

文章编号: 2095-2163(2019)05-0058-05

中图分类号: TP18

文献标志码: A

遗传算法优化支持向量机的光伏阵列故障诊断研究

郭浩然, 李泽滔

(贵州大学 电气工程学院, 贵阳 550025)

摘要: 研究了支持向量机(SVM)方法在光伏阵列故障诊断上的运用,对光伏阵列的输出特性以及故障类型进行了分析总结。支持向量机由于存在惩罚因子系数与核函数系数,在选用径向基核函数后通过遗传算法对其参数进行寻优,通过Matlab仿真实验得到数据,利用寻优后的参数建立模型训练与验证。研究结果表明:支持向量机使用通过遗传算法优化得到的参数在光伏阵列故障诊断上有较高的准确度。

关键词: 光伏阵列; 故障分类; 遗传算法; 支持向量机

Research on fault diagnosis of photovoltaic array based on Support Vector Machine optimized by Genetic Algorithm

GUO Haoran, LI Zetao

(The Electrical Engineering College, Guizhou University, Guiyang 550025, China)

[Abstract] The application of Support Vector Machine (SVM) in photovoltaic array fault diagnosis is studied. The output characteristics and fault types of photovoltaic arrays are analyzed and summarized. Because of the existence of penalty factor coefficient and kernel function coefficient, the Support Vector Machine is optimized by Genetic Algorithm after choosing the radial basis function, and the model is established by using the optimized parameters. The data are obtained through Matlab simulation experiment and the model is trained and validated. The results show that the Support Vector Machine optimized by Genetic Algorithm has high accuracy.

[Key words] photovoltaic array; fault classification; Genetic Algorithm; Support Vector Machine

0 引言

光伏发电系统中,其光伏阵列作为太阳能辐射的接收装置,在工作时对光照的吸收效果直接决定了光伏发电系统的发电效率。而大规模的光伏阵列往往是由多个光伏板单元根据供电要求串、并联在一起的。数量颇多且范围广的阵列由于会经历长时间的阳光暴晒、风吹雨淋,容易出现光伏板材料异常老化、裂块、异物及阴影遮蔽、开路和短路情况,从而影响经济效益。尤其是阴影遮蔽情况,实际应用中由于云层、灰尘、异物遮挡,原本的电源在没有足够辐照度的情况下会成为电路负载。近年来国内外的光伏装机容量增长迅速,对光伏阵列发生的故障进行预警、检测、分类就显得十分重要。目前常规的故障诊断方法有电信号检测法、红外图像检测法^[1]等。而近年来智能算法的应用日益成熟,不少学者提出利用智能算法来检测并对光伏阵列进行故障诊断。目前,常见的智能检测方法主要有决策树^[2]、神经网络^[3]、模糊控制^[4]等。

支持向量机在故障诊断领域有着广泛的应用,是一种有着较深厚理论基础的小样本学习方法,可以快速地从小样本的训练过程得到样本的预报,同时避开从归纳到演绎的过程。对样本中的非支持向量进行增加、删除对模型没有影响,因而具有较好的鲁棒性^[5]。本文利用支持向量机对光伏阵列的故障进行检测与分类,在应用中,支持向量机中参数的选取对于精确度影响很大,为选取最优参数利用遗传算法对其进行寻优后带入。本研究在仿真模型搭建后,模拟光伏阵列各种可能出现的故障,得到样本数据,从中提取输入特征向量并进行处理,处理后再利用遗传算法寻优得到的最优参数建立模型。本文对此拟展开研究论述如下。

1 支持向量机

支持向量机(SVM)是一种旨在求得结构风险最小化的算法,使得在样本以及实际数据上的分类都尽量准确。算法基本思路是将无法线性划分的低维数据映射到高维空间后,在空间寻找一个有着最

作者简介: 郭浩然(1994-),男,硕士研究生,主要研究方向:计算机控制技术;李泽滔(1960-),男,博士,教授,主要研究方向:计算机控制技术、自适应控制、鲁棒控制等。

收稿日期: 2019-06-28

哈尔滨工业大学主办 ◆ 学术研究与应用

大间隔的最优分类面,如图 1 所示。

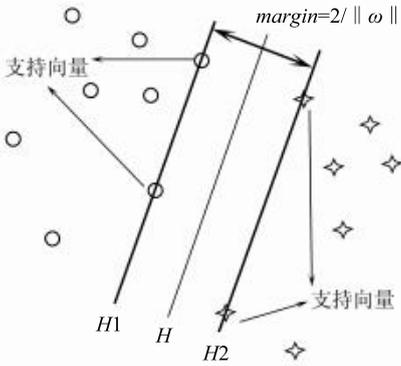


图 1 SVM 超平面

Fig. 1 SVM hyperplane

图 1 中, H 是分类面, $H1$ 与 $H2$ 为平行于 H 的面, 将 2 类样本有效分开。同时, 使得分类间隔 $margin$ 最大的就是最优分类面。设有线性可分的样本集 $(x^i, y^i), i = 1, \dots, n, x \in R^d, y \in \{1, -1\}$, 则在 d 维空间中线性判别式可表示为:

$$g(x) = \omega x + b, \quad (1)$$

其中, ω 是权值向量, b 是偏置。

此时分类面为 $\omega x + b = 0$ 。归一化后使所有样本数据都满足 $|g(x)| > 1$ 。距离分类面最近的样本 $|g(x)| = 1$, 分类间隔为 $2/\|\omega\|$ 。为使分类间隔最大, 就应满足式(2):

$$y_i(\omega x_i + b) - 1 \geq 0, (i = 1, \dots, n), \quad (2)$$

当 $\|\omega\|^2$ 最小且满足式(2)时的分类面就是最优分类面。而在 2 类样本中距离分类面最近、且同时平行于超平面 $H1, H2$ 上的样本数据称作支持向量。同时, 由于有些情况硬间隔最大策略效果不佳。此时, 考虑软间隔最大化引入松弛变量 ξ 来允许样本点落在间隔边界内, 甚至允许误分类样点存在, 分类面的最优化解问题可以转换成式(3)的约束优化问题:

$$\begin{cases} \min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m \xi_i, \\ \text{s. t. } y_i(\omega^T x_i + b) + \xi_i \geq 1, \xi_i \geq 0. \end{cases} \quad (3)$$

其中, C 是惩罚因子, 决定着 ξ 的大小。构造拉格朗日对偶函数将其转换为对偶问题求解:

$$\begin{cases} \min_{\alpha} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^m \alpha_i, \\ \text{s. t. } \sum_{i=1}^m \alpha_i y_i = 0, 0 \leq \alpha_i \leq C. \end{cases} \quad (4)$$

其中, α 为朗格朗日乘子, 通过二次规划方法求解得到各系数对应的最优解, 同时将低维数据映射

到高维时应确定核函数, 则最优分类函数即为式(5):

$$f(x) = \text{sgn}(\hat{\omega} \varphi(x) + b) = \text{sgn} \left[\sum_{i=1}^l \hat{\alpha}_i y_i K(x_i, x) + \hat{b} \right], \quad (5)$$

其中, $\varphi(x)$ 为 x 映射到高维空间后的特征向量, $K(x_i, x)$ 为核函数。常用核函数有多项式函数、径向基函数(RBF)、多层感知器函数等, RBF 函数可以在实现映射且参数较少时适当降低模型复杂度^[6]。本文光伏阵列的故障阵列选用 RBF 核函数建立分类器, RBF 函数可写作如下数学形式:

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2). \quad (6)$$

其中, γ 为核函数参数。后文将对惩罚因子 C 与核函数参数 γ 利用遗传算法进行寻优。

2 遗传算法优化的 SVM

遗传算法 (Genetic Algorithm, GA) 基于达尔文的进化论。其思想为模拟自然界的适者生存过程, 通过物竞天择不断朝着最优解进化。一般设计流程为: 先随机生成初始种群, 本文中需要对参数为惩罚因子 C 与核函数参数 γ 进行寻找优化, 用二进制编码来对其进行编码并组合在一起, 经过选择、交叉、变异产生新种群, 不断循环进化, 最后进化为一个最佳种群、即为问题的最优解^[7]。

在算法运算过程中不停地计算单个染色体的适应度, 使用交叉验证准确率作为适应度: 当个体的适应度越高, 则其存活概率就越大。交叉算子为单点交叉算子, 并用自适应变异概率产生变异基因, 随机地选出将发生变异的基因。遗传算法对 SVM 的参数进行优化的设计流程如图 2 所示。

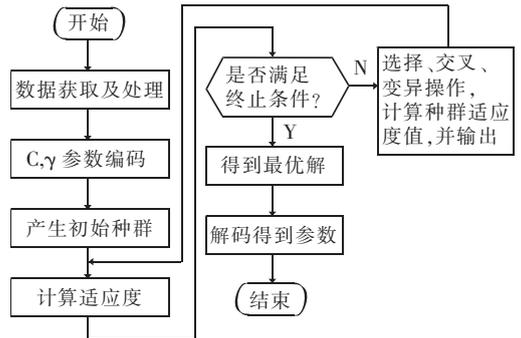


图 2 遗传算法流程图

Fig. 2 Flow chart of Genetic Algorithm

在此基础上, 研究给出了 GA-SVM 的设计步骤具体如下。

Step 1:初始化种群,对个体进行编码,基因序列为惩罚因子 C 与核参数 γ ,生成随机种群。

Step 2:设置终止迭代次数后,利用样本数据进行训练测试,根据规则计算个体适应度。

Step 3:当种群最优个体达到满足条件或达到终止迭代次数时退出寻优过程,得到优化解。跳转至 Step 5,否则进入下一步。

Step 4:对当代存活的种群执行选择、交叉、变异得到下一代种群,返回 Step 3 判断。

Step 5:得到最优核参数 γ 与惩罚因子 C 。并代入到 SVM 中通过训练样本训练得到诊断模型。

3 光伏阵列故障分析

为模拟仿真光伏阵列在各种环境下的故障特性并获得数据,在 Matlab/Simulink 内基于光伏电池的数学特性仿真建立了 3×5 阵列^[8]。即为 3 路支路并联,每路支路由 5 个光伏模块串联,如图 3 所示。

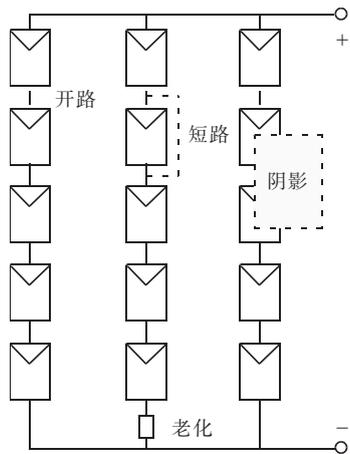


图3 光伏阵列故障模拟图

Fig. 3 Fault simulation of photovoltaic arrays

单个模块的参数相同,其中涉及的参数数值为:最大功率点电流 4.95 A,最大功率点电压 35.2 V,开路电压 44.2 V,短路电流 5.2 A。参考光照 $1\ 000\ \text{W}/\text{m}^2$,参考温度 $25\ ^\circ\text{C}$ 。模拟过程中光伏模块的辐照度设置为 $900 \sim 1\ 000\ \text{W}/\text{m}^2$,温度范围为 $25 \sim 45\ ^\circ\text{C}$ 。

在实际的并网运行中,常见的故障成因有:光伏模块间的线路由于恶劣天气等原因造成的绝缘层腐蚀、电池内部材料损坏等造成的组件短路;外界因素带来的连接线路断开造成组件开路;异物遮挡造成的遮阴故障,进而引发“热斑效应”;以及过久的使用年限导致组件老化^[9]。

对光伏阵列可能发生的故障,将其细化并对该

模型下的故障总结后分为以下几类,再通过仿真模拟,当阴影遮挡时设置辐照度为 $100 \sim 900\ \text{W}/\text{m}^2$,组件老化时串联电阻为 $1 \sim 10\ \Omega$ 。仿真得到总计 800 组数据,其中正常数据 400 组,8 类故障、各 50 组,详见表 1。

表 1 故障类别表

Tab. 1 Fault category table

序号	状态	样本数(占比)
Normal	正常	400(50%)
Fault 1	一个组件短路(通过短路开关组件模拟实现)	50(6.25%)
Fault 2	多个组件短路(通过短路开关组件模拟实现)	50(6.25%)
Fault 3	一个支路开路(通过断路开关组件模拟实现)	50(6.25%)
Fault 4	多个支路开路(通过断路开关组件模拟实现)	50(6.25%)
Fault 5	一个组件遮挡(通过减小组件光照模拟实现)	50(6.25%)
Fault 6	多个组件遮挡(通过减小组件光照模拟实现)	50(6.25%)
Fault 7	一个组件异常老化(通过串联电阻模拟实现)	50(6.25%)
Fault 8	多个组件异常老化(通过串联电阻模拟实现)	50(6.25%)

在不同故障下光伏阵列三列总特性如图 4 所示。

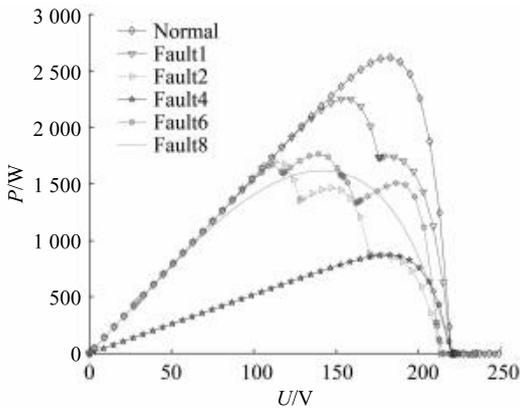
与正常工作状态进行比较,由故障状态的 $I - V$ 、 $P - V$ 曲线可知故障发生时的最大功率点均小于正常情况,且在遮阴与局部短路情况下呈多峰状态,从获取数据中提取特征参数(最大功率点电流 I_{mpp} 、最大功率点电压 U_{mpp} 、短路电流 I_{sc} 、开路电压 U_{oc}),其数据的最大功率点 $P - V$ 分布图如图 5 所示。

由图 5 可见,故障 2、6、8 分布较广,且分布交叉。故障 7、5 与正常数据贴近。部分故障分布差异较大,但部分故障数据分布集中,人为分类较困难。

4 仿真实验

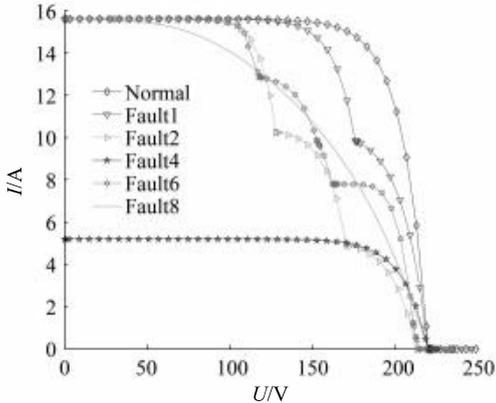
不失一般性,将所有样本分为训练集与测试集。研究选定了测试集为 150 个样本,其中 8 类故障类别各 15 份,正常数据 30 份。剩余 650 份数据为训

练集。使用遗传算法对支持向量机核函数参数 γ 与惩罚因子 C 进行寻优, 算法设定参数为: 初始种群数目 $pop = 20$, 遗传代数 $gen = 50$, 代沟 $ggap = 0.9$, 惩罚因子范围为 $(0, 100]$, 核函数参数范围为 $[0, 100]$ 。训练过程中用交叉验证方式得到的准确率即为适应度, 拥有分类准确率最高的 γ 与 C 即为最优参数。使用遗传算法对其进行参数寻优后, 经过 50 次迭代, 绘制成图, 可得运行结果如图 6 所示。由图 6 可知, 最佳准确率有过 2 次提升, 且最后稳定在 98.778%, 此时 $\gamma = 14.774, C = 75.607$ 。当不对参数寻优的情况下, 选用 $\gamma = 1, C = 1$ 参数对 SVM 模型进行建立, 再用模型对测试集进行正确率测试, 仅能正确分类样本 128 条, 准确率为 85.333%, 如图 7 (a) 所示。将训练后的模型对测试集进行测试, 正确分类样本 147 条, 准确率为 98%, 分类结果如图 7 (b) 所示。



(a) 光伏阵列故障 $P-V$ 特性曲线

(a) $P-V$ characteristic curve of photovoltaic array faults



(b) 光伏阵列故障 $I-V$ 特性曲线

(b) $I-V$ characteristic curve of photovoltaic array faults

图 4 不同故障下光伏阵列三列总特性

Fig. 4 three-column total characteristics of photovoltaic arrays under different faults

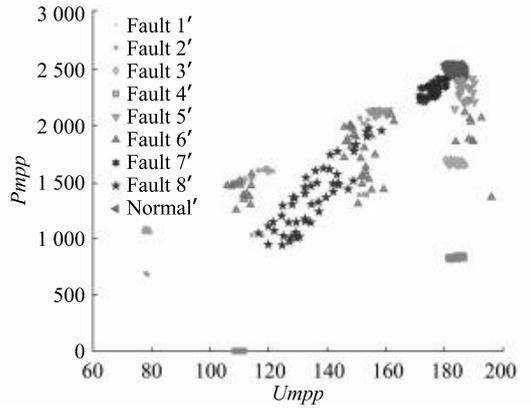


图 5 最大功率点 $P-V$ 分布图

Fig. 5 Maximum power point voltage-power distribution

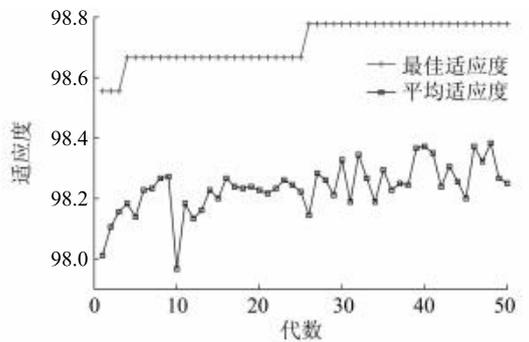


图 6 遗传算法迭代图

Fig. 6 Genetic Algorithms iteration diagram

由图 7 可见, 分类模型在故障类别 1、2、3、4、8 中均能准确识别分类, 但是对于故障类别 5 识别有误。故障类别 5 为当一个组件遮挡, 此时总体最大功率点电压电流变化不大, 当组件遮挡光照比较小时, 其特征值与正常值类似。与 BP 神经网络进行性能比较, 建立 BP 神经网络, 其网络输入层节点数为 4, 隐含层节点数为 9, 输出层节点数为 9, 学习速率为 0.1, 目标误差为 0.000 1。神经网络使用同样的训练集与测试集, 训练结束后, 准确识别测试集样本为 144 个, 准确率 96%。最终可得分类准确率的数值比较结果见表 2。

表 2 光伏阵列分类准确率

Tab.2 Classification accuracy of photovoltaic arrays %

	未进行优化的 SVM	参数优化后的 SVM	BP 神经网络
正常	100	100	100
故障 1	100	100	100
故障 2	100	100	100
故障 3	100	100	100
故障 4	100	100	100
故障 5	40	93.3	93.3
故障 6	26.7	93.3	86.7
故障 7	93.3	93.3	86.7
故障 8	93.3	100	93.3
总正确率	85.333	98.0	96.0

由表2可见,SVM分类器在进行参数寻优后准确率上有了较大提升,尤其在故障类别5、6上的精度提升很明显。在总准确度上要优于普通的BP神经网络。

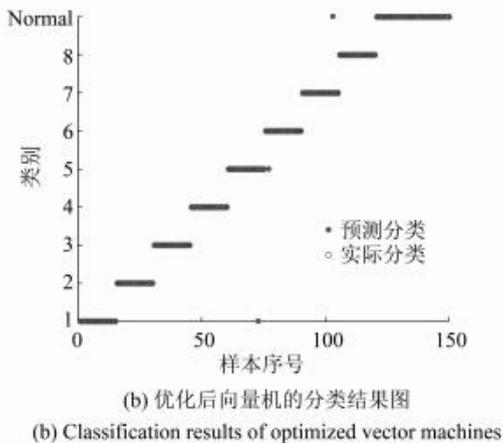
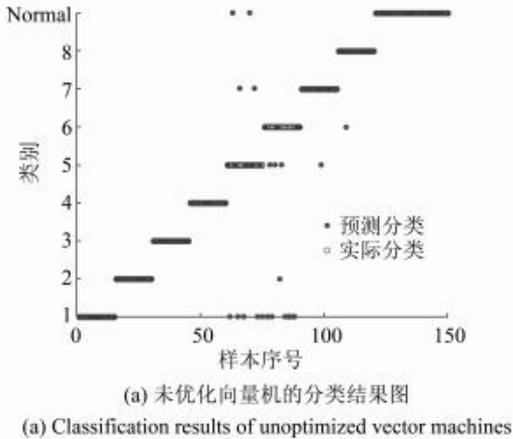


图7 优化前、化后向量机的分类结果

Fig. 7 Classification results of vector machines before and after optimization

(上接第57页)

算法对经典洗出算法的参数进行全局优化,搜索全局最优的参数,在有限的平台上充分地提升平台的回移速度和飞行员的体验感觉,尽可能地减少错误的暗示,使得优化后的飞行器将具有更高的逼真度。

参考文献

- [1] 王辉,朱道扬,傅孝靖. 飞行模拟器洗出算法优化设计及仿真[J]. 系统仿真学报, 2017,29(2):418-429.
- [2] 延皓,李洪人,姜洪洲,等. 六自由度运动模拟器的动感模拟算法研究[J]. 机械工程师, 2003(11):25-28.
- [3] 高健. 飞行模拟器动感模拟系统逼真度研究[D]. 哈尔滨:哈尔滨工业大学, 2013.
- [4] 刘国军. 六自由度运动模拟平台的分析及结构参数的优化

5 结束语

本文通过仿真获取光伏阵列在不同可能故障下的数据并提取特征值,利用遗传算法对支持向量机的参数进行寻优,再用优化后的参数建立模型,对光伏阵列进行故障诊断。实验表明,优化参数对分类准确率有较大提高,文中的支持向量机在利用遗传算法优化参数后对光伏阵列的故障诊断取得了较高准确率。

参考文献

- [1] 王培珍,郑诗程. 基于红外图像的太阳能光伏阵列故障分析[J]. 太阳能学报, 2010,31(2):197-202.
- [2] 董美辰. 基于模糊决策树的光伏阵列故障诊断方法研究[D]. 南宁:广西大学, 2018.
- [3] 李云桥. 基于改进BP神经网络的光伏阵列多传感器故障检测定位方法研究[D]. 西安:西安理工大学, 2018.
- [4] LIU Shengyang, LEI Dong, LIAO Xiaozhong, et al. Photovoltaic array fault diagnosis based on Gaussian Kernel Fuzzy C-Means clustering algorithm[J]. Sensors (Basel, Switzerland), 2019,19(7):1520.
- [5] 郭新辰. 最小二乘支持向量机算法及应用研究[D]. 长春:吉林大学, 2008.
- [6] 林培杰,陈志聪,吴丽君,等. 一种 PSO-SVM 的光伏阵列故障检测与分类[J]. 福州大学学报(自然科学版), 2017,45(5):652-658.
- [7] FANG Yihui, CHEN Xingwei, CHENG Niansheng. Estuary salinity prediction using a coupled GA-SVM model: A case study of the Min River Estuary, China [J]. Water Science & Technology: Water Supply, 2017,17(1):52-60.
- [8] NKURIKIYIMFURA I, SAFARI B, NSHINGABIGWI E K. A Simulink model of photovoltaic modules under varying environmental conditions [J]. IOP Conference Series: Earth and Environmental Science, 2018,159(1):012024.
- [9] APPIAH A Y, ZHNAG Xinghua, AYAWLI B B K, et al. Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques [J]. International Journal of Photoenergy, 2019,2019:6953530.

[D]. 哈尔滨:哈尔滨工业大学, 2014.

- [5] MENG Qiang, ZHANG Tao, HE Jingfeng, et al. Improved model-based control of a six-degree-of-freedom Stewart platform driven by permanent magnet synchronous motors [J]. Industrial Robot: An International Journal, 2012, 39(1):47-56.
- [6] PRADIPTA J, KLÜNDER M, WEICKGENANNT M, et al. Development of a pneumatically driven flight simulator Stewart platform using motion and force control [C]// IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Wollongong, NSW, Australia: IEEE, 2013:158-163.
- [7] 熊海国. 飞行模拟器发动机系统建模与面向对象的仿真研究[D]. 哈尔滨:哈尔滨工业大学, 2010.
- [8] 张立民. 飞行模拟器视觉仿真系统设计与关键技术研究[D]. 天津:天津大学, 2004.