴正半轴的夹角,方位角 $\varphi \in [0, 2\pi]$ 为原点和目标的连线从 x 轴正半轴逆时针转过的角度,目标位置为 $P(r, \theta, \varphi)$ 。

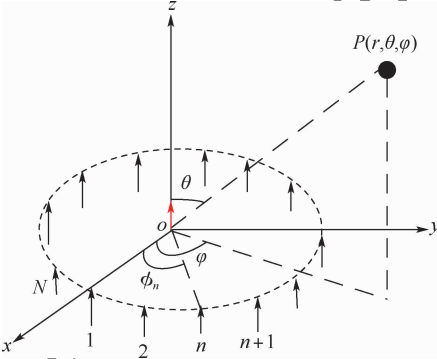


图 1 基于 MISO 的涡旋电磁波引信探测几何模型

Fig.1 MISO-based geometry model of vortex electromagnetic wave fuze detection

所有发射阵元馈入的脉冲信号为

$$s(t) = \sum_{u=1}^Z e^{i2\pi f_0(t-uT_r)} \text{rect}\left(\frac{t-uT_r}{T}\right) \quad (1)$$

式中: u 为馈入的脉冲信号个数, $u = 1, 2, \dots, Z$, Z 为 OAM 模态数; f_0 为信号载频; T 为脉冲宽度; T_r 为脉冲重复周期; $\text{rect}(t/T)$ 为矩阵函数, 即

$$\text{rect}\left(\frac{t}{T}\right) = \begin{cases} 1 & \left| \frac{t}{T} \right| \leq \frac{1}{2} \\ 0 & \text{其他} \end{cases} \quad (2)$$

为了产生 OAM 模态为 l 的涡旋电磁波, 依次给每个阵元添加相移 $l\phi_n = 2\pi nl/N, n = 1, 2, \dots, N$ 。涡旋电磁波引信目标探测通过遍历不同模态的涡旋电磁波照射待观测目标, 即每个发射脉冲的涡旋电磁波 OAM 模态不同, 则每隔脉冲重复周期 T_r , 须改变各个阵元的添加相移。假设发射脉冲 OAM 模态分别为 l_1, l_2, \dots, l_Z , 则阵列发射信号如图 2 所示。

则单个阵元发射脉冲信号为

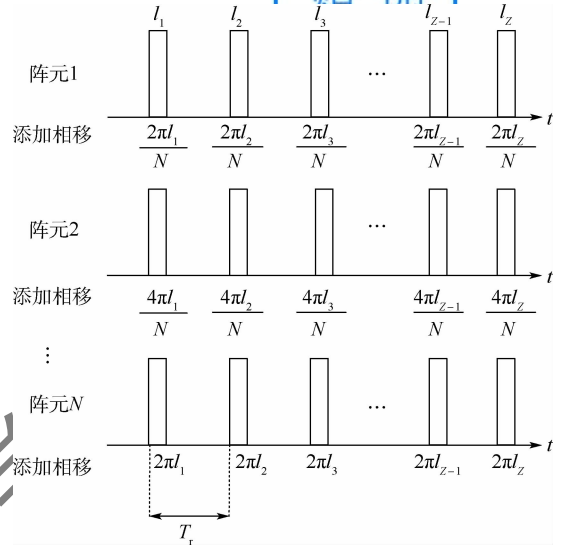


图 2 阵列发射信号

Fig.2 Array emission signal

$$S_0(t, l) = \sum_{u=1}^Z e^{i l_u \phi_n} e^{i2\pi f_0(t-uT_r)} \text{rect}\left(\frac{t-uT_r}{T}\right) \quad (3)$$

均匀圆阵发射脉冲信号为

$$S_1(t, l) = \sum_{n=1}^N S_0(t, l) = \sum_{n=1}^N \sum_{u=1}^Z e^{i l_u \phi_n} e^{i2\pi f_0(t-uT_r)} \text{rect}\left(\frac{t-uT_r}{T}\right) \quad (4)$$

接收端采用位于圆心并独立于发射天线的单一天线作为接收天线, 则单一点目标回波信号为

$$S_r(t, l) = S_1(t - \tau - \tau_n, l) = \sum_{n=1}^N \sum_{u=1}^Z e^{i l_u \phi_n} e^{i2\pi f_0(t-\tau-\tau_n-uT_r)} \text{rect}\left(\frac{t-\tau-\tau_n-uT_r}{T}\right) \approx \sum_{u=1}^Z \sum_{n=1}^N e^{i l_u \phi_n} e^{-i2\pi f_0 \tau_n} e^{i2\pi f_0(t-\tau-uT_r)} \text{rect}\left(\frac{t-\tau-uT_r}{T}\right) \quad (5)$$

式中: $\tau = 2r/c$ 为发射阵元在圆阵中心参考点处回波时延, $c = 3 \times 10^8$ m/s 为光速; $\tau_n = a \sin \theta \cos(\varphi - \phi_n)/c$ 为第 n 个发射阵元相对于参考点处的时延。

则式(5)为

$$S_r(t, l) \approx s(t - \tau) \frac{N}{2\pi} \int_0^{2\pi} e^{i l_u \phi_n} e^{-i2\pi f_0 \frac{a \sin \theta \cos(\varphi - \phi_n)}{c}} d\phi_n = Ns(t - \tau) J_{l_u}(k a \sin \theta) e^{i l_u \varphi} e^{i l_u \pi/2} \quad (6)$$

$$s(t - \tau) = \sum_{u=1}^Z e^{i2\pi f_0(t-\tau-uT_r)} \text{rect}\left(\frac{t-\tau-uT_r}{T}\right) \quad (7)$$

式中: $J_{l_u}(k a \sin \theta)$ 为 l_u 阶第一类贝塞尔函数, $k = 2\pi f_0/c$ 为信号的波数。

从式(6)可以看出, 涡旋回波包络受第一类贝塞尔函数调制, 目标距离信息与 $s(t - \tau)$ 项中的时延有关, 而方位角信息与 OAM 模态数调制

在 $e^{iu\varphi}$ 项中,故在 OAM 域进行解调便能得到目标的方位角信息。

2 基于 PM 算法的精确测向方法

2.1 PM 算法原理

当满足条件 $J_l(k \sin \theta) \gg 1$ 且 $J_l(k \sin \theta) \gg l$ 时,贝塞尔函数可以近似如下:

$$J_l(k \sin \theta) \approx \sqrt{\frac{2}{\pi k \sin \theta}} \cos\left(k \sin \theta - \frac{l\pi}{2} - \frac{\pi}{4}\right) \quad (8)$$

将回波信号的相位项 $e^{iu\pi/2}$ 进行补偿,补偿后回波信号通过式(8)的近似变形可表示为

$$S'_r(t, l) = Ns(t - \tau) e^{iu\varphi} J_l(k \sin \theta) = Ns(t - \tau) e^{iu\varphi} \sqrt{\frac{2}{2\pi k \sin \theta}} \cos\left(k \sin \theta - \frac{l\pi}{2} - \frac{\pi}{4}\right) = Ns(t - \tau) \cdot \begin{bmatrix} e^{iu(\varphi - \frac{\pi}{2})} & e^{iu(\varphi + \frac{\pi}{2})} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2\pi k \sin \theta}} e^{i(k \sin \theta - \frac{\pi}{4})} \\ \sqrt{\frac{1}{2\pi k \sin \theta}} e^{-i(k \sin \theta - \frac{\pi}{4})} \end{bmatrix} \quad (9)$$

对涡旋回波在 OAM 域进行采样,回波幅度归一化后构造接收信号矩阵为

$$X(t) = \begin{bmatrix} x(l_1, t) \\ x(l_2, t) \\ \vdots \\ x(l_Z, t) \end{bmatrix} = \begin{bmatrix} e^{iu_1(\varphi - \frac{\pi}{2})} & e^{iu_1(\varphi + \frac{\pi}{2})} \\ e^{iu_2(\varphi - \frac{\pi}{2})} & e^{iu_2(\varphi + \frac{\pi}{2})} \\ \vdots & \vdots \\ e^{iu_Z(\varphi - \frac{\pi}{2})} & e^{iu_Z(\varphi + \frac{\pi}{2})} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2\pi k \sin \theta}} e^{i(k \sin \theta - \frac{\pi}{4})} s'(t - \tau) \\ \sqrt{\frac{1}{2\pi k \sin \theta}} e^{-i(k \sin \theta - \frac{\pi}{4})} s'(t - \tau) \end{bmatrix} + \begin{bmatrix} n(l_1, t) \\ n(l_2, t) \\ \vdots \\ n(l_Z, t) \end{bmatrix} = AS(t) + N(t) \quad (10)$$

式中: $S(t) \in \mathbf{C}^{2 \times K}$ 为发射阵元位于参考点时接收信号矢量, K 为快拍数; $s'(t - \tau) = e^{i2\pi f_0(t - \tau)} \cdot \text{rect}\left(\frac{t - \tau}{T}\right)$; $N(t) \in \mathbf{C}^{Z \times 2}$ 为接收到的高斯白噪声向量; $A \in \mathbf{C}^{Z \times 2}$ ($Z > 2$) 为信号回波模型的导向矢量阵,且

$$A = \left[a\left(\varphi - \frac{\pi}{2}\right), a\left(\varphi + \frac{\pi}{2}\right) \right] \quad (11)$$

其中:

$$a\left(\varphi - \frac{\pi}{2}\right) = \left[e^{iu_1(\varphi - \frac{\pi}{2})}, e^{iu_2(\varphi - \frac{\pi}{2})}, \dots, e^{iu_Z(\varphi - \frac{\pi}{2})} \right]^T \quad (12)$$

$$a\left(\varphi + \frac{\pi}{2}\right) = \left[e^{iu_1(\varphi + \frac{\pi}{2})}, e^{iu_2(\varphi + \frac{\pi}{2})}, \dots, e^{iu_Z(\varphi + \frac{\pi}{2})} \right]^T \quad (13)$$

对接收信号矩阵进行协方差运算,得到协方差矩阵 R 为

$$R = E[X(t)X^H(t)] = AR_sA^H + \rho_{\text{noise}}^2 I \quad (14)$$

式中: $R_s = E[S(t)S^H(t)]$ 为发射天线位于参考点时接收信号的协方差矩阵; ρ_{noise}^2 为空间环境中的噪声功率; I 为 $Z \times Z$ 单位矩阵。

传播算子 P 是使式(15)成立的唯一线性算子:

$$P^H A_1 = A_2 \quad (15)$$

式中: A_1 为导向矢量阵 A 的前 2 行; A_2 为导向矢量阵 A 的后 $Z - 2$ 行。

则存在矩阵 $Q = [P^H, -I_{Z-2}]$, 满足:

$$Q^H A = [P^H, -I_{Z-2}] \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = 0 \quad (16)$$

式(16)表明, Q 与 A 正交,则 Q 组成的子空间包含在噪声子空间中,故可以构造如下估计函数:

$$P_{\text{PM}} = \frac{1}{a^H(\varphi) Q Q^H a(\varphi)} \quad (17)$$

式中: $a(\varphi) = [e^{iu_1(\varphi)}, e^{iu_2(\varphi)}, \dots, e^{iu_Z(\varphi)}]^T$ 为谱搜索导向矢量。

搜索谱峰即可估计出目标的水平方位角。

通过阵列的协方差矩阵 R 求传播算子 P , 对 R 分块如下:

$$R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \quad (18)$$

式中: R_1 为 R 的前 2 行; R_2 为 R 的后 $Z - 2$ 行,与式(15)相同,协方差矩阵也满足式(19):

$$R_2 = R_1 P \quad (19)$$

传播算子 P 的估计可以用最小二乘法进行求解,即

$$J_{\text{csm}}(P) = \|R_2 - R_1 P\|_F^2 \quad (20)$$

进而求得 P 的最优解为

$$P = (R_1 R_1^H)^{-1} R_1 R_2^H \quad (21)$$

只要求得 P , 就可以根据式(16)得到 Q 。

综上,基于 PM 算法对水平方位角谱估计的完整步骤如下:

步骤 1 根据式(21)估计传播算子 P 。

步骤 2 构造矩阵 $Q = [P^H, -I_{Z-2}]$ 。

步骤 3 通过对式(17)谱函数进行谱峰搜索

得到水平方位角估计。

2.2 基于 PM 的涡旋电磁波引信测向方法分析

假设 L 为模态数, OAM 域采样间隔为 1, K 为快拍数, 谱峰搜索范围为 $[0^\circ, 180^\circ]$, 搜索步长为 0.1° , 则本文 PM 算法复杂度为 $O(KL^2 + 4L + 2L^2 + 8 + 1800(2L(L-2) + L))$; 文献[14]中空间平滑 MUSIC 算法复杂度为 $O((2P_e + 1)N_e^2 + N_e^3 + 1800(N_e(N_e - 2) + N_e(N_e - 2 + 1)))$, 其中, P_e 为子阵数, $N_e = L - P_e + 1$ 。图 3 为 $K = 100$ 、模态数 $L = 10 \sim 50$ 时本文算法与文献[14]中空间平滑 MUSIC 算法的复杂度对比。可以看出, PM 算法的复杂度小于空间平滑 MUSIC 算法, 且随着模态数的增大, 算法复杂度差距越大, 考虑到引信的运算成本与实时性, 若采用最大时钟频率 640 MHz 的美国 Xilinx 公司 XC7VX550t 型号芯片实现 PM 算法, 则算法复杂度所对应的测向时间小于 2 ms, 可满足引信对实时性的要求。

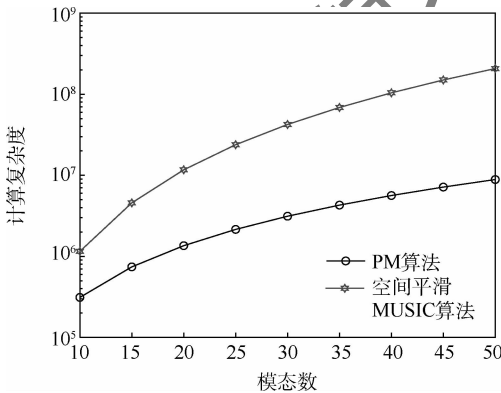


图3 算法复杂度与模态数的关系

Fig. 3 Relation between algorithm complexity and mode number

3 仿真实验与性能分析

设置仿真实验条件为: 目标位置为 $(20 \text{ m}, 10^\circ, 60^\circ)$, 均匀圆阵半径为 0.15 m , 发射脉冲信号中心频率为 30 GHz , 脉冲重复周期 $T_r = 10 \mu\text{s}$, 脉冲宽度 $T = 1 \mu\text{s}$ 。

图 4 给出了 PM 算法在阵元数 $N = 10$ 、模态 $l = [-4, 4]$ 、快拍数 $K = 100$ 和信噪比 $\text{SNR} = 20 \text{ dB}$ 时对目标进行 200 次 Monte Carlo 仿真实验水平方位角估计结果。可以看出, 相位补偿前在距离真实目标 $\pm 90^\circ$ 位置处出现谱峰, 并非对应真实目标处; 相位补偿后在距离真实目标 180° 出现虚假谱峰, 虚假谱峰不能通过滤波方法消除, 只能通过其他先验信息进行排除。这是由于式(11)中的导向矢量为 $e^{i(\varphi - \frac{\pi}{2})}$ 和 $e^{i(\varphi + \frac{\pi}{2})}$, 1 个目标对应于 2 个其真实位置偏移 90° 的谱峰, 而非真实目标

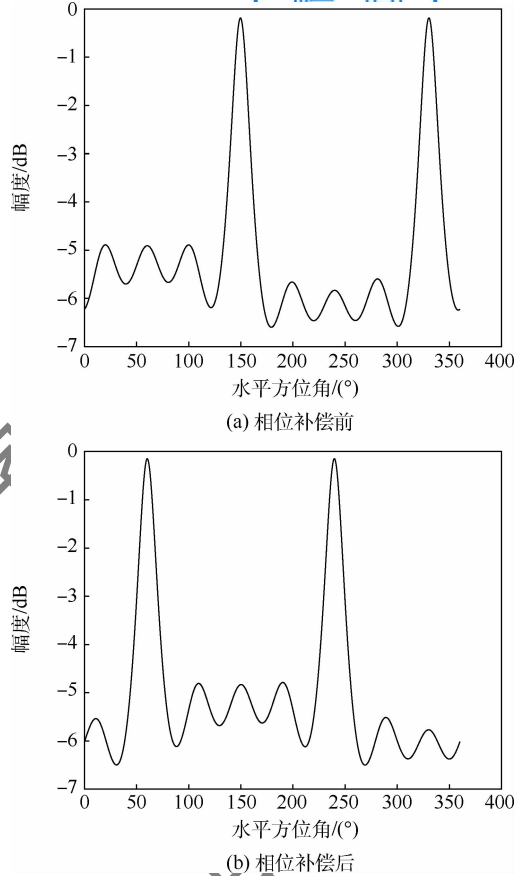


图4 PM 算法水平方位角估计结果

Fig. 4 Results of horizontal azimuth estimation of PM algorithm

值, 通过对结果进行相位补偿, 谱峰往右平移 90° , 分别对应于目标真实值和偏移 180° 的虚假谱。

图 5 给出了不同算法在阵元数 $N = 10$ 、模态 $l = [-4, 4]$ 、快拍数 $K = 100$ 和信噪比 $\text{SNR} = -20 \sim 30 \text{ dB}$ 情况下进行 200 次 Monte Carlo 仿真实验水平方位角估计性能对比, 包括本文 PM 算法、文献[13]的 FFT 算法及文献[14]的空间平滑 MUSIC 算法。

从图 5(a) 可以看出, 在阵元数为 10 时, FFT 算法的水平方位角均方根误差高达 48° , 这是由于 FFT 算法需要对大量 OAM 模态进行采样, 测向精度与模态数大小有关, 而均匀圆阵产生涡旋电磁波模态数与圆阵数满足 $l_{\max} < N/2$, 故 FFT 算法测向精度受引信尺寸限制。

对比图 5(b), 很明显可以看出, 本文 PM 算法与文献[14]算法在少阵元下测向性能远远好于文献[13]的 FFT 算法。图 5(b) 表明, 在 $\text{SNR} \leq -10 \text{ dB}$ 情况下, PM 算法的水平方位角均方根误差小于空间平滑 MUSIC 算法, 即在较低信噪比情况下, PM 算法测向性能优于空间平滑 MUSIC 算法; 随着信噪比增加, 各算法的水平方位角均方根误差逐渐减小, PM 算法测向性能逐渐次于空间

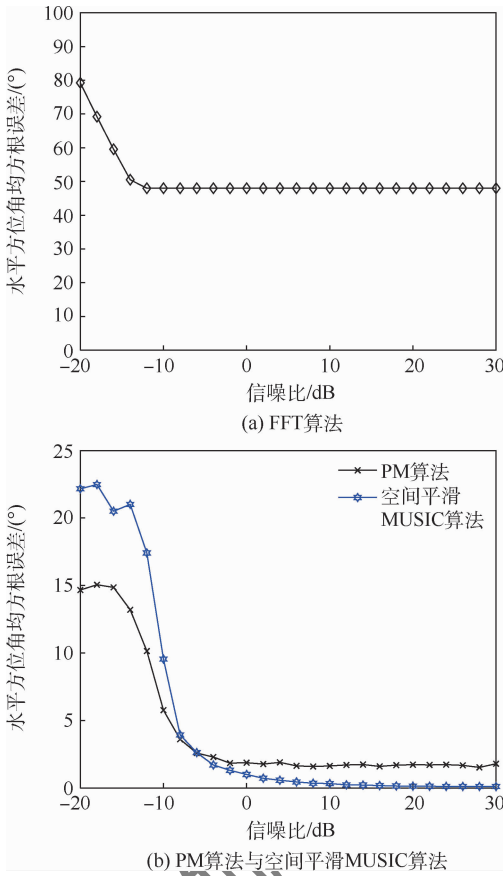


图5 不同信噪比下各算法对水平方位角估计性能对比
Fig.5 Comparison of horizontal azimuth estimation performance of different algorithms at different SNRs

平滑 MUSIC 算法,但测向误差仍稳定在 2°左右,远远满足引信对测向的需求,即能实现对目标较高精度的水平方位角估计。

图6给出了信噪比 SNR = -10 dB、快拍数 K = 100 和阵元数 N = 6 ~ 64 时不同算法进行 200 次 Monte Carlo 仿真实验水平方位角估计性能对比。可以看出,PM 算法和空间平滑 MUSIC 算法水平方位角估计性能随阵元数的增加而逐渐增强,但在阵元数小于 10 时,文献[14]的空间平滑 MUSIC 算法的测向性能急剧恶化,而本文 PM 算法测向误差相对更稳定。故本文 PM 算法可以在保证相同测向精度的同时减少圆阵的阵元数,从而满足引信尺寸小的要求。

从上述分析可知,涡旋电磁波引信测向性能与引信所发射的涡旋电磁波模态数有关,模态数越多,测向精度越高,但算法运算复杂度越高,实时性越差。因此,为了兼顾无线电引信实时性与测向精度,可在保证引信所需测向精度的基础上,取较少的模态数,但均匀圆阵产生的涡旋电磁波最大模态受限于阵元数,即 $l_{max} \leq N/2 - 1$,模态数越多,所需阵元数越多,则发射天线阵列尺寸越大,考虑到引信的尺寸,阵元数选取能产生所需模

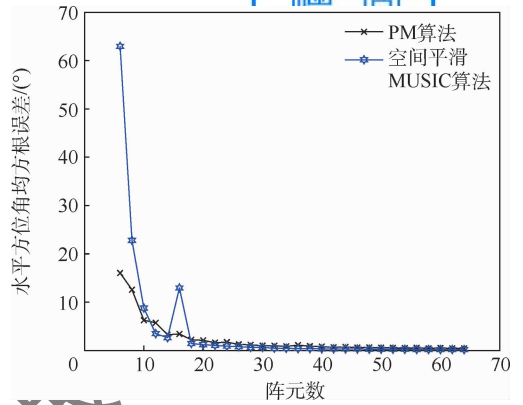


图6 不同阵元数下不同算法水平方位角估计性能对比
Fig.6 Comparison of horizontal azimuth estimation performance of different algorithms with different numbers of array elements

4 结论

本文基于 MISO 模型推导了涡旋电磁波回波方程,利用 OAM 和水平方位角的对偶关系,提出了一种具有空间超分辨率的 PM 算法,得出如下结论:

- 1) 涡旋电磁波引信的目标回波仍是发射信号的延时,但幅度受贝塞尔函数的调制,且增加 $e^{i\varphi}$ 相位项,故可在不影响测距的情况下提高测向精度。
- 2) PM 算法估计性能随信噪比和阵元数的增加而增强,在 -10 dB 的低信噪比、阵元数为 10 的情况下测向精度仍能高于 10°,实现精确测向,且与空间平滑 MUSIC 算法相对比,PM 算法的计算复杂度更低,实时性更好。

参考文献 (References)

[1] 张彤,阳世清,徐松林,等. 国外防空导弹战斗部技术研究进展[J]. 国防科技,2006(9):19-22.
ZHANG T, YANG S Q, XU S L, et al. Research progress of foreign air defense missile warhead technology[J]. National Defense Science & Technology, 2006(9):19-22 (in Chinese).

[2] 韦道知,何广军,夏训辉,等. 防空导弹无线电引信目标识别优化分析[J]. 火箭与制导学报,2009,29(1):159-161.
WEI D Z, HE G J, XIA X H, et al. Analysis on radio fuse target identification optimization of ground-to-air missile[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2009, 29(1): 159-161 (in Chinese).

[3] 李仁杰,任光亮,艾霞,等. 一发三收天线无线电引信定向探测[J]. 制导与引信,2011,32(4):4-7.
LI R J, REN G L, AI X, et al. Radio fuse directional detection to one transmitting antenna and three receiving antennae[J]. Guidance & Fuze, 2011, 32(4):4-7 (in Chinese).

[4] 张东洋,任宏滨,简金蕾. 定向探测引信技术分析[J]. 战术

- 导弹技术,2002(4):34-37.
- ZHANG D Y, REN H B, JIAN J L. Analysis of technique for directional detecting fuze [J]. Tactical Missile Technology, 2002 (4):34-37 (in Chinese).
- [5] YUAN T, CHENG Y, WANG H, et al. Mode characteristics of vortical radio wave generated by circular phased array: Theoretical and experimental results [J]. IEEE Transactions on Antennas and Propagation, 2017, 65(2):688-695.
- [6] LIU K, LI X, CHENG Y Q, et al. OAM-based multitarget detection: From theory to experiment [J]. IEEE Microwave and Wireless Components Letters, 2017, 27(8):760-762.
- [7] CHEN R, LONG W, GAO Y, et al. Orbital angular momentum-based two-dimensional super-resolution targets imaging [C] // 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Piscataway: IEEE Press, 2018:1243-1246.
- [8] ZENG Y, WANG Y, CHEN Z, et al. Two-dimensional OAM radar imaging using uniform circular antenna arrays [C] // 14th European Conference on Antennas and Propagation (EuCAP). Piscataway: IEEE Press, 2020:1-4.
- [9] YUAN T, CHENG Y, WANG H, et al. Beam steering for electromagnetic vortex imaging using uniform circular arrays [J]. IEEE Antennas and Wireless Propagation Letters, 2017, 16:704-707.
- [10] LIU K, CHENG Y, GAO Y, et al. Super-resolution radar imaging based on experimental OAM beams [J]. Applied Physics Letters, 2017, 110(16):164102.
- [11] 刘康,黎湘,王宏强,等. 涡旋电磁波及其在雷达中应用研究进展 [J]. 电子学报, 2018, 46(9):2283-2290.
- LIU K, LI X, WANG H Q, et al. The advances of vortex electromagnetic wave in radar applications [J]. Acta Electronica Sinica, 2018, 46(9):2283-2290 (in Chinese).
- [12] 郭桂蓉,胡卫东,杜小勇. 基于电磁涡旋的雷达目标成像 [J]. 国防科技大学学报, 2013, 35(6):71-76.
- GUO G R, HU W D, DU X Y. Electromagnetic vortex based radar target imaging [J]. Journal of National University of Defense Technology, 2013, 35(6):71-76 (in Chinese).
- [13] LIU K, CHENG Y, YANG Z, et al. Orbital-angular-momentum-based electromagnetic vortex imaging [J]. IEEE Antennas and Wireless Propagation Letters, 2015, 14:711-714.
- [14] LIN M Y, GAO Y, LIU P G, et al. Super-resolution orbital angular momentum based radar targets detection [J]. Electronics Letters, 2016, 52(13):1168-1170.
- [15] MARCOS S, MARSAL A, BENIDIR M. Performances analysis of the propagator method for source bearing estimation [C] // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 1994, 4:237-240.
- [16] 张小飞,张立岑,陈未央,等. MIMO 阵列中基于 PM 和降维变换的高效 DOA 估计算法 [J]. 数据采集与处理, 2014, 29(3):373-377.
- ZHANG X F, ZHANG L C, CHEN W Y, et al. Computationally efficient DOA estimation for MIMO array using propagator method and reduced-dimension transformation [J]. Journal of Data Acquisition and Processing, 2014, 29(3):373-377 (in Chinese).

PM based super-resolution method of azimuth detection for electromagnetic vortex wave fuze

SUN Lihua¹, YAN Xiaopeng^{1,*}, LIU Qiang², HAO Xinhong¹, ZHANG Hongyun¹

(1. Science and Technology on Electromechanical Dynamic Control Laboratory, Beijing Institute of Technology, Beijing 100081, China;

2. Ordnance Science Institute of China, Beijing 100089, China)

Abstract: To improve the azimuth detection accuracy of air defense missile radio fuze at low signal-to-noise ratio (SNR) background, a super-resolution method of azimuth detection for electromagnetic vortex wave fuze based on propagator method (PM) is proposed. Vortex electromagnetic waves with different orbital angular momentum modes are sequentially emitted to irradiate the target by a phase-controlled uniform circular array, and the target echo mathematical model is derived based on the multiple-input single-output (MISO) model. The received signal matrix is constructed through the approximate deformation of the echo equation. Then, PM is used to construct the spatial spectrum function, and spectral peak searching is performed to estimate the target azimuth. The effectiveness of the proposed method is verified by simulation. Results show that the proposed method can accurately detect the azimuth of a target in the case of low SNR and few elements.

Keywords: electromagnetic vortex wave; orbital angular momentum; azimuth detection of fuze; multiple-input single-output (MISO); propagator method (PM)

Received: 2021-01-14; Accepted: 2021-02-22; Published online: 2021-03-09 18:22

URL: kns.cnki.net/kcms/detail/11.2625.V.20210308.1713.001.html

Foundation items: National Natural Science Foundation of China (61973037,61871414)

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

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

DOI: 10.13700/j.bh.1001-5965.2021.0004



结构引导的图像修复

胡凯¹, 赵健^{2,*}, 刘昱¹, 牛余凯², 姬港²

(1. 天津大学 微电子学院, 天津 300072; 2. 军事科学院 系统工程研究院, 北京 100089)

摘 要:针对粗网络引入先验知识较少使得补全的内容存在明显视觉伪影问题,提出了基于边缘结构生成器的两段式图像修复方法。采用边缘结构生成器对输入的图像边缘和色彩平滑信息进行特征学习,生成缺失区域的结构内容,以引导精细网络重构高质量的语义图像。通过在公开的图像修复基准数据集 Paris Street-View 上进行实验测试,结果表明,所提模型可对掩膜占比达 50% 的图像进行补全。在客观的量化评价指标上,峰值信噪比、结构相似度系数、 L_1 和 L_2 均值误差等数值整体优于 EC、GC、SF 等方法,其中,掩膜占比为 0% ~ 20% 时,峰值信噪比指数达到 33.40 dB,优于其他方法 2.37 ~ 6.57 dB,结构相似度系数提高了 0.006 ~ 0.138。同时,补全的图像纹理更清晰,视觉质量更高。

关键词: 图像修复; 语义内容; 视觉伪影; 边缘结构生成器; 粗网络; 精细网络

中图分类号: TN911.73

文献标志码: A **文章编号:** 1001-5965(2022)07-1269-09

图像修复(也称图像补全或图像小孔填充)是一种对图像缺失区域的高维信息进行补全的技术,其使生成的图像内容语义信息合理和主观视觉真实。近年来,基于深度学习的图像修复技术快速发展,被广泛应用于文物保护、影视制作、老照片修复和图像目标移除等领域。

在基于深度学习方法的图像修复技术中,两段式网络^[1-6]是目前主要采用的修复策略,包含粗网络和精细网络 2 个阶段。其中,粗网络主要生成缺失区域所需的基本内容,并将生成的结果作为先验知识提供给精细网络,以此重构图像的语义信息并补全图像所需的纹理。然而,粗网络提供了较少有意义的先验知识^[4,6],且容易引入较多误导性信息^[2-3,5],从而造成精细网络补全的内容存在较为明显的视觉伪影。例如,Yu 等^[3]在粗网络阶段采用重构损失来约束生成器,对缺失区域的内容生成模糊纹理,然而生成的模糊纹理

包含较多的高维信息,导致精细网络在重构图像纹理时引入了较多伪影内容。Liu 等^[2]提出语义相干注意层来修复图像的破损区域,其粗网络主要是基于 U 型网络^[7](U-Net)结构,直接生成图像的缺失信息(纹理和色彩),也引入了较多的高维信息,导致补全的图像存在视觉伪影。Nazeri 等^[4]提出边缘生成器来幻想缺失区域的边缘,然而其忽略了纹理重构过程中色彩信息的重要性,导致生成的图像色彩不连续、梯度局部信息模糊。Ren 等^[6]引入结构自适应的外形流来补全图像,在粗网络阶段提出基于色彩平滑信息的结构生成器,生成图像的全局结构,虽然采用了图像的低维信息对缺失区域进行补全,但其结构真值图存在过度平滑的问题,导致引入有意义的先验知识较少,重构的图像存在语义信息不完整。

由以上分析可知,若对图像缺失区域生成高质量、语义正确的内容,则粗网络阶段输出的图像

收稿日期: 2021-01-06; 录用日期: 2021-03-05; 网络出版时间: 2021-03-16 15:27

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210316.1506.004.html

基金项目: 国家自然科学基金(62006244)

* 通信作者。E-mail: zhaojian90@u.nus.edu

引用格式: 胡凯, 赵健, 刘昱, 等. 结构引导的图像修复[J]. 北京航空航天大学学报, 2022, 48(7): 1269-1277.

HU K, ZHAO J, LIU Y, et al. Images inpainting via structure guidance[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1269-1277 (in Chinese).

信息至关重要。针对现有基于两段式网络的图像修复方法存在的视觉伪影问题,本文提出了一个新颖的基于边缘结构生成器的两段式图像修复方法。在粗网络阶段,采用边缘结构生成器对图像的边缘和色彩平滑信息进行感知与理解,为缺失区域生成更多有意义的结构特征;在精细网络阶段,基于粗网络输出的结构特征和待修复图被提取的有效像素区域特征,重构高质量的语义图像。

1 相关工作

经典的两段式图像修复方法^[4],如边缘连接模型,其网络结构如图1所示。在粗网络阶段,生成器采用编码解码器结构,输入掩膜图、待修复的边缘图和灰度图,经过2次降采样处理,由8个残

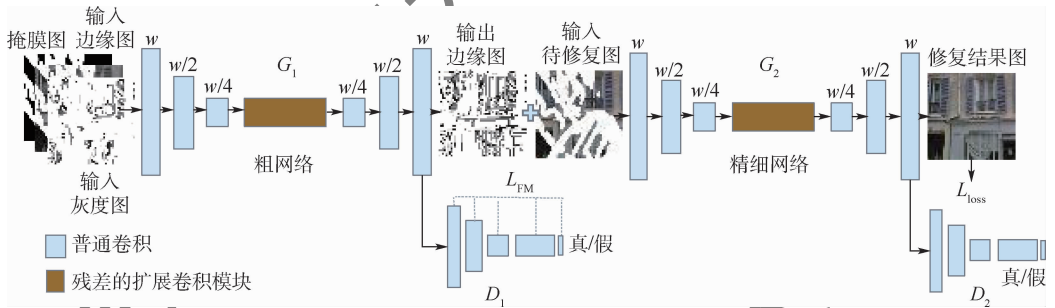


图1 边缘连接方法的两段式修复模型

Fig. 1 Two-stage restoration model of edge-connection method

差在精细网络阶段,将粗网络提供的边缘信息和待修复图像有效像素区域的信息作为先验知识,通过编码器学习缺失区域的纹理特征。编码器对图像特征的提取仍采用残差的扩展卷积模块,为了补全缺失区域的纹理信息,通过解码器重构并输出修复的图像。精细网络用于联合优化的损失函数定义为

$$L_{G_2} = \lambda_{l_1} L_{l_1} + \lambda_{adv,2} L_{adv,2} + \lambda_p L_{perc} + \lambda_s L_{style} \quad (2)$$

式中: G_2 为精细网络的生成器; λ_{l_1} 、 $\lambda_{adv,2}$ 、 λ_p 、 λ_s 为正则化参数; L_{l_1} 为修复结果图和真值图之间的绝对差值的总和; L_{perc} 和 L_{style} 分别为感知损失函数^[12]和风格损失函数^[13]; $L_{adv,2}$ 为对抗损失函数,本文采用hinge函数。

两段式网络的图像修复方法发展至今,性能已经获得了很大的提升。然而,在粗网络阶段对图像缺失区域补全的内容较少,主要是因为:其仅对图像中的高级信息(边缘)或低级信息(平滑的色彩)进行学习,引入单一的内容,使得精细网络缺少图像重构需要的相关特征,导致重构的语义图像出现错误,引发了图像的视觉伪影。

差的扩展卷积模块^[8](每个残差的扩展卷积模块由一个扩展因子为2的扩展卷积层组成)提取图像缺失区域的边缘信息,并由解码器对提取的特征进行重构,训练过程中采用特征匹配损失 L_{FM} ^[9]来约束网络生成边缘信息,最终输出补全的边缘图,为精细网络提供缺失的边缘信息;鉴别器采用谱归一化的块生成对抗网络(spectral-normalization PatchGAN, SN-PatchGAN)^[10]。粗网络阶段训练的联合损失优化函数为

$$\min_{G_1} \max_{D_1} L_{G_1} = \min_{G_1} (\lambda_{adv,1} \max_{D_1} (L_{adv,1}) + \lambda_{FM} L_{FM}) \quad (1)$$

式中: G_1 和 D_1 分别为网络的生成器和鉴别器; $L_{adv,1}$ 为对抗损失函数,本文中采用hinge函数^[11]; $\lambda_{adv,1}$ 、 λ_{FM} 为正则化参数。

2 本文方法

针对第1节所述图像修复方法所存在的视觉伪影问题,本文提出了一种新颖的基于边缘结构生成器的两段式图像修复方法,如图2所示,主要由粗网络和精细网络两部分组成。在粗网络阶段,基于边缘和色彩平滑信息的互补特性,设计了一个边缘结构生成器网络,对输入图内有效像素区域融合的边缘和色彩平滑特征充分学习,生成高质量的全局结构图,为精细网络引入更多有意义的信息;在精细网络阶段,纹理重构器参考文献[3,5,14]的网络结构,引入门卷积,仅对全局图像的有效特征进行提取,减少了无效特征的引入;同时网络中还引入上下文注意模块来构造编码器子网络,并利用其构造的子网络和扩展卷积层的子网络来提取图像特征,用于重构缺失区域的纹理。

图2中提出的两段式网络皆采用生成对抗网络(generative adversarial network, GAN)^[11]模型,对此本文分别用 G_1 和 D_1 来命名粗网络的生成器和鉴别器,用 G_2 和 D_2 来命名精细网络的生成器和鉴别器。

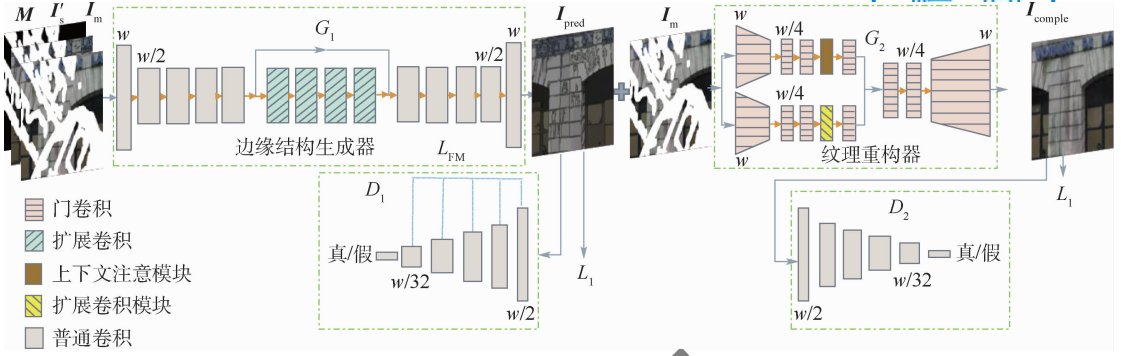


图 2 本文提出的图像修复模型
Fig. 2 Proposed image inpainting model

2.1 结构图的求解

图像的内容由纹理和结构信息组成,为了移除其中的高频信息,保留边缘和低频结构,本文采用基于相对总变差模型 (relative total variation, RTV) 的算法^[15-16]对图像的纹理和结构特征进行分解,输出结构图。该算法定义如下:

$$\arg \min_S \sum_p (S_p - I_p)^2 + \lambda \left(\frac{\partial_x(p)}{\psi_x(p) + \varepsilon} + \frac{\partial_y(p)}{\psi_y(p) + \varepsilon} \right) \quad (3)$$

$$\partial_x(p) = \sum_{q \in R(p)} g_{p,q} \cdot |(\partial_x S)_q| \quad (4)$$

$$\partial_y(p) = \sum_{q \in R(p)} g_{p,q} \cdot |(\partial_y S)_q| \quad (5)$$

$$\psi_x(p) = \left| \sum_{q \in R(p)} g_{p,q} \cdot (\partial_x S)_q \right| \quad (6)$$

$$\psi_y(p) = \left| \sum_{q \in R(p)} g_{p,q} \cdot (\partial_y S)_q \right| \quad (7)$$

其中: S 为输出的结构图, I 为输入图像, S_p 为 S 内 p 点的像素值, I_p 为 I 内 p 点的像素值; $\partial_x S$ 和 $\partial_y S$ 分别为 S 在 x 和 y 方向的偏导数; ε 为无限小的正数,可避免除数为0; λ 为权值参数; $q \in R(p)$, R 为正方形的像素区域, q 属于 R 内的其中一个像素值, p 为 R 区域的中心点; $g_{p,q}$ 为权值函数,并被定义为

$$g_{p,q} \propto \exp \left[- \frac{(x_p - x_q)^2 + (y_p - y_q)^2}{2\sigma^2} \right] \quad (8)$$

式中: σ 为标准差。

式(3)主要由非线性项和二次项组成,其中 x 方向的二次项可分解为

$$\sum_p \frac{\partial_x(p)}{\psi_x(p) + \varepsilon} = \sum_q \sum_{p \in R(q)} \frac{g_{p,q}}{\left| \sum_{q \in R(p)} g_{p,q} \cdot (\partial_x S)_q \right| + \varepsilon} |(\partial_x S)_q| \approx \sum_q \sum_{p \in R(q)} \frac{g_{p,q}}{\psi_x(p) + \varepsilon} \cdot \frac{1}{|(\partial_x S)_q| + \varepsilon_s} (\partial_x S)_q^2 =$$

$$\sum_q u_{x,q} w_{x,q} (\partial_x S)_q^2 \quad (9)$$

式中: $(\partial_x S)_q^2$ 为 x 方向的二次偏导数; ε_s 为无限小的正数。

重新构造二次项的非线性部分 $u_{x,q} w_{x,q}$,分别表示为

$$u_{x,q} = \sum_{p \in R(q)} \frac{g_{p,q}}{\psi_x(p) + \varepsilon} = \left(G_\sigma * \frac{1}{|G_\sigma * \partial_x S| + \varepsilon} \right)_q \quad (10)$$

$$w_{x,q} = \frac{1}{|(\partial_x S)_q| + \varepsilon_s} \quad (11)$$

式中: G_σ 为标准差 σ 的高斯核函数;“*”为卷积符号。

同理, y 方向的二次项分解为

$$\sum_q \frac{\partial_y(p)}{\psi_y(p) + \varepsilon} = \sum_q u_{y,q} w_{y,q} (\partial_y S)_q^2 \quad (12)$$

式中: $(\partial_y S)_q^2$ 为 y 方向的二次偏导数。

对非线性部分 $u_{y,q} w_{y,q}$ 重新进行构造,分别表示为

$$u_{y,q} = \left(G_\sigma * \frac{1}{|G_\sigma * \partial_y S| + \varepsilon} \right)_q \quad (13)$$

$$w_{y,q} = \frac{1}{|(\partial_y S)_q| + \varepsilon_s} \quad (14)$$

将上述算子代入,式(3)中的RTV模型可用如下矩阵形式表示为

$$(\mathbf{v}_s - \mathbf{v}_l)^T (\mathbf{v}_s - \mathbf{v}_l) + \lambda (\mathbf{v}_s^T \mathbf{C}_x^T \mathbf{U}_x \mathbf{W}_x \mathbf{C}_x \mathbf{v}_s + \mathbf{v}_s^T \mathbf{C}_y^T \mathbf{U}_y \mathbf{W}_y \mathbf{C}_y \mathbf{v}_s) \quad (15)$$

式中: \mathbf{v}_s 和 \mathbf{v}_l 分别为图像 S 和 I 的2个列矢量; \mathbf{C}_x 和 \mathbf{C}_y 分别为 x 方向和 y 方向的前向差分梯度算子的特普利兹矩阵; \mathbf{U}_x 、 \mathbf{U}_y 、 \mathbf{W}_x 、 \mathbf{W}_y 为对角矩阵,对角上的值分别为 $\mathbf{U}_x[i, i] = u_{x,i}$, $\mathbf{U}_y[i, i] = u_{y,i}$, $\mathbf{W}_x[i, i] = w_{x,i}$, $\mathbf{W}_y[i, i] = w_{y,i}$ 。

对式(15)求导可得

$$(\mathbf{1} + \lambda \mathbf{L}') \cdot \mathbf{v}_s^{t+1} = \mathbf{v}_l \quad (16)$$

式中: $t \in [0, 2]$; $\mathbf{L}' = \mathbf{C}_x^T \mathbf{U}_x \mathbf{W}_x \mathbf{C}_x + \mathbf{C}_y^T \mathbf{U}_y \mathbf{W}_y \mathbf{C}_y$,为用于计算结构向量 \mathbf{v}_s^t 的权值矩阵; $\mathbf{1} + \lambda \mathbf{L}'$ 为对

称的正定拉普拉斯矩阵, $\mathbf{1}$ 为单位矩阵。

通过求解上述线性方程, 即可输出结构图 S 。

2.2 边缘结构生成器

本文的边缘结构生成器模型主要基于两层的编码解码器结构。由于缺失区域的面积一般较大, 中心区域的内容不能通过卷积层提取到有效的特征, 采用一个残差的多因子扩展卷积模块来提取图像的边缘结构特征。所采用的扩展因子分别为 2、4、8 和 16, 扩展卷积模块^[17]通过级联多层卷积来逐步增大卷积核的感受野, 从而提取到特征图内较远区域的特征, 为缺失区域提供更多有效的结构信息。

通过实验验证, 发现像素为 128×128 的特征图仍能较好地保留图像的纹理和色彩信息, 因此仅对编码器做一次卷积降采样。同时, 为了避免网络引入图像的无效特征, 在编码解码器结构中间的连接部分, 本文未采用跳跃连接。针对边缘结构生成器重构的语义信息问题, 本文权衡了输入的高维和低维特征, 对高斯平滑滤波器的标准差 σ 进行调节, 以此控制 Canny 边缘检测器的灵敏度, 输出最适合的边缘信息。

网络的输入由掩膜图 M ^[18]、破损的待修复图 I_m 和边缘结构图 I_s 在通道上拼接而成。其中, 掩膜图 M 为掩膜形状不规则的黑白图(像素值被归一化为 0 和 1), 如图 3 所示。待修复图 $I_m = I_g \odot (1 - M)$, I_g 为真值图, 破损的边缘结构图 I_s' 由图像的边缘和色彩平滑信息组成:

$$I_s' = [S \odot (1 - M)] \odot [I_e \odot (1 - M)] \quad (17)$$

式中: \odot 为 Hadamard 积; I_e 为图像边缘图, 由 Canny 边缘检测算法^[19-22]计算而得, 根据实践经验 $\sigma \approx 2$; S 为色彩梯度平滑结构图, 由 2.1 节边缘保留的 RTV 算法计算得到, 其 σ 值被设置为 3, 则输入的边缘结构真值图为 $I_s = S \odot I_e$ 。



图3 输入的掩膜样本

Fig. 3 Input mask samples

生成器输出的结构预测图定义为

$$I_{\text{pred}} = G_1(I_m, I_s', M) \quad (18)$$

生成器的生成损失主要由 $L_{I_1}^1$ 损失函数和 L_{FM}^1 特征匹配损失函数^[9]构成, 其分别表示为

$$L_{I_1}^1 = \|I_s - I_{\text{pred}}\|_1 \quad (19)$$

$$L_{FM}^1 = E \left[\sum_i \frac{1}{N_i} \|D_1^{(i)}(I_s) - D_1^{(i)}(I_{\text{pred}})\|_1 \right] \quad (20)$$

式中: N_i 为判别器模型在第 i 层的特征通道数目; 上标“1”表示粗网络阶段; $L_{I_1}^1$ 损失函数主要用来度量 I_{pred} 和 I_s 之间的距离, 可直接求出两图像之间的内容误差, 包括纹理和边缘结构信息; L_{FM}^1 损失函数为求解 I_{pred} 和 I_s 之间基于 SN-PatchGAN 鉴别器网络输出的每一层激活特征图损失的累加和均值, 主要用于约束生成器的生成过程, 使其生成的边缘特征更接近真值。

为了更好地鉴别生成器输出结果在不同位置不同语义内容间的真假, 同时也为了使训练更快更稳定, 本文采用 SN-PatchGAN 为鉴别器模型。鉴别器的对抗损失函数为最小二乘损失^[17], 其定义为

$$L_{\text{adv}}^1 = \frac{1}{2} E_{I_g} [(D_1(I_s, M) - 1)^2] + \frac{1}{2} E_{I_{\text{pred}}} [D_1(G_1(I_m, I_s', M), M)^2] \quad (21)$$

为了联合训练生成器和鉴别器, 本文提出一个全局损失的优化函数, 其具体形式为

$$\min_{G_1} \max_{D_1} L^1(G_1, D_1) = \lambda_{L_{\text{adv}}}^1 L_{\text{adv}}^1 + \lambda_{L_{I_1}}^1 L_{I_1}^1 + \lambda_{L_{FM}}^1 L_{FM}^1 \quad (22)$$

式中: $\lambda_{L_{\text{adv}}}^1$ 、 $\lambda_{L_{I_1}}^1$ 、 $\lambda_{L_{FM}}^1$ 为正则化参数, 根据实践经验将其分别设置为 $\lambda_{L_{\text{adv}}}^1 = 1$, $\lambda_{L_{I_1}}^1 = 4$, $\lambda_{L_{FM}}^1 = 4$ 。

2.3 纹理重构器

在精细网络阶段, 纹理重构器将输入的待修复图和粗网络输出的边缘结构图 I_{pred} 作为其重构的先验知识, 其中缺失区域的内容由结构信息补充, 由此可得精细网络的输入图为

$$I_{\text{in}} = I_m + I_{\text{pred}} \odot M \quad (23)$$

采用编码解码器作为纹理重构器的网络模型, 该模型基于全卷积的网络结构, 包括降采样和上采样过程。其中网络卷积的过程皆采用掩膜可进行学习和自动更新的门卷积策略^[5], 门卷积整体结构主要包含门控机制和卷积两部分。在图像修复过程中, 门控机制主要采用可以更新门阈值(掩膜)的软门策略, 利用卷积层对掩膜进行特征学习, 使得掩膜特征图可以随着卷积层的变化而保持更新。

纹理重构器依赖并行的上下文注意力网络^[3]和多因子扩展卷积子网络^[17]来提取图像有效区域内的纹理和色彩特征。其中, 上下文注意力网络采用上下文注意模块, 其基于特征匹配过程可学习的特性, 主要将图像缺失区域的内容与输入的先验知识建立长依赖, 利用图像上下文的特征来进行匹配填充; 多因子扩展卷积子网络采用的扩展因子与边缘结构生成器相同, 主要用于捕获特征图多尺度的上下文信息, 能在特征图分辨率保持不变的情况下扩大感受野, 且不引入额外的参数, 解决了图像缺失部分的中心区域不能

提取到有效特征的问题。将编码器提取的特征进行通道拼接,以实现同层特征图在通道层面的信息融合,并将融合后的特征通过反卷积上采样的方式进行语义内容重构。同时结合训练的样本数据对图像的特征分布进行学习,使得补全的纹理具有多样性。最终对图像补全的过程可表示为

$$I_{\text{comple}} = G_2(I_{\text{in}}) \quad (24)$$

纹理重构器模型通过联合 $L_{l_1}^2$ 损失函数和 L_{adv}^2 对抗损失函数进行训练,即

$$\min_{G_2} \max_{D_2} L^2(G_2, D_2) = \lambda_{L_{\text{adv}}}^2 L_{\text{adv}}^2 + \lambda_{l_1}^2 L_{l_1}^2 \quad (25)$$

式中:正则化参数 $\lambda_{L_{\text{adv}}}^2 = 1, \lambda_{l_1}^2 = 4$, 上标“2”表示精细网络阶段。

$L_{l_1}^2$ 损失函数定义为

$$L_{l_1}^2 = \|I_{\text{gt}} - I_{\text{comple}}\|_1 \quad (26)$$

式中: I_{gt} 为真值图。

采用 SN-PatchGAN 鉴别器模型,对抗损失函数采用 hinge 损失,其具体形式如下:

$$L_{\text{adv}}^2 = E_{I_{\text{gt}}}[\text{ReLU}(1 - D_2(I_{\text{gt}}, M))] + E_{I_{\text{comple}}}[\text{ReLU}(1 + D_2(G_2(I_{\text{in}}), M))] \quad (27)$$

式中:ReLU 为激活函数。

3 实验与结果分析

本文基于公开的基准数据集 Paris Street-View 对所提的模型进行评估,其中 Paris 数据集包含 14 900 张训练图像和 100 张测试图像,图像大小为 256 像素 \times 256 像素。实验过程中,为了更好地验证该模型的性能,每张图像所需的掩膜图从 PConv^[18] 公开的掩膜测试集中随机选取,该测试集内包含 12 000 张不规则的掩膜图,并且每张图像内的掩膜占比不同。同时,为了使输入的掩膜图像与待修复的图像大小相同,其像素被调

整为 256 \times 256。本文模型训练测试的系统环境如下:Ubuntu 18.04, Tensorflow v1.8, CUDA v9.0, CUDNN v7.0, python 3.6.2。电脑硬件配置为:1 块 Intel(R) Xeon(R) E5-2680 v4 @ 2.40 GHz CPU, 2 块 NVIDIA(R) 2080TI GPU(11 GB)的显卡。

网络训练阶段,每批次输入 8 张图像,模型采用自适应矩估计(adaptive moment estimation, Adam)优化器^[23]($\beta_1 = 0.5, \beta_2 = 0.999$)来优化损失函数。初始过程,分别对粗网络和精细网络阶段的边缘结构生成器进行训练,其中 2 个生成器 G_1, G_2 的学习率分别被设置为 1×10^{-4} 。最终移除鉴别器,将两段式网络的生成器 G_1, G_2 进行端到端的联合微调训练,采用的学习率为 1×10^{-5} 。

3.1 定量对比分析

为了准确反映本文模型的性能,采用以下量化评价指标:峰值信噪比(peak signal to noise ratio, PSNR)、结构相似度系数(structural similarity index measurement, SSIM)、 L_1 均值误差和 L_2 均值误差,PSNR 和 SSIM 数值越大性能越好,对应的, L_1 和 L_2 数值越小性能越好。为使量化指标更好地反映图像存在的伪影,也使数值矩阵显示的结果更真实,本文中假想图像的像素彼此无关联。实验测试过程中,将本文方法在数据集 Paris Street-View 上与目前图像修复中性能较好的上下文注意^[3](contextual attention, CA)、结构流^[6](StructureFlow, SF)、边缘连接^[4](EdgeConnect, EC)和门卷积^[5](GatedConv, GC)等方法进行对比。所有方法模型均采用 100 张测试图进行验证,使用的掩膜图像含有不同的掩膜占比,如 0% ~ 20%、20% ~ 40%、40% ~ 60%,并用其分别对所有方法进行测试。

由表 1 可知,本文的量化评价指数整体优于

表 1 不同方法在数据集 Paris Street-View 上的对比结果

Table 1 Comparison results of different approaches over Paris Street-View dataset

方法	掩膜占比/%	PSNR/dB	SSIM	L_1 /%	L_2 /%
SF	0 ~ 20	28.71	0.953	2.49	2.07
	20 ~ 40	25.41	0.895	3.96	3.65
	40 ~ 60	22.32	0.756	5.26	5.03
EC	0 ~ 20	31.03	0.963	2.12	1.21
	20 ~ 40	26.07	0.876	3.43	3.04
	40 ~ 60	23.45	0.721	6.33	5.89
GC	0 ~ 20	28.26	0.932	3.30	2.30
	20 ~ 40	24.83	0.821	4.21	4.26
	40 ~ 60	22.61	0.650	7.21	6.13
CA	0 ~ 20	26.83	0.831	8.12	3.84
	20 ~ 40	23.81	0.694	10.2	5.49
	40 ~ 60	20.26	0.535	11.32	7.81
本文方法	0 ~ 20	33.40	0.969	1.31	1.07
	20 ~ 40	28.65	0.883	3.10	3.01
	40 ~ 60	24.51	0.762	5.09	4.75

其他方法。其中,当掩膜占比为 0% ~ 20% 时, PSNR 指数达到 33.40 dB,相比其他方法提高了 2.37 ~ 6.57 dB;SSIM 指数为 0.969,提高了 0.006 ~ 0.138,修复图和真值图间的 L_1 误差为 1.31%,差值较小,近乎接近于真值图。当掩膜占比为 20% ~ 40% 时,SF 方法的 SSIM 指数最优,主要是由于 SF 方法在粗网络阶段采用图像的低维结构信息进行补全,使得图像重构过程中易生成缺失区域的纹理,然而该方法由于输入的图像缺失高维信息,边缘特征较少,使得输出的修复结果整体性能稍弱。同时,由表 1 可知,本文模型能对掩膜率 50% 以上的图像进行修复,并且修复的图像质量是目前最优。

3.2 定性对比分析

为了更形象直观地展示本文方法的修复效果,对本文方法与其他方法的修复结果进行可视化对比,结果如图 4 和图 5 所示。在图 4 中,将本文方法与 EC、SF 方法在粗网络阶段的输出结果进行可视化对比。可知,EC 方法在粗网络阶段输出的边缘图出现结构信息不完整,边缘内容缺失,同时其输出的图像为仅含 0 和 1 像素的黑白图,缺少图像重构所必需的色彩信息,易导致生成的图像出现语义内容错误;SF 方法在粗网络阶段输出的结构图色彩信息过平滑,缺失较多边缘结构

信息,尤其在结构相似的区域,易生成语义错误的内容,同时还伴随着明显的视觉伪影,影响了生成的图像质量。本文提出的边缘结构生成器可以对缺失区域生成复杂多样的结构信息,包含边缘和低维的色彩特征,输出图无伪影。



图 4 不同方法在粗网络阶段时输出结果的可视化对比
Fig. 4 Visual comparison of output results for different methods in the coarse network stage



图 5 不同方法在数据集 Paris Street-View 上的可视化对比结果

Fig. 5 Comparison of visual results of different approaches in Paris Street-View dataset

在图 5 中,将本文方法与 CA、SF、EC、GC 方法一同进行测试,并将测试结果做可视化输出,发现不同方法修复的结果存在较大差异。通过用矩形框对图中部分不相同区域的内容进行标注,可以看出,本文方法可以对细节特征较为微弱的区域进行修复,同时修复的结果更清晰,伪影较少,更符合人眼的视觉观感。相比之下,采用 CA 方法输出的图像视觉伪影较为明显,可观性不强,同时其输出图像的语义内容不完整;采用 SF 方法虽然可以对图像缺失内容较多的区域进行修复,但其修复的结果还遗留掩膜的残影,输出的修复图存在模糊、色彩差异等问题;采用 EC 方法也可以对缺失内容较多的区域进行修复,然而其修复的图像存在模糊、纹理不清晰、色彩信息缺失和语义内容不完整等问题;采用 GC 方法修复的图像也存在明显的视觉伪影,部分细节区域的内容重构不完整。

综上所述,与其他方法相比,本文方法的图像修复结果为当前最优,细节的修复效果较为显著,整体在视觉观感上更清晰、真实,语义内容更完整,同时无视觉伪影产生。

3.3 扩展实验

为了验证输入不同的结构信息对图像修复效果的影响,本文使用 Canny 边缘检测器和 RTV 模型分别获取测试图像的边缘图和色彩平滑图。同时,测试图像所需的掩膜图从掩膜测试集中随机

选取,并将计算得到的待修复图像作为修复网络模型的输入,以此开展网络训练,再通过数据集 Paris Street-View 对训练好的模型分别进行测试,并采用量化评价指标对测试的结果进行直观展示,如表 2 所示。

由表 2 可知,通过本文提出的两段式修复模型,将图像的边缘和色彩平滑信息联合输入,得到的量化评估指数整体优于边缘信息或色彩平滑信息的单独使用,基于两者的互补特性可以重构更高质量的语义图像。同时,为了更直观地展示图像边缘和色彩平滑信息的联合使用带来的图像修复效果,本文将其输出结果采用了可视化的方式,如图 6 所示。与本文两者信息皆采用的输入方式相比,输入图像的边缘信息得到的输出结果图,虽然整体的边缘结构存在,然而细节模糊,缺失色彩信息,生成的图像存在部分错误的语义结构;同时,针对输入图像的色彩平滑信息得到的修复图,整体结构缺乏完整性,缺失区域被修复后仍留下

表 2 不同输入信息通过网络测试输出的修复结果
Table 2 Image inpainting results of different input information outputs through network test

结构信息	PSNR/dB	SSIM
边缘信息	28.25	0.842
色彩平滑信息	24.637	0.767
两者皆有(本文)	30.23	0.971



图 6 输入不同结构信息得到的可视化修复结果

Fig. 6 Visual restoration results obtained from different input structure information

掩膜的残影,图像的视觉伪影较为明显,影响了图像的整体视觉观感。

4 结论

本文提出了一个新颖的基于边缘结构生成器的两段式图像修复方法,基于融合的图像边缘和色彩平滑信息采用边缘结构生成器补全缺失区域的结构内容,针对粗网络输出的全局结构图采用纹理重构器生成语义内容完整的图像。通过数据集 Paris Street-View 对本文方法进行实验论证,结果如下:

1) 本文方法在量化的评价指标 PSNR、SSIM、 L_1 和 L_2 上,整体数值优于其他两段式的图像修复方法,并可对掩膜占比 50% 的图像进行修复。其中,当掩膜占比为 0% ~ 20% 时,PSNR 指数达到 33.40 dB,SSIM 指数为 0.969。

2) 相较于其他方法,本文修复的图像在视觉上更真实,纹理更清晰,语义内容更完整,并无图像伪影产生。

3) 对本文方法输入的图像结构和色彩平滑信息进行实验对比,发现输入两者信息融合后的图像,可以为缺失区域引入更多有意义的特征,使得重构完成的图像质量更高,细节部分的纹理更完整清晰。

参考文献 (References)

- [1] SHAO H, WANG Y X, FU Y H, et al. Generative image inpainting via edge structure and color aware fusion [J]. *Signal Processing: Image Communication*, 2020, 87: 115929.
- [2] LIU H Y, JIANG B, XIAO Y, et al. Coherent semantic attention for image inpainting [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 4169-4178.
- [3] YU J H, LIN Z, YANG J M, et al. Generative image inpainting with contextual attention [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5505-5514.
- [4] NAZERI K, NG E, JOSEPH T, et al. EdgeConnect: Generative image inpainting with adversarial edge learning [EB/OL]. (2019-01-01) [2020-12-31]. <https://arxiv.org/abs/1901.00212>.
- [5] YU J H, LIN Z, YANG J M, et al. Free-form image inpainting with gated convolution [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 4471-4480.
- [6] REN Y R, YU X M, ZHANG R N, et al. StructureFlow: Image inpainting via structure-aware appearance flow [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 181-190.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-assisted Intervention. Berlin: Springer, 2015: 234-241.
- [8] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 472-480.
- [9] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8798-8807.
- [10] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [EB/OL]. (2018-02-16) [2020-12-31]. <https://arxiv.org/abs/1802.05957>.
- [11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C] // Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014: 2672-2680.
- [12] GATYS L A, ECKER A S, BETHGS M. Image style transfer using convolutional neural networks [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2414-2423.
- [13] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution [C] // European Conference on Computer Vision. Berlin: Springer, 2016: 694-711.
- [14] SONG Y, YANG C, LIN Z X, et al. Contextual-based image inpainting: Infer, match, and translate [C] // European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.
- [15] XU L, LU C, XU Y, et al. Image smoothing via L_0 gradient minimization [J]. *ACM Transactions on Graphics*, 2011, 30 (6): 174.
- [16] XU L, YAN Q, XIA Y, et al. Structure extraction from texture via relative total variation [J]. *ACM Transactions on Graphics*, 2012, 31 (6): 139.
- [17] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2015-11-23) [2020-12-31]. <https://arxiv.org/abs/1511.07122>.
- [18] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C] // European Conference on Computer Vision. Berlin: Springer, 2018: 85-100.
- [19] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks [C] // 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2794-2802.
- [20] LIU Y, CHENG M M, HU X, et al. Richer convolutional features for edge detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5872-5881.
- [21] CANNY J. A computational approach to edge detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, PAMI-8 (6): 679-698.
- [22] DOLLAR P, TU Z, BELONGIE S. Supervised learning of edges and object boundaries [C] // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2006: 1964-1971.
- [23] KINGMA D P, BA J. Adam: A method for stochastic optimization [EB/OL]. (2014-12-22) [2020-12-31]. <https://arxiv.org/abs/1412.6980>.

Images inpainting via structure guidance

HU Kai¹, ZHAO Jian^{2,*}, LIU Yu¹, NIU Yukai², JI Gang²

(1. School of Microelectronics, Tianjin University, Tianjin 300072, China;

2. Institute of Systems Engineering, Academy of Military Sciences, Beijing 100089, China)

Abstract: Aiming at the problem of obvious visual artifacts in the content of rough network with less prior knowledge, a two-stage image inpainting method based on an edge structure generator is proposed. The edge structure generator is used to perform feature learning on the input image edge and color smoothing information, and generate the missing structural contents so as to guide the fine network to reconstruct high-quality semantic images. The mentioned method has been tested on the public benchmark datasets such as Paris Street-View. The experimental results show that the proposed approach can complete the hole images with the mask rate of 50%. The quantitative evaluation indicators: PSNR, SSIM, L_1 and L_2 errors respectively surpass current images inpainting algorithms with excellent performance, such as EC, GC, SF, etc. Among them, when the mask rate is 0% - 20%, the PSNR index reaches 33.40 dB, which is an increase of 2.37 - 6.57 dB compared to other methods; the SSIM index is increased by 0.006 - 0.138. Meanwhile, the completed images get clearer texture and higher visual quality.

Keywords: image inpainting; semantic contents; visual artifacts; edge structure generator; coarse network; fine network

Received: 2021-01-06; **Accepted:** 2021-03-05; **Published online:** 2021-03-16 15:27

URL: kns.cnki.net/kcms/detail/11.2625.V.20210316.1506.004.html

Foundation item: National Natural Science Foundation of China (62006244)

* **Corresponding author.** E-mail: zhaojian90@u.nus.edu

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

DOI: 10.13700/j.bh.1001-5965.2021.0014

基于 SVC 视频流的低复杂度多播组分解算法



杨静璇, 徐桢*

(北京航空航天大学 电子信息工程学院, 北京 100083)

摘 要: 多播是一种高效率利用带宽资源的技术,可以有效缓解多媒体传输过程中的带宽压力,但传统的多播技术会带来“瓶颈用户”问题,限制多播组内用户的数据速率。多播组分解技术将多播组划分为若干子组并以不同速率接收数据,可以有效解决瓶颈用户带来的速率限制。构建了面向用户端的视频多播传输方案,将可伸缩视频编码(SVC)的分层特点和组分解技术相结合,各多播子组根据实际接收能力解调得到不同质量的 SVC 视频数据,在保证用户基本视频数据传输的基础上,实现总系统速率最大化。提出了面向资源公平调配的低复杂度多播组分解算法,在改进低复杂度分组(LCS)算法过程中考虑 SVC 视频层限制,并引入常值向量抑制资源分配不公的情况。经过实验数据模拟和性能评估,所提算法在带宽资源和用户数量变化时,均可以稳定地保持较高的系统速率、频谱效率及系统公平性,且计算复杂度较低,能够实际应用于 4G 和 5G 网络架构下的视频传输。

关键词: 可伸缩视频编码(SVC); 多速率多播; 资源分配; 无线网络; 视频流

中图分类号: TN92; TN919.8

文献标志码: A

文章编号: 1001-5965(2022)07-1278-09

移动通信网络技术的不断革新,大幅度提高了系统的频谱效率、传输可靠性和终端接入能力,用户端的业务需求也随之越来越多样化,逐渐从语音文本服务转变为视频流服务,增强型多媒体广播多播业务(e-MBMS)的需求也与日剧增^[1]。在增强移动带宽(eMBB)场景中,高质量多媒体业务的传输与分发需要网络提供更高的数据速率、更大的数据容量及更为严格的时延特性^[2]。然而,用户在复杂多变的网络环境中流畅地接收质量满意的视频仍是一个巨大的挑战。

无线多播技术利用无线信道的广播特性,使基站能够同时给多个用户传输相同的数据,有效提高基站的传输效率^[3],并且无线多播高效利用现有资源的特点,尤其适用于带宽压力大的多媒体数据传输。但是,传统的多播算法会带来“瓶

颈用户”问题,因为多播组中的用户信道衰落情况不同,如果以较高的速率传输数据,信道条件差的用户则无法正确解调出信号,所以传统多播算法为了保证所有用户均能接收数据,基站往往会选择较低信息速率的调制编码方式,但是也会限制信道条件较好用户的服务质量。为了解决此类“瓶颈用户”问题,科研人员提出了多速率多播算法^[4],允许多播组内用户根据自身处理能力接收不同速率的多播数据流,从而缓解“瓶颈用户”带来的影响。

多速率多播算法主要分为流分解和组分解两大类。组分解技术是将多播组根据准则分解成多个子组,再根据子组内用户的信道质量确定各子组的传输速率。文献[5]提出了一种自适应多播算法,将多播组划分为更小的子组,并为每个子组分配子载波来最大化总数据速率,相比于传统多

收稿日期: 2021-01-11; 录用日期: 2021-03-26; 网络出版时间: 2021-04-09 09:20

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210408.1435.001.html

基金项目: 国家自然科学基金(91638301,91738301)

* 通信作者. E-mail: xuzhen@buaa.edu.cn

引用格式: 杨静璇, 徐桢. 基于 SVC 视频流的低复杂度多播组分解算法[J]. 北京航空航天大学学报, 2022, 48(7): 1278-1286.
YANG J X, XU Z. Low computational-cost multicast subgrouping for SVC streams[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1278-1286 (in Chinese).

播算法,该算法有效提高了系统速率和资源利用率。文献[6]提出了一种基于星地网络的协作多播组分解算法,在保证可靠传输的同时提高了系统吞吐量,由于组分解算法中子组划分和资源分配的计算复杂度都较高,很难直接应用于实际系统中。Araniti等^[7]提出了低复杂度子组(LCS)算法,能够以较低的计算量得到接近最佳性能的子组划分方案,但是设计目标是最大化系统吞吐量,忽略了用户角度的公平性和服务质量需求。组分解算法还可以和其他技术相结合进一步优化通信传输性能,具有良好的应用价值。Araniti等^[8]将划分子组的思想运用到应用层编码技术中,优化了应用层编码的性能,从而提高了数据在不同信道中传输的稳定性和吞吐量。

多媒体数据的多播传输常采用流分解技术,其可以将高速率数据流分解成多个低速子数据流,用户根据信道条件接收不同的子数据流。可伸缩视频编码(SVC)技术运用了流分解的思想,其将视频流编码成一个基本视频层(basic layer, BL)和多个增强视频层(enhancement layer, EL),BL层解码后的视频帧率和清晰度都较低^[9],但是可以基本满足用户的最低要求,EL层可以补充和完善BL层的视频信息。文献[10]分析了SVC视频数据在移动通信网络中的多播传输,提出了高效利用资源的多播视频传输算法,但是组分解过程的计算复杂度较高。文献[11]提出了卫星多播子组分层算法,通过创建子组和视频分层的方式实现多播传输,可以有效提高资源利用率和数据速率,但是没有考虑多播组内资源的公平性。

本文研究了如何在地面移动网络的架构下构建面向用户端的视频多播传输方案。首先,根据用户信道条件将多播组划分为多个子组;然后,各子组根据实际接收能力解调得到不同质量的SVC视频数据,在保证用户基本视频数据传输的基础上,实现系统速率最大化。同时,提出了满足用户需求和系统资源公平的低复杂度多播资源分配算法,在改进LCS算法过程中考虑SVC视频层限制,并引入常值向量抑制资源分配不公的情况。

1 视频多播传输方案

1.1 子组划分模式

假设视频多播服务于地面移动网络,主要由基站和移动用户构成。与之相对应的航空通信网

络包括无人机和飞行器等节点,具有移动速度快、范围大的特点,如果视频多播服务于航空通信网络,则需要着重考虑信息传输的有效性和连续性。如图1所示,视频数据来源于因特网,经由地面基站多播到各用户,由于多播组内用户信道质量不同,需要将多播组划分为多个子组。多播传输资源分配的过程是将时频资源划分为多个资源块(RB),并根据信道质量指标(CQI)动态地调整调制编码方案(MCS),CQI值越大,MCS调制阶数越高^[12]。

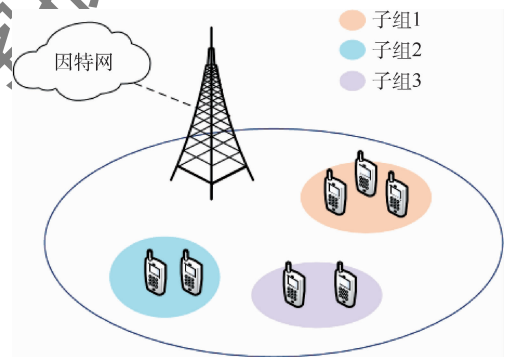


图1 系统模型

Fig. 1 System model

多播视频流是在小区内发送的,小区内 K 个用户构成一个多播组,在进行多播传输时,不同用户的CQI值用 $m(1 \leq m \leq M, m$ 为整数)来表示, M 为CQI的最大值。

基站首先根据用户反馈的CQI值将多播组分解成 M 个子组元,具有相同CQI值的用户在一个子组元中。各子组元的用户数由向量 $\mathbf{U} = [u_1, u_2, \dots, u_m, \dots, u_M]$ 来表示, u_m 为CQI值为 m 的用户数目。基站为每个子组元分配信道资源,用向量 $\mathbf{R} = [r_1, r_2, \dots, r_m, \dots, r_M]$ 来表示, r_m 为CQI值为 m 的子组元中的RB数目,如果 $r_m \neq 0$,则代表子组元是有效存在的。

随着通信系统的不断发展,MCS调制阶数越来越高,组分解最初划分的子组元数目也不断增加,而组分解方案需要为各子组元分配资源,因此会增加系统复杂度,同时资源碎片化也会导致系统速率下降^[13]。针对上述问题,子组元需要进行适度合并,最终形成 $S(S \leq M)$ 个多播子组。第 $s(1 \leq s \leq S)$ 个子组的CQI值 $m(1 \leq m \leq M)$ 由子组内信道质量最差的用户来决定,子组中用户的信息速率 d_s 为

$$d_s = b_m r_s \quad (1)$$

式中: b_m 为在CQI值为 m 的通信环境下,传输单位RB所需的数据速率, b_m 根据表1即可计算得到^[14]; r_s 为子组 s 所分配的资源数。

表1 CQI-MCS映射表^[14]
Table 1 CQI-MCS Mapping^[14]

CQI 编号	调制方式	码率 * 1 024	效率
1	QPSK	78	0.152 3
2	QPSK	120	0.234 4
3	QPSK	193	0.377
4	QPSK	308	0.601 6
5	QPSK	449	0.877
6	QPSK	602	1.175 8
7	16QAM	378	1.476 6
8	16QAM	490	1.914 1
9	16QAM	616	2.406 3
10	64QAM	466	2.730 5
11	64QAM	567	3.322 3
12	64QAM	666	3.902 3
13	64QAM	772	4.523 4
14	64QAM	873	5.115 2
15	64QAM	948	5.554 7

通常,由特定的目标函数来确定最终形成的 S 个多播子组,子组的移动用户分配向量和资源分配向量由 $\mathbf{U}_s = [u_1, u_2, \dots, u_s]$ 和 $\mathbf{R}_s = [r_1, r_2, \dots, r_s]$ 来表示。在视频多播方案中,主要采取最大化聚合数据速率(ADR)的方法,ADR 即为多播组中所有用户的数据速率的总和。

1.2 SVC 视频多播传输

多播组中的 SVC 视频流可以编码成一个 BL 层和多个 EL 层,如果移动用户能够解调的视频层数越多,获得的视频质量也就越高。因此,多播组中用户的数据传送速率和所接收的视频质量可以不再受到信道条件最差用户的限制^[15]。

如图2所示,不同子组所能接收的 SVC 视频层数也会根据数据速率的不同而有所差异。为了确保多播组中的用户都能接收到视频信息,每个子组至少需要传输 BL 层数据,即 CQI 值最低的

子组信道传输速率要高于 BL 层数据速率。根据资源分配向量 \mathbf{R}_s 得到每个子组的信道带宽,确定各子组所能传输的 SVC 视频层数。

在确定视频多播分组方案时,除了要最大化整体系统吞吐量,还要分析不同子组在传输视频数据时的频谱效率,以及不同用户之间的资源公平性,从而可以更好地满足用户的个性化视频服务需求。视频多播传输方案相关的评价指标包括系统速率、系统公平性及频谱效率。

1) 系统速率。系统速率由多播组中所有用户的数据速率总和来表示,系统速率越高,整体服务质量越好,用户的满意等级也越高。系统速率表示为

$$\text{SystemRate} = \sum_{s=1}^S d_s u_s \quad (2)$$

式中: d_s 为子组中用户的信息速率; u_s 为子组的用户数目。

2) 系统公平性。系统公平性衡量的是不同用户之间资源分配的公平性,可以由多种方式来表示,本文提出的视频多播传输方案主要参照文献[16]。系统公平性表示为

$$\text{SystemFairness} = \frac{\left(\sum_{s=1}^S \partial_s \right)^2}{S \sum_{s=1}^S \partial_s^2} \quad (3)$$

$$\begin{cases} r_1 : r_2 : \dots : r_s = \gamma_1 : \gamma_2 : \dots : \gamma_s \\ \partial_s = r_s \gamma_s \end{cases} \quad (4)$$

式中: r_1, \dots, r_s 为子组分配的资源数; $\gamma_1, \dots, \gamma_s$ 为一组确保多播组内比例公平性的预设值,其与各子组的信道条件有关。

采用式(3)来进行公平性度量,既能够对子组中信道质量好的用户匹配相应多的资源,又能有效抑制对其资源分配过多而过度损害其他用户利益的情况,可以均衡分配资源。由于 $0 \leq \text{SystemFairness} \leq 1$, 当取到最大值 1 时,系统中用户之间的资源分配达到最公平的状态。

3) 频谱效率。频谱效率是多播用户实际接收的 SVC 视频数据量和多播传输的信道带宽之间的比值,该指标可以反映系统资源应用到视频多播服务中的效率。子组的频谱效率表示为

$$\text{SpectralEfficiency} = \frac{\sum_{s=1}^S \text{eff}_s \cdot u_s}{K} \quad (5)$$

$$\text{eff}_s = \frac{v_{\text{video}_s}}{r_s} \quad (6)$$

式中: v_{video_s} 为子组 s 实际接收的 SVC 视频的信息

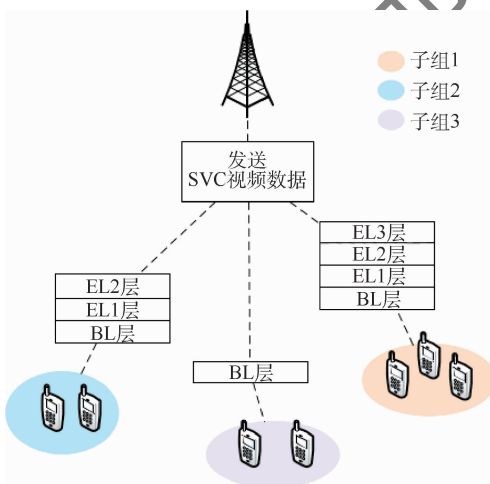


图2 视频层传输模型

Fig. 2 Video layer transmission model

速率; r_s 为子组 s 的资源数; K 为小区内的用户数。

由于子组内用户信道条件不同,频谱效率存在一定差异,系统的频谱效率表示为各子组频谱效率的加权平均值。

1.3 优化目标建立

视频多播传输方案需要考虑的是采用复杂度较低的方法将多播组中的用户划分为最合适的 S 个子组,并将信道资源分配到不同的子组,同时还要考虑各子组是否能保证 SVC 视频的有效传输,最终通过相应的评价指标分析得到更优化的组分解方案。求解组分解方案的过程可以抽象为优化问题,在满足一定限制条件下,采取合适的目标函数,即可得到最终的分组结果。

在视频多播传输方案中的系统限制条件主要包括资源限制和 SVC 视频层限制。资源限制是指至少为每个子组分配 1 个资源块,并且为各子组分配的资源块之和不能超过小区总资源 N ; SVC 视频层限制是指多播组中的用户至少能接收到 BL 层数据。

视频多播传输方案采用 ADR 作为优化目标,目标函数为

$$R = \arg \max_R \sum_{s=1}^S d_s u_s$$

$$\text{s. t. } \sum_{s=1}^S r_s = N \quad \forall s | r_s > 0, \forall s | d_s \geq v_{\text{video_BL}}$$

(7)

式中: $v_{\text{video_BL}}$ 为视频流 BL 层的数据速率; N 为小区内总资源数。

2 低复杂度多播组分解算法

2.1 面向资源公平调配的视频多播分组算法

文献[7]提出了 LCS 算法,可以降低计算处理复杂度。LCS 算法首先初始化相关参数,得到用户分配向量 U 和资源分配向量 R ,然后进入迭代合并阶段,不断地将 2 个子组(元)合并,直到目标值 ADR 不再增加时停止迭代,最终得到组分解方案。但是 LCS 算法仅考虑系统吞吐量,而忽略了用户角度的公平性和视频服务质量需求,不适合直接应用于视频多播环境中。因此针对这一问题,本文提出了考虑公平性的视频多播分组算法 VMS_F。

VMS_F 算法在初始化阶段根据用户 CQI 值将多播组划分为 M 个子组元,同时为了方便后续子组元的合并,将子组个数 S 初始化为 M ,并建立一个可变动的子组集合 S_{set} 。各子组元的资源预分配公式为

$$\alpha_m = \frac{b_m u_m}{\sum_{i=1}^M b_i u_i} \tag{8}$$

$$r_m = 1 + \left\lfloor \frac{\alpha_m}{\sum_{i=1}^M \alpha_i} (N - M) \right\rfloor \tag{9}$$

当子组元中的用户数和 CQI 值越大时,为其分配的权重 α_m 就越大,根据权重值可以得到子组元的初始化资源分配向量 R 。根据子组的 CQI 值和分配的资源数可以得到各子组的信息速率。根据多播组传输的 SVC 视频流信息,得到保证 BL 视频层传输的最小资源块数 $r_{\text{video_BL}}$ 。

初始化结束后,VMS_F 算法从子组集合中选择 2 组进行合并,子组在合并过程中将重新计算合并后所需的资源向量,而不是直接将 2 组的资源块相加,避免了因信道质量差的子组资源分配不足而导致系统公平性大幅下降的问题出现。VMS_F 算法通过引入常值向量来平衡不同用户之间的资源分配,确保信道质量好的用户可以匹配相应多的资源,而信道质量差的用户资源不被过度占用。因此,常值向量的取值和表 1 中不同 CQI 值对应的调制方式直接相关,具体由调制过程中每个符号承载的比特数来表示,比特数越多则调制阶数越高,通信传输效率越高,分配更多的资源,并且不同 CQI 值对应的比特数相差不大,可以有效抑制用户资源分配差距过大的情况。例如,QPSK 中每个符号承载的比特数为 2,以此类推,可以得到用户在不同 CQI 等级下调制方式对应的比特数,最终得到当前通信环境下的常值向量 W 表示为

$$W = [2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6] \tag{10}$$

根据常值向量 W 重新计算迭代合并阶段各子组的资源数目,可得到资源分配公式为

$$\alpha_s = \frac{w_s u_s}{\sum_{i=1}^s w_i u_i} \tag{11}$$

式中: w_s 和 w_i 分别为常值向量 W 的第 s 和第 i 个元素。

$$r_s = 1 + \left\lfloor \frac{\alpha_s}{\sum_{i=1}^s \alpha_i} (N - |S_{\text{set}}|) \right\rfloor \tag{12}$$

引入常值向量可以有效减少因信道条件不同而资源分配差距过大的情况,从而有效提升系统公平性。比较合并前后的 ADR 目标值,直到目标值 ADR 不再增加并且所有子组均可以承担 BL 视频层传输时才能结束迭代,最终得到系统公平

性提高并且满足用户基本视频数据传输的多播组分解方案。

2.2 复杂度分析

VMS_F 算法在改进 LCS 算法的基础上,主要增加了资源块分布的重新计算,即每次迭代时都会更新资源向量。相比于子组合时直接将 2 组资源块相加的方式,VMS_F 算法在子组合中仅增加了一些简单的加减乘除运算,但是由于没有改变迭代次数,不影响整体算法的计算复杂度,2 个算法的时间复杂度都是 $O(M^2)$ 。一些组分解算法的迭代过程会与总资源数目 N 或者多播组内用户数 K 相关,如文献[10]中算法的计算复杂度为 $O(GNK)$,其中, G 为多播组个数。而 VMS_F 算法复杂度仅与 CQI 最大值 M 相关,在应用场景中 N 值和 K 值一般会大于 M ,因此 VMS_F 算法可以有效降低组分解方案的计算复杂度,并且算法还综合考虑了组内用户资源公平性和 SVC 视频传输的基本约束,可以实际应用于视频多播传输系统中。

VMS_F 算法伪代码如下:

- 1: 定义 $M, S, U = [u_1, u_2, \dots, u_s]$
- 2: $S_set = \{s | u_s > 0, 1 \leq s \leq S\}$
- 3: 根据式(9)计算 R
- 4: 根据式(1)计算 d
- 5: 计算 $u_{\text{video_BL}}, u_{\text{video_EL}}$
- 6: 计算 $\Omega^t = \sum_{s=1}^S d_s u_s$
- 7: $t = 2$
- 8: while $t < M$ do
- 9: Find $i, j \in S_set$ that maximize Ω^t
- 10: $\Omega^t = \Omega^{t-1} - d_i u_i - d_j u_j + b_n r_n^* u_n^*$
- 11: where $n = \arg \min(b_i, b_j), u_n^* = u_i + u_j$
- 12: 根据式(12)计算得到 r_n^*
- 13: if $(\Omega^t > \Omega^{t-1}) \vee (\exists s | d_s < v_{\text{video_BL}})$ then
- 14: update $S_set = S_set - \{i, j\} + \{n\}$
- 15: $\text{del} = \arg \max(b_i, b_j)$
- 16: $u_n = u_i + u_j, u_{\text{del}} = 0$
- 17: 根据式(12)计算得到 $r_n, r_{\text{del}} = 0$
- 18: $t = t + 1$
- 19: else
- 20: Stop
- 21: end if
- 22: end while

3 结果与分析

3.1 仿真环境

本节通过数据模拟,对本文提出的考虑公平

性的视频多播算法进行实验仿真和性能分析,仿真实验在 Visual Studio 中实现。主要考虑由基站到多用户的无线多播传输场景,假定多播用户均匀分布在半径为 2 km 的小区中,主要仿真参数如表 2 所示。

仿真中,采用文献[10]中的 4 个标准视频测试序列: CREW、FOREMAN、BUS、ICE,表 3 为每层 SVC 视频流的源速率。

表 2 主要仿真参数

Table 2 Main simulation assumptions

参数	数值
每个 RB 的带宽/kHz	180
小区半径/km	2
可用 MCS 等级	15
噪声功率密度/(dBm · Hz ⁻¹)	-174
路径损耗/dB	128.1 + 37.6lg d
穿透损耗/dB	20
基站发射功率/dBm	28

注: d 为无线信号传播距离。

表 3 视频流源速率

Table 3 Source rates of video stream

视频流	源速率/(Mbit · s ⁻¹)			
	BL	EL1	EL2	EL3
CREW	0.306	0.578	0.814	1.184
FOREMAN	0.170	0.407	0.589	0.890
BUS	0.185	0.390	0.567	0.857
ICE	0.277	0.548	0.767	1.123

3.2 组分解算法性能评估

在系统速率、公平性及频谱效率方面对考虑公平性的视频多播分组算法 VMS_F 进行性能评估,为了更好地比较算法的性能,还引入了传统多播算法 CMS、比例公平资源分配算法 PF^[17] 及 LCS 算法。采用 4 种算法进行视频多播传输,并比较不同算法的性能。LCS 算法和 VMS_F 算法都是以 ADR 为优化目标函数,而 PF 算法则是通过最大化用户数据速率的对数和来实现。CMS 算法是传统多播方案,多播组中用户不再细化分组,统一以信道质量最差用户的速率进行传输。在对 4 种算法进行评估时,考虑 3 种不同的仿真环境,分别为信道带宽资源保持不变、信道带宽资源可变及传输不同的视频流情况。

3.2.1 信道带宽不变

假设仿真环境中的信道带宽资源为 6 MHz,此时 RB 数为 30,SVC 视频流经由地面基站多播到小区内的各用户,小区内所有用户构成一个多播组,组内传输的视频内容为 BUS 视频流,由表 3

可以看到视频流编码成 1 个 BL 层和 3 个 EL 层。同时设定用户数以 10 的步长从 10 人到 100 人依次递增,比较 4 种多播算法的系统速率、系统公平性及频谱效率。

如图 3 所示,VMS_F 算法具有最高的系统速率,然后依次是 LCS 算法、PF 算法和 CMS 算法。由于 VMS_F 算法相比其他算法综合考虑了视频传输及系统公平性的因素,其系统速率最高。CMS 算法没有进行子组划分,受到瓶颈用户的影响系统速率较低,并且用户越多越明显。如图 4 所示,由于 CMS 算法没有进行组分解,也没有进行资源分配,系统公平值为 1,即最公平的状态,可以将 CMS 算法作为其他算法系统公平性(公平指数)的参考。相比于 LCS 算法和 PF 算法,VMS_F 算法在用户资源分配方面具有更高的公平性,其公平指数基本在 0.7 以上。由于考虑了视频传输的因素,需要比较系统资源应用到视频多播服务中的效率,如图 5 所示,可以看到 VMS_F 算法具有更高的频谱效率,然后依次是 LCS 算法、PF 算法和 CMS 算法。经过改进的视频多播算法 VMS_F 考虑了视频传输的基本限制,因此可以更好地将系统资源转化到实际视频传输中。

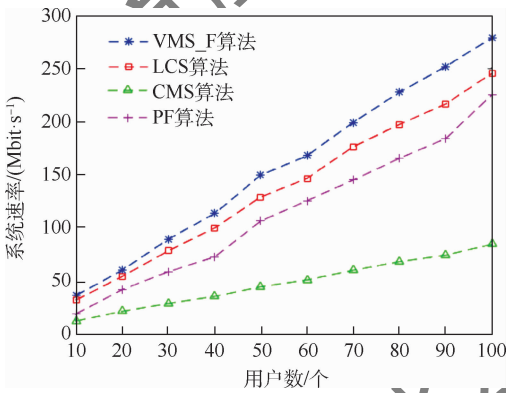


图 3 信道带宽不变时的系统速率比较

Fig.3 System rate comparison with constant channel bandwidth

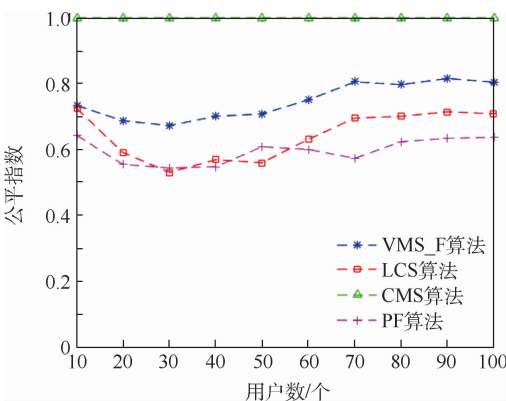


图 4 信道带宽不变时的系统公平性比较

Fig.4 System fairness comparison with constant channel bandwidth

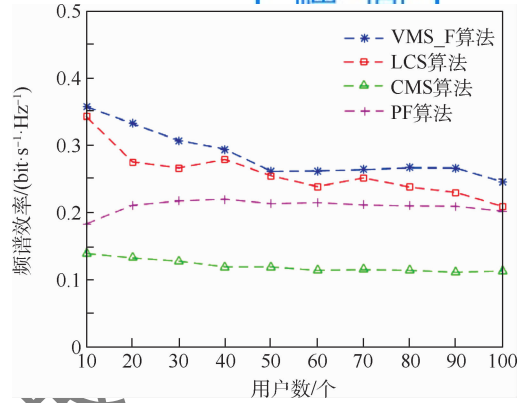


图 5 信道带宽不变时的频谱效率比较

Fig.5 Spectral efficiency comparison with constant channel bandwidth

3.2.2 信道带宽可变

从 3.2.1 节的性能评估中可以大致看到,VMS_F 算法在 3 个指标中都保持良好的性能,其次是 LCS 算法。然而在不同的通信环境中,信道资源会存在差异,因此需要比较在分配不同资源情况下各多播算法的性能。在此仿真环境中,假定多播组内用户数为 50,组内传输的视频内容仍为 BUS 视频流,信道带宽从 5 MHz 变化到 20 MHz,为其分配的 RB 数目从 25 变化到 100,比较 4 种算法的系统速率、公平指数和频谱效率。

如图 6~图 8 所示,VMS_F 算法和 LCS 算法在系统速率中性能相近,当资源数较少时,VMS_F 算法具有更好性能,随着资源数目增加,两者的性能逐渐持平。在系统公平性方面,CMS 算法由于未划分子组,其值为 1。而在其他 3 种算法中,VMS_F 算法具有更稳定的性能,因为该算法着重考虑了公平性因素,在资源块变化过程中其公平指数基本能保持在 0.7 以上。频谱效率也与系统速率指标相似,VMS_F 算法和 LCS 算法均具有较高的频谱效率,而在资源块较多时 LCS 算法性能更好。

3.2.3 不同视频流

3.2.1 节和 3.2.2 节中多播组传输的是 BUS 视频流,因此本节主要比较不同视频流下各算法的性能,并且设定 RB 数为 30,用户数为 50。如图 9 所示,可以看到不同视频流的系统速率存在差异,由于 CREW 的视频流速率较大,在分组过程中需要确保各子组均能接收到 BL 层数据,相比于其他视频流,高速率的视频数据在分组过程中受到的限制更多,在优化过程中不能更好地逼近目标值,系统速率相对较低。即使各视频流系统速率存在差异,VMS_F 算法依然具有很大优势,在不同视频流传输环境中保持良好的性能。

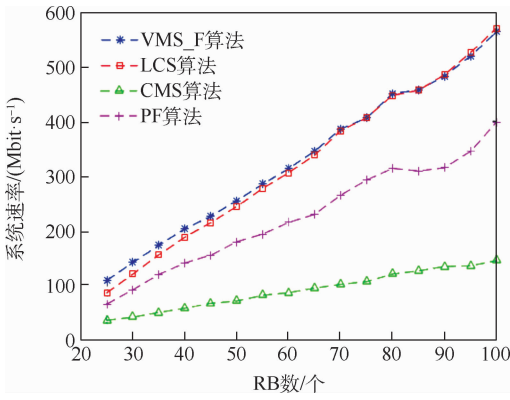


图6 信道带宽可变时的系统速率比较

Fig.6 System rate comparison with variable channel bandwidth

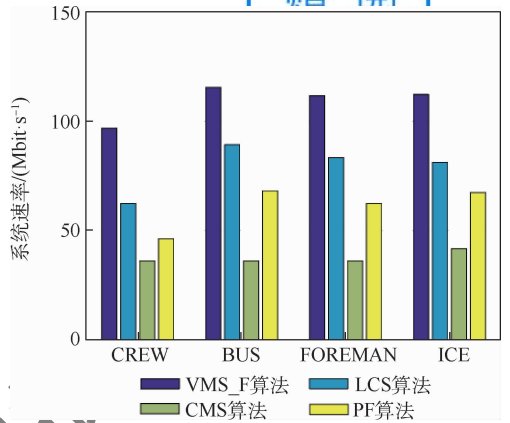


图9 不同视频流的系统速率比较

Fig.9 System rate comparison of different video streams

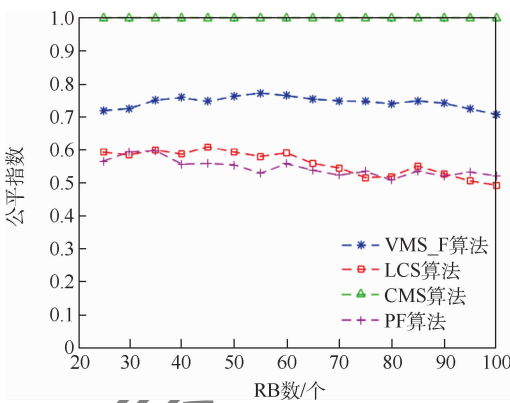


图7 信道带宽可变时的系统公平性比较

Fig.7 System fairness comparison with variable channel bandwidth

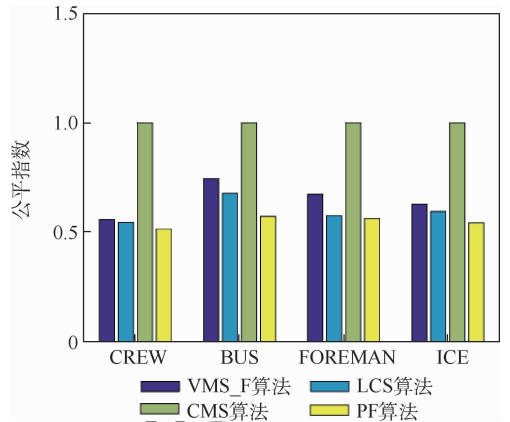


图10 不同视频流的系统公平性比较

Fig.10 System fairness comparison of different video streams

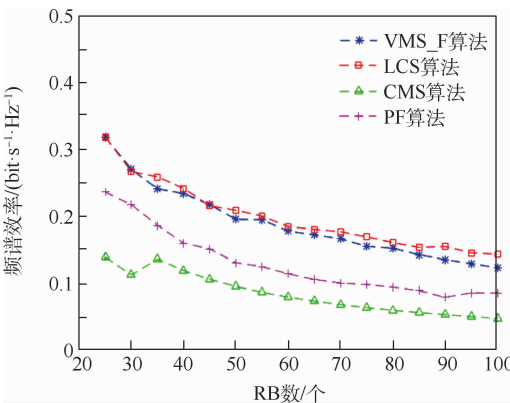


图8 信道带宽可变时的频谱效率比较

Fig.8 Spectral efficiency comparison with variable channel bandwidth

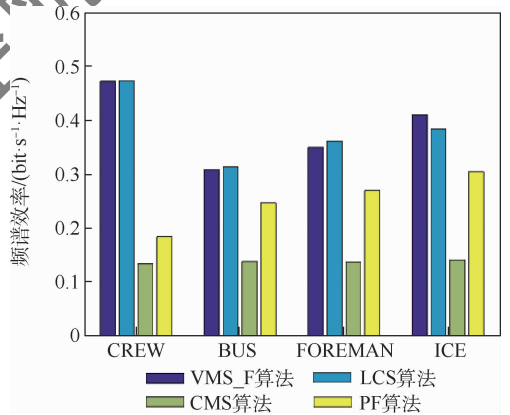


图11 不同视频流的频谱效率比较

Fig.11 Spectral efficiency comparison of different video streams

在系统公平性方面,如图10所示,不同视频流的公平指数差异不大,VMS_F算法的多播算法公平指数更高。如图11所示,CREW和ICE视频流的频谱效率更高,即系统资源转化为视频传输的效率更高,并且可以看到在不同视频流传输环境中,VMS_F算法和LCS算法具有更高的频谱效率。

4 结论

对如何构建面向用户的视频多播传输方案进行了研究,并提出了满足用户需求和系统资源公平的低复杂度多播资源分配算法VMS_F。

1) 提出的视频多播方案是对用户进行分组,重点考虑用户端的服务质量,在最大化系统速率

的目标下,得到不同子组接收SVC视频数据的最佳匹配结果。

2) 在带宽资源和用户数量变化时,相比于CMS算法、PF算法、LCS算法,VMS_F算法的系统速率和频谱效率均可以保持在很高的水平,表明其具有很好的服务质量,并且可以高效地将系统资源转化为视频传输。在系统公平性方面,VMS_F算法仅次于CMS算法,相比与LCS算法有了很大的提升。

3) 多播组中传输不同的视频内容时,VMS_F算法在各指标下依然保持稳定的优势,适合应用于复杂的通信环境中。

4) VMS_F算法具有较低的计算复杂度,可以实际应用于视频多播传输系统中。

参考文献 (References)

- [1] FENG S, LIU C, SHEN C, et al. An effective and efficient dynamic e-MBMS multicast grouping scheduling algorithm in MBSFN for public safety scenarios [J]. IEEE Access, 2020, 8: 105701-105712.
- [2] MONTALBAN J, SCOPELLITI P, FADDA M, et al. Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques [J]. IEEE Communications Magazine, 2018, 56(3): 91-95.
- [3] ABOLHASSAN B, TADROUS J, ERYILMAZ A. Wireless multicasting for content distribution: Stability and delay gain analysis [C] // IEEE INFOCOM 2019-IEEE Conference on Computer Communication. Piscataway: IEEE Press, 2019: 1-9.
- [4] STRICCOLI D, PIRO G, BOGGIA G. Multicast and broadcast services over mobile networks: A survey on standardized approaches and scientific outcomes [J]. IEEE Communications Surveys & Tutorials, 2018, 21(2): 1020-1063.
- [5] TAN C K, CHUAH T C, TAN S W. Adaptive multicast scheme for OFDMA based multicast wireless systems [J]. Electronics Letters, 2011, 47(9): 570-572.
- [6] WANG X, LI H, TONG M, et al. Network coded cooperative multicast in integrated terrestrial-satellite networks [C] // IEEE Symposium on Computers and Communications (ISCC). Piscataway: IEEE Press, 2019: 1-6.
- [7] ARANITI G, CONDOLUCI M, ORSINO A, et al. Effective resource allocation in 5G-satellite networks [C] // IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2015: 844-849.
- [8] ARANITI G, BISIO I, DESANCTIS M, et al. Joint coding and multicast subgrouping over satellite-eMBMS networks [J]. IEEE Journal on Selected Areas in Communications, 2018, 36(5): 1004-1016.
- [9] 赵一杰. 基于非正交多址接入的多组播可伸缩视频传输的研究 [D]. 西安: 西安电子科技大学, 2019: 20-25.
ZHAO Y J. Research on multicast scalable video transmission based on non-orthogonal multiple access [D]. Xi'an: Xidian University, 2019: 20-25 (in Chinese).
- [10] CONDOLUCI M, ARANITI G, MOLINARO A, et al. Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks [J]. IEEE Transactions on Vehicular Technology, 2016, 65(5): 2907-2921.
- [11] ARANITI G, IERA A, MOLINARO A, et al. Exploiting multicast subgrouping for multi-layer video services in 5G-satellite networks [C] // IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2018: 1-6.
- [12] ARANITI G, ORSINO A, COSMAS J, et al. A low computational-cost subgrouping multicast scheme for emerging 5G-satellite networks [C] // IEEE Broadband Multimedia Systems and Broadcasting (BMSB). Piscataway: IEEE Press, 2016: 1-6.
- [13] ARANITI G, CONDOLUCI M, COTRONEI M, et al. A solution to the multicast subgroup formation problem in LTE systems [J]. IEEE Wireless Communications Letters, 2015, 4(2): 149-152.
- [14] ARANITI G, BISIO I, SANCTIS M D, et al. Multimedia content delivery for emerging 5G-satellite networks [J]. IEEE Transactions on Broadcasting, 2016, 62(1): 10-23.
- [15] ZHOU H, JI Y, WANG X, et al. Joint resource allocation and user association for svc multicast over heterogeneous cellular networks [J]. IEEE Transactions on Wireless Communications, 2015, 14(7): 3673-3684.
- [16] SHARMA N, MADHUKUMAR A S. Genetic algorithm aided proportional fair resource allocation in multicast OFDM systems [J]. IEEE Transactions on Broadcasting, 2015, 61(1): 16-29.
- [17] ARANITI G, CONDOLUCI M, PETROLINO A. Efficient resource allocation for multicast transmissions in satellite-LTE networks [C] // IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2013: 3023-3028.

Low computational-cost multicast subgrouping for SVC streams

YANG Jingxuan, XU Zhen*

(School of Electronic Information Engineering, Beihang University, Beijing 100083, China)

Abstract: Multicast is widely used in multimedia transmission because it can utilize channel bandwidth efficiently. However traditional multicast scheme has bottleneck rates problem, which may limit the multicast transmission rate. Subgroup formation can solve the problem, which involves splitting and classifying the multicast group into smaller sub-groups based on intra-group users' channel qualities. Firstly, this paper proposed a user-oriented video multicast transmission scheme, that combined scalable video coding (SVC) technology with subgroup formation. Each subgroup demodulated SVC streams of different layers according to actual data receiving capability, during which the scheme maximized the total system rate on the basis of user video transmission. Moreover, a low computational-cost subgroup algorithm for fair resource allocation is proposed. This proposal considered SVC layer limitation based on low-complexity subgrouping (LCS) algorithm and adopted a constant vector to suppress unfair resource allocation. Finally, the results show that the proposed algorithm achieves well performance in terms of system rate, spectrum efficiency and system fairness when the bandwidth resource and the number of users change. In addition, the computational complexity of the algorithm is low, thus it can be applied to video transmission under 4G and 5G network architectures.

Keywords: scalable video coding (SVC); multi-rate multicast; resource allocation; wireless network; video streaming

Received: 2021-01-11; **Accepted:** 2021-03-26; **Published online:** 2021-04-09 09:20

URL: kns.cnki.net/kcms/detail/11.2625.V.20210408.1435.001.html

Foundation items: National Natural Science Foundation of China (91638301,91738301)

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

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

DOI: 10.13700/j.bh.1001-5965.2021.0003

小天体环的轨道动力学

吴太泷, 王悦*

(北京航空航天大学 宇航学院, 北京 102206)

摘 要: 随着天文观测的深入,发现不仅土星、木星等气态巨行星环绕有光环,某些小天体也环绕有光环。针对目前发现环的半人马型小天体女凯龙星(10199 Chariklo),研究其引力场中粒子(视为质点)的轨道动力学,分析赤道捕获率和小天体自转对环中粒子轨道运动的影响,通过庞加莱截面中的KAM环迭代,得到第一类周期轨道和1:3共振周期轨道,并通过对比和分析得到小天体光环位置与小天体自转平运动共振的关系。研究表明:女凯龙星内环中的粒子最可能处在满足径向振荡幅值范围的第一类周期轨道及其附近的准周期轨道上,但不能排除处在1:3共振周期轨道及其附近的准周期轨道上;外环中的粒子不可能处在1:3共振周期轨道上,只可能处在第一类周期轨道及其附近的准周期轨道上。

关键词: 小天体环; 女凯龙星; 庞加莱截面; 平运动共振; 共振周期轨道

中图分类号: V412.4*1

文献标志码: A **文章编号:** 1001-5965(2022)07-1287-10

行星环是指在行星引力作用下,物质绕其轨道运动而形成的环状结构,也被称为行星光环。现已知道,太阳系中土星、木星、天王星和海王星4个气态巨行星环绕有光环。随着近年来天文观测的进展,人们发现星环并不专属于行星,某些小天体也拥有星环。

女凯龙星(10199 Chariklo)是目前已知最大的半人马型小天体,直径约250 km。2014年,天文学家通过分析掩星观测数据,确认女凯龙星拥有2条狭窄但密集的星环,女凯龙星成为太阳系内第一颗被发现拥有环的小行星^[1]。女凯龙星的内环宽约7 km,环中心轨道半径约(390.6 ± 3.3) km,外环宽约3 km,环中心轨道半径约(404.8 ± 3.3) km,两环的间隙为(8.7 ± 0.4) km(两环中心距约为(14.2 ± 0.2) km)。女凯龙星相比巨行星质量极小,引力微弱,星环内物质难以稳定存在^[2]。随着天文探测的继续深入,另一个小天体环也被人类发现。2017年,对矮行星妊

神星(136108 Haumea)的掩星观测结果表明,妊神星也有一个光环^[3],其也因此成为海王星轨道外首个发现带有星环的天体。

自2014年发现女凯龙星环后,国内外很多学者对小天体环进行了研究。有学者研究了小天体环的起源问题。Ortiz等^[4]认为女凯龙星环可能来自小天体受撞击后的喷散物质,也可能来自小卫星破裂或撞击后的喷散物质。Hyodo等^[5]认为女凯龙星曾与巨行星近距离相遇,潮汐作用导致了女凯龙星环的形成。有的学者试图利用女凯龙星环可能具有偏心率的事实,探究其是否如天王星环那样通过环内粒子的相互引力来补偿女凯龙星非球形引力带来的进动^[6]。然而,内环的偏心率范围目前并没有很好的估计^[7],因此,女凯龙星环和天王星环的相似性尚待证实。一些学者利用环中物质的相互引力和碰撞仿真进行数值计算开展研究。Gupta等^[8]采用N体仿真的方法,将环中粒子限制在一个小的微团内,除了中心天体的

收稿日期: 2021-01-05; 录用日期: 2021-03-28; 网络出版时间: 2021-04-26 09:39

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210425.1806.002.html

基金项目: 国家自然科学基金(11872007); 中央高校基本科研业务费专项资金

* 通信作者。E-mail: ywang@buaa.edu.cn

引用格式: 吴太泷, 王悦. 小天体环的轨道动力学[J]. 北京航空航天大学学报, 2022, 48(7): 1287-1296.

WU T L, WANG Y. Orbital dynamics of rings of small bodies [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1287-1296 (in Chinese).

引力外,微团中的粒子还存在相互引力和碰撞等作用,根据该方法研究了非球形天体,特别是三轴椭球体附近的轨道动力学,对比了土星环和女凯龙星环的性质,证明了非球形引力势会影响环的动力学性质。还有一些学者从平运动共振的角度研究小天体环的动力学。Winter 等^[9]利用庞加莱截面方法分析了妊神星环所在区域内单个粒子的轨道动力学,研究了 2 类周期轨道的特征,发现第一类周期轨道的偏心率较小,能够符合妊神星环的特征,而 1:3 共振周期轨道偏心率较大,与妊神星环的特征不符。Sicardy 等^[10-11]研究了非轴对称天体的环动力学,以女凯龙星和妊神星为对象,分别研究赤道上的地形地貌特征及扁率、扁率对环的影响,进而分析多种比例平运动共振的位置与环的关系,还对环内粒子相互作用引起的阻尼及长时间的演化问题进行了研究。

本文主要利用庞加莱截面方法分析女凯龙星的赤道扁率和自转对环中粒子轨道运动的影响,并进一步分析女凯龙星的光环位置与其自转平运动共振的关系。由于小天体环位于小天体 Hill 半径以内,且距离小天体相对较近,忽略其他天体的引力摄动,在轨道动力学中只考虑小天体的引力。另外,本文将小天体环中尘埃和颗粒物简化为可视为质点的粒子进行研究,考虑小天体三轴椭球体的形状和自转,建立小天体固连坐标系中单个粒子的轨道动力学模型,根据平运动共振现象估算 1:3 共振周期轨道的位置。通过数值方法得到女凯龙星环附近轨道的庞加莱截面相图,并通过 KAM 环迭代方法得到第一类周期轨道和 1:3 共振周期轨道。通过研究 2 类周期轨道的特性和进一步的对比分析,得出女凯龙星环与这 2 类周期轨道的联系。

1 轨道动力学模型

1.1 坐标系

如图 1 所示,本文的轨道动力学建立在小天体固连坐标系 $oxyz$ 中,其定义为:坐标系原点位于小天体质心 o , x 轴、 y 轴、 z 轴分别指向小天体的最小、中间、最大惯量轴,构成右手系。由于小天体存在自转,固连坐标系是一种质心旋转坐标系。此外,本文研究将用到质心惯性坐标系 $OXYZ$,其定义为:坐标系原点位于小天体质心 O , X 轴、 Y 轴、 Z 轴分别与初始时刻固连坐标系 $oxyz$ 的 x 轴、 y 轴、 z 轴重合。相比固连坐标系 $oxyz$,惯性坐标系不随小天体的自转而旋转。

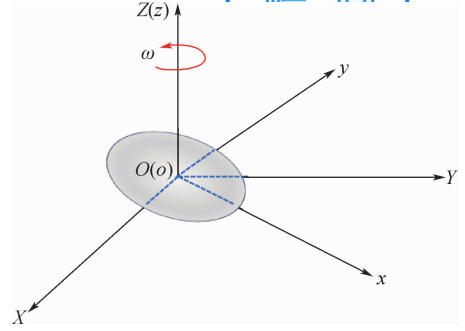


图 1 固连坐标系和惯性坐标系
Fig. 1 Body-fixed frame and inertial frame

1.2 引力场模型

对于本文研究的接近三轴椭球体的小天体,球谐函数法是描述小天体引力场较为适合的方法。根据文献[12],采用二阶二次球谐函数表示的小天体引力势 U 的表达式为

$$U = \frac{GM}{r} + \frac{GMR_e^2}{r^3} \left(-\frac{C_{20}}{2} + 3C_{22} \cos(2\lambda) \right) \quad (1)$$

式中: G 为万有引力常量; M 为小天体的质量; R_e 为小天体的参考半径; r, δ, λ 为质点位置矢量在球坐标中的 3 个分量,即 r 为到原点的距离, δ 为纬度, λ 为经度; C_{20}, C_{22} 为球谐系数。

由于小天体环分布于赤道平面,厚度非常小,可以仅考虑在赤道平面 oxy 内的运动,如图 2 所示。

在笛卡尔坐标系中,有 $r = \sqrt{x^2 + y^2}$, $\cos \lambda = x/r$, $\sin \lambda = y/r$ 。因此,对于半长轴分别为 a, b, c ($a > b > c$) 的三轴椭球形状的小天体,最终化简得到其引力势为

$$U(x, y) = \frac{GM}{r} \left\{ 1 - \left(\frac{R_e}{r} \right)^2 \left[\frac{C_{20}}{2} - \frac{3C_{22}}{r^2} (x^2 - y^2) \right] \right\} \quad (2)$$

式中: $R_e = \sqrt[3]{abc}$; $C_{20} = (2c^2 - a^2 - b^2)/(10R_e^2)$; $C_{22} = (a^2 - b^2)/(20R_e^2)$ 。

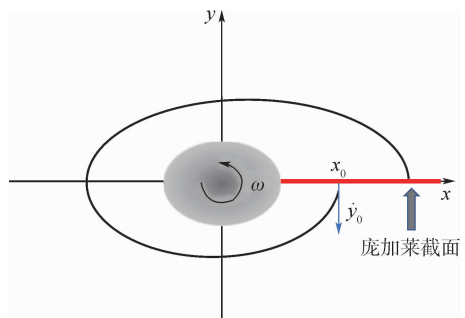


图 2 粒子在固连坐标系 oxy 平面内的运动及庞加莱截面
Fig. 2 Motion of particles in oxy plane of body-fixed frame and Poincaré section

1.3 固连坐标系中的轨道动力学

在固连坐标系中,粒子的轨道运动方程为

$$\ddot{\mathbf{r}} + 2\boldsymbol{\omega} \times \dot{\mathbf{r}} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = \nabla U \quad (3)$$

仅考虑粒子在固连坐标系平面 oxy 内的运动,单个粒子的运动满足以下方程:

$$\begin{cases} \ddot{x} - 2\omega\dot{y} - \omega^2x = \frac{\partial U}{\partial x} \\ \ddot{y} + 2\omega\dot{x} - \omega^2y = \frac{\partial U}{\partial y} \end{cases} \quad (4)$$

运动方程与时间无关,存在首次积分,即反映粒子能量的 Jacobi 常数:

$$C = \omega^2(x^2 + y^2) + 2U(x, y) - \dot{x}^2 - \dot{y}^2 \quad (5)$$

1.4 平运动共振与共振周期轨道

在经典的平面圆型限制性三体问题中,粒子的周期轨道可分为第一类周期轨道和第二类周期轨道^[9]。第一类周期轨道接近圆轨道,而当粒子处于第二类周期轨道时,粒子运行在椭圆轨道上,并与引力场的周期变化产生平运动共振,也就是所谓的共振周期轨道^[9]。对于当前研究的旋转三轴椭球体和粒子构成的限制性两体问题,粒子平运动与中心天体自转之间也存在共振周期轨道^[9]。

将引力势用偏率 ε 和扁率 f 表示为

$$U(r) = \frac{GM}{r} + \frac{GMR_c^2}{r^3} \left(\frac{f}{5} + \frac{3\varepsilon}{10} \cos(2\lambda) \right) \quad (6)$$

可见,由于存在偏率 ε ,引力势依赖于方位角,也就是经度 λ 。

对式(6)在方位角 λ 上进行平均化处理,得到距离小天体 r 处的平均引力势为

$$\bar{U}(r) = \frac{GM}{r} \left[1 + \frac{f}{5} \left(\frac{R_c}{r} \right)^2 \right] \quad (7)$$

在赤道平面内运动的粒子有2个运动自由度,它们与2个基本频率相关:平运动角速度(mean motion) n 和径向振荡频率(radial epicyclic frequency) κ 。环中粒子在惯性坐标系中的平运动角速度 n 和径向振荡频率 κ 分别为^[10]

$$\begin{cases} n^2(r) = -\frac{1}{r} \cdot \frac{d\bar{U}(r)}{dr} = \frac{GM}{r^3} \left[1 + \frac{3f}{5} \left(\frac{R_c}{r} \right)^2 \right] \\ \kappa^2(r) = -\frac{1}{r^3} \cdot \frac{d(r^4 n^2)}{dr} = \frac{GM}{r^3} \left[1 - \frac{3f}{5} \left(\frac{R_c}{r} \right)^2 \right] \end{cases} \quad (8)$$

由于 n 是在惯性坐标系中的平运动角速度,在固连坐标系中粒子运动的基本频率为 $n - \omega$ 和 κ 。根据文献[10],粒子的轨道运动与小天体自转之间可以产生2种共振:一种满足 $n = \omega$,被称为同步共振;另一种满足以下关系:

$$j\kappa = m(n - \omega) \quad (9)$$

式中: m 为整数; j 为正整数,意味着粒子在 j 个固连坐标系周期 $\frac{2\pi}{n - \omega}$ 内完成 m 次径向振荡。

由于轨道拱线的进动速率(apsidal precession rate)为 $\dot{\omega} = n - \kappa$,式(9)可以写为

$$\frac{n - \dot{\omega}}{\omega - \dot{\omega}} = \frac{m}{m - j} \quad (10)$$

意味着粒子相对于轨道拱线完成 m 个轨道周期时,小天体相对于粒子的轨道拱线旋转 $m - j$ 周,也就是粒子轨道运动与小天体自转之间满足 $m/(m - j)$ 共振,其中, j 为共振阶数。

根据文献[10]的结论,由于三轴椭球形状小天体的引力势具有对称性,即方位角旋转 π 角度后引力势大小不变, m 必须为偶数,也就是所谓的“二倍”共振。那么 1:3 共振实质上最有可能是 4 阶的 2:6 共振,由此可以求出 $m = -2, j = 4$ 。根据文献[1,11]可以得到女凯龙星及其环的部分参数,将其代入式(8)和式(9)中,便可估算出女凯龙星的 2:6 轨道共振的位置 r ,如表 1 所示。可见,女凯龙星内环的位置与 2:6 共振区域非常接近。因此,本文提到的 1:3 共振实际上就是 2:6 共振,仍记为 1:3 共振。

虽然由式(8)和式(9)可知,共振发生位置与三轴椭球体扁率 f 直接相关,而与偏率 ε 无关,但正是由于存在偏率 ε ,引力势依赖于方位角 λ ,才出现平运动共振现象。

表1 女凯龙星及其环的参数和估算出的 2:6 共振位置

Table 1 Parameters of 10199 Chariklo and its rings and estimated 2:6 resonance position

参数	数值
a/km	157
b/km	139
c/km	86
参考半径 R_c/km	123.349 6
质量 M/kg	6.3×10^{18}
自转周期 T/h	7.004
偏率 ε	0.175 1
扁率 f	0.479 4
环中心位置/ km	CR1:390.6 \pm 3.3, CR2:404.8 \pm 3.3
环宽/ km	CR1:7, CR2:3
2:6 共振位置/ km	392.22

2 庞加莱截面及其相图

2.1 庞加莱截面的设定

庞加莱截面的基本思想是:假设有一个 n 维

时间连续动力学系统,在其 n 维相空间中选取一个超平面,也就是庞加莱截面,若运动与该超平面相交,便会留下一系列离散的轨迹点,这些轨迹点便可以反映原动力学系统的一些特征。因此,庞加莱截面通过降低至少一维相空间维数,增强了系统相空间的可视性^[13]。

庞加莱截面上的点可以划分为 2 个区域:①有序的、充满环状结构的区域;②无序的、充满噪点的混沌区。由动力学系统中的 KAM 理论可知,庞加莱截面中的每一个环状结构,即 KAM 环,都对应着一条准周期轨道;在一系列 KAM 环的中心存在着一条周期轨道^[14];而在环状结构外的混沌区,也被称为混沌海(chaos sea),相流对初值非常敏感。

如图 2 所示,本文选取的庞加莱截面为 $y = 0, x > 0$ 半平面。系统相流在穿越庞加莱截面时有 2 种情况:垂直于庞加莱截面的速度分量 $\dot{y} < 0$ (固连坐标系逆行轨道)和 $\dot{y} > 0$ (固连坐标系顺行轨道)。由于 1:3 共振发生在 $\dot{y} < 0$ 的情况,本文只需研究此种情况。

2.2 庞加莱截面的相图

选取合适的初始条件,对粒子轨道运动进行数值积分,利用二分法得到系统相流穿越庞加莱截面时的运动状态,最终便可得到庞加莱截面的相图。

由于本节主要探究女凯龙星环是否与 1:3 共振周期轨道相关,选取与 1:3 共振周期轨道相关的 Jacobi 常数 C ,其范围大致为 $7\ 306\ \text{m}^2/\text{s}^2 < C < 7\ 552\ \text{m}^2/\text{s}^2$,得到一些典型的庞加莱截面相图,如图 3 所示。

由于 1:3 共振实际上是 4 阶的 2:6 共振,将产生一对 1:3 共振周期轨道及其周围的准周期轨道^[9]。在相图中,这一对 1:3 共振周期轨道关于 \dot{x} 轴对称,分别生成 2 个被其准周期轨道 KAM 环所包围的不动点,因此存在互相对称的 2 对 1:3 共振周期轨道的稳定岛(一对用红色表示,另一对用蓝色表示)。而第一类周期轨道及其周围的准周期轨道则会形成一个对称的稳定岛。由于这种共振是双倍的,在相图中 2 类周期轨道稳定岛之间会有一条分界线,并产生一个混沌区域^[9]。

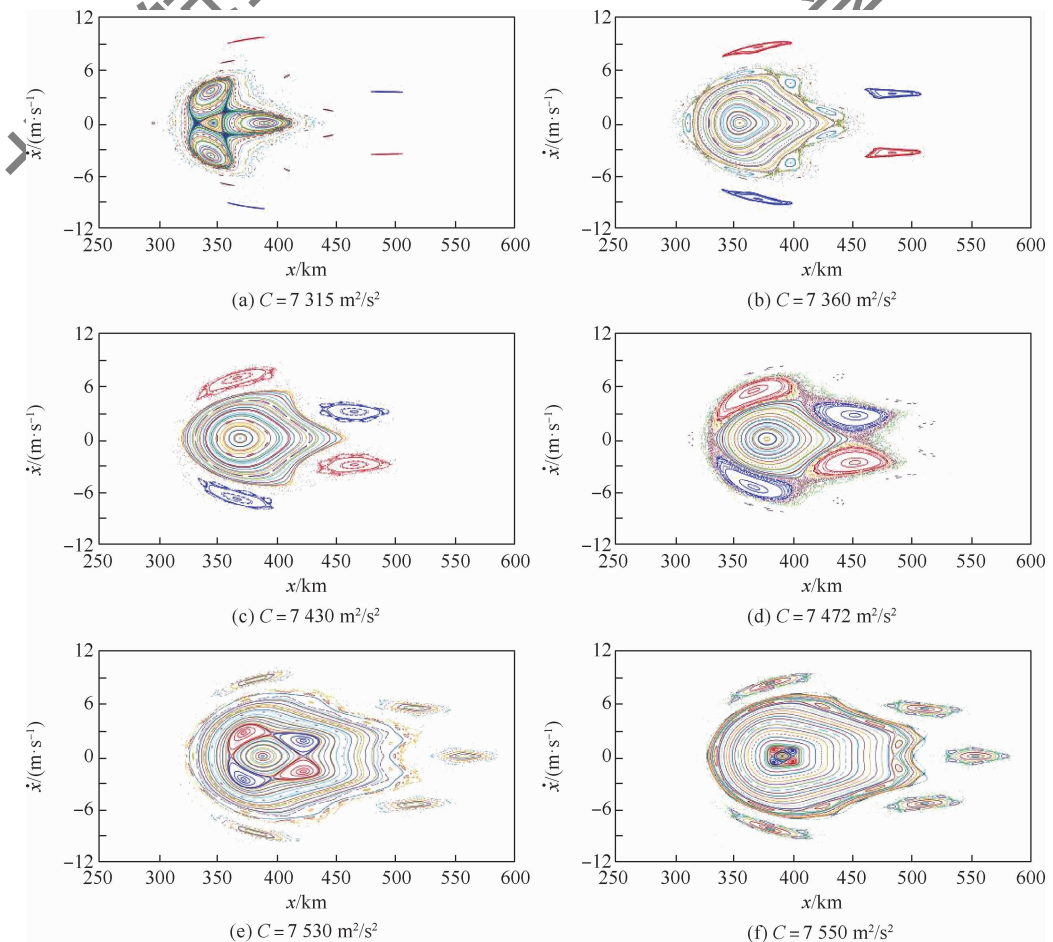


图 3 庞加莱截面的相图

Fig. 3 Phase diagram of Poincaré section

通过相图可以发现,2类周期轨道的稳定岛及之间的混沌区域随 Jacobi 常数 C 的改变而变化。在 C 较小时,第一类周期轨道的稳定岛比较大,1:3 共振周期轨道的稳定岛比较小,两者相离较远;随着 C 逐渐增大,第一类周期轨道的稳定岛逐渐右移(轨道半径不断增大)且逐渐缩小,而 1:3 共振周期轨道的稳定岛不断变大,并逐渐向第一类周期轨道靠拢。在 $C = 7\,472\text{ m}^2/\text{s}^2$ 时,1:3 共振周期轨道的稳定岛达到最大,并与第一类周期轨道的稳定岛相连;随着 C 的进一步增大,1:3 共振周期轨道的稳定岛不断缩小,并被第一类周期轨道的稳定岛包围,不断靠近第一类周期轨道,直至最终消失。

3 女凯龙星环和 2 类周期轨道

3.1 1:3 共振周期轨道的特性

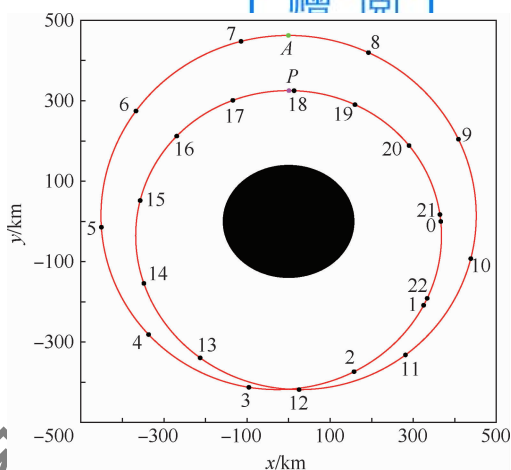
为了更好地理解这 2 类周期轨道,本节利用文献[15]提到的 KAM 环迭代方法,获取这 2 类周期轨道的精确位置来进行特性分析。

图 4 为典型情况 $C = 7\,472\text{ m}^2/\text{s}^2$ 时女凯龙星附近的 1:3 共振周期轨道在固连坐标系中的形状,其中近心点 P 和远心点 A 均已标出,数字代表经过该点所用的小时数。可以看出,每条轨道都是逆行轨道且自身关于 y 轴对称,而这 2 条轨道关于 x 轴对称。

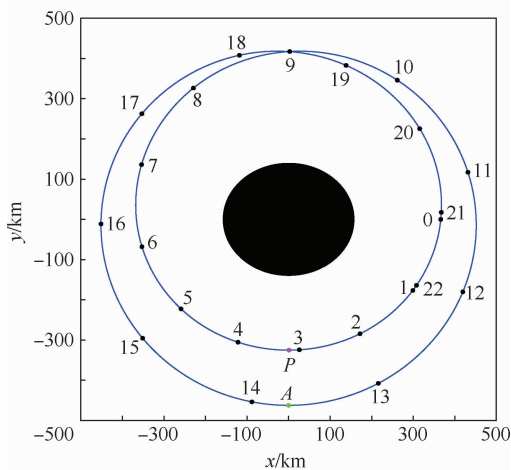
由于天文观测到的环的形状和大小是在惯性坐标系中描述的,还需要进一步研究 1:3 共振周期轨道在惯性坐标系中的特性。如图 5 所示,在惯性坐标系中,2 条 1:3 共振周期轨道都是顺行轨道(与小天体的自转方向相同),自身不具备对称性,但是这 2 条轨道是关于 x 轴对称的。

根据计算, $C = 7\,472\text{ m}^2/\text{s}^2$ 时,1:3 共振周期轨道与女凯龙星中心的距离约为 325 ~ 462 km,其径向振荡的幅度较大,而女凯龙星的 2 个星环分别处于 387.1 ~ 394.1 km 和 403.3 ~ 406.3 km 范围内。显然, $C = 7\,472\text{ m}^2/\text{s}^2$ 条件下的 1:3 共振周期轨道的径向振荡幅度不满足女凯龙星环的位置分布,其他条件下是否满足,还需要进一步分析。

图 6 为 $C = 7\,472\text{ m}^2/\text{s}^2$ 时 1:3 共振周期轨道在 3 年内的轨迹,其中惯性坐标系中的黑线是把第一个周期的轨迹作为参考。在固连坐标系中,即使存在仿真误差的情况,经过 3 年时间,轨道的形状依然保持得非常好,证实了周期轨道求解具有很高的精度。在惯性坐标系中,每完成一个轨



(a) 红色不动点



(b) 蓝色不动点

图 4 1:3 共振周期轨道在固连坐标系中的形状 ($C = 7\,472\text{ m}^2/\text{s}^2$)

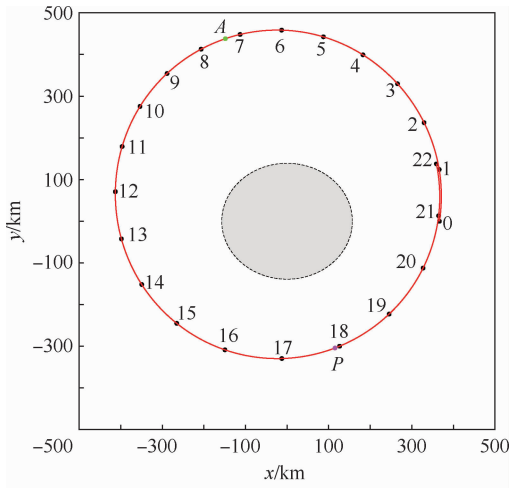
Fig. 4 Shape of 1:3 resonant periodic orbits in body-fixed frame ($C = 7\,472\text{ m}^2/\text{s}^2$)

道周期就要产生一定的相位差,3 年时间内的轨迹会被限制在由无数近心点和远心点组成的 2 条包络线构成的圆环内,粒子径向振荡幅值与一个周期内的振荡幅值保持一致。

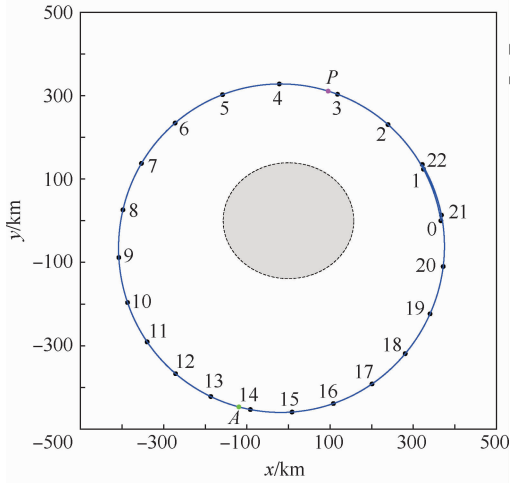
图 7 为 1:3 共振周期轨道随 C 的变化。可见,随着 C 的不断增大,1:3 共振周期轨道的径向振荡幅度逐渐缩小,惯性坐标系中的轨道偏心率不断减小,有可能存在合适的 C 使得径向振荡幅度满足女凯龙星环的位置分布范围。

3.2 第一类周期轨道的特性

图 8 为典型情况 $C = 7\,472\text{ m}^2/\text{s}^2$ 时第一类周期轨道在固连坐标系中的形状。可以看出,第一类周期轨道在固连坐标系中是一个逆行轨道(与小天体自转方向相反),其形状近似为一个以女凯龙星为中心的椭圆,其最远点与女凯龙星的长轴对齐,最近点与女凯龙星的短轴对齐。根据计算得出,此时第一类周期轨道的周期为 10.964 3 h,



(a) 红色不动点



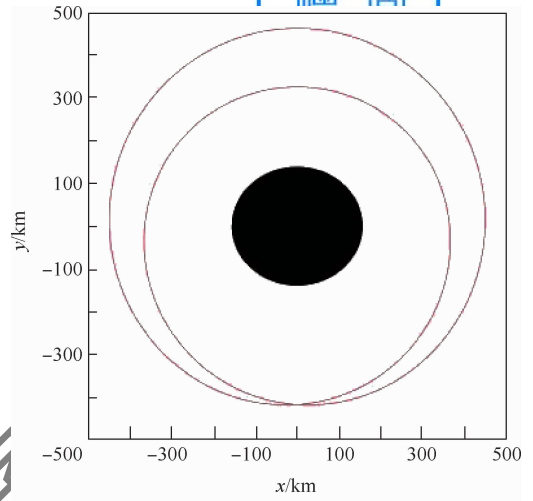
(b) 蓝色不动点

图5 1:3共振周期轨道在惯性坐标系中的形状
($C = 7\,472 \text{ m}^2/\text{s}^2$)

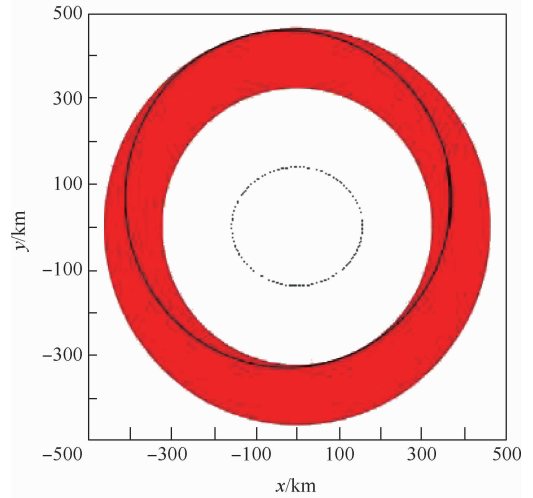
Fig. 5 Shape of 1:3 resonant periodic orbits in inertial frame ($C = 7\,472 \text{ m}^2/\text{s}^2$)

大约为女凯龙星自转周期的1.5654倍。

图9(a)为 $C = 7\,472 \text{ m}^2/\text{s}^2$ 时第一类周期轨道在惯性坐标系中的形状。可以看出,第一类周期轨道在惯性坐标系中为顺行轨道(与小天体自转的方向相同),且由图8和图9(a)进行对比可以发现,处在第一类周期轨道上的粒子在固连坐标系中完成一个周期,在惯性坐标系中大约完成其绕女凯龙星轨道的一半。为观察其在惯性坐标系中完整的一圈,画出双倍时间的轨迹,如图9(b)所示。可以看出,第一类周期轨道在惯性坐标系中的形状也近似为一个椭圆。根据计算, $C = 7\,472 \text{ m}^2/\text{s}^2$ 时,第一类周期轨道的轨道半径约为376.0~377.5 km,其径向振荡幅度很小,但女凯龙星的2个星环分别处于387.1~394.1 km和403.3~406.3 km范围内。显然, $C = 7\,472 \text{ m}^2/\text{s}^2$ 条件下,第一类周期轨道的位置不满足女凯龙星环的位置分布。其他条件下第一类周期轨道是否



(a) 固连坐标系



(b) 惯性坐标系

图6 1:3共振周期轨道在3年内的轨迹
($C = 7\,472 \text{ m}^2/\text{s}^2$)

Fig. 6 Trajectory in 3 years of 1:3 resonant periodic orbits ($C = 7\,472 \text{ m}^2/\text{s}^2$)

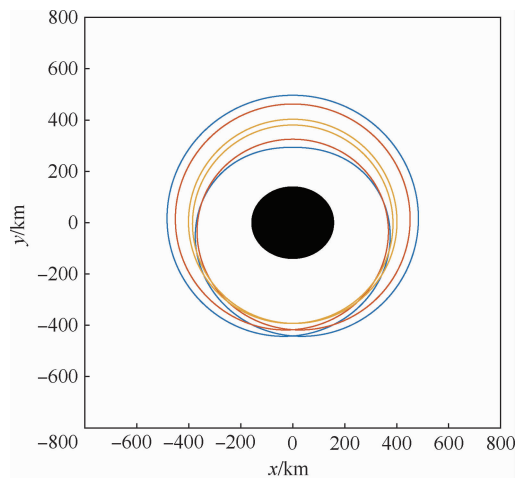
满足,还需要进一步分析。

通过观察 $C = 7\,472 \text{ m}^2/\text{s}^2$ 时第一类周期轨道在3年内的轨迹可以发现,在固连坐标系中不同 C 下第一类周期轨道是一系列同心的椭圆,其轨道半径的大小及径向振荡的幅值将在下文分析。

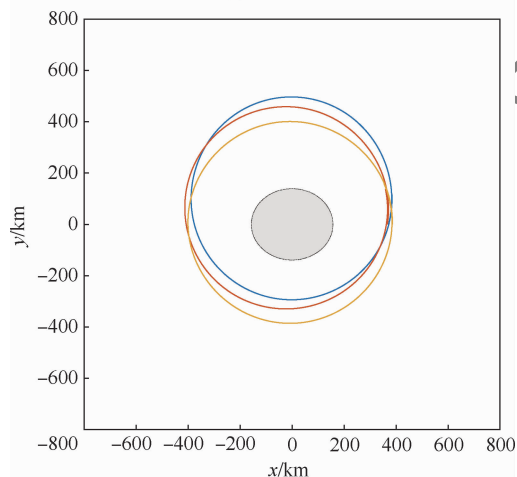
3.3 女凯龙星环的分析

区别第一类周期轨道和1:3共振周期轨道的一个重要特征就是轨道周期,通过图10不难发现,1:3共振周期轨道的周期(蓝线)只存在于女凯龙星自转周期的3倍附近,变化幅值非常小;第一类周期轨道的周期(红线)约为女凯龙星自转周期的1.5倍,且随 C 变化较快,覆盖范围很宽。可见,2类周期轨道的轨道周期有着明显的差别。

将2类周期轨道的径向振荡振幅进行比较,可以发现存在着明显的差异。下面研究女凯龙星附近2类周期轨道和女凯龙星的2个星环位置之



(a) 固连坐标系



(b) 惯性坐标系

— $C = 7\ 360\ \text{m}^2/\text{s}^2$ — $C = 7\ 472\ \text{m}^2/\text{s}^2$
 — $C = 7\ 550\ \text{m}^2/\text{s}^2$

图 7 不同 C 下的 1:3 共振周期轨道
 Fig. 7 1:3 resonant periodic orbits with different Jacobi constants

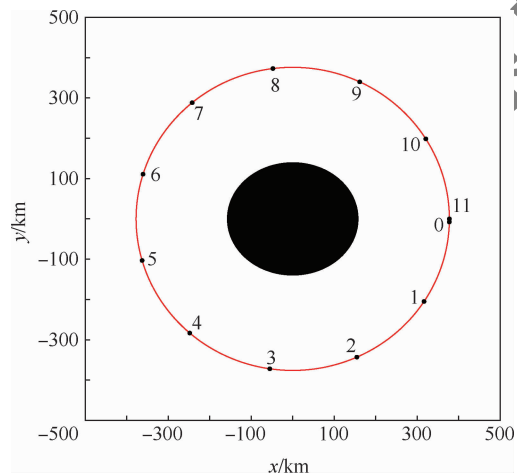
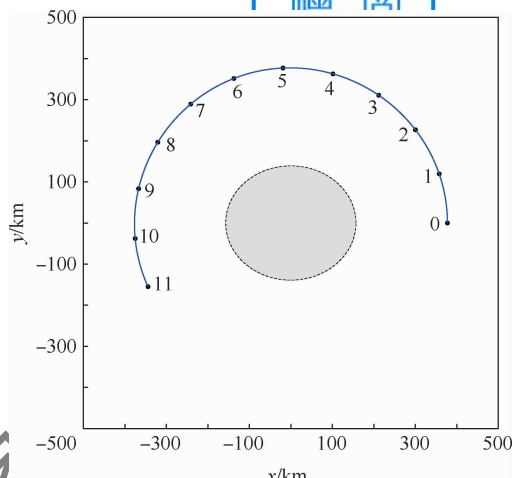


图 8 第一类周期轨道在固连坐标系中的形状
 ($C = 7\ 472\ \text{m}^2/\text{s}^2$)

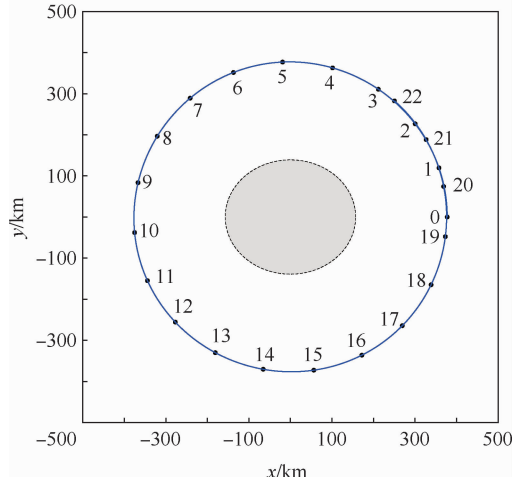
Fig. 8 Shape of the first kind of periodic orbits in body-fixed frame ($C = 7\ 472\ \text{m}^2/\text{s}^2$)

间的相关性。

通过图 11(a) 不难发现,在 C 比较小时,1:3



(a) 11 h



(b) 22 h

图 9 第一类周期轨道在惯性坐标系中的形状
 ($C = 7\ 472\ \text{m}^2/\text{s}^2$)

Fig. 9 Shape of the first kind of periodic orbits in inertial frame ($C = 7\ 472\ \text{m}^2/\text{s}^2$)

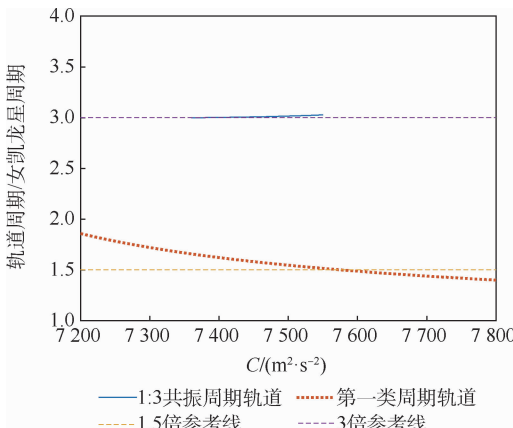


图 10 两类周期轨道的周期随 C 的变化
 Fig. 10 Change of period with C of two kinds of periodic orbits

共振周期轨道的径向振荡幅度非常大,第一类周期轨道的径向振荡幅度较小,2 类轨道均不满足女凯龙星 2 个星环的位置分布范围;随着 C 的增大,1:3 共振周期轨道的径向振荡幅度快速减小,

第一类周期轨道的径向振荡幅度依然保持在较低水平,并且位置逐渐向女凯龙星的2个星环靠拢。在约 $7\ 525\ \text{m}^2/\text{s}^2 < C < 7\ 570\ \text{m}^2/\text{s}^2$ 范围内,第一类周期轨道的位置和径向振荡幅值均能满足女凯龙星内环的位置分布范围;在约 $7\ 610\ \text{m}^2/\text{s}^2 < C < 7\ 630\ \text{m}^2/\text{s}^2$ 的范围内,第一类周期轨道的位置和径向振荡幅值均能满足女凯龙星外环的位置分布范围。

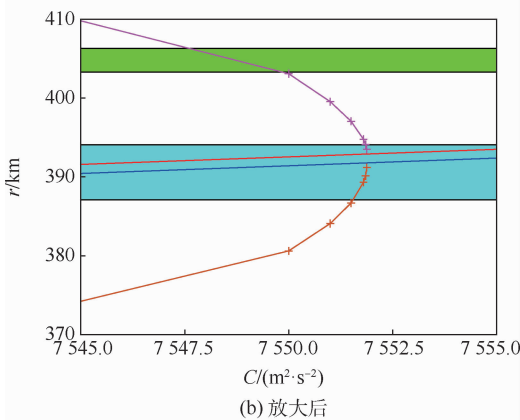
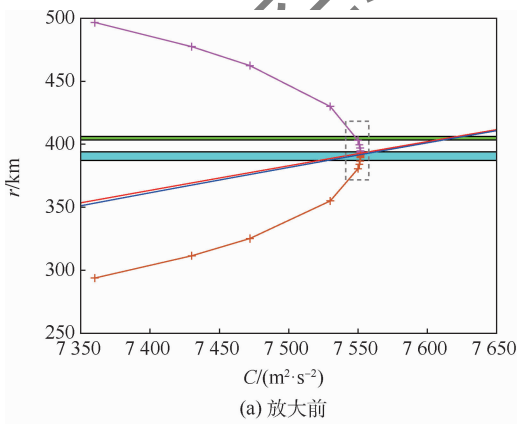
将图 11(a)虚线框中的图像放大得到图 11(b),可以发现,在 $7\ 551\ \text{m}^2/\text{s}^2 < C < 7\ 552\ \text{m}^2/\text{s}^2$ 的狭小区间内,1:3共振周期轨道的径向振荡幅值快速减小,直到进入女凯龙星内环的位置分布范围。

为了更为直观地观察轨道的径向振荡范围与女凯龙星环的关系,根据文献[9]的方法,如图 12所示,使用等效半长轴 a_{eq} 和等效偏心率 e_{eq} 来进行研究。由于在固连坐标系中第一类周期轨道的轨迹类似于以女凯龙星为中心的椭圆,而不是像开普勒椭圆轨道以中心天体为一个焦点。因此,

利用式(11)将其等效为椭圆,其近心点和远心点的轨道半径分别与第一类周期轨道的轨道半径最小值和最大值相同。这样,以女凯龙星为焦点,半长轴为 a_{eq} 且偏心率为 e_{eq} 的等效椭圆将覆盖与第一类周期轨道相同的径向振荡范围。对于 1:3 共振周期轨道,由于其在惯性坐标系中的轨迹(一个周期内)本身就类似于一个开普勒椭圆,同样适用式(11)。类似地,对于女凯龙星环,其标称位置也可由一系列符合该径向范围的等效开普勒椭圆表示。

$$\begin{cases} a_{\text{eq}} = \frac{r_{\text{min}} + r_{\text{max}}}{2} \\ e_{\text{eq}} = 1 - \frac{r_{\text{min}}}{a_{\text{eq}}} \end{cases} \quad (11)$$

将 2 类周期轨道和女凯龙星 2 个星环等效开普勒椭圆的 $(a_{\text{eq}}, e_{\text{eq}})$ 绘制在图 13 中,其中淡蓝色区域和绿色区域分别代表女凯龙星内环和外环所在区域,深蓝色色线和棕红色色线分别代表第一类周期轨道和 1:3 共振周期轨道所在位置。



- 女凯龙星的内环
- 女凯龙星的外环
- 1:3共振周期轨道的最大半径
- 1:3共振周期轨道的最小半径
- 第一类周期轨道的最大半径
- 第一类周期轨道的最小半径

图 11 两类周期轨道的轨道半径 r 随 C 的变化
Fig. 11 Change of orbit radius r with C of two kinds of periodic orbits

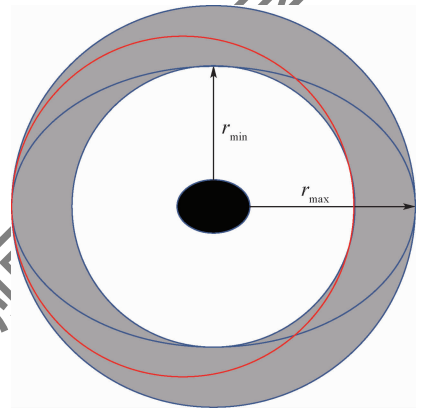


图 12 a_{eq} 和 e_{eq} 的定义示意图

Fig. 12 Schematic diagram for definition of a_{eq} and e_{eq}

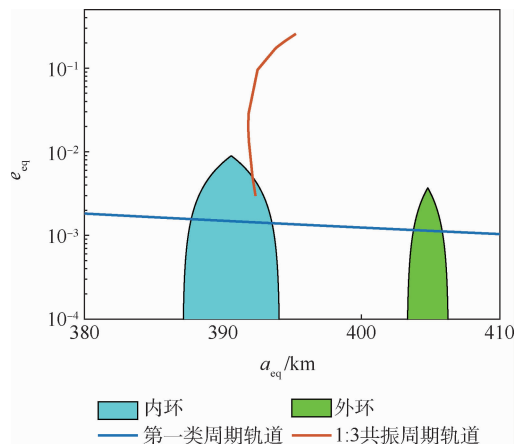


图 13 两类周期轨道和女凯龙星环的 $a_{\text{eq}}-e_{\text{eq}}$ 图
Fig. 13 $a_{\text{eq}}-e_{\text{eq}}$ map of two kinds of periodic orbits and 10199 Chariklo's rings

不难发现,在女凯龙星的2个星环附近,第一类周期轨道的最大偏心率小于1:3共振周期轨道的最小偏心率,意味着几乎所有的1:3共振周期轨道都显示出比第一类周期轨道更大的径向振荡幅值。因此,分布于女凯龙星内环范围内的1:3共振周期轨道只有非常小的区间,而第一类周期轨道可以在很大范围内满足女凯龙星内环和外环的位置分布范围。

图14为 $C=7\ 551.8\ \text{m}^2/\text{s}^2$ 时(1:3共振周期轨道的径向振荡幅值能够满足女凯龙星环的位置分布范围)2类周期轨道的对比。结合图3庞加莱截面相图的结果,可以认为1:3共振周期轨道不断趋近于第一类周期轨道,直到“退化”为第一类周期轨道或第一类准周期轨道。

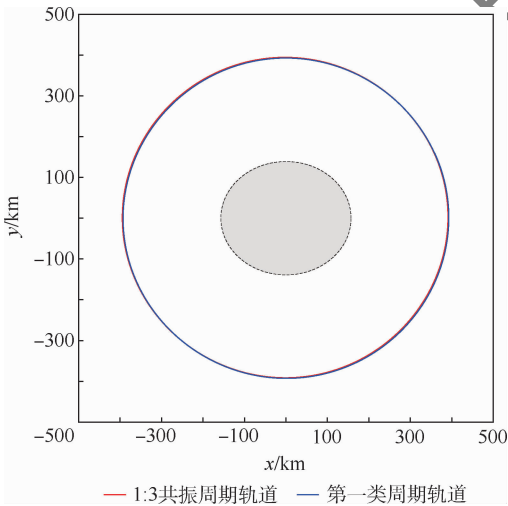


图14 极限范围内2类周期轨道在惯性坐标系中的对比
Fig. 14 Comparison of two kinds of periodic orbits in limit range in the inertial frame

而对于2类准周期轨道,随着其KAM环的不断增大,径向振荡的幅值增大,幅值的最大增量就取决于稳定岛的大小。通过对相图进行分析并计算发现,极限范围处稳定岛非常小,因此准周期轨道与其对应周期轨道的径向振荡幅值差别不大。因此,极限范围内的部分准1:3共振周期轨道和部分第一类准周期轨道径向振荡幅值可能会满足女凯龙星内环的径向范围,而部分第一类准周期轨道也会满足女凯龙星外环的径向范围。

对2类周期轨道进行稳定性的定性分析发现:①在庞加莱截面相图中,稳定岛的大小可以代表周期轨道的稳定性。因为满足内环径向振荡幅值范围的狭小 C 范围内的1:3共振周期轨道稳定岛非常小,其稳定性比较差;而满足内环径向振荡幅值范围的第一类周期轨道的 C 范围很广,稳定岛较大,稳定性较好。②由于现实中存在粒子碰

撞和其他因素带来的能量改变,而 C 反映粒子在固连坐标系中的能量,实际上粒子的 C 并不是恒定的。因为满足内环径向振荡幅值范围的1:3共振周期轨道只有很窄的 C 范围,所以如果 C 改变过大便不再满足内环的范围;而满足范围的第一类周期轨道的 C 范围很广,即使 C 发生较大改变仍旧可以满足内环和外环的范围。由此看来,第一类周期轨道相比1:3共振周期轨道的稳定性更好,女凯龙星内环中的粒子更有可能处在第一类周期轨道上。

通过上述对比和分析发现,女凯龙星外环中的粒子不可能处在1:3共振周期轨道上,只可能处在第一类周期轨道及其附近的准周期轨道上;而内环中的粒子最有可能处于满足径向振荡幅值范围的第一类周期轨道及其附近的准周期轨道上,但也不可能处于满足径向振荡幅值范围的1:3共振周期轨道及其附近的准周期轨道上。

4 结论

本文考虑小天体扁率、赤道椭率和自转,建立了小天体引力场中的轨道动力学模型,并进行数值仿真研究,利用庞加莱截面方法,得到第一类周期轨道和1:3共振周期轨道,并分别分析其轨道特性,揭示了小天体光环与各类周期轨道的内在联系,主要结论如下:

- 1) 经过估算得出1:3共振周期轨道的位置,发现女凯龙星环的位置与之非常接近,即女凯龙星环可能与1:3共振周期轨道相关。
- 2) 通过KAM环迭代,在环的位置附近得到了第一类周期轨道和1:3共振周期轨道,并分别研究了其轨道特性。
- 3) 通过对小天体附近2类周期轨道及其附近的准周期轨道进行对比和分析,最终发现,女凯龙星内环中的粒子最有可能处在满足径向振荡幅值范围的第一类周期轨道及其附近的准周期轨道上,但也不能排除处在满足径向振荡幅值范围的1:3共振周期轨道及其附近的准周期轨道上;外环中的粒子不可能处在1:3共振周期轨道上,只可能处在第一类周期轨道及其附近的准周期轨道上。

本文采用的二阶二次球谐引力场模型没有考虑小天体引力场的不规则性、环内粒子的局部黏性、相互引力作用和碰撞等其他因素的影响。此外,本文只关注了粒子在较短时间内的轨道动力学问题,没有涉及大时间尺度下的长期轨道演化问题。在实际中,小天体环粒子经受着更加复杂

的动力学环境,会表现出更加复杂的动力学特征,需要后续更加深入的研究。

参考文献 (References)

- [1] BRAGA-RIBAS F, SICARDY B, ORTIZ J L, et al. A ring system detected around the Centaur (10199) Chariklo [J]. *Nature*, 2014, 508 (7494) : 72-75.
- [2] KONDRATYEV B P. Dynamics of Centaur Chariklo and evolution of its rings [J]. *Astrophysics and Space Science*, 2016, 361 (12) : 389.
- [3] ORTIZ J L, SANTOS-SANZ P, SICARDY B, et al. The size, shape, density and ring of the dwarf planet Haumea from a stellar occultation [J]. *Nature*, 2017, 550 (7675) : 219-223.
- [4] ORTIZ J L, DUFFARD R, PINILLA-ALONSO N, et al. Possible ring material around Centaur (2060) Chiron [J]. *Astronomy & Astrophysics*, 2015, 576: A18.
- [5] HYODO R, CHARNOZ S, GENDA H, et al. Formation of Centaurs' rings through their partial tidal disruption during planetary encounters [J]. *Astrophysical Journal*, 2016, 828 (1) : L8.
- [6] PAN M, WU Y. On the mass and origin of Chariklo's rings [J]. *Astrophysical Journal*, 2016, 821 (1) : 18.
- [7] BÉRARD D, SICARDY B, CAMARGO J J B, et al. The structure of Chariklo's rings from stellar occultations [J]. *The Astronomical Journal*, 2017, 154 (4) : 144.
- [8] GUPTA A, NADKARNI GHOSH S, SHARMA I. Rings of non-spherical, axisymmetric bodies [J]. *Icarus*, 2018, 299: 97-116.
- [9] WINTER O C, BORDERES-MOTTA G, RIBEIRO T. On the location of the ring around the dwarf planet Haumea [J]. *Monthly Notices of the Royal Astronomical Society*, 2019, 484 (3) : 3765-3771.
- [10] SICARDY B. Resonances in nonaxisymmetric gravitational potentials [J]. *The Astronomical Journal*, 2020, 159 (3) : 102.
- [11] SICARDY B, LEIVA R, RENNER S, et al. Ring dynamics around non-axisymmetric bodies with application to Chariklo and Haumea [J]. *Nature Astronomy*, 2018, 3 (2) : 146-153.
- [12] 李俊峰, 宝音贺西, 蒋方华. 深空探测动力学与控制 [M]. 北京: 清华大学出版社, 2014: 307-315.
- [13] LI J F, BAOYIN H X, JIANG F H. Dynamics and control of interplanetary flight [M]. Beijing: Tsinghua University Press, 2014: 307-315 (in Chinese).
- [14] TUCKER W. Computing accurate Poincaré maps [J]. *Physica D Nonlinear Phenomena*, 2002, 171 (3) : 127-137.
- [15] MOTTA G B, WINTER O C. Poincaré surfaces of section around a 3-D irregular body: The case of asteroid 4179 Toutatis [J]. *Monthly Notices of the Royal Astronomical Society*, 2017, 474 (2) : 2452-2466.
- [16] 吴晓杰, 王悦, 徐世杰. 火卫一周期准卫星轨道及入轨分析 [J]. *北京航空航天大学学报*, 2020, 46 (6) : 1133-1141.
- [17] WU X J, WANG Y, XU S J. Periodic quasi-satellite orbits around Phobos and their injections [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2020, 46 (6) : 1133-1141 (in Chinese).

Orbital dynamics of rings of small bodies

WU Tailong, WANG Yue*

(School of Astronautics, Beihang University, Beijing 102206, China)

Abstract: With the development of astronomical observations, it has been found that, not only Saturn, Jupiter and other giant planets, but also some small bodies are surrounded by rings. Focussing on rings of the Centaur 10199 Chariklo, we study the orbital dynamics of particles in the gravitational field of Chariklo, and analyze the impacts of equatorial ellipticity and rotation of Chariklo on the orbital motion of particles. The first kind of periodic orbits and 1:3 resonant periodic orbits are obtained by using KAM torus iteration on the Poincaré sections. The relationship between the position of rings and the mean motion resonances are revealed by analyzing the properties of the periodic orbits. The results show that, the particles in the inner ring of Chariklo are most likely associated with the first kind of periodic orbits and their quasi-periodic orbits, but are also likely associated with the 1:3 resonant periodic orbits and their quasi-periodic orbits. The particles in the outer ring of Chariklo are not associated with the 1:3 resonant periodic orbits, but are associated with the first kind of periodic orbits and their quasi-periodic orbits.

Keywords: rings of small bodies; 10199 Chariklo; Poincaré section; mean motion resonance; resonant periodic orbit

Received: 2021-01-05; Accepted: 2021-03-28; Published online: 2021-04-26 09:39

URL: kns.cnki.net/kcms/detail/11.2625.V.20210425.1806.002.html

Foundation items: National Natural Science Foundation of China (11872007); the Fundamental Research Funds for the Central Universities

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

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

DOI: 10.13700/j.bh.1001-5965.2021.0015

基于平面变换的高精度相机标定方法

彭妍¹, 郭君斌^{1,*}, 于传强¹, 柯冰²

(1. 火箭军工程大学 导弹工程学院, 西安 710025; 2. 火箭军装备部, 北京 100089)

摘 要: 针对传统方法未考虑透视偏差的存在, 提取椭圆圆心作为真实物理圆心投影点造成相机标定误差的问题, 提出了一种基于平面变换的高精度相机标定方法。提取标定板内外边框上的角点, 对标定板进行平面变换, 将标记点由椭圆投影成近似的标准圆; 利用图像矩提取标准圆圆心的坐标, 投影回原标定板平面, 得到标记点实际圆心的像素坐标; 根据圆形标记点实际圆心的坐标, 利用张正友标定法完成相机标定。实验结果表明: 与传统方法相比, 所提方法将相机标定的误差降低了66.169%, 有效提高了相机标定的精度。

关键词: 相机标定; 张正友标定; 圆形标记点; 圆心检测; 透视变换

中图分类号: TP391

文献标志码: A

文章编号: 1001-5965(2022)07-1297-07

相机标定建立起三维空间的世界坐标与二维图像的像素坐标之间的成像模型, 在计算机视觉领域发挥着十分重要的作用。目前, 相机标定方法主要有自标定法、主动视觉标定法和传统标定法。其中, 自标定法^[1-3]不需要标定物, 比较灵活, 但鲁棒性弱, 精度较差; 主动视觉标定法^[4]原理简单, 易于线性求解, 但受限于相机运动参数未知的应用场景; 传统标定法精度较高, 广泛应用于高精度测量和三维重建领域, 但需要标定物, 典型方法有直接线性变换标定法^[5]、Tsai 两步标定法^[6]和张正友标定法^[7]等。

张正友标定法仅需要一个棋盘格图案的平面标定板来完成相机标定, 在实际使用中具有操作简便、标定精度高等优点, 因而成为当今高精度工业测量领域使用最广泛的标定方法之一。但棋盘格角点检测精度低, 角点提取精度对噪声和图像质量比较敏感^[8], 相反, 圆形特征不仅对噪声的抑制性比较强, 而且检测精度高^[9], 因此以圆为标记点的平面标定板在高精度相机标定中被广泛采纳^[8-12]。

利用以圆为标记点的平面标定板进行相机标定时, 圆形标记点圆心坐标的提取精度决定着相机标定的精度^[13]。由于透视偏差的存在, 在标定板中提取到的椭圆圆心并不是真实物理圆心的投影点^[14]。而传统方法通常提取椭圆圆心作为真实物理圆心的投影点, 因此基于传统方法的相机标定方法必然会引入误差。针对传统方法存在圆心提取偏差的问题, 卢晓冬等^[10]提出了一种基于圆心真实图像坐标计算的高精度相机标定方法, 通过多次迭代计算解决透视投影图像椭圆中心偏差使相机标定产生误差的问题; 吴建霖等^[15]提出了一种基于三同心圆的圆形标志投影偏心差补偿算法, 根据3组椭圆拟合中心坐标解算偏心差模型进行计算; 解则晓和王晓敏^[16]提出了一种基于对偶二次曲线几何特性的圆心提取算法。针对传统方法存在圆形标记点圆心提取偏差造成相机标定误差的问题, 本文提出了一种基于平面变换的高精度相机标定方法。首先, 提取标定板内外边框上的角点, 通过透视变换使标定板平面和相机成像平面平行, 将标记点由椭圆投影成近似的标

收稿日期: 2021-01-11; 录用日期: 2021-04-09; 网络出版时间: 2021-04-19 17:33

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210419.1445.002.html

基金项目: 国家自然科学基金(61501470)

* 通信作者。E-mail: gjb_202@163.com

引用格式: 彭妍, 郭君斌, 于传强, 等. 基于平面变换的高精度相机标定方法[J]. 北京航空航天大学学报, 2022, 48(7): 1297-1303. PENG Y, GUO J B, YU C Q, et al. Calibration method for high precision camera based on plane transformation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1297-1303 (in Chinese).

准圆;然后,利用图像矩提取标准圆圆心的坐标,并反透视变换回原标定板平面,获取标记点实际圆心的像素坐标;最后,根据提取到的圆形标记点实际圆心的坐标,利用张正友标定法完成相机标定。实验结果表明,与传统方法相比,本文方法有效提高了相机标定的精度。

1 相机标定原理

在理想情况下,相机成像模型本质上是一个针孔模型。假设空间中一点 P 的三维坐标为 $[X_w, Y_w, Z_w]^T$,在图像上投影点 p 的像素坐标为 $[u, v]^T$,对应齐次坐标分别为 $\tilde{P} = [X_w, Y_w, Z_w, 1]^T$ 和 $\tilde{p} = [u, v, 1]^T$,则成像模型表示为

$$\begin{cases} s\tilde{p} = K[R \ t] \tilde{P} \\ K = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \end{cases} \quad (1)$$

式中: s 为任意比例因子; K 为内参矩阵; R 和 t 分别为从世界坐标系到相机坐标系的旋转和平移矩阵,共同构成外参矩阵; (u_0, v_0) 为图像的主点坐标; f_x 和 f_y 分别为图像在 x 和 y 轴上的有效焦距; γ 为倾斜因子。

在实际情况中,透镜在制造及组装过程中不可避免地会引入几何畸变,相机成像模型不可能达到完全理想的线性模型,为了准确表达相机成像模型,必须考虑透镜畸变的影响。其中,非线性的畸变模型表示为

$$\begin{cases} x_u = x_d + \delta_x(x_d, y_d) \\ y_u = y_d + \delta_y(x_d, y_d) \end{cases} \quad (2)$$

式中: (x_d, y_d) 为理想情况下的成像点坐标, (x_u, y_u) 为实际发生畸变的成像点坐标; $\delta_x(x_d, y_d)$ 和 $\delta_y(x_d, y_d)$ 分别表示投影点坐标为 (x_d, y_d) 时在 x 和 y 轴上发生的畸变量。

透镜的畸变主要分为径向畸变、切向畸变和薄透镜畸变等。其中,径向和切向畸变影响最为显著。因此,畸变量主要考虑径向畸变和切向畸变。

$$\begin{cases} \delta_x(x_d, y_d) = \underbrace{x_d(k_1 r^2 + k_2 r^4 + k_3 r^6)}_{\text{径向畸变}} + \underbrace{[2p_1 x_d y_d + p_2 (r^2 + 2x_d^2)]}_{\text{切向畸变}} \\ \delta_y(x_d, y_d) = \underbrace{y_d(k_1 r^2 + k_2 r^4 + k_3 r^6)}_{\text{径向畸变}} + \underbrace{[2p_2 x_d y_d + p_1 (r^2 + 2y_d^2)]}_{\text{切向畸变}} \end{cases} \quad (3)$$

式中: k_1, k_2, k_3 为径向畸变系数; p_1, p_2 为切向畸变系数; $r^2 = x_d^2 + y_d^2$ 。

除了透镜畸变,相机成像模型不可避免会受到噪声的影响,为了得到最优的相机标定参数,张正友标定法通过构造重投影误差的目标函数,利用极大似然估计提升所有参数的精度。根据提取到的特征点坐标,将式(1)得到的理想情况下的内外参数矩阵,以及式(2)和式(3)利用最小二乘法得到的透镜畸变系数,作为优化函数的初始值,避免参数陷入局部最优值。其中的目标函数为

$$\sum_{i=1}^n \sum_{j=1}^m \|p_{ij} - \hat{p}_{ij}(K, D, R_i, t_i, P_j)\|^2 \quad (4)$$

式中: n 为标定图像的数量; m 为每张标定图像上特征点的数量; D 为相机的畸变系数; R_i 和 t_i 为每幅图像对应的外参矩阵; p_{ij} 和 $\hat{p}_{ij}(K, D, R_i, t_i, P_j)$ 分别为 i 张图像上第 j 个特征点实际的投影点和经过相机成像模型得到的投影点。

2 高精度相机标定方法

本文提出了一种基于平面变换的高精度相机标定方法,整体流程如下:

步骤1 采集标定图像,对图像进行预处理。

步骤2 提取标定板内外边框上的角点,对标定板平面进行透视变换,使标定板平面与相机成像平面平行。

步骤3 利用图像矩提取标准圆圆心的像素坐标。

步骤4 对标定板平面进行反透视变换,提取圆形标记点实际圆心的像素坐标。

步骤5 利用张正友标定法完成相机标定。

2.1 图像采集及预处理

本文采用由7行7列共49个等直径的黑色圆形标记点构成的点阵标定板,如图1所示。将相机固定不动,变换标定板的姿态和位置,拍摄多张标定板在不同视角下的图片。对采集到的图片依次进行预处理:首先将图像转换为灰度图像,然后利用高斯滤波对图像进行去噪,最后利用最大类间方差法对图像进行二值化。预处理结果如图2所示。最大类间方差法是一种自动确定分割阈值的二值化算法,对图像的亮度和对比度不敏感,在图像处理中应用广泛,但对噪声敏感,故在应用最大类间方差法前需要对图像去噪。

2.2 平面变换

透视变换的本质是利用透视变换矩阵将图像由一个视平面变换到一个新视平面上,表示为

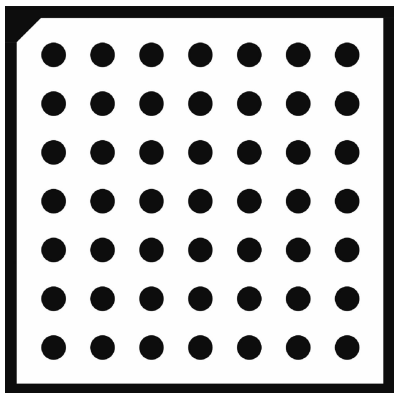
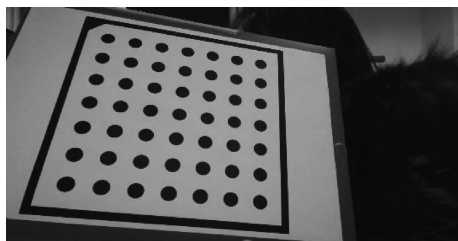
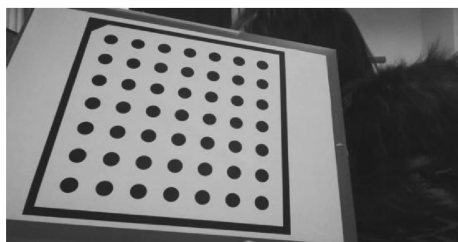


图 1 点阵标定板

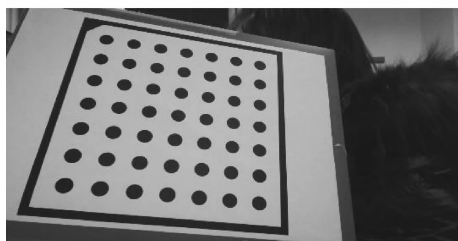
Fig. 1 Lattice calibration plate



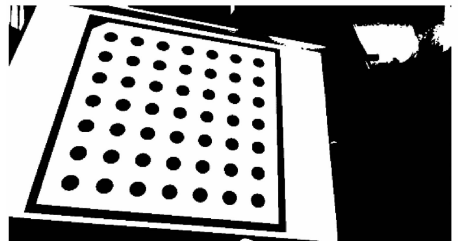
(a) 原图



(b) 灰度图



(c) 去噪图



(d) 二值化图

图 2 图像预处理

Fig. 2 Image pre-processing

$$\begin{cases} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = T \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\ T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \end{cases} \quad (5)$$

式中: T 为透视变换矩阵; (u, v, w) 和 (x, y, z) 分别为图像经过透视变换前后的坐标。

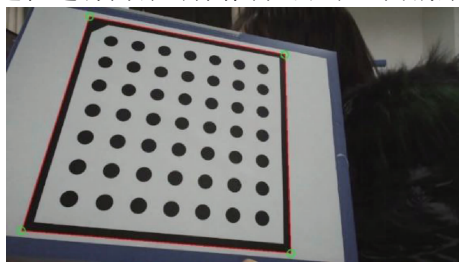
点阵标定板分别采用一个五边形和四边形作为标定板黑色边界的内外边框,边框是一种线特征,不会像圆一样存在透视偏差的问题。另外,透视偏差的大小与椭圆中心提取精度无关,只取决于相机成像平面与标定板平面之间的夹角^[10]。本文通过定位到标定板的内外边框,利用内外边框上的角点,通过透视变换使标定板平面与成像平面平行,避免透视偏差对圆形标记点实际圆心坐标提取的干扰。

假设图像经过透视变换前后的像素坐标分别为 (u', v') 和 (x', y') , 对式(5)进行求解得到

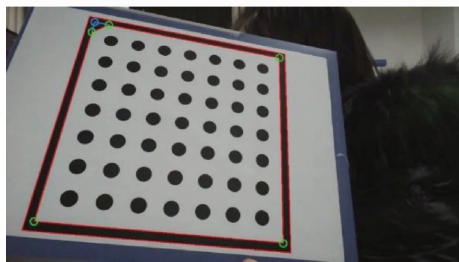
$$\begin{cases} x' = \frac{x}{z} = \frac{a_{11}u' + a_{21}v' + a_{31}}{a_{13}u' + a_{23}v' + a_{33}} = \frac{a_{11}u' + a_{21}v' + a_{31}}{a_{13}u' + a_{23}v' + 1} \\ y' = \frac{y}{z} = \frac{a_{12}u' + a_{22}v' + a_{32}}{a_{13}u' + a_{23}v' + a_{33}} = \frac{a_{12}u' + a_{22}v' + a_{32}}{a_{13}u' + a_{23}v' + 1} \end{cases} \quad (6)$$

由式(6)可知,通过 4 对像素坐标就可以得到透视变换矩阵,由于图像可能存在噪声、角点误提取等问题,一般需要 4 对以上的像素坐标求解出透视变换矩阵,进而完成平面变换。

首先,对图像的轮廓进行检测,综合轮廓的面积、长度约束进行筛选,定位到标定板的外边框上;然后,利用 Shi-Tomasi 算法对外边框的角点进行提取,为使提取到的角点坐标更加准确,进一步提高平面变换的准确性,将角点坐标精确至亚像素级,如图 3(a)所示。为了进一步提高平面变换的鲁棒性,需要获取更多的角点坐标。对标定板的内边框进行同样的操作,检测到 5 个角点,但一



(a) 提取外边框的角点



(b) 提取内边框的角点

图 3 提取边框的角点

Fig. 3 Extraction of border corners

般五边形的角度信息未知,因此将五边形的2条边进行延长形成一个四边形,交点处得到一个新的角点,如图3(b)所示。

由于图像可能存在噪声、角点误提取等问题,得到的亚像素级角点坐标可能并不都是准确的,利用随机抽样一致(random sample consensus, RANSAC)算法,通过迭代的思想,从内外边框包含的8个亚像素级的角点坐标中估算出最优的透视变换矩阵,通过透视变换使标定板平面与成像平面平行,结果如图4所示。可以看到,标记点由椭圆投影为近似的标准圆。

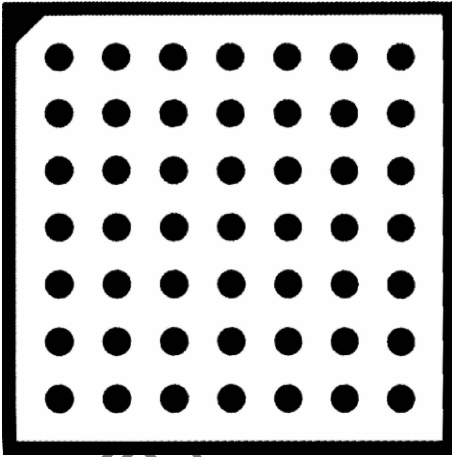


图4 经过透视变换的标定板平面

Fig. 4 Calibration plate plane after perspective transformation

2.3 提取标准圆的圆心

图像矩本质上是一种通过对图像的灰度值进行加权来描述图像特征的算子。一幅 $h \times l$ 的数字图像的 $p+q$ 阶矩表示为

$$m_{pq} = \sum_{i=1}^h \sum_{j=1}^l u^p v^q f(u, v) \quad (7)$$

式中: $f(u, v)$ 为图像在像素坐标 (u, v) 处的灰度值。

根据式(7),利用图像的零阶矩 m_{00} 和一阶矩 m_{10} 、 m_{01} ,得到图像的质心坐标 (\bar{u}, \bar{v}) :

$$\begin{cases} \bar{u} = \frac{m_{10}}{m_{00}} \\ \bar{v} = \frac{m_{01}}{m_{00}} \end{cases} \quad (8)$$

经过平面变换后,标记点近似为标准圆,利用图像矩计算标记点的质心,进而近似得到标准圆圆心的像素坐标,如图5所示。

2.4 提取圆形标记点的实际圆心

反透视变换就是透视变换的逆过程,本质是将图像由新视平面变换到原视平面上,表示为

$$\begin{cases} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = T^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ T_{inv} = T^{-1} \end{cases} \quad (9)$$

式中: T 和 T_{inv} 分别为透视变换和反透视变换矩阵。

根据提取到的亚像素级的角点坐标,利用RANSAC算法得到最优的透视变换和反透视变换矩阵。提取完标准圆圆心的坐标后,利用反透视变换矩阵,将圆心投影回原标定板平面,得到圆形标记点实际圆心的像素坐标,如图6所示。

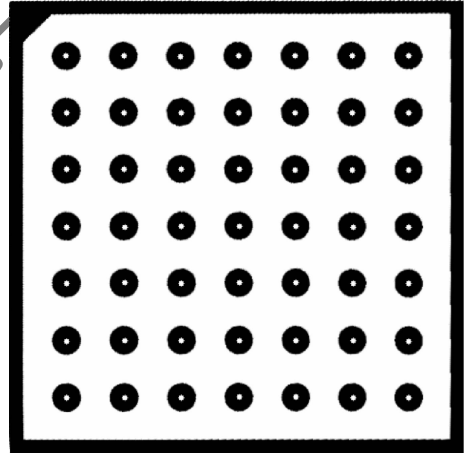


图5 提取标准圆的圆心

Fig. 5 Extraction of center of standard circle

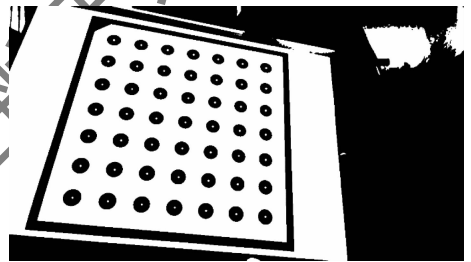


图6 提取圆形标记点的圆心

Fig. 6 Extraction of center of circular markers

2.5 相机标定

利用标定板的五边形内边框来自动确定标定板世界坐标系和像素坐标系的轴线分布,使得每个圆形标记点实际圆心的世界坐标与像素坐标一一对应,再利用张正友标定法完成相机标定。

像素坐标系的轴线分布如图7所示。首先,对标定板平面进行透视变换,将标定板外边框的4个角点 G 、 H 、 I 、 J 作为变换后平面的4个顶点,选取其中距离 AB 最近的角点 G 作为像素坐标系的原点,保证内边框最短边长 AB 始终在标定板平面的左上方;然后,从距离 AB 最近的圆心点 P 开始,从左到右,从上到下,按顺序得到各个标准圆圆心的像素坐标;最后,通过反透视变换按顺序得到各个圆形标记点实际圆心的像素坐标。

世界坐标系的轴线分布如图 8 所示。假设标定板在世界坐标系为 $z=0$ 的平面上,选择距离内

边框最短边长 AB 最近标记点的圆心 Q 作为世界坐标系的原点,水平向右和垂直向下分别建立 x 、 y 坐标系,已知标定板规格,从左到右,从上到下,按顺序得到各个标记点圆心的三维坐标。

3 实验结果与分析

本文实验采用的相机分辨率为 $1\ 280 \times 720$ 像素,点阵标定板打印在 A4 纸上,标定板的规格如表 1 所示。实验环境在 CPU 为 Intel core i5 1.80 GHz,操作系统为 Windows 10 的 PC 上的 Visual Studio 2019 下,配合图像处理开源库 OpenCV 3.2.0。

表 1 点阵标定板规格

Table 1 Specification of lattice calibration plate

参数	数值
标定板尺寸/mm	200 × 200
圆点数量/个	7 × 7
圆点直径/mm	12.5
圆心距/mm	25
外边框长度/mm	200
内边框长度/mm	187.5

将相机固定不动,对不同姿态和位置下的标定板进行拍摄,一共拍摄 16 幅不同的图像。对拍摄的 16 幅图像,先分别通过本文方法和传统方法提取圆形标记点的圆心,再利用张正友标定法分别进行标定,得到的标定结果如表 2 所示。可以看出,本文方法得到的主点坐标为 (616.318, 333.515) 像素,比传统方法更接近于图像实际的主点坐标 (640, 360) 像素, x 和 y 轴上的有效焦距分别为 (920.942, 918.769) 像素,同样比传统方法更接近于图像实际的焦距 (960, 960) 像素,并且径向畸变系数和切向畸变系数都不大,在理论值范围之内。

在相机标定中,通常采用重投影误差来判定相机标定精度。重投影误差是指利用标定得到的相机内外参数和畸变系数构建相机成像模型,对

表 2 相机标定结果对比

Table 2 Comparison of camera calibration results

参数	传统方法	本文方法	
焦距	f_x /像素	904.829	920.942
	f_y /像素	905.350	918.769
主点坐标	u_0 /像素	604.321	616.318
	v_0 /像素	344.552	333.515
镜头畸变系数	k_1	0.016 2	0.003 4
	k_2	0.010 2	0.002 0
	k_3	-0.103 4	-0.028 9
	p_1	0.001 3	-0.000 8
	p_2	-0.003 2	0.001 2

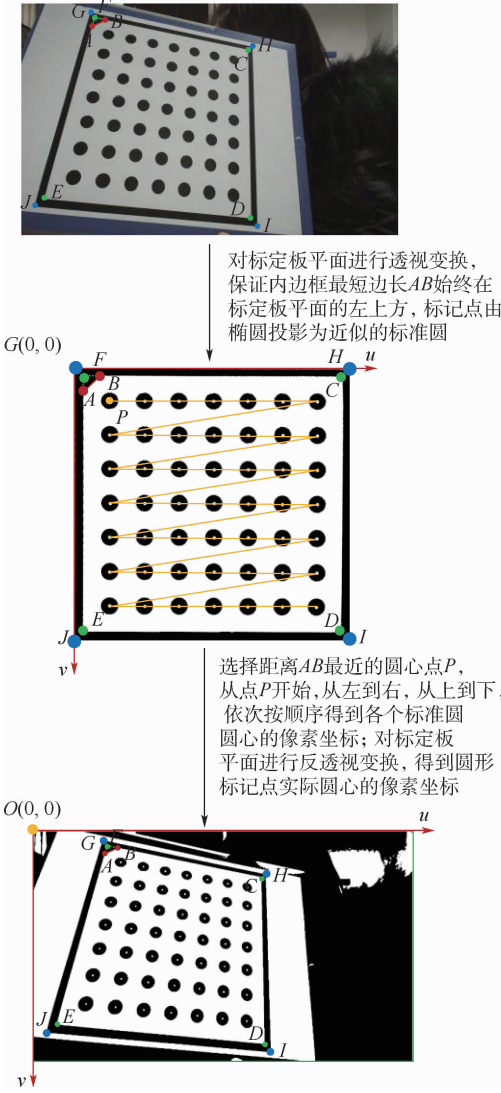


图 7 点阵标定板的像素坐标系轴线分布

Fig. 7 Axis distribution of pixel coordinate system of lattice calibration plate

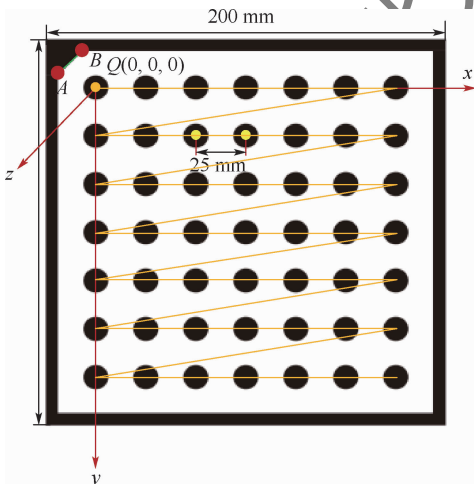


图 8 点阵标定板的世界坐标系轴线分布

Fig. 8 Axis distribution of world coordinate system of lattice calibration plate

空间的三维点重新投影,得到的投影点坐标与实际成像点坐标之间的偏差。一般重投影误差越小,相机标定的精度越高。本文方法和传统方法对于每幅图像的重投影误差和对于所有图像的平均重投影误差分别如表3和表4所示。由表3可知,与传统方法相比,本文方法的重投影误差降低了54.921%~75.410%,并且均低于传统方法。由表4可知,本文方法的平均重投影误差为0.034 0,与传统方法相比,降低了66.169%。综合表3和表4,说明本文方法极大提高了相机标定的精度,同时也证明了其可行性和有效性。

表3 相机标定的重投影误差对比

Table 3 Comparison of re-projection errors in camera calibration

图像序号	重投影误差/像素	
	传统方法	本文方法
1	0.077 992 668	0.027 848 8
2	0.160 459 9	0.050 742 9
3	0.099 213 23	0.024 396 3
4	0.139 651 5	0.050 264 5
5	0.066 711 65	0.023 466 5
6	0.125 555 3	0.045 840 3
7	0.097 389 74	0.037 083 4
8	0.128 29	0.057 831 3
9	0.099 505 6	0.034 262 6
10	0.151 173	0.043 132 7
11	0.049 206	0.018 771 4
12	0.070 045 991	0.024 623 1
13	0.089 888 6	0.023 999 7
14	0.112 454	0.036 633 7
15	0.060 713 33	0.022 829 6
16	0.079 800 88	0.022 511 8

表4 相机标定的平均重投影误差对比

Table 4 Comparison of average re-projection errors in camera calibration

标定方法	平均重投影误差/像素
传统方法	0.100 5
本文方法	0.034 0

4 结 论

1) 针对圆形标记点存在透视偏差造成圆心提取有误的问题,提出了一种基于平面变换的圆心检测方法,在此基础上,结合张正友标定法,提出了一种基于平面变换的高精度相机标定方法。

2) 在拍摄的16张标定图像下,本文提出的标定方法与传统方法进行了实验,本文方法的平均重投影误差为0.034 0,与传统方法相比降低了66.169%,证明了本文方法的可行性和有效性。

参考文献 (References)

- [1] FAUGERAS O D, LUONG Q T, MAYBANK S J. Camera self-calibration: Theory and experiments[J]. Lecture Notes in Computer Science, 1992, 588(12): 321-334.
- [2] MAYBANK S J, FAUGERAS O D. A theory of self-calibration of a moving camera[J]. International Journal of Computer Vision, 1992, 8(2): 123-151.
- [3] LI J, YANG Y, FU G. Camera self-calibration method based on GA-PSO algorithm[C] // Proceedings of IEEE International Conference on Cloud Computing and Intelligence Systems. Piscataway: IEEE Press, 2011: 149-152.
- [4] 王宪, 谭建平, 陈国强, 等. 基于主动视觉的测量装置精度校准方法[J]. 中南大学学报(自然科学版), 2014, 45(4): 1071-1076.
WANG X, TAN J P, CHEN G Q, et al. Active vision-based accuracy calibration technology for measurement devices[J]. Journal of Central South University (Science and Technology), 2014, 45(4): 1071-1076 (in Chinese).
- [5] ABDEL-AZIZ Y I, KARARA H M. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry[J]. Photogrammetric Engineering and Remote Sensing, 2015, 81(2): 103-107.
- [6] TSAI R Y. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses[J]. IEEE Journal on Robotics and Automation, 2003, 3(4): 323-344.
- [7] ZHANG Z Y. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [8] 秦凡路, 刘建军, 任鑫, 等. 基于圆形标志点的深空探测全景相机标定方法[J]. 光学学报, 2013, 33(11): 139-145.
WU F L, LIU J J, REN X, et al. Deep space exploration panoramic camera calibration technique based on circular markers[J]. Acta Optica Sinica, 2013, 33(11): 139-145 (in Chinese).
- [9] 汪首坤, 赵金枝, 姜明, 等. 基于圆形阵列标定板的张氏相机标定法[J]. 北京理工大学学报, 2019, 39(8): 92-96.
WANG S K, ZHAO J Z, JIANG M, et al. Zhang's camera calibration method based on circular array calibration board[J]. Transactions of Beijing Institute of Technology, 2019, 39(8): 92-96 (in Chinese).
- [10] 卢晓冬, 薛俊鹏, 张启灿. 基于圆心真实图像坐标计算的高精度相机标定方法[J]. 中国激光, 2020, 47(3): 234-241.
LU X D, XUE J P, ZHANG Q C. High camera calibration method based on true coordinate computation of circle center[J]. Chinese Journal of Lasers, 2020, 47(3): 234-241 (in Chinese).
- [11] 段振云, 王宁, 赵文辉, 等. 基于点阵标定板的视觉测量系统的标定方法[J]. 光学学报, 2016, 36(5): 143-151.
DUAN Z Y, WANG N, ZHAO W H, et al. Calibration method based on lattice calibration plate in vision measurement system[J]. Acta Optica Sinica, 2016, 36(5): 143-151 (in Chinese).
- [12] 李中伟, 王从军, 史玉升. 3D测量系统中的高精度摄像机标定算法[J]. 光电工程, 2008, 35(4): 58-63.

- LI Z W, WANG C J, SHI Y S. High precision camera calibration algorithm for 3D measurement system[J]. Opto-Electronic Engineering, 2008, 35(4): 58-63 (in Chinese).
- [13] 孔玮琦, 刘京南, 达飞鹏, 等. 基于一般成像模型的微小物体测量系统标定方法[J]. 光学学报, 2016, 36(9): 187-198.
KONG W Q, LIU J N, DA F P, et al. Calibration method based on general imaging model for micro-object measurement system [J]. Acta Optica Sinica, 2016, 36(9): 187-198 (in Chinese).
- [14] 魏振忠, 张广军. 透视投影变换中椭圆中心畸变误差模型及其仿真研究[J]. 仪器仪表学报, 2003, 24(2): 160-164.
WEI Z Z, ZHANG G J. A distortion error model of the perspective projection of ellipse center and its simulation [J]. Chinese Journal of Scientific Instrument, 2003, 24(2): 160-164 (in Chinese).
- [15] 吴建霖, 蒋理兴, 王安成, 等. 圆形标志投影偏心差补偿算法 [J]. 中国图象图形学报, 2018, 23(10): 1549-1557.
WU J L, JIANG L X, WANG A C, et al. Eccentricity error compensation for circular targets [J]. Journal of Image and Graphics, 2018, 23(10): 1549-1557 (in Chinese).
- [16] 解则晓, 王晓敏. 平面标定靶标标记点的圆心提取 [J]. 光学精密工程, 2019, 27(2): 440-449.
XIE Z X, WANG X M. Research on extraction algorithm of projected circular centers of marked points on the planar calibration targets [J]. Optics and Precision Engineering, 2019, 27(2): 440-449 (in Chinese).

Calibration method for high precision camera based on plane transformation

PENG Yan¹, GUO Junbin^{1,*}, YU Chuanqiang¹, KE Bing²

(1. Missile Engineering Institute, Rocket Force University of Engineering, Xi'an 710025, China;

2. Equipment Department of Rocket Force, Beijing 100089, China)

Abstract: Traditional methods do not consider the existence of perspective deviation, and usually extract the center of the ellipse as the projection point of a real physical center, resulting in camera calibration errors. A calibration method for high precision cameras was proposed based on plane transformation. The corner points on the inner and outer borders of the calibration plate were extracted to carry out the plane transformation of the calibration plate, and the marked points were projected from ellipses into approximate standard circles. The coordinates of the center of the standard circle were extracted by image moments, and were projected back to the plane of the original calibration plate to obtain the pixel coordinates of the actual center of the marked points. According to the coordinates of the actual center of the circular markers, the Zhang Zhengyou calibration method was used to complete the camera calibration. Experimental results show that compared thus the traditional method, the proposed method reduced the camera calibration error by 66.169%, thus effectively improving the camera calibration accuracy.

Keywords: camera calibration; Zhang Zhengyou calibration; circular marker; circular center detection; perspective transformation

Received: 2021-01-11; Accepted: 2021-04-09; Published online: 2021-04-19 17:33

URL: kns.cnki.net/kcms/detail/11.2625.V.20210419.1445.002.html

Foundation item: National Natural Science Foundation of China (61501470)

* Corresponding author. E-mail: gjb_202@163.com

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

DOI: 10.13700/j.bh.1001-5965.2021.0457

基于注意力机制的光伏组件热斑检测算法



樊涛¹, 孙涛^{2,*}, 刘虎¹

(1. 国家电网有限公司, 北京 100031; 2. 国网数字科技控股有限公司, 北京 100053)

摘 要: 热斑现象是造成光伏组件发电能力下降的重要原因之一,热斑检测是光伏电站运维必不可少的工作。然而分布式光伏电站的规模普遍较小、选址分散、环境复杂多样,使用传统的热斑检测算法需要投入大量的人力资源。基于此,提出了一种基于注意力机制的热斑检测算法 HSNet。通过图像分割消除反光影响,结合通道注意力机制,学习通道间的特征信息,增强目标区域的重要性,采用自定义锚点的方法提高检测速度,使用焦点损失激活函数和基于物体先验概率的类别预测方式改善训练目标样本不均衡导致的分类准确性低的问题,通过回归方法获取准确的目标位置。实验表明:设计的目标检测算法在窗体回归精度和分类准确性方面均有明显的优势,边界框平均精度和准确率分别提升了3.18%和2.42%。

关键词: 热斑检测; 目标检测; 残差网络; 注意力机制; 分布式光伏电站

中图分类号: TP391

文献标志码: A **文章编号:** 1001-5965(2022)07-1304-10

在光伏设备运行的过程中,难免会出现飞鸟、落叶、尘土等堆积在光伏组件上,这样的遮挡会导致该区域成为发电设备的负载,造成局部发热严重、发电量降低等问题。更严重的是,随着温度的积聚,会使电池组件上的焊点熔化并毁坏栅线,从而导致整个太阳电池组件的损毁。分布式光伏电站具有分布范围广、位置偏远、规模小等特点,因此运维人员对光伏设备热斑效应检测的难度更大、运维难度更高。

传统的热斑检测依赖于人工巡检的方式,即通过工作人员在光伏组件处实地校验每个光伏设备的状态来判断是否发生了热斑现象。这种方法需要大量的人力物力和时间成本,效率很低。例如,使用电子元件的电压电流特性对热斑的发生进行检测,如基于 $I-V$ 特性曲线的热斑检测方法^[1]、等效电路法^[2]、旁路二极管在线监测方法^[3]等。这些方法存在的问题也很明显,不同设备因为其规格、材料、品牌的不同, $I-V$ 特性曲线有

较大差异,同时,在节点上安装采集器、汇总传输到检测中心,安装的成本高、线路设计更为复杂。

近年来,人工智能和深度学习技术迅速发展,热斑检测问题可以通过无人机采集红外照片,再通过图像处理的方式来解决。深度学习在图像处理的人脸识别、目标跟踪、图像分类、行人重识别等应用均取得了较好的效果。例如,谢彭宇和徐新^[4]使用改进的 Faster R-CNN 实现对 SAR 图像的飞机检测;李广帅等^[5]使用深度学习方法实现了多尺度的行人重识别算法;廖华年和徐新^[6]使用深度学习结合注意力机制实现了对不同分辨率的行人重识别。这些方法的共同点是:先使用卷积神经网络的强大的特征提取能力,再对提取的特征进行分类,达到对目标进行识别的目的。本文将深度学习目标检测算法应用于红外图像的热斑检测中去,实现了对热斑的检测。

目前,光伏热斑检测面临以下问题:①各光伏组件之间排列紧密、数量众多,给目标检测算法提

收稿日期: 2021-08-11; 录用日期: 2022-01-03; 网络出版时间: 2022-01-25 20:27

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220124.1849.011.html

基金项目: 国家重点研发计划(2018YFB1500800); 国家电网有限公司科技项目(SGTJDK00DYJS2000148)

* 通信作者. E-mail: suntao@sgcc.sgcc.com.cn

引用格式: 樊涛, 孙涛, 刘虎. 基于注意力机制的光伏组件热斑检测算法[J]. 北京航空航天大学学报, 2022, 48(7): 1304-1313.
FAN T, SUN T, LIU H. Hot spot detection algorithm of photovoltaic module based on attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1304-1313 (in Chinese).

出了很高的要求。②由于光伏组件表面光滑,无人机拍摄时会发生太阳光的反射形成光斑,目标检测算法需要对光斑形成的原因进行分辨,增加了目标检测的难度。③发生热斑现象的光伏组件数量远少于正常的光伏组件数量,给训练带来很大的难度,处理不好往往会导致总体的准确率很高,但是对于某一类别的准确率却很低的现象发生。

针对以上问题,本文提出了一种面向光伏红外图像的热斑检测算法 HSNet。该算法将热斑检测过程分为反光消除处理、图像特征提取、自定义目标推荐、多目标分类4个步骤。算法以图像分割的方式检测出图像中的反光部分,并进行消除,降低后续检测工作的难度。对于特征提取模块,引入了注意力机制^[7],利用在不同区域的权重信息,构建出目标特征丰富的权重信息表,帮助检测系统忽略权重较轻的目标。在目标推荐模块,结合光伏组件的特点,使用自定义目标推荐策略,构建出特定大小的检测区域,在尽量不损失精度的情况下,缩小了检测模型的大小。在多目标分类阶段,基于物体先验概率获得类别预测结果,并引入 Focal Loss 函数^[8]计算损失值,用以解决检测中不能种类样本数量失衡的问题。

1 相关工作

自从卷积神经网络(CNN)被提出之后,各种深度学习模型被应用于通用物体识别领域。在这些深度学习模型中,主要划分为以下几个模块,分别为特征提取网络、区域推荐网络(region proposal network, RPN)和分类回归处理。

对于图像特征的提取最早可以追溯到 Lecun 等提出的 LeNet^[9]。2012年 Hinton 团队使用 AlexNet 获得了 ILSVRC 比赛的冠军,并开启了深度学习的时代, AlexNet 的论文中指出,深度学习模型的深度对于提高性能至关重要^[10]。之后,神经网络变得越来越深,由此也带来了新的问题:随着深度的不断增加,深度学习模型的性能反而越来越差,出现了梯度消失和梯度爆炸的问题^[11-13]。针对该问题,研究人员提出了很多解决方法,如采取无监督逐层训练方法^[14]、梯度剪切^[15]、使用更合适的激活函数^[16-19]、批规范化(BatchNorm, BN)^[20],直至残差网络(ResNet)模型^[21]的提出才解决了这一问题。Residual 结构将原来的输入改成了输入和上一层的输出残差之和,避免了网络深度增加带来的梯度爆炸和梯度消失的问题,使得网络的深度可以变得更深,性能

更好。之后的特征提取网络模型大多在 ResNet 的基础上做出改进^[22-24]。

传统的生成检测框普遍采用滑动窗口的方法,非常耗时^[25-26],而 Faster R-CNN 抛弃了滑动窗口的方法,采用 RPN 网络生成检测框^[27],极大地加快了检测框的生成速度,提高了模型的执行速度。RPN 网络使用生成的特征图中的特征值作为关键点,针对每个关键点对应不同比例和尺寸的9个矩形框,这些矩形框被称为锚点(anchor)。RPN 网络分别对 anchor 分类和回归操作。分类的目的是为了判断当前 anchor 选中的框内是否有物体,而回归操作为了获取物体相对当前 anchor 的偏移量。该层根据分类和回归的结果输出有可能存在物体的回归后的 anchor 框。

回归算法是目标检测模型的通用后处理机制,目的是为了对候选区域中的位置进行精确定位和修正。使用分类算法的目的是为了判断目标区域是否为所检测的目标。Softmax 算法将线性模型输出的实数域映射到 $[0, 1]$ 的空间,表示二分类算法不同种类的概率,通常用来作为分类模型的激活函数。然而,在热斑检测中,训练集中目标数量较少,大多数目标推荐的区域均为背景,因此需要对分类算法进行改进。

2 光伏组件的热斑检测框架

2.1 问题定义

本文用到的相关符号及其含义如表1所示。设训练集为 D , 训练集的大小为 N , I_i 表示第 i 张图像,其中 $i \in \{1, 2, \dots, N\}$, 第 i 张图像中包含 K_i

表1 符号表示

Table 1 Symbolic representation

符号	含义
D	训练集
I_i	第 i 张图像
y_{ij}	第 i 张图像的第 j 个光伏组件的类别
x	层运算输入
F	卷积等层计算公式
L_{cls}	分类损失函数
PL(pt)	概率为 pt 的 Focal Loss 损失函数
$bbox_{cls}$	回归损失函数
SE	SE 处理模块
GT	Ground Truth 人工标注框
M	训练模型
ReLU	ReLU 激活函数
θ_F	初始化网络模型参数
batch_size	训练图像批量大小
FM	提取的特征图
$(O_x, O_y), O_w, O_h$	输出目标的位置及目标的宽度和高度

个光伏组件,第 j 个光伏组件表示为 I_{ij} ,其中 $j \in \{1, 2, \dots, K_i\}$, y 表示光伏组件是否含有热斑, $y_{ij} = 0$ 表示第 i 张图像的第 j 个光伏组件不包含热斑, $y_{ij} = 1$ 表示第 i 张图像的第 j 个光伏组件包含热斑。训练模型为 M ,训练输出结果为 (O_x, O_y) , O_w, O_h ,表示检测目标回归后的位置及目标的宽度和高度, L_{cls} 表示分类损失函数。

2.2 反光消除处理

在热斑检测问题中,图像采集时的反光现象会对检测效果造成较大的影响。如图1所示,绿色框、黑色框和蓝色框选中的分别为含有热斑的光伏组件、正常的光伏组件和反射造成光斑的正常组件。

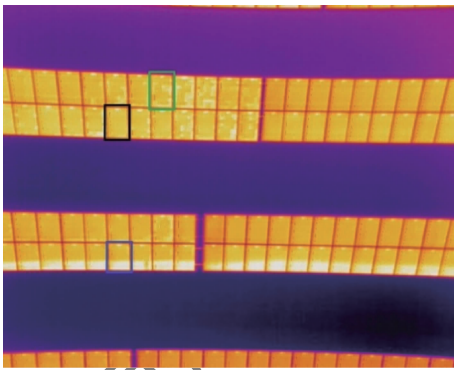


图1 无人机拍摄的光伏阵列

Fig. 1 Photovoltaic arrays photographed by a UAV

为了避免反光部分对热斑检测的干扰,设计了图像分割网络进行反光部分检测。网络采用类U-Net的结构,并根据应用需求进行了简化,如图2所示。其中,卷积单元由2层卷积层组成,使用ReLU作为激活函数;下采样单元、上采样单元都包含2层卷积层,分别使用最大池化操作和双线性插值法进行采样。对应的上、下采样单元之间使用张量拼接的方式进行连接,使网络的浅层特征能够直接到达上采样层,增强图像纹理信息的传递。

训练时,反光检测网络作为独立的模块进行训练。在测试时,输入图像在经过缩放、归一化等操作后输入反光检测网络,得到反光部分掩码。根据掩码,将图像相应部分的像素值置为0,从而完成反光区域消除的处理。

2.3 热斑检测网络

光伏组件上除了散乱排列的热斑外,还有规则排列的器件固有的亮斑,二者极易使检测算法混淆,因此在提取特征时需要丰富各个亮斑的表达语义。图3为热斑检测模型处理框架,本文提出在特征提取阶段使用ResNet结构结合注意力

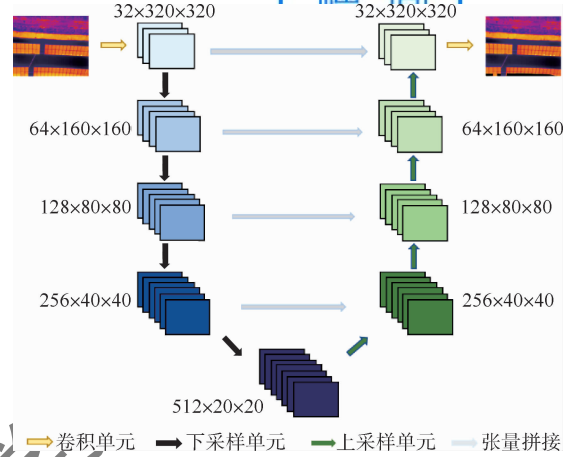


图2 反光消除网络结构

Fig. 2 Network structure of reflection elimination

机制。更深的网络能够获得更加丰富的语义信息,抽象能力更强,ResNet解决了更深层次网络的梯度消失、爆炸的问题,同时,注意力机制的引入使得网络模型在训练的过程中学习到图像中更感兴趣的区域的特征(更有可能出现目标)。

在特征提取阶段,首先,使用大小为 7×7 的卷积核对图像进行初步提取,为了加快处理进程,设置步长为2,输出通道为64。然后,使用更小的 3×3 的卷积核做进一步卷积提取特征,连接4个SE模块,每个模块输出通道分别为256、512、1 024和2 048,通道数越多,提取出的抽象网络信息越多。

然后,使用RPN将得到的特征图映射回原图像,根据预设的不同大小的anchor初步生成区域提案。最终区域提案输入到分类回归模块,对其进行分类和回归分析,得到更加精确的目标位置和类别。

2.4 光伏组件热斑的特征提取

热斑特征提取模块采用了Residual结构,即在原来的卷积模块上增加了一个分支,用来传递上一次卷积的结果。Residual结构如图4所示。

该Residual结构单元在每2个权重层前后开辟了一个通路,将前一个Residual结构的输出与本层Residual结构的输出相加,即上一个结构单元的输出误差为 x ,则经过2个权重层与ReLU函数的计算输出误差为 $F(x)$,将2个误差合并为 $F(x) + x$ 作为本层的残差输出。从函数层面来说,即使 $F(x)$ 取最小值为0,更深层次的网络输出 $F(x) + x$ 也会大于或等于输入 x ,则浅层次的热斑特征在更深层次上得到了继承和表达,从而避免了梯度消失的现象发生。

2.5 注意力机制

注意力机制能够在网络训练的过程中学习出

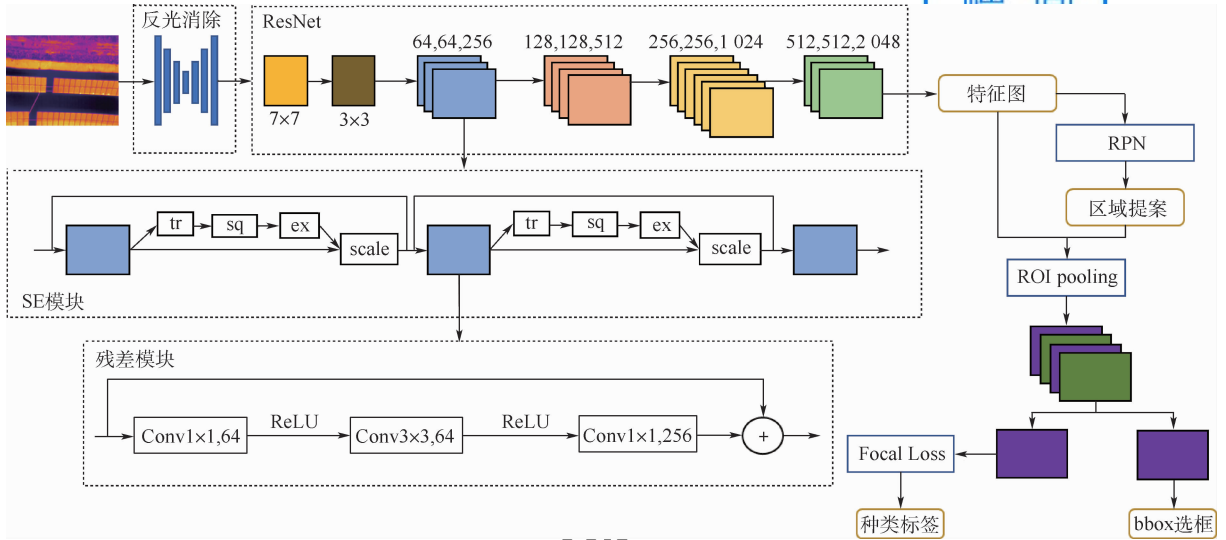


图 3 热斑检测模型处理框架 HSNet

Fig. 3 Hot spot detection framework of HSNet

图像中更感兴趣的区域,热斑的分布不规则,利用注意力机制可以获得更加准确的热斑特征信息。根据 Vaswani^[7] 和 Hu^[28] 等的论文,注意力机制的核心思想是通过网络为不同的特征图学习权重,再通过权重的大小增强感兴趣区域的表达,并抑制权重较小的区域。本文借鉴 SeNet 结构^[28],使用挤压/激励模块,模块结构如图 5 所示。

SE 模块包含 4 个组成部分,分别为 tr 操作、squeeze(sq)操作、excitation(ex)操作和 scale 操作,如图 6 所示。tr 操作为标准卷积公式,其定义为

$$F_{tr}: X \rightarrow U \quad X \in \mathbf{R}^{W' \times H' \times C'}, U \in \mathbf{R}^{W \times H \times C} \quad (1)$$

$$u_c = v_c X = \sum_{s=1}^{C'} v_c^s x^s \quad (2)$$

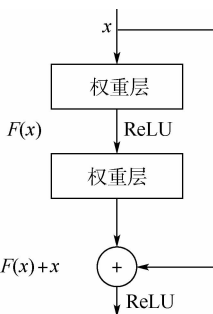


图 4 Residual 单元结构

Fig. 4 Structure of Residual unit

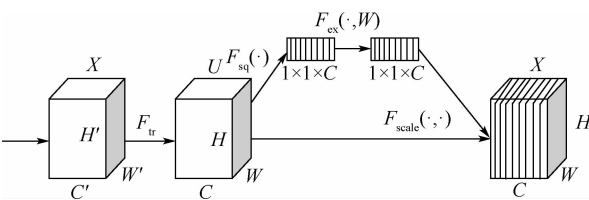


图 5 SE 模块结构

Fig. 5 Structure of SE block

式中: X 为输入数据,高、宽和通道数分别为 H' 、 W' 和 C' 的热斑图像的中间特征图; U 为输出,高、宽和通道数分别为 H 、 W 和 C ; u_c 为输出的第 c 通道; v_c 为第 c 个卷积核; x^s 为第 s 个输入。输出 U 即为大小为 $H \times W$ 的 C 个特征图。

squeeze操作目的是将tr操作的结果转化为长度为 C 的一维向量,如下:

$$z_c = F_{sq}(U_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H U_c(i, j) \quad (3)$$

式中: U_c 为特征图 U 的第 c 通道; H 和 W 分别为特征图 U 的高和宽。式(3)表明,squeeze操作是将tr操作的输出热斑特征图做平均池化,取每个通道的平均值作为该通道的squeeze输出。全局平均池化可以使注意力模块获得全局的感受野,学习到全局信息的热斑分布信息。

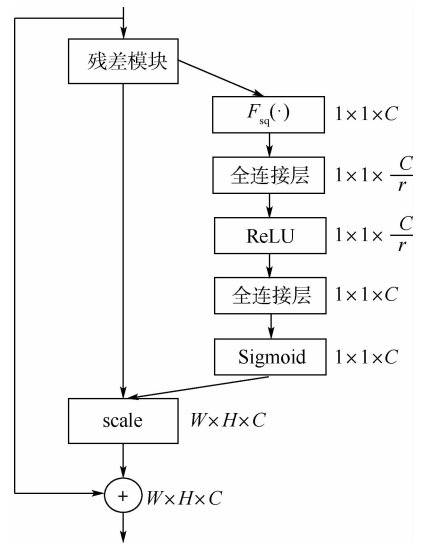


图 6 SE-Residual 模块结构

Fig. 6 Structure of SE-Residual block

excitation 操作目的是为了学习不同通道的权重信息,本文使用全连接层来融合不同通道的热斑特征图的信息,使用 Sigmoid 激活函数来刻画不同通道之间的非线性关系。excitation 模块的公式为

$$s = F_{\text{ex}}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (4)$$

$$W_1 \in \mathbf{R}^{r \times c}, W_2 \in \mathbf{R}^{c \times r} \quad (5)$$

式中: $\delta = \text{ReLU}()$; $\sigma = \text{sigmoid}()$; r 为计算的通道数,为了降低通道数量,简化计算复杂度,本文 r 选择 16。因此,第 1 个全连接层的输出为 $1 \times 1 \times C/r$ 。ReLU 函数不改变通道数量,第 2 个全连接层输出改为 $1 \times 1 \times C$ 。 z 为 squeeze 层的输出特征信息。

scale 层的设计是为了将残差模块的卷积特征图与注意力机制的输出相结合,scale 模块是将残差模块的第 c 个通道的值分别乘以 Sigmoid 层输出的第 c 个权重。这样,便可以将注意力机制中学习到的权重作用于学习中,计算公式为

$$x_c = F_{\text{scale}}(u_c, s_c) = s_c u_c \quad (6)$$

式中: u_c 和 s_c 分别为残差模块的第 c 个通道和 Sigmoid 层输出的第 c 个通道值。

2.6 自定义锚点

Ren 等^[27]提出的 Faster R-CNN 模型的设计目的是为了检测出更多的目标,因此在 Faster R-CNN 的 anchor 设计中采用了 9 种不同尺度和宽高比的 anchor。从图 1 中可以看出,不同光伏组件的大小与宽高比都相似,采用原始 anchor 大小设置方式一方面会产生大量的无用 anchor,延长检测时间,另一方面不合适的 anchor 大小会影响模型的检测效果。

本文采用 k -means 聚类的方式确定生成 anchor 的大小。标准的 k -means 采用 2 个标注框的宽和高的欧氏距离作为度量,这样会导致较大的候选框之间的距离比小的候选框之间的距离大,因此本文选择 2 个标注框的重合面积比 (IOU) 作为度量。 k -means 公式为

$$\text{IOU}_{(\text{box}, \text{anchor})} = \frac{\text{intersection}}{\text{union}} \quad (7)$$

$$d_{(\text{box}, \text{anchor})} = 1 - \text{IOU}_{(\text{box}, \text{anchor})} \quad (8)$$

式中: box 和 anchor 分别为中心簇的矩形框和其他矩形框; intersection 为 2 个矩形框的相交部分的面积; union 为 2 个矩形框相并覆盖的面积,在计算时 2 个矩形框的左上角重合; d 为 k -means 的度量,因为这 2 个矩形框 IOU 越大,其面积就越相似,所以计算距离时取 $(1 - \text{IOU})$ 的值。

2.7 基于物体先验的类别预测

在 R-CNN 分类阶段,一般仅使用分类器得到建议区域的类别预测结果,但类别预测结果无法反映区域预测本身的准确性,准确度较低的结果框可能有较高的分类得分。

本文将类别预测表示为物体先验与具体类别条件概率的结合,物体先验可以由 RPN 的输出得到。设 $O_i \in \{0, 1\}$ 表示第 i 个建议区域是否为物体, C_i 表示第 i 个建议区域在通过 R-CNN 后的类别预测,则最终类别预测可表示为

$$P(C_i) = P(C_i | O_i = 1)P(O_i = 1) \quad (9)$$

尽管测试流程发生了变化,整体的训练过程是无需变化的,从极大似然的角度看,类别预测等价于:

$$\ln P(C_i) = \ln P(C_i | O_i = 1) + \ln P(O_i = 1) \quad (10)$$

用 $P(\text{bg})$ 表示建议区域是背景 (bg) 而非物体的概率,则极大似然则可以表示为

$$\ln P(\text{bg}) = \ln(P(\text{bg} | O_i = 1)P(O_i = 1) + P(O_i = 0)) \quad (11)$$

由 Jensen 不等式可得

$$\ln P(\text{bg}) \geq \ln P(O_i = 0) \quad (12)$$

由此可得,背景类别极大似然的优化等价于优化其下界,也就是提高 RPN 背景类预测概率。同理,可得另一个下界为

$$\ln P(\text{bg}) \geq P(O_i = 1) \ln P(\text{bg} | O_i = 1) \quad (13)$$

式(13)表明同样需要优化 R-CNN 阶段的背景类预测的极大似然。再结合式(10)可以看出,在更改预测流程后,原训练过程与现预测流程是等价的,无需对损失函数做出任何改动。

2.8 Focal Loss 损失函数

在目标检测领域一般存在着正负样本不平衡的问题,给模型训练带来了极大的困难。在训练的样本中,本文将图像的标注框分成了 4 个不同的种类,分别为正常光伏组件 (normal)、脏污型热斑 (dirty)、斑点型热斑 (hotspot)、长线型热斑 (hotline)。不同热斑标注框的数量如表 2 所示。

如表 2 所示,不同种类的样本数量相差较多,使得模型的训练难度增大,难以准确地对结果进

表 2 不同标注框的数量

Table 2 Number of different kinds of labeled boxes

种类	图像数量	标注框数量	占比/%
normal	440	24 846	89.32
dirty	286	2 736	9.84
hotspot	28	109	0.39
hotline	98	125	0.45

行分类。而在热斑检测任务中,可将不同种类的热斑融合为一种,即将 hotspot、hotline、dirty 融合为一种 hotspot,但即便如此,正常样本与热斑样本的比例依然有 9:1。为了克服样本不均衡的问题,本文将 Faster R-CNN 分类阶段的分类函数 Softmax 改换成 Focal Loss 函数。Focal Loss 函数在不同类别的 loss 上分别施加不同的权重,使得注意力能够更多地放在分类错误的类别上。Focal Loss 的公式为

$$p_i = \begin{cases} p & y = 1 \\ 1 - p & \text{其他} \end{cases} \quad (14)$$

$$PL(p_i) = -\alpha_i \ln(1 - p_i)^\gamma \ln(p_i) \quad (15)$$

式中: α_i 为训练批次中这种样本的比例的倒数; $p_i = p$ 表示样本为热斑的概率; γ 为超参数,一般等于 2,用于控制 PL 对大量样本惩罚力度。在训练中,由于热斑样本较小,则相对应的 $\ln(1 - p_i)$ 的值较大,对 PL 影响不大,而当遇到正常样本时,该种类概率较大,相对应的 $\ln(1 - p_i)$ 的值较小,降低了大概率样本对 PL 值的贡献。

以上分别对数据预处理、网络模型结构、损失函数等进行了介绍。在此基础上,本文提出的热斑检测算法如算法 1 所示。

算法 1 基于注意力机制的热斑检测算法。

输入: 无人机拍摄的光伏组件红外图像 I 、测试光伏组件红外图像 I_{test} 。

输出: 测试光伏组件红外图像中光伏组件位置 (L_x, L_y) 、宽度 L_w 、高度 L_h 和光伏组件的类别 y_i 。

步骤 1 初始化网络参数 θ_F 。

步骤 2 从训练图像中选取 batch_size 数量的光伏图像,缩放到固定的 320×320 大小。

步骤 3 对选取的光伏图像进行反光消除处理。

步骤 4 使用 SENet 模块提取缩放后的光伏图像特征图 FM。

步骤 5 对特征图中的每个关键点设置多个 anchor,再将 anchor 划分为背景和光伏组件,将包含光伏组件的 anchor 输出,输出为 O_x 、 O_y 、 O_w 、 O_h 和标记向量 O_l 。

步骤 6 计算人工标注框 GT 与 O_x 、 O_y 、 O_w 、 O_h 的损失函数 bbox_{cls} ,通过反向传播优化使得 bbox_{cls} 达到最小。

步骤 7 使用 Focal Loss 函数计算标签损失函数 L_{cls} ,通过计算两者的欧氏距离来对输出标签向量进行优化。

步骤 8 重复步骤 1 ~ 步骤 7 进行训练,直至 bbox_{cls} 和 L_{cls} 的值收敛,得到热斑检测模型 M 。

步骤 9 将测试数据 I_{test} 输入到热斑检测模型 M 中去,获得 O_x 、 O_y 、 O_w 、 O_h 和标签 O_l ,再结合 RPN 输出得分得到最终类别预测分数,获得目标光伏组件的位置和光伏组件种类(是否含有热斑)。

3 实验结果与分析

3.1 数据集

实验中使用的数据集来源于江苏某分布式光伏电站。数据集中的图像均是使用无人机低空采集的红外图像,整个数据集包含 560 张红外图像,27 816 个光伏组件,其中包含热斑的光伏组件 2 970 个,占比 10.7%,不含热斑的正常的光伏组件的数量为 24 846,占比 89.3%。实验过程中,将训练集和测试集按照 4:1 的比例划分,训练集包含 448 张图像,测试集包括 112 张图像。在模型训练过程中,采用水平和垂直翻转进行数据扩增。

3.2 参数设置

为了加速模型训练过程,模型训练之前统一将采集的训练图像缩放到 320×320 大小。在 SE-Residual 模块选择 16 作为 r 的值,即全连接层选择原始通道数的 $1/16$ 作为中间层输出, C 为 256 作为残差网络模块的输出层通道数。输入批量大小为 32,输入图像之前将红外图像转化为 RGB 图像处理。为了扩展数据集大小,在图像训练时选择随机旋转增广数据。模型训练周期为 24,初始学习率为 0.003,为防止过拟合,设置权重衰减为 0.000 1。

3.3 模型评估

在二分类项目中, P - R 曲线常用来评估分类结果的好坏, P 为查准率, R 为查全率。两者的计算公式为

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$R = \frac{TP}{TP + FN} \quad (17)$$

式中:TP 表示真正例;FP 表示假正例;FN 表示假反例。一般情况下,查准率和查全率是一对相互矛盾的度量单位,查准率高时,查全率往往较低,而在不同种类的数量相差很大时,单独大的查准率或者查全率则不能准确描述模型性能的好坏。因此,引入平均检测精度 AP,公式为

$$AP(c_i) = \int_0^1 p(R) dR \quad (18)$$

式中:AP(c_i) 为第 c_i 种目标的平均检测精度; $p(R)$ 为查全率为 R 时的准确率。

对于多目标分类项目,一般使用 mAP 表示各

类别 AP 的平均值:

$$mAP = \frac{1}{m} \sum_{i=1}^m AP(i) \quad (19)$$

式中: m 为种类的数量; $AP(i)$ 表示种类为 i 的 AP 值。

对于多目标检测项目,还需要对检测框的准确性进行度量,一般使用 bbox_mAP 检测框的平均检测精度。

3.4 结果分析

图 7 为反光消除处理的效果。可以看到,该处理能够较好地抹除图像中的反光区域。经过处理的图像只包含热斑及设备固有亮斑,大大降低了后续热斑检测的难度。

在反光消除的基础上,本文分别对经典的深度学习模型 Faster R-CNN、RetinaNet、SSD300、YOLOv3 在相同的数据集上进行了训练,模型结果均已达到最优。将以上 4 种深度学习模型的训练结果与本文模型结果进行对比,结果如表 3 所示。

从表 3 中可以看出,传统的 Faster R-CNN 的各方面统计数据均不如 SSD300、RetinaNet 模型的表现最差。与 SSD 相比,本文算法 HSNet 选择的先验框的长宽尺度数量较少,而 SSD 选择的尺度数量较多,造成了在候选框精度上的不足。另外,本文在分类阶段采用了 Focal Loss 算法,提高了

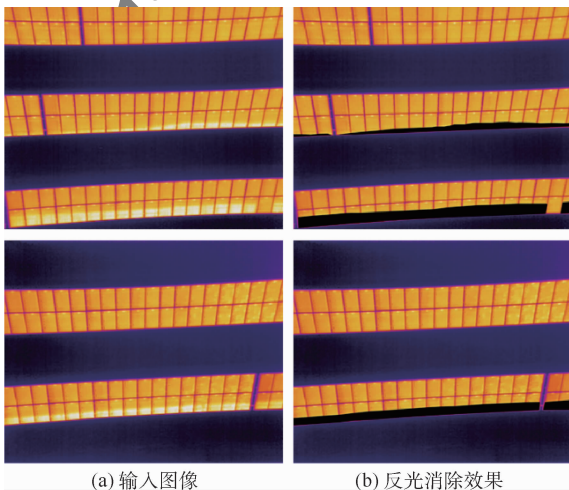


图 7 反光消除网络效果

Fig. 7 Qualitative results of reflection elimination

表 3 不同模型的结果对比

Table 3 Comparison of results with different models

算法	bbox_mAP/%	mAP/%	准确率/%
HSNet	55.97	70.95	98.22
Faster R-CNN	50.64	67.77	95.80
RetinaNet	20.98	38.17	43.92
SSD300	66.57	67.96	96.50
YOLOv3	48.78	59.48	91.32

HSNet 在分类问题上的准确率, mAP 值较高。本文算法是在 Faster R-CNN 的基础上改进而来,从结果上来看,改进的算法对检测结果有效。改进后的模型在 bbox_mAP、mAP 和准确率上均有不同程度的提高,其中 mAP 和准确率 2 个指标达到了实验的最高值。

图 8 为不同类型的检测模型对于热斑 (hotspot) 样本类型的 $P-R$ 曲线比较。可以看出,改进后的模型在查全率相同时拥有更高的查准率,说明特征提取模块的改进、loss 计算方法的改进和注意力机制的添加对热斑检测起到了一定的作用,提高了检测模型的性能。

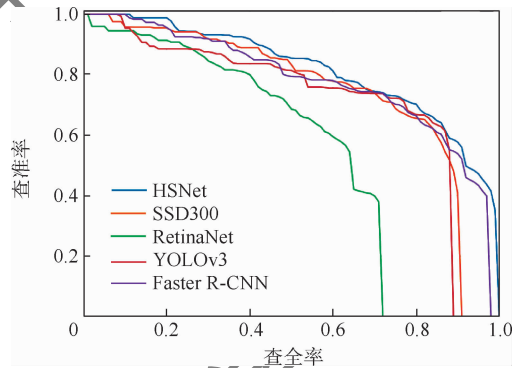


图 8 不同模型的 $P-R$ 曲线比较

Fig. 8 Comparison of $P-R$ curves of different models

图 9 为不同目标检测算法对于热斑这种类别的数据在 4 张红外测试图像上的检测结果可视化显示。从检测结果上来看,5 种检测模型在检测框的回归效果上差异不大,但是检测出的结果数量有很大不同。

对图 9 第 1 行图像的检测中,HSNet 算法检测的准确率最高,与人工标注的相比漏检了 1 个含有热斑的光伏组件,多检测出了 1 个正常的光伏组件,可能原因是热斑部位不明显,标注不准确。RetinaNet 算法检测出了 1 块正常的光伏组件,效果最差。

对图 9 第 2 行图像的检测中,本文算法 HSNet 与 Faster R-CNN、SSD、YOLOv3 算法均有较好的表现,而 RetinaNet 效果最差,仅检测出了下方的 4 块光伏组件。第 2 行图中包含多块热斑光伏组件且热斑类型与反射型正常光伏组件交织在一起形成干扰,算法难以分辨含有热斑的光伏组件。

对图 9 第 3 行的图像检测中,光伏阵列分为上中 2 组,本文算法在上组阵列比其他算法均多检测出了 1 个含有热斑的光伏组件。

对图 9 第 4 行的图像检测中,热斑分为脏污型热斑 (dirty) 和长线型热斑 (hotline),所有算法在检测脏污型热斑的表现相似,均能够检测出包含脏污型热斑的光伏组件。

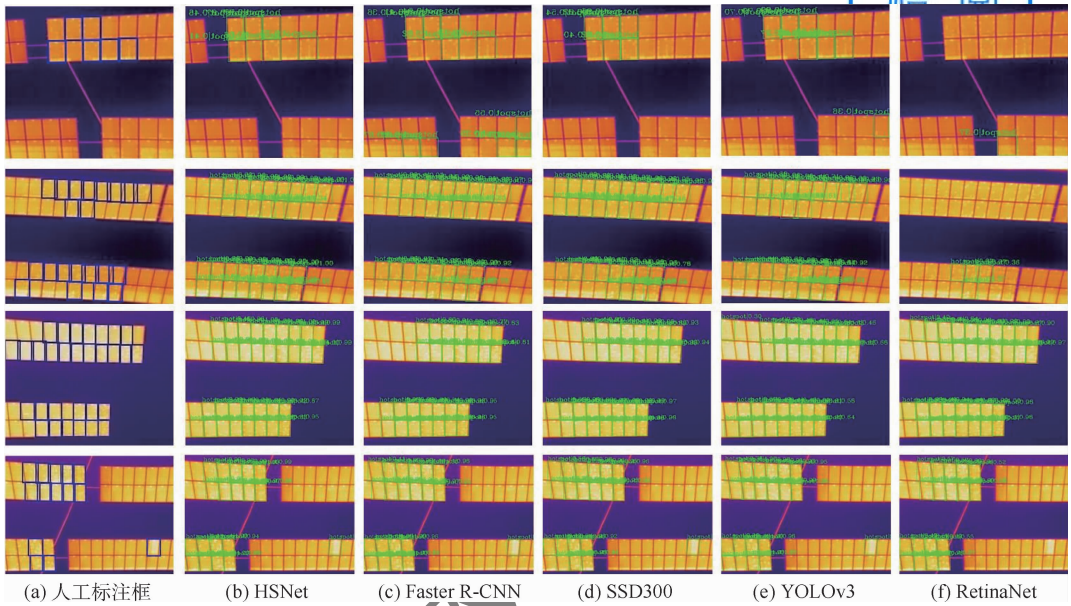


图 9 不同检测算法的结果比较

Fig. 9 Comparison of detection results with different detection algorithms

4 消融实验

为了验证本文算法对原始 Faster R-CNN 算法的改进是否都有效,设计了消融实验。消融实验使用控制变量法,所有算法的训练周期都为 24,训练集与测试集相同。F-RCNN 表示原始的 Faster R-CNN 算法,UN 表示使用反光消除预处理模块,RN 表示使用 ResNet50 模块替换 VGG16 特征提取,SE 表示使用 SE 模块替换 VGG16,SE 模块比 Residual 模块增加了注意力机制,FL 表示将分类阶段的 Softmax 激活函数替换为 Focal Loss 激活函数,PRI 表示使用物体先验的类别预测。

表 4 为使用不同模块替换后的 Faster R-CNN 算法在光伏热斑检测中的表现。通过对比可知,ResNet 模块和 SE 模块能够显著提升 bbox_mAP 的数值,而对检测精度也能有一定的提高效果。RPN 物体先验概率(PRI)的引入同样能提高检测精度。同时,Focal Loss 对原始 Softmax 损失函数

表 4 消融实验结果

Table 4 Results of ablation experiments

算法	bbox_mAP/%	mAP/%	准确率/%
F-RCNN	50.64	67.77	95.80
F-RCNN + UN	51.72	68.86	96.92
F-RCNN + UN + RN	53.68	68.92	96.99
F-RCNN + UN + SE	55.38	69.18	96.43
F-RCNN + UN + FL	52.21	69.30	96.95
F-RCNN + UN + RN + FL	54.23	69.58	97.34
F-RCNN + UN + SE + FL	55.85	70.50	98.01
F-RCNN + UN + SE + FL + PRI	55.97	70.95	98.22

的替换,能够显著提高检测的精度,而对 bbox_mAP 几乎没有什么影响。

从表 4 中可知,最佳的搭配是采用前处理 UN + SE 模块 + Focal Loss 激活函数 + PRI 先验概率预测的组合,该组合能够将 bbox_mAP、mAP 和检测准确率分别提升 5.33%、3.18%、和 2.42%。实验表明,添加了 SE 模块和 Focal Loss 损失函数,并结合物体先验的类别预测后的算法,比原始的模型检测算法在样本数量不均衡的光伏热斑检测任务中,达到了更好的效果,证明了残差模块、注意力机制、物体先验预测和调节 loss 对实验的有效性。

5 结论

针对分布式光伏电站运维中的光伏组件热斑检测问题,本文提出了一种基于注意力机制的热斑检测算法,能够克服传统检测算法依靠人力的缺陷。

1) 针对图像中反光部分对热斑检测的干扰问题,本文设计了反光消除网络,并将其应用在预处理环节中。网络能够以图像分割的形式消除反光部分,消除其对热斑检测的影响。

2) 针对图片中光伏组件中含有热斑和不含热斑的目标差异不大、检测困难的问题,本文提出将特征提取层数加深,使用 ResNet 结构避免梯度爆炸的现象,同时在 Residual 模块的基础上添加注意力机制,加强了特征提取的有效性,使得检测精度进一步提高。

3) 在 R-CNN 阶段使用 Focal Loss 函数对原来的 Softmax 函数进行替换,在分类阶段对样本类别不均衡问题进行探索,同时结合 RPN 阶段的物体先验概率进行类别预测。实验表明,以上方法在检测存在热斑的光伏组件问题上有效。

本文在实验中受实际应用条件的限制,数据集包含 560 张红外图像,数量偏少。在今后的研究中如果能够获取到更加丰富的数据集,可以进一步提升光伏热斑检测的效果。

参考文献 (References)

- [1] 云平,张志祥,刘恒. 基于 I - V 特性的多晶硅光伏组件故障及失效研究[J]. 太阳能,2020(1):61-69.
YUN P, ZHANG Z X, LIU H. Study on failure mechanism of polycrystalline silicon PV module base on I - V characteristics [J]. Solar Energy, 2020(1):61-69 (in Chinese).
- [2] 温浚铎,郭珂,王瀚笠,等. 光伏阵列热斑故障在线检测方法研究[J]. 电子制作,2019(15):55-58.
WEN J D, GUO K, WANG H L, et al. Photovoltaic array hot spot fault online detection method research [J]. Practical Electronics, 2019(15):55-58 (in Chinese).
- [3] 陈斯,戴新宇,杨增汪,等. 光伏旁路二极管在线监测系统[J]. 电测与仪表,2012,49(10):96-100.
CHEN S, DAI X Y, YANG Z W, et al. A bypass diode of photovoltaic on-line monitoring system [J]. Electrical Measurement & Instrumentation, 2012, 49(10):96-100 (in Chinese).
- [4] 谢彭宇,徐新. 基于多尺度联合学习的行人重识别[J]. 北京航空航天大学学报,2021,47(3):613-622.
XIE P Y, XU X. Multi-scale joint learning for person re-identification [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(3):613-622 (in Chinese).
- [5] 李广帅,苏娟,李义红. 基于改进 Faster R-CNN 的 SAR 图像飞机检测算法[J]. 北京航空航天大学学报,2021,47(1):159-168.
LI G S, SU J, LI Y H. An aircraft detection algorithm in SAR image based on improved Faster R-CNN [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1):159-168 (in Chinese).
- [6] 廖华年,徐新. 基于注意力机制的跨分辨率行人重识别[J]. 北京航空航天大学学报,2021,47(3):605-612.
LIAO H N, XU X. Cross-resolution person re-identification based on attention mechanism [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(3):605-612 (in Chinese).
- [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Advances in Neural Information Processing Systems, 2017:5998-6008.
- [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017:2999-3007.
- [9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6):84-90.
- [11] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks, 1994, 5(2):157-166.
- [12] DUVENAUD D, RIPPEL O, ADAMS R P, et al. Avoiding pathologies in very deep networks [EB/OL]. (2016-07-08) [2021-08-01]. <https://arxiv.org/abs/1402.5836>.
- [13] O'BRIEN A E, PITKOW X. Skip connections eliminate singularities [EB/OL]. (2018-03-04) [2021-08-01]. <https://arxiv.org/abs/1701.09175>.
- [14] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786):504-507.
- [15] ZHANG J Z, HE T X, SRA S, et al. Why gradient clipping accelerates training: A theoretical justification for adaptivity [EB/OL]. (2020-02-10) [2021-08-01]. <https://arxiv.org/abs/1905.11881>.
- [16] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [C] // Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011:315-323.
- [17] MAAS A L. Rectifier nonlinearities improve neural network acoustic models [C] // Proceedings of the 30th International Conference on Machine Learning, 2013.
- [18] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs) [C] // Proceedings of the 4th International Conference on Learning Representations, 2015.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [C] // 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015:1026-1034.
- [20] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C] // Proceedings of the 32nd International Conference on Machine Learning, 2015:448-456.
- [21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016:770-778.
- [22] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017:2261-2269.
- [23] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017:5987-5995.
- [24] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks [C] // 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017:764-773.
- [25] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature

- hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [26] GIRSHICK R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.
- [27] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [28] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

Hot spot detection algorithm of photovoltaic module based on attention mechanism

FAN Tao¹, SUN Tao², LIU Hu¹

(1. State Grid Corporation of China, Beijing 100031, China; 2. State Grid Digital Technology Holding Co., Ltd., Beijing 100053, China)

Abstract: The hot spot phenomenon is one of the important reasons for the reduction of power generation capacity of photovoltaic panels, and the detection of hot spots is an essential task for operation and maintenance personnel. The scale of distributed photovoltaic power plants is generally small, the site is scattered, the environment is complex and diverse, and the operation and maintenance personnel need to invest a lot of human resources to detect hot spots using traditional hot spot detection methods. In this paper, we propose a new hot spot detection algorithm HSNet. Firstly, the influence of reflection is eliminated through image segmentation. Secondly, the feature information between channels is learned in combination with the channel attention mechanism to enhance the importance of the target area. The method of user-defined anchor points is used to improve the detection speed. Then, the focus loss activation function and the category prediction method based on the prior probability of objects are used to improve the problem of low classification accuracy caused by the imbalance of training target samples. Finally, the accurate target position is obtained by regression method. Experiments show that the target detection algorithm designed in this paper has significant advantages over other algorithms in terms of window regression accuracy and classification accuracy, and the mean accuracy and accuracy of the bounding box are improved by 3.18% and 2.42%, respectively.

Keywords: hot spot detection; object detection; residual networks; attention mechanism; distributed photovoltaic plants

Received: 2021-08-11; **Accepted:** 2022-01-03; **Published online:** 2022-01-25 20:27

URL: kns.cnki.net/kcms/detail/11.2625.V.20220124.1849.011.html

Foundation items: National Key R & D Program of China (2018YFB1500800); Technology Project of State Grid Corporation of China (SGTJDK00DYJS2000148)

* **Corresponding author.** E-mail: suntao@sgec.sgcc.com.cn

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

DOI: 10.13700/j.bh.1001-5965.2021.0571

一种新的气动伺服弹性失稳模式的机理分析

姜宇, 杨超, 吴志刚*

(北京航空航天大学 航空科学与工程学院, 北京 100083)

摘 要: 某超声速导弹在飞行试验中发生了气动伺服弹性失稳导致的结构解体, 通过对飞行试验数据的分析发现, 气动伺服弹性失稳的振动频率高于弹体一阶弯曲模态频率, 对导弹进行气动伺服弹性稳定性频域分析, 并未发现该频率段发生气动伺服弹性失稳。针对该问题, 建立了一种可以考虑数字式飞控系统采样过程影响的气动伺服弹性稳定性仿真分析方法, 并对该导弹进行了建模分析, 数值结果再现了该导弹的失稳现象。讨论了这一新型失稳现象发生的原因, 包括连续结构滤波器离散化带来的移频现象和频率混叠问题。给出了对应的改进措施和相关的结论。

关键词: 气动伺服弹性; 离散化; 结构滤波器; 采样率; 频率混叠

中图分类号: V211.47

文献标志码: A **文章编号:** 1001-5965(2022)07-1314-10

气动弹性这一涉及空气动力与弹性结构之间相互作用的问题, 对飞行器的结构安全与飞行性能起着重要的影响作用^[1]。自 20 世纪 50 年代起, 自动控制技术大量应用于导弹, 气动弹性与控制系统的耦合问题开始出现^[2]。由于真实飞行器的弹性效应, 飞行控制系统(简称飞控系统)接收的惯性敏感元件信号不仅包含飞行器刚体运动信息, 也包含飞行器结构的弹性振动信息^[3]。一般来说, 这将会导致原本将飞行器视为刚体而进行设计的飞控系统稳定裕度下降, 严重时还会发生失稳, 即所谓的气动伺服弹性(aeroservoelasticity, ASE)^[3]失稳问题。

目前, 一般的 ASE 系统建模方法^[4-6]就是将各环节单独建模并根据输入输出关系进行组合。而在 ASE 系统稳定性分析方面, 对于单输入/单输出系统, 工程上一般采用经典控制理论中的 Nyquist 方法^[7]进行稳定性分析; 对于多输入/多输出系统, 可供选择的分析方法有基于结构奇异值的 ASE 鲁棒分析方法^[8-9]、最小奇异值方法^[6]

和广义 Nyquist 方法^[10]。对于 ASE 失稳问题, 在飞控系统中加入适当的结构滤波器可以降低其结构模态频率附近的系统响应幅值, 从而尽可能地减小弹性模态的不利影响, 这是一种工程上普遍采用的解决方法^[11]。在实际的结构滤波器设计流程中, 一般先设计连续结构滤波器并对其进行离散化处理, 得到数字式结构滤波器并最终应用到飞控系统中^[12]。因此, 包含离散结构滤波器的数字式飞控系统, 其特性也与结构滤波器离散化处理过程中选用的离散化方法^[12]和采样频率^[13]有关。此外, 在飞控系统中加入低通滤波器, 对高频段系统响应幅值进行抑制也是常用的方法。

近期, 某超声速导弹在飞行试验中发生了 ASE 失稳导致的结构解体。通过对飞行试验数据的分析发现, 导弹 ASE 失稳的振动频率高于弹体一阶弯曲模态频率。且对导弹进行传统 ASE 稳定性频域分析之后, 也并未发现在飞行试验失稳的振动频率处发生 ASE 失稳。

本文针对这种新的 ASE 失稳模式, 建立了一

收稿日期: 2021-09-26; 录用日期: 2021-12-19; 网络出版时间: 2022-01-05 12:18

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220104.1937.005.html

* 通信作者。E-mail: wuzhigang@buaa.edu.cn

引用格式: 姜宇, 杨超, 吴志刚. 一种新的气动伺服弹性失稳模式的机理分析[J]. 北京航空航天大学学报, 2022, 48(7): 1314-1323. JIANG Y, YANG C, WU Z G. Mechanism analysis of a new aeroservoelastic instability mode[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1314-1323 (in Chinese).

种可以考虑数字式飞控系统采样过程影响的 ASE 稳定性仿真分析方法,并应用该方法对导弹进行了建模与数值分析,成功复现了这种新的 ASE 稳定性失稳模式,并对这种新的 ASE 失稳模式的机理进行研究,提出了相应的改进措施。

1 传统 ASE 稳定性仿真分析方法

ASE 系统分为弹性结构机体、惯性敏感元件、舵机及飞控系统等多个环节。分别对每个环节建立传递函数,串联相乘得到 ASE 系统的开环传递函数^[14-15]用于稳定性分析。

以飞行器的纵向通道为例,带有飞控系统的弹性飞行器 ASE 闭环系统框图^[16]如图 1 所示。图中: q 为弹性结构机体模态向量; δ 为纵向通道控制面刚体偏转变量; η 为控制指令。ASE 系统建模与频域稳定性分析方法在下文中介绍。

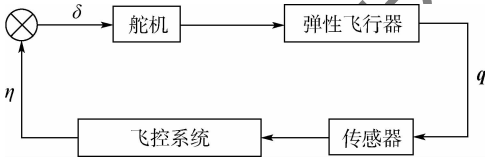


图 1 弹性飞行器 ASE 系统框图

Fig. 1 Block diagram of flexible flight vehicle aeroservoelasticity system

1.1 ASE 系统

1) 弹性结构机体

考虑非定常气动力,弹性结构机体纵向通道气动弹性运动方程^[6]一般可以写为

$$[M_{qq} \quad M_{q\delta}] \begin{bmatrix} \ddot{q} \\ \ddot{\delta} \end{bmatrix} + [C_{qq} \quad 0] \begin{bmatrix} \dot{q} \\ \dot{\delta} \end{bmatrix} + [K_{qq} \quad 0] \begin{bmatrix} q \\ \delta \end{bmatrix} = \frac{1}{2}\rho V^2 [Q_{qq} \quad Q_{q\delta}] \begin{bmatrix} q \\ \delta \end{bmatrix} \quad (1)$$

式中: ρ 为大气密度; V 为飞行速度; M_{qq} 和 $M_{q\delta}$ 为广义质量矩阵; C_{qq} 为广义阻尼矩阵; K_{qq} 为广义刚度矩阵; Q_{qq} 和 $Q_{q\delta}$ 为广义非定常气动力系数矩阵; 0 为元素都是 0 的矩阵。

引入谐振振荡运动条件,即对式(1)等号两边进行傅里叶变换,可以得到弹性结构机体纵向通道的频域气动弹性方程为

$$\left[-\omega^2 M_{qq} + i\omega C_{qq} + K_{qq} - \frac{1}{2}\rho V^2 Q_{qq}(\omega) \right] q(\omega) = \left[\omega^2 M_{q\delta} + \frac{1}{2}\rho V^2 Q_{q\delta}(\omega) \right] \delta(\omega) \quad (2)$$

式中: ω 为圆频率; $q(\omega)$ 和 $\delta(\omega)$ 为频域中的变量。

解式(2)可得弹性结构机体纵向通道的传递

函数矩阵 G 为

$$q(\omega) = G\delta(\omega) = A^{-1}B\delta(\omega) \quad (3)$$

式中:

$$A = -\omega^2 M_{qq} + i\omega C_{qq} + K_{qq} - \frac{1}{2}\rho V^2 Q_{qq}(\omega)$$

$$B = \omega^2 M_{q\delta} + \frac{1}{2}\rho V^2 Q_{q\delta}(\omega)$$

2) 惯性敏感元件与舵机

根据模态叠加原理,弹性结构机体的运动可以表示为固有模态的线性叠加形式。导弹的纵向通道采用的反馈信号包括俯仰角 θ 和俯仰角速度 $\dot{\theta}$,传感器安装在导弹的弹身上。若弹身的垂直面内弯曲振动位移用 $w(x,t)$ 表示, x 为弹身纵轴坐标,则有

$$w(x,t) = \sum_{i=1}^n f_i(x) q_i(t) \quad (4)$$

式中: $f_i(x)$ 为第 i 阶导弹固有振动模态,与广义坐标 $q_i(t)$ 对应。

同样,引入谐振振荡条件,则有如下频域的表达式:

$$\theta = \frac{\partial w(x,t)}{\partial x} = \sum_{i=1}^n f_i'(x) q_i(\omega) \quad (5)$$

$$\dot{\theta} = \frac{d}{dt} \left(\frac{\partial w(x,t)}{\partial x} \right) = \sum_{i=1}^n i\omega f_i'(x) q_i(\omega) \quad (6)$$

代入传感器所在位置的弹身纵轴坐标并合并为矩阵形式,则弹性结构机体运动输出的反馈信号向量^[17]可以表示为

$$\begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} = S q(\omega) = \begin{bmatrix} 1 & 0 \\ 0 & i\omega \end{bmatrix} \begin{bmatrix} f_s \\ f_s \end{bmatrix} q(\omega) \quad (7)$$

式中: f_s 为惯性敏感元件所在位置处的模态行阵; S 为惯性敏感元件环节传递函数矩阵。

弹性飞行器纵向通道控制面的偏转与飞控系统控制指令 η 之间有如下关系:

$$\delta(\omega) = T(i\omega)\eta \quad (8)$$

式中: $T()$ 为控制面对应舵机环节的传递函数,以频率 ω 为自变量。

3) 飞控系统

飞控系统根据惯性敏感元件采集的弹性结构机体反馈信号进行解算,给出控制指令 η 。飞控系统框图^[16]如图 2 所示。

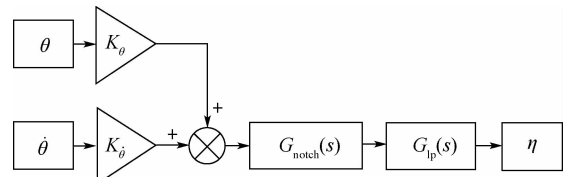


图 2 飞控系统框图

Fig. 2 Block diagram of flight control system

对应的内回路控制律频域传递函数矩阵 K 为

$$\eta = K \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} = G_{\text{notch}}(s) G_{\text{lp}}(s) [K_{\theta} \quad K_{\dot{\theta}}] \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} \quad (9)$$

式中: $G_{\text{notch}}(s)$ 为结构滤波器; $G_{\text{lp}}(s)$ 为低通滤波器; K_{θ} 为俯仰角增益; $K_{\dot{\theta}}$ 为俯仰角速度增益。

当飞行器 ASE 系统响应较大时, 通常会在控制系统中加入结构滤波器或低通滤波器对系统响应进行抑制。

工程中常常采用的二阶结构滤波器, 其传递函数的设计公式^[18]为

$$G_{\text{notch}}(s) = \frac{s^2 + 2\xi_1\omega_1s + \omega_c^2}{s^2 + 2\xi_2\omega_2s + \omega_c^2} \quad (10)$$

式中: ξ_1, ξ_2 为结构滤波器阻尼系数; ω_1, ω_2 为结构滤波器带宽系数; ω_c 为二阶结构滤波器中心频率, rad/s。二阶结构滤波器的中心频率衰减幅度为 $-20\lg(\xi_2\omega_2/\xi_1\omega_1)$, 参数 $\xi_2\omega_2$ 的大小决定了二阶结构滤波器的阻带宽度。结构滤波器在减小弹性模态响应的同时会引起中心频率前的相位滞后^[19], 使原本将飞行器视为刚体设计的飞控系统带宽降低, 并且对低频段的稳定性产生不利影响^[20]。此外, 当通过离散化方法和离散采样时间对连续结构滤波器进行离散化之后, 得到的数字结构滤波器相比较连续结构滤波器会发生移频现象^[12]。

二阶低通滤波器的传递函数设计公式为

$$G_{\text{lp}}(s) = \frac{A_0\omega_0^2}{s^2 + \xi\omega_0s + \omega_0^2} \quad (11)$$

式中: A_0 为零频增益; ω_0 为截止频率; ξ 为阻尼系数, 决定了低通滤波器幅频特性曲线有无峰值, 或者谐振峰的高低。

1.2 频域 ASE 稳定性分析

在舵机输入端断开闭环系统, 将各环节的传递函数串联相乘, 则可求得 ASE 系统的开环传递函数 L 为

$$L(i\omega) = -KSGT(i\omega) \quad (12)$$

式中: L 以圆频率 ω 为自变量。

对于单输入/单输出系统, 工程上一般采用经典控制理论中的 Nyquist 方法^[7]进行稳定性分析, 并给出稳定裕度评估。

1.3 算例结果

本文中的算例是某超声速导弹, 该导弹由细长旋成单身和 4 片 X 型分布的全动尾舵组成。需要说明的是, 算例所用的数据已经过一定处理, 但不损失所要展示的 ASE 系统的稳定性特性。

1) 无滤波器情况的稳定性分析。选取该型

导弹的前二阶弹性模态进行 ASE 稳定性分析, 其中弹体一阶弯曲模态频率为 50 Hz, 弹体二阶弯曲模态频率为 120 Hz。飞控系统的纵向通道以俯仰角和俯仰角速度采集信号作为输入, 增益皆为 -0.025 , 初始飞控系统并未串联结构滤波器和低通滤波器。通过对该型导弹的纵向通道进行传统频域方法 ASE 稳定性分析发现, 在频率为 50 Hz 和 120 Hz 处有较大的系统响应。在未串联结构滤波器和低通滤波器的情况下, 53.4 Hz 处和 118.4 Hz 处的 ASE 稳定裕度已经不足, 分别为 -1.7 dB 和 -11.1 dB。得到的开环传递函数曲线如图 3 中黑色曲线所示。

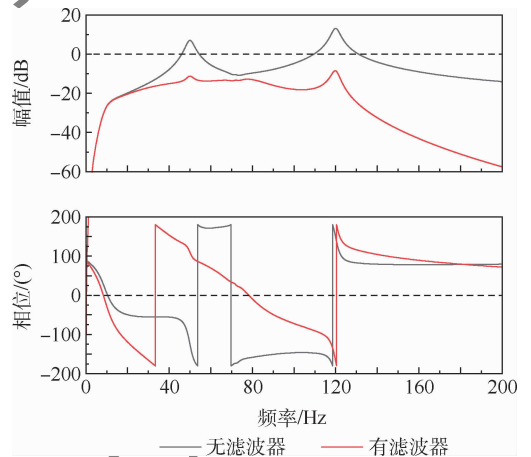


图 3 有无滤波器导弹的开环传递函数曲线
Fig. 3 Open-loop transfer function curves of missile with and without filter

2) 滤波器设计。由于飞行器在频率 50.0 Hz 和 120.0 Hz 处有较大的系统响应, 需要分别设计相应的连续结构滤波器和低通滤波器去抑制这 2 个频率段的系统响应。

利用式(10)设计中心频率为 50.0 Hz、衰减幅度为 20 dB 的连续结构滤波器传递函数为

$$G_{\text{notch}}(s) = \frac{s^2 + 37.68s + 98\,596.0}{s^2 + 376.8s + 98\,596.0} \quad (13)$$

利用式(11)设计截止频率为 50.0 Hz 的二阶连续低通滤波器传递函数为

$$G_{\text{lp}}(s) = \frac{98\,596.0}{s^2 + 376.8s + 98\,596.0} \quad (14)$$

3) 含滤波器情况的稳定性分析。在飞控系统中串联加入设计的连续结构滤波器和二阶连续低通滤波器, 并再次进行传统频域方法 ASE 稳定性分析, 得到的 ASE 系统开环传递函数曲线如图 3 中红色曲线所示。可以看到, 在 50 Hz 处和 120 Hz 处的系统响应都得到了较好的抑制, 且 ASE 稳定裕度都大于 6 dB, 满足了稳定裕度要求。

1.4 传统仿真分析方法的问题

通过传统 ASE 稳定性分析方法得到的系统开环传递函数曲线可以看出,在加入了适当的连续结构滤波器和低通滤波器后,原本不稳定的 ASE 系统满足了稳定裕度的要求。可是根据导弹飞行试验的结果,相同的条件下发生了 ASE 失稳的事故。通过对导弹飞行试验数据的分析,导弹发生 ASE 失稳时系统响应的频率约为 80 Hz,并不是传统分析方法结果关注的频率区间,这是一种新的 ASE 稳定性失稳模式。说明传统的 ASE 稳定性频域分析方法不能考虑数字式飞控系统采样过程对飞行器 ASE 稳定性的影响,无法通过仿真分析发现这种失稳模式的潜在危险。

2 考虑数字飞控系统的 ASE 稳定性仿真分析

为了完整、准确地反映出数字式飞控系统采样率对飞行器 ASE 稳定性特性的实际影响,本文根据实际的 ASE 试验工作流程,建立了一种考虑数字式飞控系统的 ASE 稳定性仿真分析方法。本节算例基于 MATLAB 软件的 Simulink 模块进行数值仿真,需要说明的是,该方法并不限制时域仿真分析软件的选择。

2.1 工作原理

根据第 1 节推导出的弹性结构机体、惯性敏感元件、舵机等各环节的传递函数,得到对应的状态空间方程,并搭建各环节对应的 Simulink 模块。

对于飞控系统,应按照飞行器采样过程实际采用的采样率进行离散化处理,搭建对应的 Simulink 模块。常见的离散化方法有双线性变换法、脉冲响应不变法、阶跃响应不变法和零极点匹配映射法等^[21]。本文采用的离散化方法为双线性

变换法(Tustin 方法)。

双线性变换法把整个模拟频率范围压缩至区间 $-\pi < \omega < \pi$,把 s 平面的左半平面映射到 z 平面单位圆的内部,避免了多值映射问题,当生成离散时间传输函数时,保持了连续时间传递函数的幅度频响特性。双线性变换法的变换公式为

$$G(z) = G(s) \Big|_{s = \frac{z-1}{z+1}} \quad (15)$$

式中: T 为离散采样时间。

对于实际飞行器的采样过程,可以在 Simulink 仿真模型中对应位置加入零阶保持器模块进行模拟。模块参数变量为采样时间,当采样时间设定为 0.002 5 s 时,代表仿真采样率为 400 Hz;当采样时间设定为 0.005 s 时,代表仿真采样率为 200 Hz。

按照图 1 所示框图进行模块组合,将闭环系统在数字式飞控系统与舵机系统之间断开,完成包含数字式飞控系统的弹性飞行器 ASE 开环系统的建模工作,搭建完成的飞行器俯仰通道 ASE 系统 Simulink 开环仿真模型如图 4 所示。

图 4 所示的开环仿真模型中,在舵机系统的输入信号(控制指令)处,加载由 MATLAB 软件生成的步进正弦激励信号,并用设定好的采样率进行采样得到开环仿真模型的输入激励信号;同时,在控制系统的解算指令信号处进行同采样率的采样,得到开环仿真模型的输出响应信号。最终得到开环仿真模型的输入、输出信号后,即可通过频响函数获取^[22]方法得到开环仿真模型的频响特性。频响函数获取方法包括傅里叶变换谱分析法、最小二乘法^[23]和相关分析法^[24]等,对于步进正弦扫频信号,在单一的频率点上有充足的数据,考虑到傅里叶变换谱分析法对噪声敏感,本文采用最小二乘法计算系统在每个频率点上的幅值和相位。

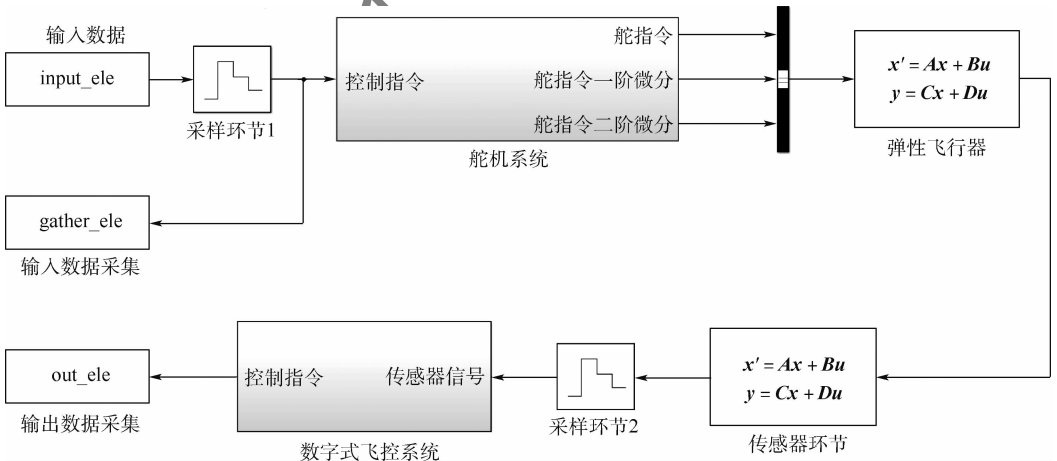


图 4 有数字式飞控系统的飞行器 ASE 系统开环 Simulink 模型

Fig. 4 Simulink model of aircraft ASE open-loop system with digital flight control system

按照图 1 所示框图进行组合,完成包含数字式飞控系统的弹性飞行器 ASE 闭环系统的建模

工作,搭建完成的飞行器俯仰通道 ASE 系统 Simulink 闭环仿真模型如图 5 所示。

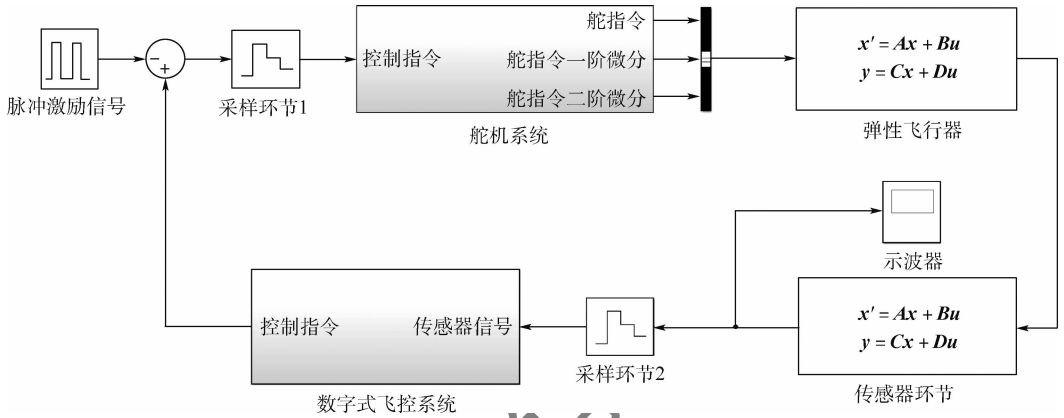


图 5 有数字式飞控系统的飞行器 ASE 系统闭环 Simulink 模型

Fig. 5 Simulink model of aircraft ASE close-loop system with digital flight control system

图 5 所示的闭环仿真模型中,在舵机系统的输入信号(控制指令)处,给予一个脉冲信号并进行时域仿真。仿真结束后,观察示波器的显示结果,如果示波器信号呈发散状,说明系统发生了 ASE 稳定性失稳;如果示波器信号呈收敛状,说明没有发生 ASE 稳定性失稳。此外,如果 ASE 系统发生失稳,通过观察发散信号的频率可以确定系统发生失稳的频率位置。

2.2 数值仿真

输入步进正弦信号的幅值为 1,频率范围为 10 ~ 200 Hz,每个频率点步进 0.1 Hz 且持续时间 2 s,频率点之间间隔 1 s,信号总时长为 5 713 s,总仿真时间为 6 000 s。

零阶保持器的采样时间设置为 0.002 5 s,即采样率为 400 Hz,飞控系统未串联结构滤波器和低通滤波器,得到的输入激励信号和响应信号的开环系统仿真结果如图 6 所示,应用最小二乘法得到的开环传递函数曲线如图 7 所示。由图 6 可

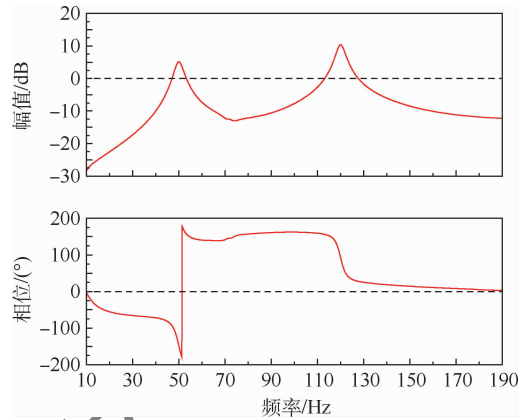


图 7 无滤波器开环传递函数曲线(采样率 400 Hz)

Fig. 7 Open-loop transfer function curves without filter

(sampling rate 400 Hz)

知,在 1 175 s 和 3 300 s 处,系统响应信号的幅值比较大,幅值已经大于 1。这 2 个时间点对应的输入激励信号的频率分别为 49.2 Hz 和 120 Hz。

采样频率保持为 400 Hz,在数字式飞控系统加入以 400 Hz 采样率离散化(Tustin 方法)后的结构滤波器和低通滤波器为

$$G_{\text{notch}}(z) = \frac{0.738 2z^2 - 1.04z + 0.680 1}{z^2 - 1.04z + 0.418 4} \quad (16)$$

$$G_{\text{lp}}(z) = \frac{0.094 67z^2 + 0.189 3z + 0.094 67}{z^2 - 1.04z + 0.418 4} \quad (17)$$

得到的输入激励信号和响应信号的开环系统仿真结果如图 8 所示,应用最小二乘法得到的开环传递函数曲线如图 9 所示。由图 9 可知,在整个仿真频域范围内,系统没有发生 ASE 失稳,且幅值稳定裕度大于 6 dB。

设置飞控系统采样率为 200 Hz,零阶保持器的采样时间设置为 0.005 s,也即采样率为 200 Hz。在数字式飞控系统加入以 200 Hz 采样率离散

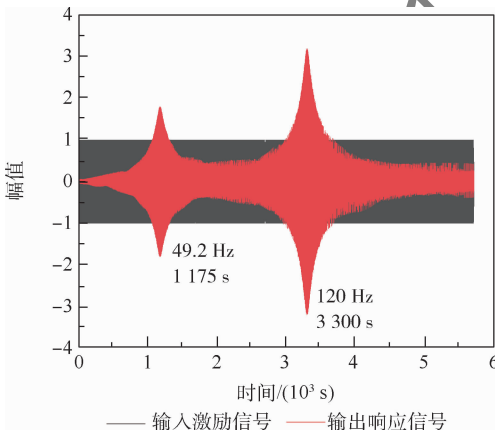


图 6 无滤波器开环仿真结果(采样率 400 Hz)

Fig. 6 Open-loop simulation results without filter

(sampling rate 400 Hz)

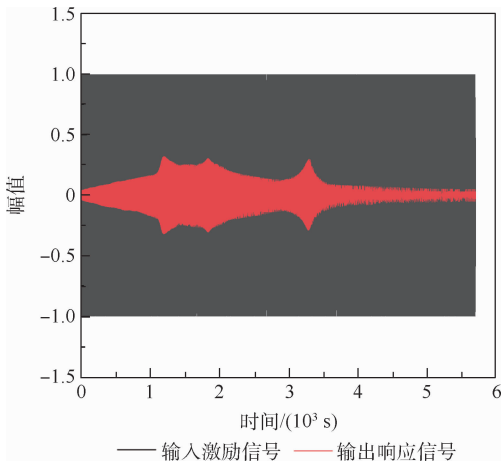


图 8 有滤波器开环仿真结果(采样率 400 Hz)
Fig. 8 Open-loop simulation results with filter (sampling rate 400 Hz)

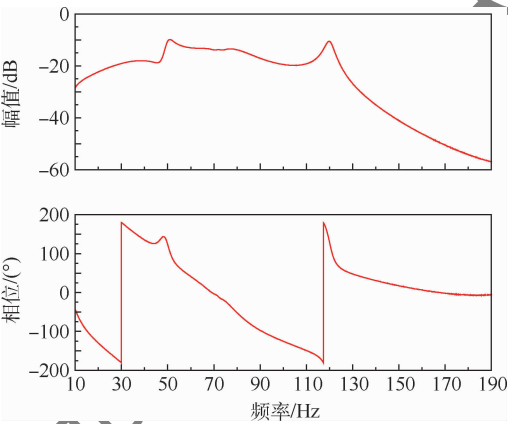


图 9 有滤波器开环传递函数曲线(采样率 400 Hz)
Fig. 9 Open-loop transfer function curves with filter (sampling rate 400 Hz)

结构滤波器(式(18))和低通滤波器(式(19))。时域仿真结束后,系统的闭环时域响应如图 12 所示。可知,惯性敏感元件的时域响应幅值很快发散。观察 0.8 ~ 1.0 s 之间,响应曲线共有 16 个周期,换算之后得到系统的闭环时域响应振荡频率为 80 Hz。

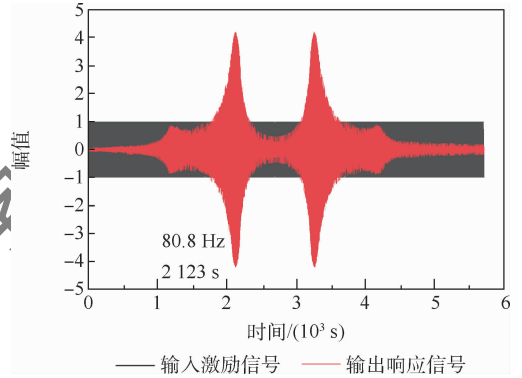


图 10 有滤波器开环仿真结果(采样率 200 Hz)
Fig. 10 Open-loop simulation results with filter (sampling rate 200 Hz)

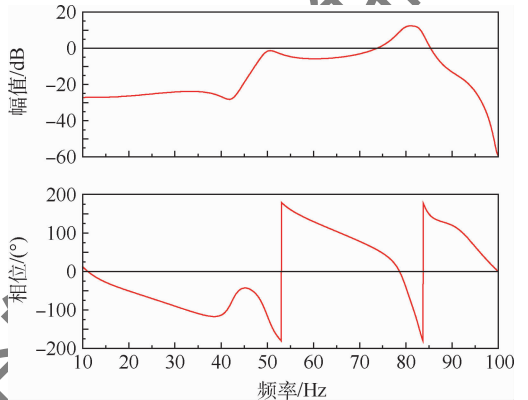


图 11 有滤波器开环传递函数曲线(采样率 200 Hz)
Fig. 11 Open-loop transfer function curves with filter (sampling rate 200 Hz)

化(Tustin 方法)后的结构滤波器和低通滤波器为

$$G_{notch}(z) = \frac{0.667 6z^2 - 0.299 5z + 0.593 7}{z^2 - 0.299 5z + 0.261 3} \quad (18)$$

$$G_{lp}(z) = \frac{0.240 5z^2 + 0.480 9z + 0.240 5}{z^2 - 0.299 5z + 0.261 3} \quad (19)$$

得到的输入激励信号和响应信号的开环系统仿真结果如图 10 所示,最小二乘法得到的开环传递函数曲线如图 11 所示。由图 10 可知,在 2 123 s 处,系统响应信号的幅值比较大,幅值已经大于 1。这个时间点对应的输入激励信号的频率为 80.8 Hz。此外,由于采样率为 200 Hz,已经小于输入激励信号中最高频率的 2 倍,导致响应信号的时域仿真结果关于 2 685.7 s 对称,这个时间点对应的输入激励信号的频率为 99.5 Hz。

在图 5 所示的有数字式飞控系统的飞行器 ASE 系统闭环 Simulink 模型中,设置零阶保持器的采样时间为 0.005 s,也即采样率为 200 Hz。数字式飞控系统中加入以 200 Hz 采样率离散化后的

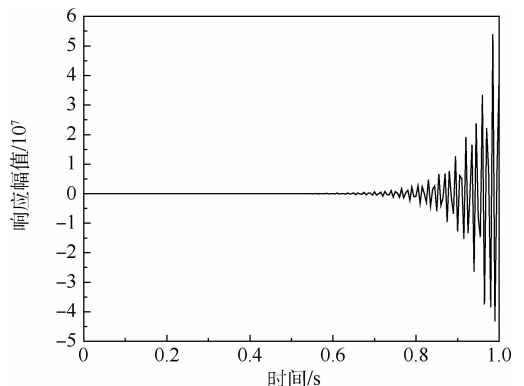


图 12 系统闭环时域响应(采样率 200 Hz)
Fig. 12 Closed-loop time domain response of system (sampling rate 200 Hz)

2.3 结果分析

通过数值仿真可以看出,在采样率设置为 400 Hz 时,采样率大于输入激励信号的最大频率

的2倍,考虑数字式飞控系统的 ASE 稳定性仿真分析的结果与传统频域 ASE 稳定性分析方法的分析结果是一致的,串联了以 400 Hz 采样率离散化后的结构滤波器和低通滤波器较好地抑制了 50 Hz 处和 120 Hz 处的系统响应。验证了考虑数字式飞控系统的 ASE 稳定性仿真分析方法的有效性。

当采样率设置为 200 Hz 时,采样率不足输入激励信号的最大频率的 2 倍。开环系统数值仿真结果表明,系统在 80 Hz 频率附近响应幅值较大,稳定裕度已经不足,且在 50 Hz 频率附近,虽然还未发生失稳,但是稳定裕度已经比较小。闭环系统时域仿真结果表明,系统在 80 Hz 频率附近发生了 ASE 失稳现象。这与该型导弹在飞行试验中出现的 ASE 失稳事故的现象一致,对应的失稳频率也一致。说明考虑数字式飞控系统的 ASE 稳定性仿真分析方法可以完整、准确地考虑数字式飞控系统采样率对飞行器 ASE 稳定性特性的影响。

3 新型失稳模式的机理探讨

通过该型导弹的飞行试验和数值算例可知,这一新型 ASE 失稳模式是在控制系统采样率为 200 Hz 时,在 80 Hz 频率附近发生了 ASE 失稳现象;在 50 Hz 频率附近,虽然还未发生失稳,但是稳定裕度已经较小。

3.1 频率混叠问题

频率混叠是指由于采样信号频谱发生变化,而出现高、低频成分发生混淆的一种现象。采样时频率不够高,抽样出来的点既代表了信号中的低频信号的样本值,也同时代表高频信号样本值,在信号重建时,高频信号被低频信号代替,2 种波形完全重叠在一起,形成严重失真。频率混叠是数字信号处理中的一个重要概念,它是数字信号处理中的特有现象,是由数字信号中离散采样引起的。凡是等步长离散采样,一定会产生频率混叠现象。频率混叠会产生假频率、假信号,严重影响测量结果。

根据时域抽样定理,当采样率 f_s 不能满足如下要求时:

$$f_s \geq 2f_{\max} \quad (20)$$

采样得到的信号及其频谱就会发生混叠^[25]。根据混叠机理可以得出分析信号的混叠频率的计算公式^[26]。设实际信号频率为 f ,采样率为 f_s ,并且 $f_s < 2f$,经采样分析得到的混叠后频率为 f_A ,折叠

频率 $f_c = f_s/2$,则有混叠频率为

$$f_A = |f - nf_s| \quad (21)$$

式中: $n = \text{Int}\left(\frac{f}{f_s} + 0.5\right)$ 。

根据式(21),当采样率为 200 Hz 时,系统中 120 Hz 处的响应信号会混叠到 80 Hz 处,而在理论设计阶段并未关注该频率段。这就导致了在导弹的 ASE 系统稳定性特性中,80 Hz 频率处的系统响应并未得到有效的抑制,甚至发生了 ASE 稳定性裕度已经不足的情况。

3.2 结构滤波器的移频现象

同一个连续结构滤波器,通过双线性变换法在不同的采样率下进行离散化,得到离散结构滤波器频响特性的幅值衰减幅度相同,但是会发生移频现象。

利用双线性变换法的变换式(15)对式(10)进行变换,可以得到双线性变换法以采样时间 T 离散化后的离散结构滤波器为

$$G(z) = \frac{A_1 z^2 + Bz + C_1}{A_2 z^2 + Bz + C_2} \quad (22)$$

式中: $z = e^{i\omega T}$; $A_1 = \omega_c^2 T^2 + 4T\xi_i \omega_i + 4$; $B = 2\omega_c^2 T^2 - 8$; $C_1 = \omega_c^2 T^2 - 4T\xi_i \omega_i + 4$ 。

令式(22)对 ω 进行求导并令导数为 0,可以得到连续结构滤波器通过双线性变换法离散化后的中心频率与采样时间的关系为

$$\omega_z = \frac{\arctan(4\omega_c T, -\omega_c^2 T^2 + 4)}{T} \quad (23)$$

式中: ω_z 为离散结构滤波器的中心频率,rad/s; ω_c 为连续结构滤波器中心频率,rad/s。

对于式(13)设计的连续结构滤波器,当采样率为 400 Hz 时,离散结构滤波器的中心频率变至 47.62 Hz;当采样率为 200 Hz 时,离散结构滤波器的中心频率变至 42.37 Hz,如图 13 所示。采样率越低,结构滤波器的移频现象越明显。

在采样率为 200 Hz 而未经过中心频率修正时,离散结构滤波器对导弹在 50 Hz 处的系统响应抑制效果已经不足。这最终导致了当数字式飞控系统的采样率变为 200 Hz 时,导弹在 50 Hz 处系统响应增大,ASE 稳定性裕度变小,存在发生 ASE 稳定性失稳的风险。

3.3 改进措施

对于 3.1 节和 3.2 节中 2 个由数字式飞控系统采样率带来的问题,可以通过提高数字式飞控系统采样率的方法来避免。不过,当前飞行器受限于处理器性能和带宽等,飞控系统的采样率一般都没有办法设置得很高。因此,本文提出了具

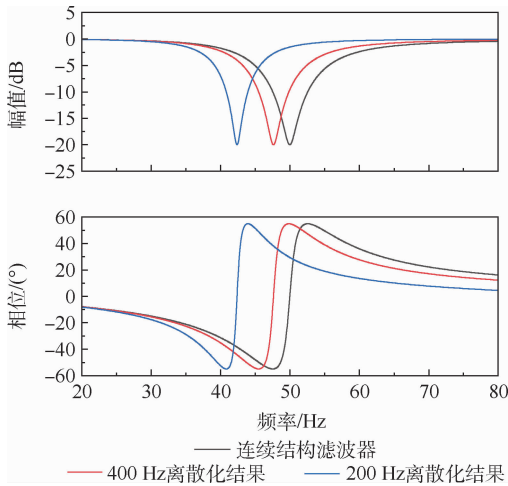


图13 不同采样率离散化的结构滤波器频响曲线
Fig. 13 Frequency response curves of discrete structured filters with different sampling rates

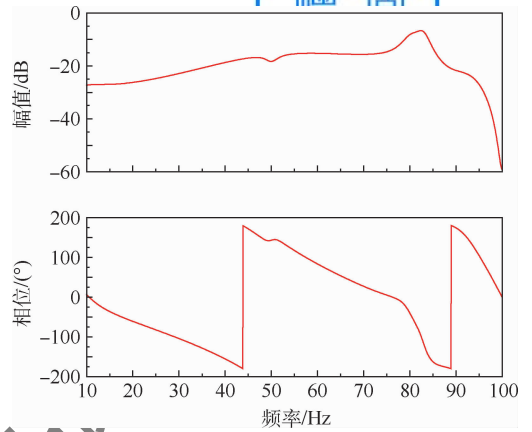


图14 系统辨识开环传递函数曲线(采样率200 Hz)
Fig. 14 Open-loop transfer function curves for system identification (sampling rate 200 Hz)

有实际应用价值的对应改进措施,并通过数值仿真验证了其效果。

针对离散结构陷波器移频现象,在已知采样率的情况下,可以通过推导的式(23)计算得出连续结构滤波器离散化后的中心频率偏移量。修改连续结构滤波器的参数对离散化后的结构滤波器中心频率进行校正,使得离散结构滤波器达到期望的效果。

经过校正,在200 Hz采样率下,离散化后中心频率为50 Hz、衰减幅度为20 dB的离散结构滤波器的差分方程为

$$G(z) = \frac{0.7109z^2 + 0.6466}{z^2 + 0.3575} \quad (24)$$

针对频率混叠问题,可以在采样过程之前添加防混叠滤波器。其本质是一种低通滤波器,可以过滤掉信号中的高频分量,避免频率混叠的问题。或者在已知数字式飞控系统采样率的情况下,通过式(24)计算得到混叠频率,并在该频率处预加设计好的结构陷波器,避免频率混叠现象对飞行器ASE稳定性造成影响。

经过校正,在200 Hz采样率下,离散化后中心频率为80 Hz、衰减幅度为20 dB的离散结构滤波器的差分方程为

$$G(z) = \frac{0.5107z^2 + 0.7386z + 0.4019}{z^2 + 0.7386z - 0.08737} \quad (25)$$

在数字式飞控系统串联了式(24)和式(25)所示的离散结构滤波器后,再次对该型导弹进行数值仿真,系统辨识得到的开环传递函数曲线如图14所示。可知,改进措施效果明显,导弹在全频率段的系统响应都比较小,满足了稳定裕度的要求。

4 结论

1) 对于采用数字式飞控系统的飞行器,较低的采样率会造成频率混叠和结构滤波器中心频率偏移这2种效应,二者均可能引起ASE稳定裕度不足或失稳。

2) 为避免出现以上情况,飞行器控制系统应采用较高的采样率。当采样率无法提高时,针对频率混叠应采用在混叠频率处预加结构滤波器或在采样过程之前串联防混叠滤波器的方法;针对结构滤波器移频问题,应采用中心频率预校正的办法。

3) 所建立的考虑数字式飞控系统的ASE仿真分析方法可以反映出数字式飞控系统采样过程对飞行器ASE稳定性的影响,可以通过仿真分析发现这种ASE失稳模式的潜在危险。

参考文献 (References)

[1] RAMSEY J K. NASA aeroelasticity handbook. Volume 2: Design guides. Part 2: NASA/TP-2006-212490 [R]. Washington, D. C.: NASA Glenn Research Center, 2006.

[2] 杨超, 黄超, 吴志刚, 等. 气动伺服弹性研究的进展与挑战 [J]. 航空学报, 2015, 36(4): 1011-1033.
YANG C, HUANG C, WU Z G, et al. Progress and challenges for aeroservoelasticity research [J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(4): 1011-1033 (in Chinese).

[3] 杨超, 吴志刚. 导弹气动伺服弹性稳定性分析 [J]. 飞行力学, 2000, 18(4): 1-5.
YANG C, WU Z G. Aeroservoelastic stability of missile [J]. Flight Dynamics, 2000, 18(4): 1-5 (in Chinese).

[4] KARPEL M. Procedures and models for aeroservoelastic analysis and design [J]. ZAMM Journal of Applied Mathematics and Mechanics, 2001, 81(9): 579-592.

[5] GUPTA K K, MEEK J L. Finite element multidisciplinary analysis [M]. 2nd ed. Reston: AIAA, 2003.

- [6] 杨超,吴志刚,万志强,等.飞行器气动弹性原理[M].北京:北京航空航天大学出版社,2011:148-167.
YANG C, WU Z G, WAN Z Q, et al. Principle of aircraft aeroelasticity[M]. Beijing: Beihang University Press, 2011: 148-167 (in Chinese).
- [7] 胡寿松.自动控制原理[M].5版.北京:科学出版社,2007:206-212.
HU S S. Automatic control principle[M]. 5th ed. Beijing: Science Press, 2007: 206-212 (in Chinese).
- [8] KARPEL M, MOULIN B, IDAN M. Robust aeroservoelastic design with structural variations and modeling uncertainties[J]. Journal of Aircraft, 2003, 40(5): 946-954.
- [9] DAI Y T, YANG C. Methods and advances in the study of aeroelasticity with uncertainties[J]. Chinese Journal of Aeronautics, 2014, 27(3): 461-474.
- [10] 蒋慰孙,叶银忠.多变量控制系统分析与设计[M].北京:中国石化出版社,1997:129-133.
JIANG W S, YE Y Z. Analysis and design of multivariable control system[M]. Beijing: China Petrochemical Press, 1997: 129-133 (in Chinese).
- [11] PITT D, HAYES B, GOODMAN C. F/A-18E/F aeroservoelastic design, analysis, and test: AIAA 2003-1880[R]. Reston: AIAA, 2003.
- [12] 高金源,等.计算机控制系统——理论、设计与实现[M].北京:北京航空航天大学出版社,2001:94-112.
GAO J Y, et al. Computer control system—Theory, design and implementation[M]. Beijing: Beihang University Press, 2001: 94-112 (in Chinese).
- [13] 肖建,徐志根.多采样率数字控制系统综述[J].信息与控制,2003,32(5):436-441.
XIAO J, XU Z G. Survey on the research of multirate digital control systems[J]. Information and Control, 2003, 32(5): 436-441 (in Chinese).
- [14] 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).
- [15] 宋晨,杨超,吴志刚.3种气动弹性状态空间建模方法的对比[J].航空学报,2007,28(S1):81-86.
SONG C, YANG C, WU Z G. Comparison of three aeroelastic state-space modeling methods[J]. Acta Aeronautica et Astronautica Sinica, 2007, 28(S1): 81-86 (in Chinese).
- [16] 楚龙飞,吴志刚,杨超,等.导弹自适应结构滤波器的设计与仿真[J].航空学报,2011,32(2):195-201.
CHU L F, WU Z G, YANG C, et al. Design and simulation of adaptive structure filter for missile[J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(2): 195-201 (in Chinese).
- [17] 陈桂彬,杨超,邹丛青.气动弹性设计基础[M].2版.北京:北京航空航天大学出版社,2010:179-189.
CHEN G B, YANG C, ZOU C Q. Aeroelastic design basis[M]. 2nd ed. Beijing: Beihang University Press, 2010: 179-189 (in Chinese).
- [18] YAMASHIRO H, STIRLING R. Reduction of flight control system/structural mode interaction: AIAA 2007-6381[R]. Reston: AIAA, 2007.
- [19] CUNNINGHAM D C, HIGGINS W T. A comparison of conventional and tracking filter systems for launch vehicle stabilization[J]. Journal of Spacecraft and Rockets, 1970, 7(8): 934-940.
- [20] WIE B, BYUN K. A new concept of generalized structural filtering for active vibration control synthesis[C]//Guidance, Navigation and Control Conference. Reston: AIAA, 1987.
- [21] 郭宝龙,闫允一,朱娟娟.工程信号与系统[M].北京:高等教育出版社,2014:312-318.
GUO B L, YAN Y Y, ZHU J J. Engineering signal and system[M]. Beijing: Higher Education Press, 2014: 312-318 (in Chinese).
- [22] ZENG J, KUKREJA S L, MOULIN B. Experimental model-based aeroelastic control for flutter suppression and gust-load alleviation[J]. Journal of Guidance, Control, and Dynamics, 2012, 35(5): 1377-1390.
- [23] 张仁嘉.飞行器气动伺服弹性若干关键问题研究[D].北京:北京航空航天大学,2015:101-102.
ZHANG R J. Extensional research on several critical aeroservoelastic problems of air vehicles[D]. Beijing: Beihang University, 2015: 101-102 (in Chinese).
- [24] 潘家保,刘慧,贾宏光,等.电动舵机伺服系统的模型辨识及其校正[J].光学精密工程,2008,16(10):1971-1976.
ZHANG J B, LIU H, JIA H G, et al. Model identification and corrector design for servo system of electromechanical actuator[J]. Optics and Precision Engineering, 2008, 16(10): 1971-1976 (in Chinese).
- [25] 李泽光.信号与系统分析和应用[M].北京:高等教育出版社,2015:145-154.
LI Z G. Signal and system analysis and application[M]. Beijing: Higher Education Press, 2015: 145-154 (in Chinese).
- [26] 应怀樵,沈松,刘进明.频率混叠在时域和频域现象中的研究[J].振动、测试与诊断,2006,26(1):1-4.
YING H Q, SHEN S, LIU J M. Study on frequency aliasing in time and frequency domains[J]. Journal of Vibration, Measurement & Diagnosis, 2006, 26(1): 1-4 (in Chinese).

Mechanism analysis of a new aeroservoelastic instability mode

JIANG Yu, YANG Chao, WU Zhigang*

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: Recently, the structure of a supersonic missile disintegrated due to aeroservoelastic instability in flight test. Through the analysis of flight test data, it is found that the vibration frequency of aerodynamic servo elastic instability is higher than the first-order bending elastic modal frequency of missile body. Based on the frequency domain analysis of aerodynamic servo elastic stability of missile, it is not found that aeroservoelastic instability occurs in this frequency band. To solve this problem, a simulation analysis method of aeroservoelastic stability considering the influence of sampling process of digital flight control system is established, and the missile is modeled and analyzed. The numerical results reproduce the instability phenomenon of the missile. The causes of this new instability phenomenon are discussed, including the frequency shift phenomenon caused by the discretization of continuous structure filter and frequency aliasing. Finally, the corresponding improvement measures and relevant conclusions are given.

Keywords: aeroservoelastic; discretization; structure filter; sampling rate; frequency aliasing

北京航空航天大学学报

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

DOI: 10.13700/j.bh.1001-5965.2021.0593

面向扑翼翼面运动参数的优化设计

吴越, 谢长川*, 杨超, 安朝

(北京航空航天大学 航空科学与工程学院, 北京 100083)



摘 要: 扑翼机的飞行依赖于扑翼翼面的运动, 经过优化的运动策略能够使特定翼面发挥最佳的气动性能。然而目前扑翼机设计中缺乏有效的运动参数优化方法, 无法针对给定机翼确定一组最优运动参数。采用非定常涡格法(UVLM)计算扑翼气动力, 与现有的实验数据进行对比, 验证了气动力计算方法的准确性。基于DIRECT(矩形分割)全局优化算法, 以最大化推进效率为特定优化目标, 对扑翼运动参数进行了迭代优化。结果表明, 通过该优化算法能够得到最优扑翼运动参数, 有效提高特定气动性能; 应用优化算法计算得到的平均推力与基准运动的平均推力相比, 在数值上有1.04倍的提高。在设计过程中, 降低气动力约束有利于扑翼运动优化, 使给定扑翼翼面具有更大的推进效率, 无气动力约束的最大推进效率与基准运动的推进效率相比提高了46.8%。

关键词: 非定常气动力; 涡格法; 全局优化; 扑翼; 飞行器设计

中图分类号: V222; TB553

文献标志码: A **文章编号:** 1001-5965(2022)07-1324-08

扑翼飞行的气动特性与扑翼运动模式密切相关。一方面, 对于特定形状结构的翼面, 不同的扑翼运动模式会导致气动特性出现显著差异。另一方面, 扑翼机往往需要先确定机械结构, 在制造完成后具有固定的扑翼运动模式。因此, 有必要在设计扑翼机结构之前确定一种高效的扑翼运动策略。Keennon等^[1]在开发微型扑翼机的过程中总结发现, 传统扑翼结构设计后, 为了确定翼面扑动和俯仰的耦合运动模式, 运动参数只能通过实验在数十个运动参数组合中进行选择, 限制了扑翼机的应用。Nick等^[2]制造了一架仿蝙蝠扑翼机, 其运动模式采用简单的平面内对称扑动, 并未采用运动优化方法来提高其气动特性。Razak和Dimitriadis^[3]通过实验探究了扑动和俯仰的耦合运动模式对扑翼气动特性的影响, 其实验给定了固定的扑动和俯仰幅度, 只能探究2种运动的相位差的气动效应, 不能综合考虑其他运动参数的

影响。Srigarom和Chan^[4]探究了大型鸟类扑翼的流场和噪声特性, 由于缺乏运动过程的相关数据, 其采用简单的锯齿函数表征扑动角度随时间变化的过程。在扑翼运动设计和扑翼机理研究领域, 能够有效进行特定机翼运动设计的优化方法迄今仍比较少见。

扑翼具有复杂的设计空间, 通常以特定气动特性为优化目标时具有多个局部最优点。Stanford和Beran^[5]研究了平飞扑翼的主动变形参数优化设计问题, 使用序列二次规划法, 得到高频和低频2种扑动频率下的最佳主动变形参数。Ghommem等^[6]在进行扑翼的形状优化研究时, 将基于梯度的优化方法与非定常涡格法(UVLM)相结合, 确定了一组最大化推进效率的翼面形状参数。同时, Ghommem等^[7]还开展了扑翼主动变形策略的优化研究, 发现采用全局优化方法能够得到与基于梯度的优化方法不同的最优参数组合。

收稿日期: 2021-10-09; 录用日期: 2021-11-14; 网络出版时间: 2021-11-22 17:16

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20211122.1307.003.html

*通信作者: E-mail: xiechange@buaa.edu.cn

引用格式: 吴越, 谢长川, 杨超, 等. 面向扑翼翼面运动参数的优化设计[J]. 北京航空航天大学学报, 2022, 48(7): 1324-1331.

WU Y, XIE C C, YANG C, et al. Optimal design of motion parameters of flapping wing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1324-1331 (in Chinese).

DIRECT 优化算法^[8]是 Lipschitzian 优化算法的改进方法,是一种为了解决有界约束和实值目标函数问题而建立的全局优化算法。DIRECT 算法的优势在于:不需要目标函数的梯度信息,而是对给定域中的点进行采样,使得 DIRECT 算法适合于“黑箱”函数的仿真模拟;其代价是:需要在给定域上进行大规模的搜索,来保证算法收敛到最优解。

扑翼运动具有低频率的特性,为了达到稳定状态需要计算若干个扑翼周期,每个计算状态需要一定的计算时间。另外,全局优化方法需要对数百个计算状态进行采样,气动力方法计算效率对运动参数优化具有重要影响。非定常涡格法计算效率高,具有中等保真度,能够以较少的计算耗费取得较好的仿真结果。鉴于这类计算方法的效率,诸多学者将其应用于飞行中的仿鸟扑翼设计^[9-12]。Persson 等^[13]通过对非定常涡格法和扑翼高保真 CFD 仿真结果的比较,证明非定常涡格法能够准确预测扑翼气动效应,甚至当存在流动分离时仍保持了趋势相关性。近年来,非定常涡格法在扑翼对象上的应用有了新的进展。Roccia 等^[14]研究了非定常涡格法在低雷诺数条件下的应用,计算结果与实验数据吻合较好,能够基于果蝇或小型鸟类的实际机翼运动进行流场仿真和升力预测。这些研究表明非定常涡格法能够计算扑翼运动引起的非定常效应。

本文首先将非定常涡格法应用于扑翼气动力计算,引入匹配相关实验的上下平移运动模型,与现有实验数据进行对比,验证气动计算工具的准确性。然后建立符合实际扑翼运动的扑动和俯仰耦合运动模型,基于参数化扑翼运动方程,应用 DIRECT 全局优化算法,在给定参数设计空间内求解最大化推进效率的扑翼运动参数。

1 气动力和优化计算方法

1.1 非定常涡格法

基于无黏流理论的非定常涡格法可以用来求解薄翼气动力和预测流场演化,气动面自由运动的尾迹由离散涡单元构成,其中涡的瞬时速度受 Biot-Savart 定律控制^[15]。由附着涡和自由涡诱导引起的诱导速度用向量 $[\dot{x}, \dot{y}, \dot{z}]$ 表示,来流的速度为 U_∞ 。在算法的时域推进过程中,每个时间间隔的自由涡位移可由式(1)计算:

$$[\Delta x, \Delta y, \Delta z] = ([\dot{x}, \dot{y}, \dot{z}] + U_\infty) \Delta t \quad (1)$$

在扑翼涡格法求解时,需要计算若干个扑动周期来保证气动力计算结果不受扑动起始过程的

影响,本文算例中,每个计算状态至少求解了5个扑动周期,并取最后一个扑动周期的数据作为该状态的最终结果。扑翼的气动网格初始化后,就可以在每个时间步计算每个网格的压力:

$$\Delta p_i = \rho \left\{ U_i \cdot \tau_x \frac{\Gamma_i - \Gamma_x}{\Delta \tau_x} + U_i \cdot \tau_y \frac{\Gamma_i - \Gamma_y}{\Delta \tau_y} + \frac{\partial}{\partial t} \Gamma_i \right\} \quad (2)$$

式中: ρ 为空气密度; τ_x 和 τ_y 分别为第 i 个涡格沿弦向和展向的单位向量; $U_i \cdot \tau_x$ 和 $U_i \cdot \tau_y$ 分别计算了自由流速在涡格沿弦向和展向的投影; Γ_i 为第 i 个涡格的涡环强度; Γ_x 和 Γ_y 分别表示在 $-\tau_x$ 和 $-\tau_y$ 方向与第 i 个涡格相邻的涡强度。在 $-\tau_x$ 和 $-\tau_y$ 方向无相邻涡格时,如前缘附着涡,则用 Γ_i 替换式(2)中的 $\Gamma_i - \Gamma_x$ 或者 $\Gamma_i - \Gamma_y$ 。因此,通过计算所有附着涡的压力积分可以得到机翼的升力和阻力为

$$L = \sum_{i=1}^N \Delta p_i \Delta S_i \cos \alpha \quad (3)$$

$$D = \sum_{i=1}^N \Delta p_i \Delta S_i \sin \alpha \quad (4)$$

式中: ΔS_i 为第 i 个附着涡单元的面积; N 为涡单元总数; α 为翼面法向量与升力方向的夹角。

假定机翼总面积为 S ,相应的升力和阻力系数的计算公式为

$$C_L = \frac{2L}{\rho U_\infty^2 S} \quad (5)$$

$$C_D = \frac{2D}{\rho U_\infty^2 S} \quad (6)$$

式中: U_∞ 为来流速度 U_∞ 的标量值。

1.2 全局优化方法

扑翼设计空间具有多个局部最优点,本文采用 DIRECT 全局优化算法求解最优扑动参数。求解扑翼气动特性的非定常涡格法采用时域推进仿真,其结果可以作为优化算法的目标函数,为采样点赋予函数值。而扑翼气动特性的满足条件可以作为约束函数,判定采样点是否符合设计规范。假设目标函数为 f ,约束函数为 f_c ,DIRECT 全局优化算法可以求解以下问题:

$$\begin{aligned} \min f(x) \\ l \leq x \leq u, f_c \leq 0 \end{aligned} \quad (7)$$

算法的初始化是将优化问题的 N 个待解参数的给定域映射到 N 维的边长为1的超立方体,令 e_i 为超立方体的第 i 维度的单位向量,在第1步优化迭代中,对超立方体的中心点 c_1 和 $c_1 \pm \frac{1}{3}e_i$ 的点进行采样,可以得到这些采样点的最小值为

$$f_{\min} = \min\left(f\left(\mathbf{c}_i \pm \frac{1}{3}\mathbf{e}_i\right), f(\mathbf{c}_i)\right) \quad 1 \leq i \leq N \quad (8)$$

然后进行第 1 步优化迭代的超立方体分割。分割的目的是将函数值最小的采样点保留在范围最大的区域中。定义：

$$w_i = \min\left(f\left(\mathbf{c}_i + \frac{1}{3}\mathbf{e}_i\right), f\left(\mathbf{c}_i - \frac{1}{3}\mathbf{e}_i\right)\right) \quad 1 \leq i \leq N \quad (9)$$

首先,将超立方体在最小的 w_i 对应的第 i 维度分割为三等份;然后,选取剩下 w_i 的最小值,并在其对应的维度上将剩下的超长方体分割为三等份。重复这一操作,将超立方体完全分割,得到一系列超立方体和超长方体。令 d_i 为单个超体的度量,通常可以定义为该超体的中心到顶点的距离,可以根据以下方法判定第 j 个超体是否包含潜在的最优解。令超体的集合为 I ,根据其度量可以分为 3 类:

$$\begin{cases} I_1 = \{i \in I; d_i < d_j\} \\ I_2 = \{i \in I; d_i > d_j\} \\ I_3 = \{i \in I; d_i = d_j\} \end{cases} \quad (10)$$

如果第 j 个超体包含潜在最优解,则满足:

$$f(\mathbf{c}_j) \leq f(\mathbf{c}_i) \quad \forall i \in I_3 \quad (11)$$

或者满足 ϵ 为大于 0 的正常数,当 $\hat{K} > 0$ 存在时

$$\max_{\forall i \in I_1} \frac{f(\mathbf{c}_j) - f(\mathbf{c}_i)}{d_j - d_i} \leq \hat{K} \leq \min_{\forall i \in I_2} \frac{f(\mathbf{c}_i) - f(\mathbf{c}_j)}{d_i - d_j} \quad (12)$$

且

$$\epsilon \leq \frac{f_{\min} - f(\mathbf{c}_j)}{|f_{\min}|} + \frac{d_j}{|f_{\min}|} \min_{\forall i \in I_2} \frac{f(\mathbf{c}_i) - f(\mathbf{c}_j)}{d_i - d_j} \quad f_{\min} \neq 0 \quad (13)$$

或者

$$f(\mathbf{c}_j) \leq d_j \min_{\forall i \in I_2} \frac{f(\mathbf{c}_i) - f(\mathbf{c}_j)}{d_i - d_j} \quad f_{\min} = 0 \quad (14)$$

将包含潜在的最优解超体重新分割,计算采样点和函数值,进入下一步的优化迭代。对于给定约束的优化问题,当采样点超出约束条件时,可以采用罚函数法,人为给定目标函数较大的值,就可以有效避免该采样点所在的区域被判定为包含潜在最优解的区域。优化问题通常分为有约束和无约束 2 种。在给定设计空间中搜索符合约束的设计点,可以采用罚函数法,当超出约束时人为指定一个较大的值,来避免该点周围的区域被判定为潜在最优空间,而无约束问题则不需要采用罚函数法。

以二维 Shubert 函数为例,其表达式为

$$f(x_1, x_2) = \left(\sum_{j=1}^5 j \cos((j+1)x_1 + j) \right) \cdot \left(\sum_{j=1}^5 j \cos((j+1)x_2 + j) \right) \quad -2 \leq x_1, x_2 \leq 2 \quad (15)$$

如图 1 所示,在区间 $-2 \leq x_1, x_2 \leq 2$ 上具有多个极小值点,最小值为 -186.83 ,用 DIRECT 全局优化算法进行求解,经过 20 个周期迭代的结果为 -186.729 ,对应 $x_1 = -1.4248, x_2 = -0.8011$ 或 $x_1 = -0.8011, x_2 = -1.4248$ 。每个周期的最小值如图 2 所示,图 3 显示了 DIRECT 全局优化算法在搜索域中的采样点,在全局最小值的邻近区域内进行了密集的采样。

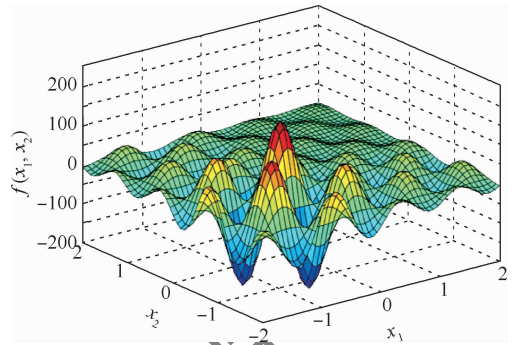


图 1 二维 Shubert 函数

Fig. 1 Plot of two-dimensional Shubert function

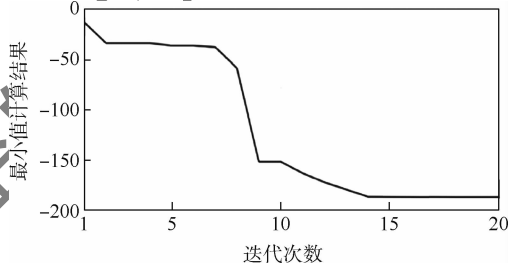


图 2 最小值计算结果随迭代次数的变化

Fig. 2 Minimum value vs. number of iterations

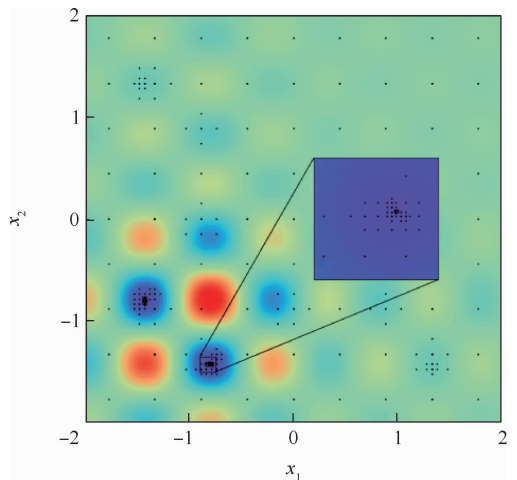


图 3 DIRECT 全局优化算法在搜索域中的采样情况

Fig. 3 Sampling of DIRECT global optimization algorithm in search domain

2 气动力方法验证和优化结果分析

2.1 气动力方法验证

为了验证非定常涡格法对扑翼气动力计算的准确性,本节根据文献[16-17]中的实验数据和数值结果,采用展长为 300 mm、弦长为 100 mm 的机翼模型进行计算对比。实验中给定了上下简谐振荡的运动形式,如图 4 所示,平移的距离为 h 。这种运动形式适合实验系统的设计和搭建。给定运动随时间变化满足:

$$h(t) = h_m \cos(2\pi ft) \quad (16)$$

式中: h_m 为扑动幅度; f 为扑动频率。

本文主要研究了三维气动力方法与全局优化方法相结合的计算框架,为了简化建模过程,选用不具有弯度的薄矩形翼模型,忽略其厚度。机翼平面形状是展长为 b 、弦长为 c 的矩形机翼。

验证算例计算了扑动幅度为 $h_a = 0.175c$, 减缩频率为 $k = \pi fc / U_\infty = 1.82$ 的状态下的升力系数,非定常涡格法生成的动态网格如图 5 所示,灰色表示翼面附着涡格,当翼面向下运动时,产生尾涡量为负,当翼面向上运动时,产生尾涡量为正。与相应的实验数据和数值结果进行对比,如图 6 所示。

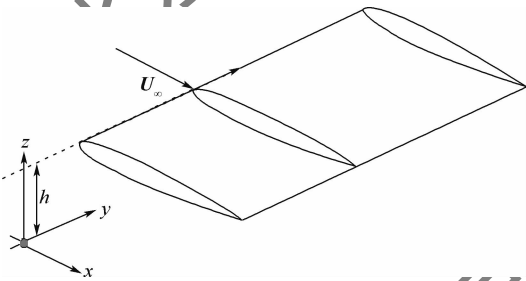


图 4 扑翼机翼上下简谐振荡的运动示意图

Fig. 4 Schematic diagram of vertical-harmonic-oscillating flapping wing motion

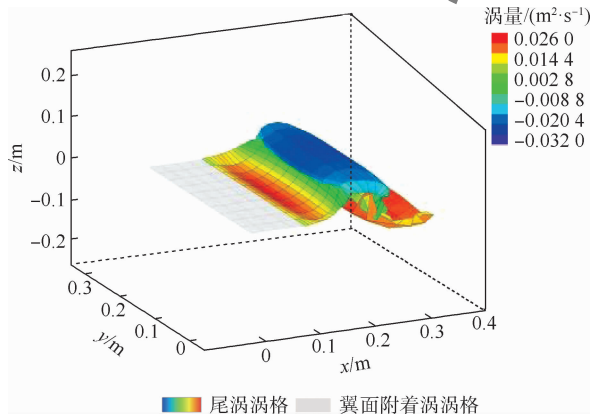


图 5 扑翼运动非定常涡格法动态网格的三维视图

Fig. 5 Three-dimensional view of UVLM dynamic mesh caused by flapping motion

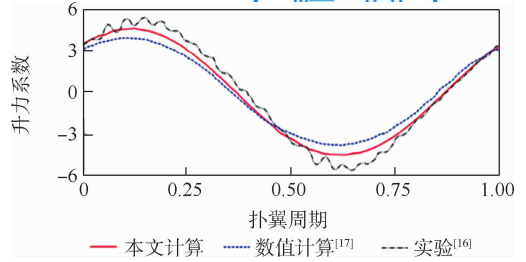


图 6 完整扑翼周期内的升力系数变化

Fig. 6 Lift coefficient caused by flapping motion in a complete cycle

从图 6 可以看出,本文采用非定常涡格法的计算结果与实验数据^[16]符合良好,与 Aono 等^[17]的数值计算结果相比,有更靠近实验数据的趋势。升力系数实验数据还存在短周期的波动,这是由于翼面模型无法做到完全刚性,引起小幅值弹性振动。

2.2 优化计算和结果分析

仿鸟类飞行器通常采用根部绕点扑动和俯仰的耦合运动,翼面每个区域既存在上下运动,也存在横侧向运动。因此,扑翼运动为图 7 所示的转动和平移相耦合的运动形式,转动中心配置于机翼根部前缘,俯仰角度为 θ ,扑动角度为 γ 。

根据 Wolf 和 Konrath^[18]对真实鸟翼的测量,本节采用与其相同形状参数的矩形机翼,展长为 0.454 m,弦长为 0.142 m。Lang 等^[19]对 Wolf 鸟类翅膀运动学建模得到的参数化运动模式,本文将其中半展长剖面运动转换到三维扑翼运动,可以得到俯仰角度 θ 和扑动角度 γ 的表达式为

$$\gamma(t) = \gamma_m \cos(26.07t) \quad (17)$$

$$\theta(t) = \theta_0 - 16.48\sin(26.07t) + 0.8752\cos(26.07t) - 0.2124\sin(52.14t) - 5.302\cos(52.14t) - 1.093\sin(78.21t) - 0.2537\cos(78.21t) - 0.2968\sin(104.28t) - 0.9419\cos(104.28t) \quad (18)$$

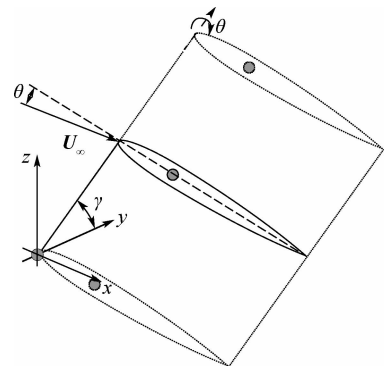


图 7 扑翼机翼绕点扑动和俯仰的耦合运动

Fig. 7 Schematic diagram of coupled flapping wing motion of flapping and pitching around a point

文献[19]中原始模型给定参数 $\gamma_m = 27.86^\circ$ 和 $\theta_0 = 5.46^\circ$, 则扑动运动的俯仰角度和平移距离如图8所示, 扑动角度 $\gamma(t)$ 符合正弦运动。

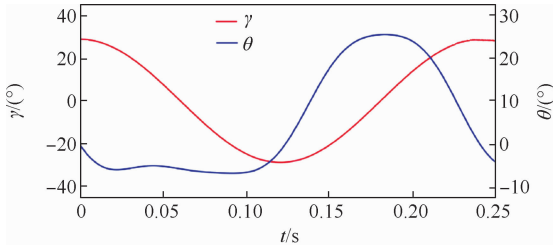


图8 仿生扑翼运动学的俯仰角度和扑动角度
Fig.8 Pitch angle and translation distance of bionic flapping motion

本节的计算状态为减缩频率 $k = 0.35$ 、来流速度 $U_\infty = 5.2895 \text{ m/s}$ 。与验证算例相比, 用于优化计算的雷诺数更高, 减缩频率更低, 其黏性效应和非定常气动效应更小, 在无前缘分离假设条件下, 该气动力计算方法能够估算扑翼的气动效应。

扑翼运动是保证和控制扑翼飞行器前飞的主要机制。扑翼运动应满足设计需求, 首先产生足够的升力支撑机体, 以及足够的推力维持向前飞行, 然后最小化扑翼克服的空气动力, 即效率最大化, 达到理想的扑翼飞行状态。为了展示本文计算框架对于有约束和无约束2种优化问题的适用性, 求解了在有升力和推力约束及无升力和推力约束2种条件下, 最大化推进效率的扑翼运动参数 θ_0 和 γ_m 的最优组合。针对实际扑翼机设计问题, 可以根据扑翼飞行器质量和所需升力确定平均升力约束, 根据飞行阻力确定平均推力约束。基于涡格法的扑翼推进效率 η 定义为平均推进功率 \bar{P}_{out} 和平均气动功率 \bar{P}_{in} 之比:

$$\eta = \bar{P}_{\text{out}} / \bar{P}_{\text{in}} \quad (19)$$

平均推进功率和平均气动功率是周期内的时间平均量, 瞬时推进功率和瞬时气动功率定义为

$$P_{\text{out}} = -DU_\infty \quad (20)$$

$$P_{\text{in}} = \sum_{i=1}^N \Delta p_i \Delta S_i \mathbf{n}_i \cdot \mathbf{v}_{\text{motion}}^i \quad (21)$$

式中: D 为机翼的阻力; \mathbf{n}_i 和 $\mathbf{v}_{\text{motion}}^i$ 分别为第 i 个翼面附着涡的法向单位向量和瞬时运动速度。当扑翼产生正推力时, 水平方向产生负阻力, 则推进功率大于零。给定大范围扰动的 θ_0 和 γ_m 的范围, 无升力和推力约束条件下优化问题可以表述为

$$\max \eta$$

$$-5^\circ \leq \theta_0 \leq 10^\circ, 0^\circ \leq \gamma_m \leq 40^\circ \quad (22)$$

对于有约束的优化问题, 本文以文献[19]中

初始运动学模型参数为比较基准, 产生的平均推力 \bar{D}_0 和平均升力 \bar{L}_0 为约束条件(平均推力和平均升力是推力和升力在拍动周期内的时间平均量), 可得 $\bar{D}_0 = -0.095 \text{ N}$, $\bar{L}_0 = 0.414 \text{ N}$ 。则有升力和推力约束条件的优化问题可以表述为

$$\max \eta$$

$$-5^\circ \leq \theta_0 \leq 10^\circ, 0^\circ \leq \gamma_m \leq 40^\circ$$

$$\bar{D} \leq \bar{D}_0, \bar{L} \geq \bar{L}_0 \quad (23)$$

当采样点超出升力和推力约束条件时, 指定罚函数为 $\eta = 0$ 。采用基于非定常涡格法的 DIRECT 全局优化算法求解以上优化问题, 仿生扑翼运动学参数的优化结果如表1所示。有气动力约束的条件下平均推力与基准运动的平均推力相比, 在数值上有1.04倍的提高。无气动力约束的最大推进效率与基准运动的推进效率相比提高了46.8%。

表1 DIRECT 全局优化算法的最优运动学结果
Table 1 Optimal kinematics results of DIRECT global optimization algorithm

项目	基准运动参数	有气动力约束	无气动力约束
$\theta_0 / (^\circ)$	5.46	5.28	-0.17
$\gamma_m / (^\circ)$	27.86	35.73	25.32
η	0.222	0.235	0.326
\bar{D} / N	-0.095	-0.194	-0.111
\bar{L} / N	0.414	0.414	-0.006

令推力系数 $C_T = -C_D$, 3种运动参数的升力系数和推力系数随时间的变化如图9所示。本节的优化计算在有升力和推力约束条件下进行全局搜索, 以找到一个合适的运动学参数组合来满足设计规范。在基准运动参数下, 给定优化问题的平均推力 \bar{D}_0 和平均升力 \bar{L}_0 约束边界, 虽然是多参数优化, 但2个设计变量所表示的扑翼运动仍较少, 导致了有升力和推力约束的最优运动学结果接近基准运动参数结果, 在之后的研究中可以尝试增加运动设计变量。无升力和推力约束的优化问题是为了探究该运动模型能够达到的最大推进效率, 同时比较升力和推力约束条件对优化求解的影响。无升力和推力约束结果仍是一个扑动和俯仰组合的运动, 无升力和推力约束推进效率明显大于有升力和推力约束的结果。

在有升力和推力约束最优参数状态下, 平均升力系数与基准运动参数状态接近, 而平均推力系数有了较大的提升, 也提高了推进效率。有升力和推力约束最优参数的 γ_m 与 θ_0 与基准运动参数相似。无升力和推力约束最优参数状态下, 平均升力系数和平均推力系数均小于基准运动参数

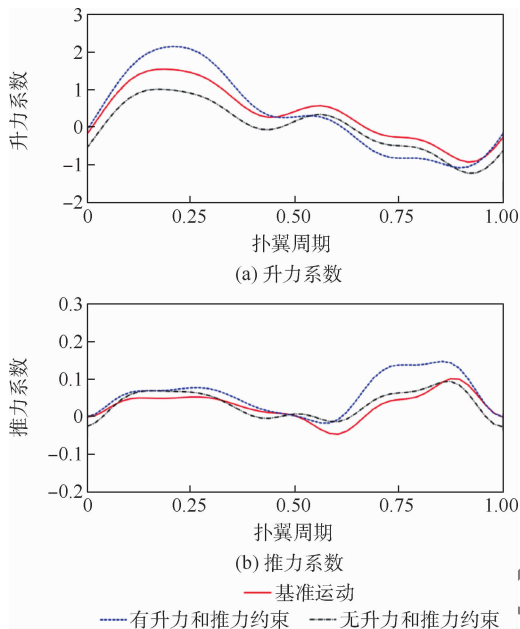


图 9 三种运动参数下扑翼运动的升力系数和推力系数
Fig. 9 Lift coefficient and thrust coefficient under flapping motion with three sets of parameters

状态,而推进效率有了较大的提升,该状态的平均升力系数几乎为零,尽量多地将克服气动力做功转化为推力,使得推进效率比基准运动提高了约 50%。扑动角度 $\gamma(t)$ 是正弦运动,前半周期为下扑阶段,后半周期为上扑阶段。从图 9 可以看出,有升力和推力约束的升力系数变化与基准运动参数状态接近,主要区别在于上扑时产生了更大的推力,这是通过改变 γ_m 和 θ_0 调节上扑时的速度和有效迎角实现的。有升力和推力约束与无升力和推力约束相比,升力差异主要发生在下扑阶段,推力差异主要发生在上扑阶段,升力的峰值均出现在下扑阶段。如果下扑时有效迎角减小则会导致升力减小,无升力和推力约束的 $\theta_0 = -0.17^\circ$ 最小,使下扑的绝对迎角更大,气动力在推力方向投影比例更大,能够增加推进效率。受运动模型的限制,无升力和推力约束增加了下扑时的绝对迎角,相应的,上扑时的绝对迎角就会减小,导致上扑时推进效率减小。但升力峰值位于下扑阶段,在该阶段选择增加推进效率更合适。图 10 显示了单个周期内扑翼运动在优化前后气动力分布的变化。

比较图 9 的力峰值位置和图 10 的气动力分布可以看出,有升力和推力约束的优化运动在下扑阶段,外翼段前缘出现气动力增强的现象,这能够解释为什么下扑时出现明显大于基准运动的升力峰值;在上扑阶段,在外翼段前缘出现气动力增强的现象,气动力方向朝翼面下方,而绝对迎角为正,对应上扑时明显大于基准运动的推力峰值,然

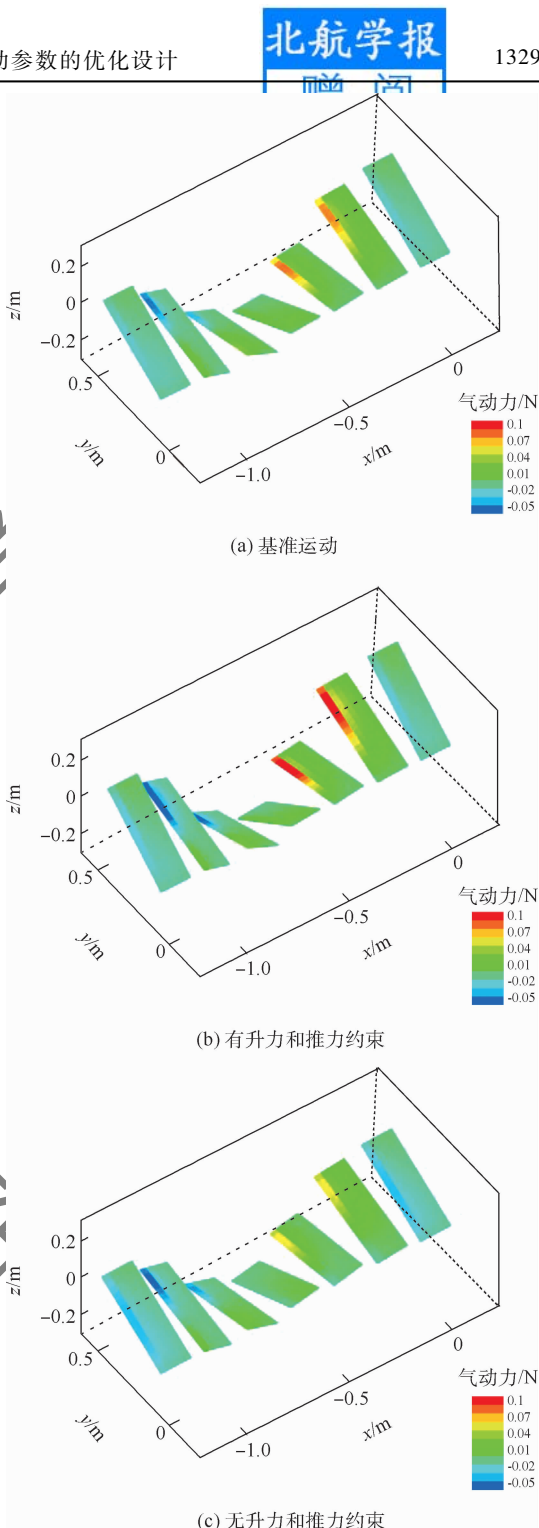


图 10 单个周期的气动力分布

Fig. 10 Time slice of aerodynamic distribution in a period

而对升力产生负效应,使得上扑时升力比基准运动小。无升力和推力约束的优化运动在下扑时外翼段前缘的气动力比基准运动小,使得升力减小,但其俯仰角低于基准运动,俯仰角带来的推力增益更大,使得推力比基准运动大;在上扑时外翼段前缘的负方向气动力比基准运动大,导致升力减小,上扑时绝对俯仰角在正负间切换,对推力造成不一致的影响。

对于每个优化问题计算了 20 个 DIRECT 迭

代周期,每个周期的最大推进效率如图 11 所示。无升力和推力约束优化共计算了 331 组不同的运动参数,有升力和推力约束优化共计算了 591 组不同的运动参数,可见 DIRECT 优化能够快速收敛到最优解。

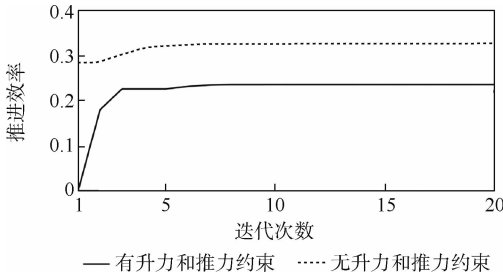


图 11 最大推进效率随优化迭代次数的变化
Fig. 11 Maximum propulsion efficiency vs. number of optimization iteration

3 结论

本文着眼于仿生扑翼运动的多参数优化问题,用非定常涡格法对扑翼气动效应和推进效率进行快速计算,构造了适用于 DIRECT 全局优化算法的全局优化问题,根据计算结果得出以下结论:

1) 非定常涡格法作为一种低计算耗费的计算方法,用较少的计算时长就能获得收敛结果,可以用于扑翼运动优化设计中,而且在较低的流速下仍与实验数据符合良好。

2) DIRECT 全局优化算法对扑翼运动参数优化问题能够快速收敛到最优解,作为一种全局优化方法,不依赖于初始采样点的选择,在有升力和推力约束和无升力和推力约束 2 种条件下能够找到最优运动参数组合,有效提升扑翼推进效率。

3) 升力和推力约束会限制最大推进效率,对扑翼飞行器进行增升减阻设计,降低升力和推力边界,能够为扑翼运动优化带来设计优势,有利于获得更节能的扑翼运动方案。

参考文献 (References)

[1] KEENNON M, KLINGEBIEL K, WON H. Development of the nano hummingbird: A tailless flapping wing micro air vehicle: AIAA 2012-0588 [R]. Reston: AIAA, 2012.
[2] NICK T P S, TAI Y C, HO C M, et al Microbat: A palm-sized electrically powered ornithopter [C] // Proceedings of the NASA/JPL Workshop on Biomimetic Robotics, 2001: 14-17.
[3] RAZAK N A, DIMITRIADIS G. Experimental study of wings

undergoing active root flapping and pitching [J]. Journal of Fluids and Structures, 2014, 49: 687-704.
[4] SRIGRAROM S, CHAN W L. Flow field of flapping albatross-like wing and sound at low Reynolds number [J]. Journal of Unmanned System Technology, 2013, 1 (2): 69-72.
[5] STANFORD B K, BERAN P S. Analytical sensitivity analysis of an unsteady vortex-lattice method for flapping-wing optimization [J]. Journal of Aircraft, 2010, 47 (2): 647-662.
[6] GHOMMEM M, COLLIER N, NIEMI A H, et al. On the shape optimization of flapping wings and their performance analysis [J]. Aerospace Science and Technology, 2014, 32 (1): 274-292.
[7] GHOMMEM M, HAJJ M R, MOOK D T, et al. Global optimization of actively morphing flapping wings [J]. Journal of Fluids and Structures, 2012, 33: 210-228.
[8] GABLONSKY J M. Modifications of the DIRECT algorithm [D]. Raleigh: North Carolina State University, 2001.
[9] GHOMMEM M, COLLIER N, NIEMI A H, et al. Shape optimization and performance analysis of flapping wings [C] // Proceedings of the Eighth International Conference on Engineering Computational Technology, 2012.
[10] ELDRIDGE J D, JONES A R. Leading-edge vortices: Mechanics and modeling [J]. Annual Review of Fluid Mechanics, 2019, 51: 75-104.
[11] SMITH M, WILKIN P, WILLIAMS M. The advantages of an unsteady panel method in modelling the aerodynamic forces on rigid flapping wings [J]. The Journal of Experimental Biology, 1996, 199 (Pt. 5): 1073-1083.
[12] VEST M S, KATZ J. Unsteady aerodynamic model of flapping wings [J]. AIAA Journal, 1996, 34 (7): 1435-1440.
[13] PERSSON P O, WILLIS D J, PERAIRE J. Numerical simulation of flapping wings using a panel method and a high-order Navier-Stokes solver [J]. International Journal for Numerical Methods in Engineering, 2012, 89 (10): 1296-1316.
[14] ROCCIA B A, PREDIKMAN S, MASSA J C, et al. Modified unsteady vortex-lattice method to study flapping wings in hover flight [J]. AIAA Journal, 2013, 51 (11): 2628-2642.
[15] KATZ J, PLOTKIN A. Low-speed aerodynamics [M]. Cambridge: Cambridge University Press, 2001.
[16] HEATHCOTE S, WANG Z, GURSUL I. Effect of spanwise flexibility on flapping wing propulsion [J]. Journal of Fluids and Structures, 2008, 24 (2): 183-199.
[17] AONO H, CHIMAKURTHI S, CESNIK C, et al. Computational modeling of spanwise flexibility effects on flapping wing aerodynamics: AIAA 2009-1270 [R]. Reston: AIAA, 2009.
[18] WOLF T, KONRATH R. Avian wing geometry and kinematics of a free-flying barn owl in flapping flight [J]. Experiments in Fluids, 2015, 56 (2): 1-18.
[19] LANG X Y, SONG B F, YANG W Q, et al. Aerodynamic performance of owl-like airfoil undergoing bio-inspired flapping kinematics [J]. Chinese Journal of Aeronautics, 2021, 34 (5): 239-252.

Optimal design of motion parameters of flapping wing

WU Yue, XIE Changchuan*, YANG Chao, An Chao

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: The flight of an ornithopter depends on the motions of the flapping wing. The optimal aerodynamic characteristics of a specific flapping wing will be obtained when using an optimized motion strategy. Furthermore, it provides a design basis for the transmission mechanism of a flapping wing aircraft. However, there is currently a lack of effective method for motion optimization in design stage to determine a set of optimal motion parameters for a given wing. In this paper, the unsteady vortex lattice method (UVLM) is applied to calculate the aerodynamic effect caused by the flapping motion. To verify accuracy of the aerodynamic calculation method, the result is correctly compared with existing experiment data. Then based on the DIRECT (divide rectangle) global optimization algorithm, the flapping kinematics parameters are iteratively optimized to maximize the propulsion efficiency. The results show that the optimization method can effectively solve the optimal parameters of flapping kinematics parameters and improve specific aerodynamic performance. The average thrust calculated by the optimization algorithm in this paper has a 104% numerical improvement compared to that of the baseline motion. Besides, it indicates that reducing the lift and thrust constraints are beneficial to the optimization to achieve a higher propulsion efficiency in the design process. The maximum propulsion efficiency without aerodynamic constraints in this paper is improved by 46.8% compared to that of the baseline motion.

Keywords: unsteady aerodynamic force; vortex lattice method; global optimization; flapping wing; aircraft design

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

DOI: 10.13700/j.bh.1001-5965.2021.0224

空间光学遥感器高精度 6-DOF 并联平台设计与实现

梁凤超*, 谭爽, 黄刚, 范建凯, 林喆, 康晓军

(北京空间机电研究所, 北京 100094)

摘 要: 高精度六自由度并联平台可精密调整次镜位姿, 实现空间光学遥感器地面光学装调及光学像差在轨主动校正。为解决研制高精度并联平台的多指标多约束结构优化设计及高分辨力驱动支链设计 2 个难点, 建立了六自由度并联机构逆解数学模型及 ADAMS 参数化模型, 确定了结构优化目标函数, 结合支链长度、铰链转角等约束进行了结构优化设计, 得到并联机构结构参数及驱动支链分辨力需达到 60 nm 的需求。基于此需求, 设计了基于“无刷直流电机+滚珠丝杠+光栅尺”的驱动支链, 采用 PI 控制律实现高精度消静差闭环伺服控制, 使驱动支链分辨力达 50 nm。对并联平台精度进行光学测试, 结果表明, 平台承载平移分辨力为 0.2 μm , 转角分辨力为 1", 满足指标要求。该平台已成功应用于空间相机的地面光学装调及像差主动校正实验, 为在轨应用奠定了理论与实践基础。

关键词: 六自由度; 并联平台; 次镜; 空间光学遥感器; 结构优化设计

中图分类号: TP242

文献标志码: A

文章编号: 1001-5965(2022)07-1332-11

随着空间光学遥感器分辨力需求的不断提高, 超长焦距、超大口径空间光学遥感器成为当前对地观测领域的重要研究方向之一。根据瑞利判据, 在轨道高度和观测谱段一定的情况下, 口径越大, 衍射极限分辨力越高。为实现 500 km 轨道亚米级地面分辨率, 空间相机口径要求达到 3 m, 而静止轨道实现米级乃至亚米级光学观测, 空间相机口径需达 8 m 以上。但受镜面加工能力和运载火箭整流罩尺寸的限制, 大口径、高分辨力空间望远镜的主镜通常采用分块可展开设计。由于主镜展开机构的精度有限, 主镜的各分块镜将存在多自由度的位姿误差, 必须由分块镜后连接的多自由度调整机构实施在轨校正, 以满足空间望远镜高分辨力成像的要求, 如詹姆斯·韦伯空间望远镜采用了六自由度并联机构调整分块镜位

姿^[1-2]。同时, 光学系统焦距增加、主次镜间距增大, 相机入轨后因力热环境变化导致主次镜间距、次镜位姿发生改变, 同样需要高精度位姿调整平台在轨调整主次镜间距和次镜位姿^[3]。

现有的六自由度位姿调整方案主要包括串联机构及并联机构方案。并联机构是在精度、刚性及操作大负载能力方面具有良好性能的闭链机械^[4-5], 适用于高精度、大载荷的场合。与传统串联机构^[6-7]相比, Stewart 构型六自由度并联机构具有精度高、刚度高、稳定性高、误差小、摩擦小及动态性好^[8-11]等优点, 六自由度并联机构是解决空间相机主镜分块镜及次镜六自由度位姿精密调整的理想机构, 詹姆斯·韦伯空间望远镜的主镜分块镜及次镜均使用了 Stewart 六自由度并联机构, 2009 年欧洲航天局研制发射的赫歇尔 (Her-

收稿日期: 2021-05-06; 录用日期: 2022-06-13; 网络出版时间: 2021-07-19 09:27

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210716.1312.002.html

基金项目: 装备发展部中俄联合研制项目 (03CN19018/2021GWKJ-5Y-LZYD)

* 通信作者。E-mail: fc.liang@qq.com

引用格式: 梁凤超, 谭爽, 黄刚, 等. 空间光学遥感器高精度 6-DOF 并联平台设计与实现[J]. 北京航空航天大学学报, 2022, 48(7): 1332-1342. LIANG F C, TAN S, HUANG G, et al. Design and implementation of a high precision 6-DOF parallel platform for a space optical remote sensor[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1332-1342 (in Chinese).

schel) 太空望远镜^[12]和 2012 年意大利投入使用的巡天地基望远镜(VST)中次镜调整均使用了 Stewart 六自由度并联机构^[13]。可见,高精度六自由度精密位姿调整并联平台是实现空间光学传感器在轨主动光学、自适应光学的关键执行机构之一。

本文解决了某空间相机高分辨力 Stewart 平台的结构优化设计及高分辨力驱动支链设计 2 个关键难题。成功研制了平移分辨力 0.2 μm、转角分辨力 1"的 Stewart 次镜平台,最终实现了空间相机地面光学装调及像差主动校正实验;给出了 Stewart 平台设计的运动学分析、数学建模、参数化建模、结构优化设计、高分辨力驱动支链设计、平台分辨力设计、带载能力设计、控制策略及平台分辨力光学测量等关键设计过程,对高分辨力并联平台的研制具有较高的参考价值。

1 Stewart 次镜平台需求

1.1 在轨像差校正原理

采用分块式拼接主镜的高分辨率大口径空间光学传感器需通过六自由度位姿调整机构调整主镜分块镜、次镜位姿,以在轨校正像差,实现高质量成像。其像差在轨主动校正系统如图 1 所示,包括分块式主镜、分块镜六自由度位姿调整机构及控制器、次镜、次镜六自由度调整机构及控制器、波前探测装置(波前传感器、波前探测、像差解算系统)及中控系统等。像差主动校正过程为:波前探测装置探测出波前误差,经解算得到分块镜的调整位姿;中控系统向分块镜六自由度机构控制器发送位姿调整指令,完成对分块镜阵列位姿调整,使得拼接后的分块主镜恰如整块反射镜那样工作;次镜六自由度机构控制器执行中控

系统指令,将次镜调整到最佳位姿,消除像差,提高成像质量。

1.2 次镜平台主要指标

某空间相机主反射镜口径大于 3 m,为整体镜面设计,与图 1 相比,不需要进行主镜分块拼接,也就不需要主镜分块镜六自由度位姿调整机构。该相机次镜口径大于 0.5 m,重约 25 kg,拟通过精密调整次镜位姿实现地面状态光学装调及模拟在轨像差主动校正,因此亟需研制一台主要指标如表 1 所示的高精度次镜平台。其中分辨力指标与德国 PI 公司同尺寸量级的 H850 平台(平移 X、Y 轴 1 μm,Z 轴 0.5 μm,转角 1")及中国科学院长春光学精密机械与物理研究所于阳等^[14]研制的次镜平台(平移 0.5 μm,转角约 1")相比,转角分辨力指标相当,平移分辨力指标要求较高,研制难度增大。

表 1 六自由度次镜平台主要指标

Table 1 Main indexes of 6-DOF secondary mirror platform

指标	数值
X, Y, Z 轴平移调整分辨力/μm	0.2
X, Y, Z 轴运动范围/mm	±2
绕 X, Y, Z 轴转角分辨力/(")	1
绕 X, Y, Z 轴转动范围/(°)	±1
承载能力/kg	25
外形包装/(mm × mm)	Φ548 × 280

2 运动学分析

并联机构的运动学分析是进行动力学分析、控制策略研究、结构优化设计的基础,因此建立准确的运动学模型对并联机构的研究至关重要。Stewart 并联机构结构如图 2 所示^[15-17],由上下 2 个

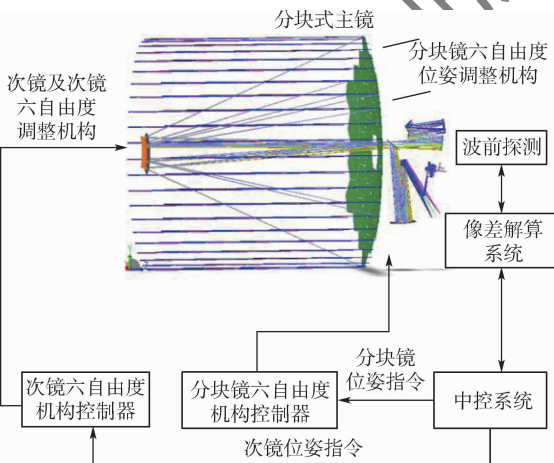


图 1 空间光学传感器在轨像差主动校正原理

Fig. 1 Active correction principle of on orbit aberrations for space optical remote sensors

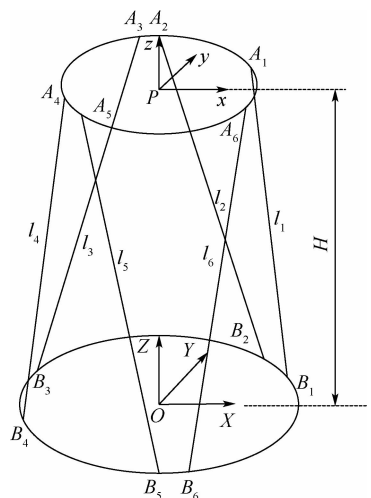


图 2 Stewart 并联机构坐标系

Fig. 2 Coordinate systems of Stewart parallel mechanism

平台和6个并联的、可独立自由伸缩的驱动支链组成,驱动支链和平台之间通过虎克铰 $A_1、A_2、A_3、A_4、A_5、A_6$ 和虎克铰 $B_1、B_2、B_3、B_4、B_5、B_6$ 连接。下平台作为静平台,以驱动支链的位移作为输入变量,则可以控制动平台(上平台)的空间位移和姿态。实际机构中,通过改变6根可以伸缩的驱动支链长度来实现并联机构动平台的运动,即沿 $X、Y、Z$ 轴平移和绕动坐标系 $X、Y、Z$ 轴转动的欧拉角 $U、V、W$ 。

为了定量表示动平台各处的坐标,在静平台铰点所在圆的圆心处建立静坐标系 $O-XYZ$,坐标原点为 O ;在动平台铰点所在圆心处建立动坐标系 $P-xyz$,坐标原点为 P ,各坐标系的方向如图2

所示。并联机构动平台的位姿用广义坐标向量 $q = [X, Y, Z, U, V, W]^T$ 表示, $[X, Y, Z]^T$ 表示动坐标系坐标原点 P 在静坐标系中的坐标向量, $[U, V, W]^T$ 表示动坐标系在静坐标系中的姿态角。这6个参数就确定了动平台的空间位姿。将欧拉定理中提到的刚体绕某一轴的有限转动分解为依一定顺序绕连体坐标轴的3次有限转动,则每次转过的角度可定义为确定刚体转动前后相对位置的3个广义坐标。按照不同的旋转顺序,可以有24种旋转方式,本文选取绕动坐标系 $Z \rightarrow Y \rightarrow X$ 坐标轴的顺序旋转,这与以相反顺序绕静坐标系固定轴旋转3次结果相同^[18]。经过3次旋转,可以得到式(1)所示的最终旋转变换矩阵:

$${}^o_p R = R(z, W) \cdot R(y, V) \cdot R(x, U) = \begin{bmatrix} \cos W \cos U & \cos W \sin V \sin U - \sin W \cos U & \cos W \sin V \cos U + \sin W \sin U \\ \sin W \cos V & \sin W \sin V \sin U + \cos W \cos U & \sin W \sin V \cos U - \cos W \sin U \\ -\sin V & \cos V \sin U & \cos V \cos U \end{bmatrix} \quad (1)$$

动平台姿态广义坐标系 $q = [X, Y, Z, U, V, W]^T$ 时,各驱动支链的长度为

$$l_i = |L_i| = \sqrt{L_i \cdot L_i} \quad (2)$$

式中: L_i 为图1所示驱动支链矢量 $\vec{A_i B_i}$; l_i 为驱动支链长度, $i = 1, 2, \dots, 6$ 。

至此,Stewart 并联机构的逆解运动学方程就建立起来了。

3 结构优化设计

六自由度并联机构结构参数的选取直接影响到机构运动性能。并联机构的优化设计是寻找一组最优的参数组合使并联机构具备较好的性能。

3.1 参数化建模

ADAMS 软件是集建模、求解、可视化技术于一体的虚拟样机软件,其提供了参数化建模功能,即将所建立模型的特征值都用 ADAMS 中的设计参数表示,这样模型就会随着设计参数的修改而自动改变。特别是在 ADAMS 中作参数分析时,对设计参数在设置区间内平均化,模型自动更新,ADAMS 自动进行系列仿真,能够观察不同参数值下样机性能。

如图2所示,Stewart 并联机构简化为静平台、动平台、6个驱动支链、6个上虎克铰链和6个下虎克铰链。六自由度并联机构的结构参数有 r (动平台铰点圆半径)、 R (静平台铰点圆半径)、 α (动平台铰点短边圆心角一半)、 β (静平台铰点短

边圆心角一半)和 H (并联机构中位时 P 与 O 的距离),确定上述5个参数,就可以完全描述 Stewart 六自由度并联机构的基本结构。

动平台铰点在动坐标系 $P-xyz$ 中的位置关系如图3所示, $\triangle A_1 A_3 A_5$ 、 $\triangle A_2 A_4 A_6$ 都为正三角形, $A_1、A_6$ 、 $A_2、A_3$ 、 $A_3、A_4$ 分别关于 x 轴对称。 $\alpha = \angle A_1 P x = \angle A_6 P x$, α 为可调参数, α 与 r 决定了动平台各铰点坐标。动平台各铰点在动坐标系的坐标如下:

$$\begin{aligned} A_1 & (r \cos \alpha, r \sin \alpha) \\ A_2 & (r \cos(2\pi/3 - \alpha), r \sin(2\pi/3 - \alpha)) \\ A_3 & (r \cos(2\pi/3 + \alpha), r \sin(2\pi/3 + \alpha)) \\ A_4 & (r \cos(4\pi/3 - \alpha), r \sin(4\pi/3 - \alpha)) \\ A_5 & (r \cos(4\pi/3 + \alpha), r \sin(4\pi/3 + \alpha)) \\ A_6 & (r \cos(-\alpha), r \sin(-\alpha)) \end{aligned}$$

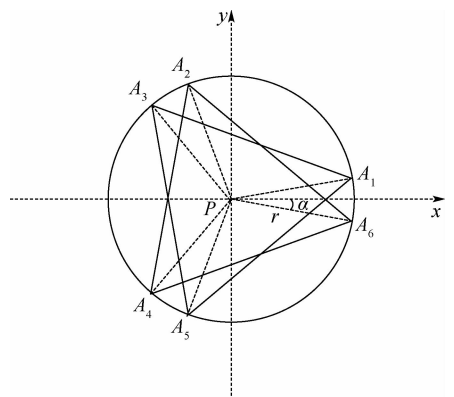


图3 动平台铰点所在圆平面

Fig. 3 Circular plane of hinge point for moving platform

同理,静平台铰点在静坐标系 $O-XYZ$ 中的位置关系如图 4 所示, $\triangle B_1B_3B_5$ 、 $\triangle B_2B_4B_6$ 都为正

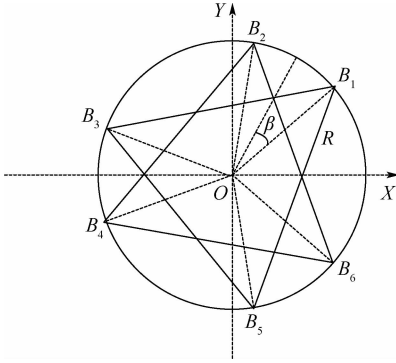


图 4 静平台铰链铰点所在圆平面

Fig. 4 Circular plane of hinge points for static platform

三角形, $B_1, B_6, B_2, B_5, B_3, B_4$ 分别关于 X 轴对称。
 $\beta = 0.5 \angle B_1O B_2$, β 为可调整参数, β 与 R 决定了静平台各铰点坐标。静平台各铰点在静坐标系的坐标如下:

$$B_1(R \cos(\pi/3 - \beta), R \sin(\pi/3 - \beta))$$

$$B_2(R \cos(\pi/3 + \beta), R \sin(\pi/3 + \beta))$$

$$B_3(R \cos(\pi - \beta), R \sin(\pi - \beta))$$

$$B_4(R \cos(\pi + \beta), R \sin(\pi + \beta))$$

$$B_5(R \cos(5\pi/3 - \beta), R \sin(5\pi/3 - \beta))$$

$$B_6(R \cos(-\pi/3 + \beta), R \sin(-\pi/3 + \beta))$$

在 ADAMS 参数化时, r, α, R, β 和 H 分别参数化为 $r_{top}, \alpha_{top}, R_{bottom}, \beta_{bottom}$ 和 H 。参数化建模参数表及模型如图 5 和图 6 所示。

Loc X	Loc Y	Loc Z
A1	(rtop * COS(Pi / 3 - topalpha))	(rtop * SIN(Pi / 3 - topalpha)) (H)
A2	(rtop * COS(Pi / 3 + topalpha))	(rtop * SIN(Pi / 3 + topalpha)) (H)
A3	(rtop * COS(Pi - topalpha))	(rtop * SIN(Pi - topalpha)) (H)
A4	(rtop * COS(Pi + topalpha))	(rtop * SIN(Pi + topalpha)) (H)
A5	(rtop * COS(Pi * 5 / 3 - topalpha))	(rtop * SIN(Pi * 5 / 3 - topalpha)) (H)
A6	(rtop * COS(topalpha - Pi / 3))	(rtop * SIN(topalpha - Pi / 3)) (H)
Bbottom1	(Rbottom * COS(bottombeta))	(Rbottom * SIN(bottombeta)) 0.0
Bbottom2	(Rbottom * COS(2 * Pi / 3 - bottombeta))	(Rbottom * SIN(2 * Pi / 3 - bottombeta)) 0.0
Bbottom3	(Rbottom * COS(2 * Pi / 3 + bottombeta))	(Rbottom * SIN(2 * Pi / 3 + bottombeta)) 0.0
Bbottom4	(Rbottom * COS(4 * Pi / 3 - bottombeta))	(Rbottom * SIN(4 * Pi / 3 - bottombeta)) 0.0
Bbottom5	(Rbottom * COS(4 * Pi / 3 + bottombeta))	(Rbottom * SIN(4 * Pi / 3 + bottombeta)) 0.0
Bbottom6	(Rbottom * COS(-bottombeta))	(Rbottom * SIN(-bottombeta)) 0.0
CO1	(k_topb * Bbottom1.loc_x + k_topt * A1.loc_x)	(k_topb * Bbottom1.loc_y + k_topt * A1.loc_y) (k_topb * Bbottom1.loc_z + k_topt * A1.loc_z)
CO2	(k_topb * Bbottom2.loc_x + k_topt * A2.loc_x)	(k_topb * Bbottom2.loc_y + k_topt * A2.loc_y) (k_topb * Bbottom2.loc_z + k_topt * A2.loc_z)
CO3	(k_topb * Bbottom3.loc_x + k_topt * A3.loc_x)	(k_topb * Bbottom3.loc_y + k_topt * A3.loc_y) (k_topb * Bbottom3.loc_z + k_topt * A3.loc_z)
CO4	(k_topb * Bbottom4.loc_x + k_topt * A4.loc_x)	(k_topb * Bbottom4.loc_y + k_topt * A4.loc_y) (k_topb * Bbottom4.loc_z + k_topt * A4.loc_z)
CO5	(k_topb * Bbottom5.loc_x + k_topt * A5.loc_x)	(k_topb * Bbottom5.loc_y + k_topt * A5.loc_y) (k_topb * Bbottom5.loc_z + k_topt * A5.loc_z)
CO6	(k_topb * Bbottom6.loc_x + k_topt * A6.loc_x)	(k_topb * Bbottom6.loc_y + k_topt * A6.loc_y) (k_topb * Bbottom6.loc_z + k_topt * A6.loc_z)
D1	(k_bottomh * Rbottom1.loc_x + k_bottomt * A1.loc_x)	(k_bottomh * Rbottom1.loc_y + k_bottomt * A1.loc_y) (k_bottomh * Rbottom1.loc_z + k_bottomt * A1.loc_z)

图 5 参数化建模参数

Fig. 5 Parametric modeling parameters

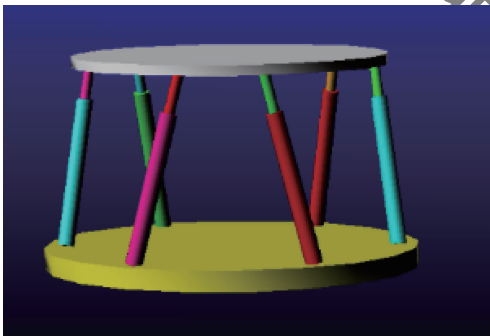


图 6 并联机构参数化模型

Fig. 6 Parametric model of parallel mechanism

3.2 目标函数与约束方程建立

在六自由度并联机构的设计中,驱动支链伸缩量(即行程)的变化是最直观的。在给定同样的运动指标时,驱动支链行程越短,驱动速度越小,能量损耗越小;在动平台位姿改变相同的情况下,驱动支链行程越小,运动时间越短,机构刚度

越好;而当驱动支链行程越较大,机构产生奇异位形的几率越大。因此,取运动中 6 个驱动支链伸缩量绝对值的最大值最小为最优目标函数,即

$$f(\mathbf{x}) = \min(\max(\max |l_i(\mathbf{x}) - l_i(0)|)) \quad (3)$$

式中: \mathbf{x} 为自变量,由设计变量 r, α, R, β, H 组成, $\mathbf{x} = [r \ \alpha \ R \ \beta \ H]$; $l_i(0)$ 为并联平台在零位时第 i 个驱动支链的长度; $l_i(\mathbf{x})$ 为动平台在复合运动中第 i 个驱动支链的长度,则第 i 个驱动支链增量 $DL_i = l_i(\mathbf{x}) - l_i(0)$ 。

确定目标函数后,就需要确定函数的约束条件。

1) 动静平台及驱动支链半径约束。对于一般的并联机构,通常静平台要比动平台具有更大的参数值。因此,动静平台的约束方程为 $r - R \leq 0$;动静平台分布角 α, β 的最小值与驱动支链半径参数有关,如果参数不当,会造成驱动支链之间

的机械干涉,也需要建立约束,考虑 20% 的裕度,约束方程为

$$\begin{cases} r - R < 0 \\ 1.2 \arcsin\left(\frac{r_{\text{strut_top}}}{r}\right) - \alpha \leq 0 \\ 1.2 \arcsin\left(\frac{r_{\text{strut_bottom}}}{R}\right) - \beta \leq 0 \end{cases} \quad (4)$$

式中: $r_{\text{strut_top}}$ 为驱动支链动杆半径; $r_{\text{strut_bottom}}$ 为驱动支链静杆半径。

2) 驱动支链长度约束。这是设置在技术要求的工作空间范围内,运行驱动支链的最大行程。次镜平台工作空间为平移 $\pm 2 \text{ mm}$, 转角 $\pm 1^\circ$, 相对较小,因此驱动支链行程小,且考虑到机构在复合运动中的空间限制,根据单驱动支链的机构初步设计结果,设定驱动支链中位长度不小于 272.3 mm , 其行程范围为 $260 \sim 290 \text{ mm}$ 。因此,驱动支链长度 l_i 的约束方程为

$$\begin{cases} 260 - \min(l_i) \leq 0 \\ \max(l_i) - 290 \leq 0 \\ l_0 \geq 272.3 \end{cases} \quad (5)$$

3) 上下铰链转角范围约束。由于实际的虎克铰都会有旋转角度范围的限制,本文设计的虎克铰旋转角度大于 3° , 预留部分裕量,设置约束角度为 5° , 上下虎克铰转角范围、约束方程为

$$\begin{cases} \max((\gamma_u)_i) - 5 \leq 0 \\ \max((\gamma_d)_i) - 5 \leq 0 \end{cases} \quad (6)$$

式中: γ_u 、 γ_d 分别为上、下虎克铰转角。

综上所述,在该六自由度并联机构结构优化设计问题中,找一组结构参数 $\mathbf{x} = [r \ \alpha \ R \ \beta \ H]$, 满足:

$$\begin{cases} \min f(\mathbf{x}) = \max(\max |l_i(\mathbf{x}) - l_i(0)|) \\ \text{s. t.} \begin{cases} 227 \leq r \leq 244 \\ 227 \leq R \leq 249 \\ 195 \leq H \\ \frac{r_{\text{strut_top}}}{r} < \sin \alpha < \frac{\pi}{6} \\ \frac{r_{\text{strut_bottom}}}{R} < \sin \beta < \frac{\pi}{6} \\ l_0 \geq 272.3 \\ r - R \leq 0 \\ l_{\min} - \min(l_i) \leq 0 \\ \max(l_i) - l_{\max} \leq 0 \\ \max((\gamma_d)_i) - 5 \leq 0 \\ \max((\gamma_u)_i) - 5 \leq 0 \end{cases} \end{cases} \quad (7)$$

式中: l_0 为驱动支链的中位长度,表达式为

$$l_0 = \sqrt{H^2 + \left[\frac{(A_{ix} - B_{ix})^2}{A_{ix}^2} + \frac{(A_{iy} - B_{iy})^2}{A_{iy}^2} \right]} \quad (8)$$

式中: $i = 1, 2, \dots, 6$; A_{ix} 、 B_{ix} 分别为铰点 A_i 、 B_i 的 x 坐标; A_{iy} 、 A_{iy} 分别为铰点 A_i 、 B_i 的 y 坐标。

依据式(7)在 ADAMS 中建立目标函数与约束方程,就可以进行结构优化了。

3.3 优化设计流程及结果

六自由度并联机构结构优化流程如图 7 所示。设计研究是研究六自由度并联机构的 5 个设计参数对机构性能的影响,研究设计参数的变化对并联机构的影响非常重要;试验设计是研究多个设计变量产生的变化,且将多个设计变量的取值组成组,研究在设计变量取不同可能组合时目标函数的取值情况,而试验设计是 ADAMS 中设计研究仅为单个设计变量产生变化;优化设计是找一组结构参数 \mathbf{x} , 满足约束条件式(7)。经过多次优化,得到满足约束条件的最终优化结果,如表 2 所示。

建立表 2 所示的 Stewart 并联机构模型,进行运动学逆解及正解分析,给出满足平台分辨力指标所需的驱动支链分辨力,再进行动力学分析,给出满足平台带载指标所需的驱动支链驱动力。根

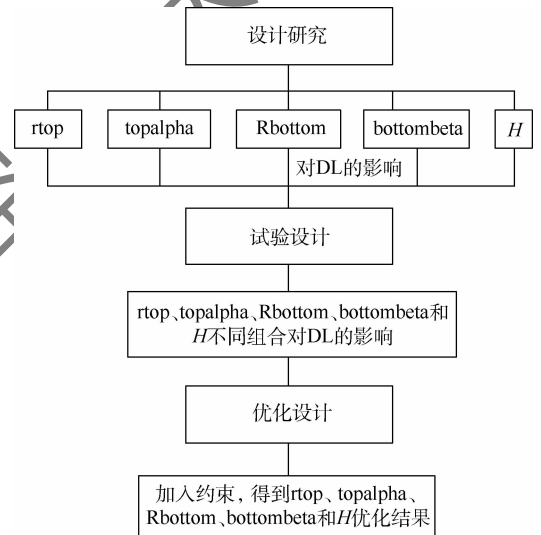


图 7 并联机构结构优化流程

Fig. 7 Structure optimization flow chart of parallel mechanism

表 2 优化后的六自由度并联机构结构参数
Table 2 Optimized structural parameters of 6-DOF parallel mechanism

结构参数	优化结果
r/mm	239.51
R/mm	249
H/mm	195.03
α/rad	0.115
β/rad	0.134

据上述分析结果,设计符合要求的高分辨力、具有一定力矩裕度的驱动支链,确保平台的带载能力及高运动分辨力。

4 Stewart 次镜平台设计

4.1 驱动支链分辨力需求

Stewart 并联机构的分辨力主要取决于于各驱动支链的分辨力。实际工程实现上,各驱动支链位移分辨力是有限的,这会造成特定结构参数平台的分辨力受限于驱动支链的分辨力。因此,在工程应用的设计时,还需要根据平台位移分辨力的要求,进行驱动支链位移分辨力的逆解运算,得到所需的驱动支链分辨力;依据工程实际,判定该分辨力的驱动支链是否可以工程实现,如果可以工程实现,可再依据驱动支链分辨力进行正解运算,计算平台的分辨力。

根据对平台平移分辨力 $0.2 \mu\text{m}$ 、旋转分辨力 $1''$ 的指标要求,进行运动学逆解运算,得到所需驱动支链的最高分辨力为 $0.06 \mu\text{m}$ 。这就需要 6 个驱动支链均具有至少 60 nm 的运动分辨力。

4.2 驱动支链力矩需求

六自由度并联机构质量为 26.74 kg ,其中动平台质量为 3.95 kg ,要求带载能力为 25 kg ,则次镜平台实际总负载为 28.95 kg 。通过 ADAMS 仿真得出 6 个驱动支链受力情况,其中最大值为 108.8 N 。因此,在地面工作状态下,驱动支链至少能提供 108.8 N 的驱动力。经计算,选择连续堵转力矩 $0.706 \text{ N} \cdot \text{m}$ 的直流伺服电机,具有 6.67 的力矩裕度,满足次镜平台带载能力及瞬时过载的需求。

4.3 高分辨力驱动支链设计

对于六自由度并联机构,如何实现高分辨力的驱动支链是其设计的重点和难点。随着滚珠丝杠机构加工装配技术的提高,直流伺服电机和滚珠丝杠的伺服驱动机构具有越来越高的刚度和精度。卢礼华等^[19]采用“直流伺服电机+滚珠丝杠”驱动机构实现了大范围的纳米定位。本文采用“无刷伺服电机+滚珠丝杠+光栅尺”的驱动机构。如图 8 所示,驱动支链由虎克铰链、丝杠导轨动力传动部分、绝对式光栅尺闭环反馈部分及锁定部分组成。由无刷直流伺服电机直驱驱动滚珠丝杠,丝杠螺母在动平台的被动约束下沿杆长方向移动。为了消除驱动支链绕杆长方向上被动转动对丝杠螺母移动距离的影响,采用测量方式,直接测量驱动支链的移动距离,反馈给控制器形

成位置闭环,驱动支链运动到指令位置,从而实现六自由度调整装置的位姿调整。图 9 为驱动支链的 50 nm 位移阶跃响应曲线。优于驱动支链至少 60 nm 运动分辨力的需求。

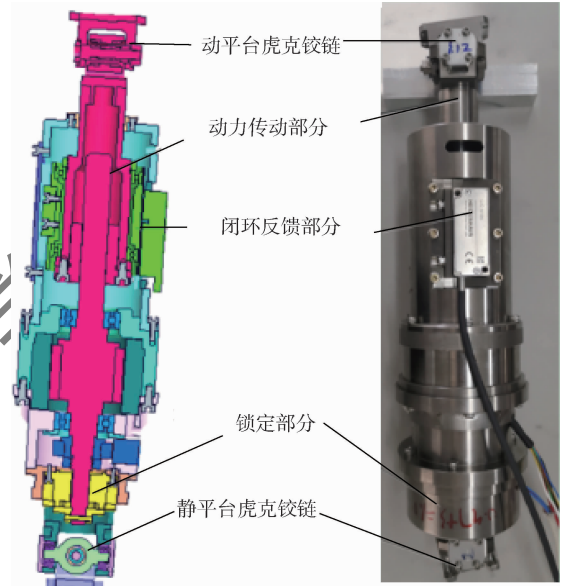


图 8 驱动支链
Fig. 8 Driving strut

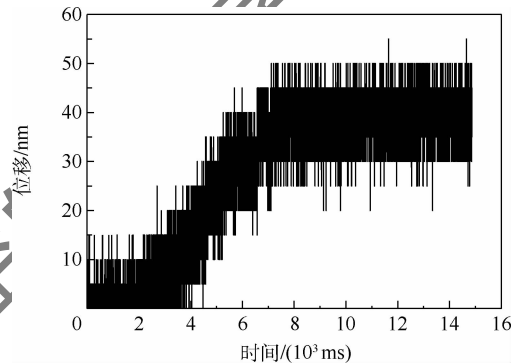


图 9 驱动支链位移分辨力

Fig. 9 Displacement resolution of drive strut

4.4 平台分辨力分析

平台分辨力分析步骤为:首先,根据各平台自由度运动分辨力指标,以单自由度最小运动增量为目标位姿,应用并联机构运动学逆解计算 6 个驱动支链的运动增量;然后,根据驱动支链的实际运动分辨力,将驱动支链运动增量进行量化,即按照驱动支链分辨力取分辨力的整数倍,这个倍数就是驱动支链位移的量化步数;最后,将量化步数乘以驱动支链分辨力,得到支链可实际运动距离,再应用并联机构运动学正解,计算出驱动支链实际位移对应的平台位姿,即可得到平台该自由度分辨力。

例如,次镜平台零位姿时,给定目标位姿 $\mathbf{q} = [0.0002 \text{ mm}, 0, 0, 0, 0, 0]^T$ 时,经并联机构运动学逆解得到 6 个驱动支链的运动增量,表示为:

-56.71, 109.86, -52.16, -52.16, 109.86, -56.71 nm,再以驱动支链实际 50 nm 分辨力量化,可得到表 3 第 1 行 X 轴平移对应的各驱动支链量化步数为 -1、2、-1、-1、2、-1。同理,给定目标位姿 $q = [0, 0.000\ 2\ \text{nm}, 0, 0, 0, 0]^T$ 时,可得到表 3 第 2 行 Y 轴平移对应各驱动支链的量化步数。以此类推,可得到表 3 所示全部 6 个自由度分辨力对应的驱动支链量化步数。

当六自由度并联机构按照表 3 所示的量化步数进行正运动学分析时,即可得到动平台 6 个自由度运动的最小增量,即平台分辨力。例如,表 3 第 1 行 X 轴平移量化步骤对应的驱动支链位移

为: -50 nm、100 nm、-50 nm、-50 nm、100 nm、-50 nm,将这 6 个位移增量作为驱动支链位移输入,运用并联机构正解即可计算出此时的平台位姿,如表 4 所示。表 4 的第 1 行表示 X 轴平移运动分辨力为 0.14 μm 。同理,可计算出平台其他 5 个自由度的分辨力。

由表 4 可知,表 2 所示的 Stewart 并联机构,如果 6 个驱动支链位移分辨力为 50 nm,则平台的分辨力为: X 轴 0.14 μm 、Y 轴 0.08 μm 、Z 轴 0.07 μm ;绕 X 轴 0.109",绕 Y 轴 0.114",绕 Z 轴 0.066"。平移分辨力优于 0.2 μm ,转角分辨力优于 1"。

表 3 平台位姿分辨力对应的驱动支链量化步数

Table 3 Driving struts quantization steps corresponding to platform pose resolution

自由度	量化					
	驱动支链 1	驱动支链 2	驱动支链 3	驱动支链 4	驱动支链 5	驱动支链 6
X 轴平移	-1	2	-1	-1	2	-1
Y 轴平移	1	0	-1	1	0	-1
Z 轴平移	1	1	1	1	1	1
绕 X 轴转动	1	1	0	0	-1	-1
绕 Y 轴转动	-1	0	1	1	0	-1
绕 Z 轴转动	1	-1	1	-1	1	-1

表 4 驱动支链量化步数对应的平台位姿分辨力

Table 4 Platform pose resolution corresponding to driving struts quantitation steps

自由度	分辨力					
	X/ μm	Y/ μm	Z/ μm	RX/(")	RY/(")	RZ/(")
X 轴平移	0.14	0	0	0	-0.007	0
Y 轴平移	0	0.08	0	0.004	0	0
Z 轴平移	0	0	0.07	0	0	0
绕 X 轴转动	0	-0.01	0	0.109	0	0
绕 Y 轴转动	-0.017 5	0	0	0	0.114	0
绕 Z 轴转动	0	0	0	0	0	0.066

由于六自由度并联机构动平台的位姿调整是通过 6 个驱动支链的长度变化来实现的,要实现高分辨力六自由度位姿并联平台,关键之一就是要有高分辨力的驱动支链。本文采用的驱动支链分辨力达 50 nm,这是平台高分辨力的有力保障之一。

4.5 机构控制系统

次镜平台控制器基于“DSP + FPGA”架构设计,DSP 负责核心控制算法的实时计算,包括运动学正解、逆解、驱动支链闭环伺服控制等,FPGA 负责光栅尺数据实时采集、力矩电机换相、PWM 波生成和与综合电子单元通信等。采用这样的设计架构可以最大限度地保证 DSP 的高效浮点运算特性和 FPGA 的并行控制能力。FPGA 内置的代码将外设映射成寄存器接口,DSP 可以通过 EMIF 接口访问这些寄存器,实现 DSP 与 FPGA 的无缝连接,并能大幅提高 DSP 访问外设的速度和

效率。

六自由度次镜平台及控制器如图 10 所示。控制器接收位姿调整命令,经并联机构运动学逆解得到各驱动支链运动位移,位移指令经驱动电

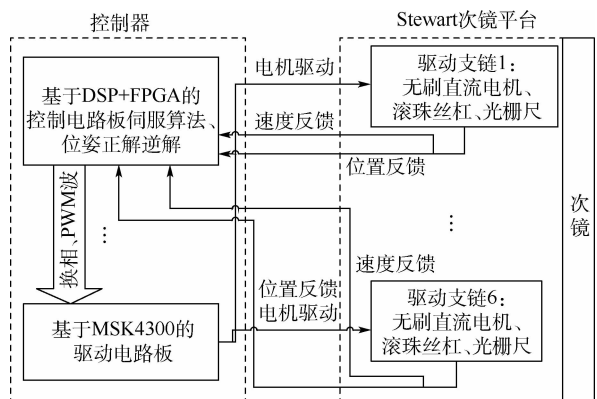


图 10 六自由度并联平台控制器原理

Fig.10 Principle of 6-DOF parallel platform controller

路功率放大后驱动 6 个驱动支链协调运动;各驱动支链的高精度光栅尺实时反馈驱动支链位置给控制电路板,控制电路板依据指令位置和当前实际位置的偏差,经简单可靠的铰点空间 PID 控制律,给出合适的控制量,继续驱动各驱动支链运动到目标位置,实现闭环控制,进而实现对次镜的六自由度位姿调整。

5 平台分辨力测量

本文研制的 Stewart 次镜平台如图 11 所示。该平台的位移及转角分辨力是实现次镜位姿精密调整的关键,这就需建立六自由度位姿测量系统,对该平台的位姿调整分辨力及调整范围进行测试。下面着重介绍平台分辨力的测试。

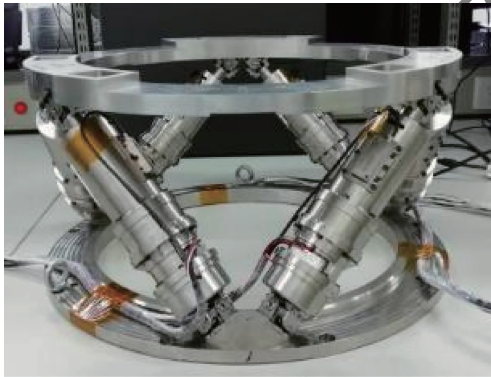


图 11 六自由度并联平台
Fig. 11 6-DOF parallel platform

5.1 分辨力测试系统

Stewart 次镜平台平移及转角分辨力测试系统如图 12 所示。系统由 RENISHAW XL-80 双频激光干涉仪、AUTOMAT 自准直仪、Stewart 次镜平

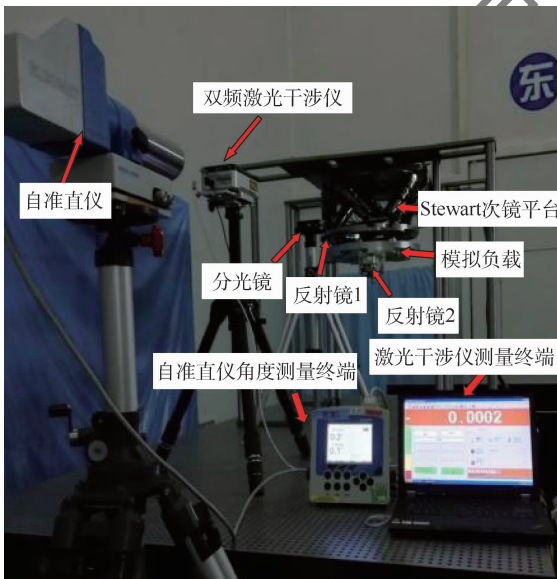


图 12 六自由度并联平台分辨力测试系统

Fig. 12 Resolution test system of 6-DOF parallel platform

台、模拟负载及并联机构控制器组成。双频激光干涉仪用于平移分辨力测试,自准直仪用于转角分辨力测试。整个测试系统搭建在恒温实验室内隔振光学平台上,可有效防止温漂引起的形变误差,隔离和衰减低频、高频振动。该测试系统平移测量分辨力达到 1 nm、平移测量行程达到 ±500 mm,角度测量分辨力达到 0.002",角度测量行程达到 ±1°,满足平移行程 ±2 mm、分辨力 0.2 μm、角度行程 ±1°、角度分辨力 1"的测量需求。

5.2 分辨力测试

平台分辨力测试采用各自由度分开测试的方法,每个自由度都是先以技术要求指标 5 倍的步距步进,停留约 20 s,使设备充分稳定下来,再步进,前进 5 步后,再后退 5 步。依此步骤逐步降低指令值,如 X 轴分辨力测试,分别 X 轴平移 1 μm、0.8 μm、0.6 μm、0.4 μm、0.2 μm、0.1 μm 的平移指令,激光干涉仪测试曲线能够很好分辨出 X 轴平移 1 μm 到 0.2 μm 的指令,继续尝试给出更小的平移指令 0.1 μm 时,测得位移曲线台阶不明显,故判定平台的 X 轴平移自由度分辨力为 0.2 μm;同理测得平台 Y、Z 轴自由度平移分辨力为 0.2 μm。X、Y、Z 轴平移分辨力测试曲线如图 13 ~ 图 15 所示。

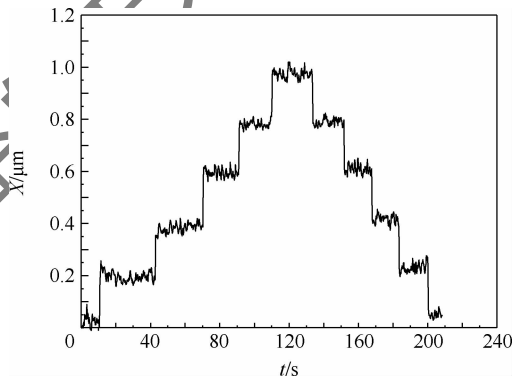


图 13 X 轴自由度平移分辨力测试曲线
Fig. 13 X-DOF translation resolution test curve

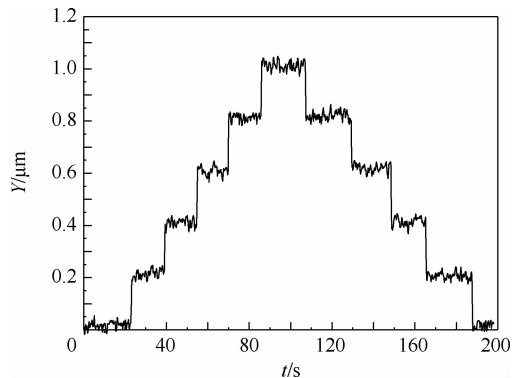


图 14 Y 轴自由度平移分辨力测试曲线
Fig. 14 Y-DOF translation resolution test curve

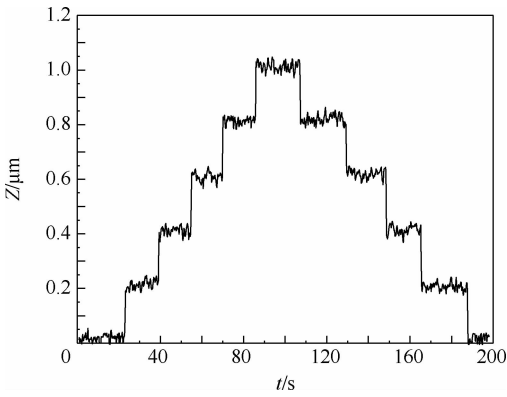


图15 Z轴自由度平移分辨力测试曲线

Fig. 15 Z-DOF translation resolution test curve

平台绕X轴旋转分辨力指标为 $1''$,依次给定 $5''$ 、 $4''$ 、 $3''$ 、 $2''$ 、 $1''$ 、 $0.5''$ 的转角步距指令,进行分辨力测试,自准直仪测试结果表明次镜平台很好地分辨出了 $5''$ 、 $4''$ 、 $3''$ 、 $2''$ 、 $1''$ 的转角指令;当指令值低于 $1''$ 时,曲线台阶不明显,台阶几乎被噪声淹没。判定平台绕X轴转角分辨力为 $1''$ 。同理测得平台绕Y、Z轴旋转的分辨力均达到 $1''$ 。测试曲线如图16~图18所示。

次镜平台的分辨力指标及测试值对比如表5所示。实验结果表明,该Stewart次镜平台的X、Y、Z轴平移分辨力为 $0.2\ \mu\text{m}$,绕X、Y、Z轴旋转分辨力为 $1''$,均达到指标要求。

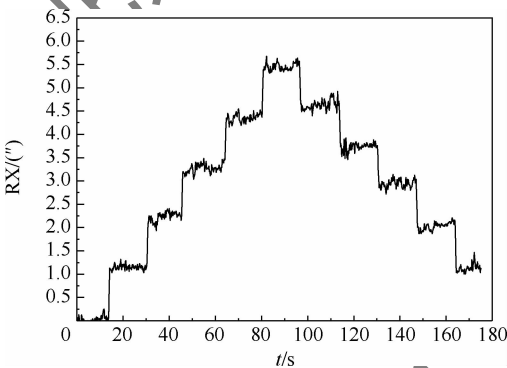


图16 RX自由度分辨力测试曲线

Fig. 16 RX-DOF resolution test curve

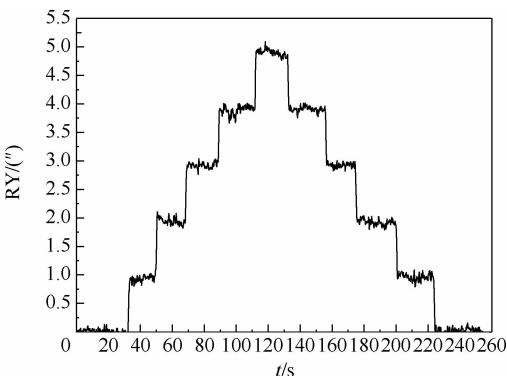


图17 RY自由度分辨力测试曲线

Fig. 17 RY-DOF resolution test curve

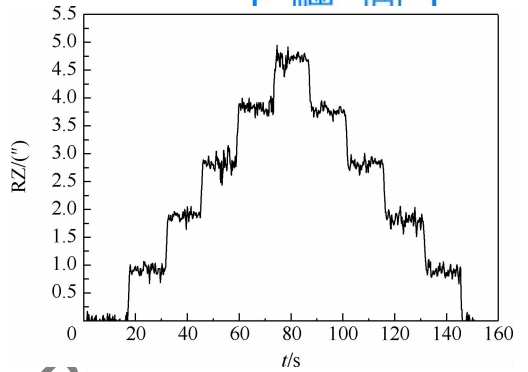


图18 RZ自由度分辨力测试曲线

Fig. 18 RZ-DOF resolution test curve

表5 平台分辨力指标与实测值对比

Table 5 Comparison between platform resolution index and measured value

自由度	指标	实测值
X轴平移/ μm	0.2	0.2
Y轴平移/ μm	0.2	0.2
Z轴平移/ μm	0.2	0.2
绕X轴旋转/ $''$	1	1
绕Y轴旋转/ $''$	1	1
绕Z轴旋转/ $''$	1	1

6 结论

研制高精度Stewart次镜平台的2个突出难点是:多参数指标多约束条件的并联机构结构优化设计和高平台位姿分辨力的工程实现。针对这2个难点,本文给出了如下设计与实现的方法:

1) 基于六自由度并联机构数学模型及参数化模型,确定了驱动支链伸长量最大值最小的目标函数,结合支链长度、铰链转角等约束完成了结构优化设计,得到了满足多指标多参数要求的Stewart次镜平台结构。

2) 针对平台分辨力对高驱动支链分辨力的需求,设计了基于“无刷伺服电机+滚珠丝杠+光栅尺”的驱动支链,以及基于“DSP+FPGA”的控制器,实现了驱动分支 $50\ \text{nm}$ 的高分辨力,确保了平台高位姿分辨力。

3) 光学测试结果表明,Stewart次镜平台带载平移分辨力 $0.2\ \mu\text{m}$,转角分辨力 $1''$,满足该空间光学遥感器地面光学装调及在轨主动光学调整任务需求。

参考文献 (References)

- [1] CLAMPIN M. The James Webb space telescope (JWST) [J]. Advances in Space Research, 2008, 41(12): 1983-1991.
- [2] WARDEN R M. Cryogenic nano-actuator for JWST [C] // Proceedings of the 38th Aerospace Mechanisms Symposium, 2006:

239-252.

- [3] 谭爽. 空间光学遥感器次镜六自由度精密控制方法研究 [D]. 北京: 中国空间技术研究院, 2015.
- TAN S. Studies on high precision control of 6-DOF second mirror of the space optical remote sensor [D]. Beijing: China Academy of Space Technology, 2015 (in Chinese).
- [4] MERLET J P. Parallel robots [M]. 2nd ed. Berlin: Springer, 2006: 66-74.
- [5] KONG X W, GOSSELIN C. Type synthesis of parallel mechanisms [M]. Berlin: Springer, 2007.
- [6] 王富国, 张景旭, 杨飞, 等. 四翼梁式次镜支撑结构的研究 [J]. 光子学报, 2009, 38 (3): 674-676.
- WANG F G, ZHANG J X, YANG F, et al. Crossed-plate type support structure of the second mirror [J]. Acta Photonica Sinica, 2009, 38 (3): 674-676 (in Chinese).
- [7] 温正方, 张景旭, 张丽敏. 五自由度次镜调整机构的研究 [J]. 工程设计学报, 2010, 17 (6): 473-478.
- WEN Z F, ZHANG J X, ZHANG L M. Research of five-degree-of-freedom adjustment mechanism for secondary mirror [J]. Journal of Engineering Design, 2010, 17 (6): 473-478 (in Chinese).
- [8] 房海蓉, 王立, 张涛强, 等. 一种含闭环支链的新型并联机构设计与分析 [J]. 北京航空航天大学学报, 2019, 45 (3): 454-463.
- FANG H R, WANG L, ZHANG T Q, et al. Design and analysis of a novel parallel mechanism with closed-loop limbs [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45 (3): 454-463 (in Chinese).
- [9] 房海蓉, 李壮壮. 新型大转角 2T2R 并联机构的设计与分析 [J]. 北京航空航天大学学报, 2020, 46 (3): 465-473.
- FANG H R, LI Z Z. Design and analysis of a new large rotation angle 2T2R parallel mechanism [J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46 (3): 465-473 (in Chinese).
- [10] 陈森, 张氢, 葛韵斐, 等. 2UPR-RRU 并联机构及其运动学分析 [J]. 北京航空航天大学学报, 2019, 45 (6): 1145-1152.
- CHEN M, ZHANG Q, GE Y F, et al. 2UPR-RRU parallel mechanism and its kinematic analysis [J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45 (6): 1145-1152 (in Chinese).
- [11] 王保兴, 孟刚, 林苗, 等. 3-PPP 型柔性并联微定位平台的设计与分析 [J]. 北京航空航天大学学报, 2020, 46 (4): 798-807.
- WANG B X, MENG G, LIN M, et al. Design and analysis of a 3-PPP compliant parallel micro-positioning stage [J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46 (4): 798-807 (in Chinese).
- [12] TOULEMONT Y, PASSVOGEL T, PILBRATT G L, et al. The 3.5-m all-SiC telescope for HERSCHEL [C] // Proc SPIE 5487, Optical, Infrared, and Millimeter Space Telescopes. Bellingham: SPIE, 2004, 5487: 1119-1128.
- [13] SCHIPANI L P, PERROTTA F, MOLFESE C, et al. The VST secondary mirror support system [C] // Proc SPIE 7018, Advanced Optical and Mechanical Technologies in Telescopes and Instrumentation. Bellingham: SPIE, 2008, 7018: 1380-1389.
- [14] 于阳, 徐振邦, 于鹏, 等. 空间光学遥感器次镜定位平台的设计与测试 [J]. 载人航天, 2016, 22 (1): 74-80.
- YU Y, XU Z B, YU P, et al. Design and test of parallel platform with precise positioning for secondary mirrors in space optical telescope [J]. Manned Spaceflight, 2016, 22 (1): 74-80 (in Chinese).
- [15] GOUGH V E, WHITEHALL S G. Universal tyre test machine [C] // Proceedings of the 9th International Technical Congress, 1962: 117.
- [16] STEWART D. A platform with six degrees of freedom [J]. Proceedings of the Institution of Mechanical Engineers, 1965, 180 (1): 371-386.
- [17] 黄真, 孔令富, 方跃法. 并联机器人机构学理论及控制 [M]. 北京: 机械工业出版社, 1997: 18-29.
- HUANG Z, KONG L F, FANG Y F. Mechanism theory and control of parallel manipulator [M]. Beijing: China Machine Press, 1997: 18-29 (in Chinese).
- [18] CRAIG J J. Introduction to robotics mechanics and control [M]. Upper Saddle River: Prentice Hall, 2005: 28-41.
- [19] 卢礼华, 梁迎春, 大刀川博之, 等. 快速无超调纳米定位 [J]. 纳米技术与精密工程, 2006, 4 (3): 212-216.
- LU L H, LIANG Y C, TACHIKAWA H, et al. Rapid nanometer positioning without overshoot [J]. Nanotechnology and Precision Engineering, 2006, 4 (3): 212-216 (in Chinese).

Design and implementation of a high precision 6-DOF parallel platform for a space optical remote sensor

LIANG Fengchao^{*}, TAN Shuang, HUANG Gang, FAN Jiankai, LIN Zhe, KANG Xiaojun

(Beijing Institute of Space Mechanics & Electricity, Beijing 100094, China)

Abstract: The high precision 6-DOF parallel platform can precisely adjust the position and pose of the secondary mirror of the space optical remote sensor, which could realize the ground optical alignment and the active correction of optical aberration on orbit. In order to solve the two difficulties, ie, multi index and multi constraint structural optimization design and high-resolution driving strut design, in the development of a high-precision parallel platform, 3 steps are taken as follows: Firstly, the inverse solution mathematical model and ADAMS parametric model of the 6-DOF parallel mechanism are established, and the structure optimization objective function is determined. The structure optimization design is carried out in combination with the constraints such as strut length and hinge angle, therefore, the structure parameters and the requirement of 60 nm driving strut resolution are obtained. Then, to get high resolution strut, the driving strut based on “brushless DC motor + ball screw + grating ruler” is designed, and PI control law is used to realize the high-precision closed-loop servo control of eliminating static error. As a result, the resolution of the driving strut reaches 50 nm. Finally, the accuracy of the parallel platform is tested, and the test results show that the resolution of the platform is 0.2 μm , and the angular resolution is 1", which meet the requirements of the index. The platform has been successfully applied to the ground optical alignment and aberration active correction experiment of the space remote sensor, which lays a solid theoretical and practical foundation for the on orbit application.

Keywords: 6-DOF; parallel platform; secondary mirror; space optical remote sensor; structure optimization design

Received: 2021-05-06; **Accepted:** 2021-06-13; **Published online:** 2021-07-19 09:27

URL: kns.cnki.net/kcms/detail/11.2625.V.20210716.1312.002.html

Foundation item: China Russia Joint Development Project of Equipment Development Department (03CN19018/2021GWKJ-5Y-LZYD)

*** Corresponding author.** E-mail: fc.liang@qq.com

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学核心期刊,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材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码,中文摘要(包括目的、方法、结果及结论),中文关键词(5~8 个),中图分类号,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词,引言,正文,参考文献。首页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投它刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经刊登,即赠送单行本。

3.4 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 北京航空航天大学学报编辑部

办公地点:北京航空航天大学办公楼 405,407,409 房间

电 话:(010)82315594,82338922,82314839,82315426

E-mail: jbuaa@buaa.edu.cn

http://bhxb.buaa.edu.cn

http://www.buaa.edu.cn

《北京航空航天大学学报》
第五届编辑委员会

主任(主编): 赵沁平

(以下按姓氏笔画为序)

副主任(副主编): 丁希仑 王少萍 孙志梅 李秋实 李焕喜 杨嘉陵
苗俊刚 相艳 徐立军 钱德沛 曹晋滨

编委: 马殿富 王琪 王聪 邓小燕 王青云 王荣明 刘宇
刘红 江洁 刘强 闫鹏 朱天乐 刘铁钢 齐铂金
陈万春 邹正平 苏东林 杨世春 沈成平 邱志平 宋知人
杨树斌 张晓林 杨晓奕 杨继萍 李惠峰 吴新开 张瑞丰
杨照华 宋凝芳 周锐 林宇震 林贵平 战强 姚仰平
胡庆雷 赵秋红 段海滨 赵巍胜 席平 郭宏 徐洁
徐世杰 郭洪波 康锐 翟锦 熊华钢

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

(原《北京航空学院学报》)

(月刊 1956年创刊)

第48卷第7期 2022年7月

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS
(JBUA)

(Monthly, Started in 1956)

Vol.48 No.7 July 2022

主管单位 中华人民共和国工业和信息化部
主办单位 北京航空航天大学
主编 赵沁平
编辑出版 《北京航空航天大学学报》编辑部
邮编 100083
地址 北京市海淀区学院路37号
印刷 北京科信印刷有限公司
发行 《北京航空航天大学学报》编辑部
发行范围 国内外发行
联系电话 (010) 82315594 82338922
82314839
电子信箱 jbuaa@buaa.edu.cn

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 Qingping
Edited and Published by Editorial Board of JBUA
Printed by Beijing Kexin Printing Co., Ltd.
Distributed by Editorial Board of JBUA
Telephone (010) 82315594 82338922
82314839
E-mail jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

中国标准连续出版物号: $\frac{\text{ISSN } 1001-5965}{\text{CN } 11-2625/V}$

国内定价: 50.00元/期

ISSN 1001-5965



9 771001 596229