

智能网格天气预报监控平台设计研究

何林^{1,2}, 任杰³, 吉庆⁴

(1. 陕西省气象信息中心, 陕西 西安 710014;

2. 秦岭和黄土高原生态环境气象重点实验室, 陕西 西安 710014;

3. 陕西师范大学 计算机科学学院, 陕西 西安 710119;

4. 陕西省气象台, 陕西 西安 710014)

摘要:近年来,全国气象部门构建了面向“国-省”的智能网格预报业务体系,将传统的站点天气预报转换为精细服务的格点天气预报。但由于业务流程繁杂,且缺少规范化的监控平台,为业务的运行维护带来很大困难。针对此问题,该文分析了当前气象部门各业务领域的监控系统研发现状,结合智能网格预报业务特点,提出将业务流程和数据流程解耦的策略。以数据流为中心,设计了对智能网格天气预报业务“全流程、一体化、可视化”的监控平台。通过采用 Kafka 消息中间件和 Logstash、Cassandra、ElasticSearch、Redis 分布式缓存等核心技术,实现了对业务监视数据的建模、采集、处理、存储和展示。目前,智能网格天气预报监控平台已在陕西省气象部门投入使用,运行稳定、可靠性高,有效提高了运维人员的监控效率,保障了智能网格预报业务的顺利开展。

关键词:智能网格;监视指标;日志事件;中间件;可视化

中图分类号:TP311.11

文献标识码:A

文章编号:1673-629X(2022)01-0197-07

doi:10.3969/j.issn.1673-629X.2022.01.034

Application Research of Intelligent Grid Weather Forecast Business Monitoring Platform

HE Lin^{1,2}, REN Jie³, JI Qing⁴

(1. Shaanxi Meteorological Information Center, Xi'an 710014, China;

2. Key Laboratory of Eco Environmental Meteorology in Qinling Mountains and Loess Plateau,
Xi'an 710014, China;

3. School of Computer Science, Shaanxi Normal University, Xi'an 710119, China;

4. Shaanxi Meteorological Observatory, Xi'an 710014, China)

Abstract: In recent years, the National Meteorological Department has constructed the national-provincial intelligent grid forecast system, which transforms the traditional weather forecast from station to grid weather forecast with fine service. However, due to the complex business process and lack of standardized monitoring platform, it is difficult for the operation and maintenance of the business. In order to solve this problem, we analyze the current situation of monitoring system development in various business areas of the meteorological department. Combined with the characteristics of the intelligent grid forecasting business, the idea of decoupling business process and data process is proposed. Taking data flow as the center, the monitoring platform of "whole process, integration and visualization" for intelligent grid weather forecasting business is designed. The platform has used Kafka message middleware, and core technologies such as Logstash, Cassandra, ElasticSearch, Redis distributed cache to realize the modeling, collection, processing, storage and display of business monitoring data. At present, the platform has been put into use in Shaanxi Meteorological Department with stable operation and high reliability, which effectively improves the monitoring efficiency of operation and maintenance personnel and ensures the smooth development of intelligent grid forecasting business.

Key words: intelligent grid; monitoring indicators; log events; middleware; visualization

收稿日期:2021-02-01

修回日期:2021-06-02

基金项目:国家自然科学基金(61902229);陕西省自然科学基金基础研究计划(2019JQ-271);秦岭和黄土高原生态环境气象重点实验室开放研究青年基金课题(2019Y-6)

作者简介:何林(1987-),男,工程师,CCF会员(F8689M),研究方向为气象数据管理与应用。

0 引言

天气预报与人们的生产生活息息相关,在交通出行、工农业生产、自然灾害防御等方面都发挥着重要作用。传统天气预报以县及以上行政区域为空间范围,针对未来几小时到几天可能发生的天气状况进行定性描述,存在时空尺度较大、服务不够精细和预报准确率较低等诸多问题。近年来,随着社会经济的发展及信息化水平的不断提高,国-省气象部门基本建成了以时空分辨率更高、要素定量化和多样化、预报精准、服务贴心为特征的智能网格预报业务体系,为政府部门和公众提供了全方位、高分辨率、定量定点的精细化气象要素预报服务^[1]。但由于智能网格预报业务流程及功能操作较为复杂,往往一个小环节的故障都会导致整个业务流发生中断与瘫痪,加之预报员与运维人员存在信息不对称、故障定位困难等问题,进一步加剧了运维难度。因此实现一体化监控、建立自动化运维体系、保障预报业务系统的稳定运行已成为制约智能网格预报业务发展亟待解决的关键问题。

建立满足业务需求的监控系统是气象行业信息化支撑能力的最重要体现^[2]。近年来,国内外气象部门纷纷开展了气象业务领域监控系统的设计研发工作,并取得一定进展。在国外,美国国家环境预报中心(NCEP)的实时数据监测系统(RTDMS)针对观测数据的数量和时效进行实时监控^[3]。欧洲中期天气预报中心(ECMWF)通过研发观测告警系统,实现对数据可用性和质量问题的监控^[4]。在国内,国家级代表性气象监控系统如文献^[5]的全国综合气象信息共享平台(CIMISS)的业务监控系统(MCP),针对气象观测数据,从数据的收集分发、加工处理、存储管理和共享服务等全生命周期进行了监视,并具备监视信息的统计分析和自动告警等功能。文献^[6-7]的综合气象观测系统运行监控平台(ASOM)则针对国家级地面自动观测气象站、探空观测气象站、新一代天气雷达等重点

观测网的设备运行状态、探测数据质量等进行监控,并改进运行监控综合评估、监控信息发布等功能。省级监控系统如文献^[8]设计了省级气象信息发布监控系统,整合现有的预警信息发布手段,实现一键式发布气象灾害预警信息、气象服务信息、气象为农信息的全程监控。文献^[9]提出适用于基层台站、基于云平台的气象数据监控服务系统,侧重于观测站数据下载、监控信息获取、主动告警服务的设计实现。国内气象业务监控的集大成者是气象综合业务实时监控系统(简称“天镜”),按照“横向集中、下沉一级、综合监控”的原则,建立横纵一体化的气象综合业务全流程监控,覆盖气象观测、信息、预报、服务、政务等气象业务全流程,监控范围覆盖业务系统的场地、网络、服务器、存储、数据流程、应用状态等,实现全栈业务的集中告警、综合监控、运维管理、大屏综合展示等功能^[10]。

该文按照“天镜”的设计理念,全面梳理智能网格天气预报业务中的关键环节,设计了智能网格预报监控平台的总体框架及功能模块。该平台以数据流向为主线,对衔接紧密的关键业务环节和监视对象进行解耦,分别建立监控指标模型,从监视信息的采集、处理、存储、展示实现分层逐级处理。通过可视化大屏设计,实现分层综合监视与集中告警,并建立与之联动的后台运维信息管理。该平台满足运维无纸化、便捷化、规范化、自动化的要求,提高了省级业务运维保障能力。

1 系统设计

1.1 数据流分析

监控平台围绕国家级-省级智能网格预报业务流程进行构建。表1分析了智能网格预报业务每个阶段的输入输出数据对应关系,将抽象的业务流程与可落地的数据产品分离,便于抽取关键数据产品,将关键业务环节的监视转换为对输入输出数据完整性和时效性的监视。

表1 智能网格天气预报业务流与数据流的对应关系

业务环节	输入数据	输出数据	关键算法
①原始数据加工处理	国内外数值模式数据	统一降尺度数据	模式解码、插值算法
②省级客观预报产品制作	统一降尺度数据	省级客观背景场	多模式交叉取优算法
③地市主观预报订正	省级客观背景场	分地市订正预报产品	人工交互主观订正
④省级预报把关拼接	分地市订正预报产品	主客观融合预报产品	多要素协同算法
⑤产品分发	主客观融合预报产品	站点预报产品	格式检查、FTP传输
⑥服务产品制作与发布	主客观融合预报产品	服务产品	模板解析算法
⑦预报产品检验评估	实况数据、预报产品	产品评分	检验指标评分算法

1.2 总体架构

监控平台的总体架构如图1所示,以数据流为中心屏蔽智能网格预报的业务流程。监视对象包括与业

务相关的各类硬件设备、网络、系统软件和应用软件等,适配相应的采集技术,获取监视源的状态、指标、事件等监视信息。存储层汇聚了监控平台所用到的各类

数据资源,包括依托以 CIMISS 为核心的省级集约化气象数据环境构建智能网格预报专题库,如基础气象数据、中间产品、预报产品和服务产品等^[11];通过采集工具获取各类监视对象的指标数据库和日志事件库;平台前端及后端设置面向不同监控对象的阈值、参数及相关元数据的配置信息库;为提高访问速度,设计数

据缓存库对热点信息进行存储。存储层与数据处理和应用层进行读写交互。业务逻辑层提供各类计算模型和管理组件,对数据进行加工处理,部分加工产品回写至存储层,部分直接提供给应用层进行调用。应用层即监控平台的前台大屏展示和后台运维信息管理,通过调用存储层的统一数据接口获取相关数据。

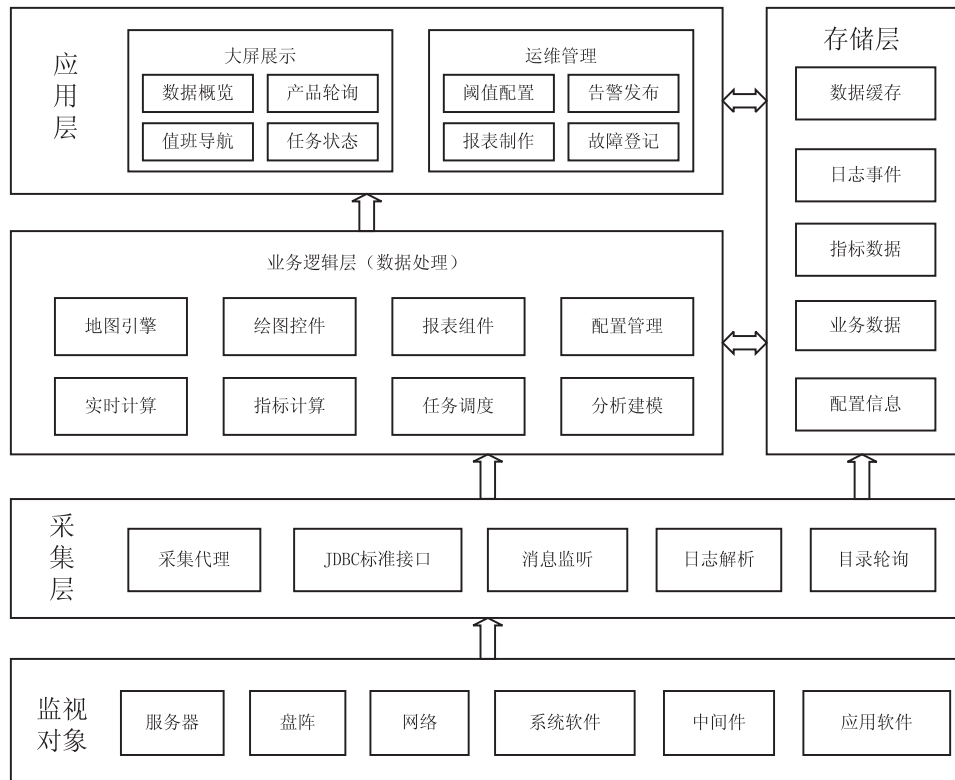


图1 智能网格天气预报监控平台架构

1.3 功能设计

包括综合信息概览、基础平台监视、数据全流程、算法与服务信息、告警配置管理五个功能模块。

(1)综合信息概览。作为系统访问的入口,综合展示当前智能网格预报业务中用到的各类气象基础数据、预报产品、服务产品的存储总量及分布,业务整体运行状况,当前所处业务流程环节、任务状态,当日业务值班信息,以及各类导航、通知及个性化推荐信息。

(2)基础平台监视。对智能网格预报业务中使用的所有服务器、存储等硬件设备,网络连通状态及流量信息、数据库和消息中间件等支撑软件的监控。包括查询单个设备节点状态,统计某时段内的流量,以及通过可视化组件对查询和统计结果进行展示。

(3)数据全流程。对智能网格预报业务中各环节输入和输出数据产品的应接收总数、实接收数、接收率、应入库总数、实入库数、入库率等以表格方式展示。并可查看接收和入库详情,包括每类数据产品记录列表、每条记录的接收时间和入库时间等。

(4)算法与服务信息。针对智能网格预报算法库

里各算法的运行及调度状况进行监视,包括运行节点、触发时间、执行是否成功。对预报数据的访问情况进行统计分析和展示,包括默认和任选时段内的数据调用次数、用户交互访问量等。

(5)告警配置管理。面向监控平台的管理人员,针对不同的监视对象,配置IP地址、监视指标、告警阈值、优先级别。定制告警发布方式,如语音、微信、短信、钉钉,以及相应的告警信息发送对象和范围。对告警信息进行统计生成报表,对影响业务运行的故障进行登记留痕等。

2 关键技术及实现

2.1 数据模型设计

智能网格监控平台的设计关键是对监视信息的采集、存储、处理及展示。对于每个监视对象的监视指标,首先需要从时间维度记录相应的监视数值,对其进行建模:

$$m_data = \text{getValue}(m_class, m_object, index1, index2, \dots, indexN) \& timestamp$$

其中, `m_class` 表示对象类别, `m_object` 表示监视对象, `index` 表示具体指标, 对于不同的监视对象, 其监视指标可以有多个, `timestamp` 表示监视时间戳。 `getValue`

函数表示采用不同采集方式获取监视数据, 将其结果记录在 `m_data` 中。

表 2 智能网格预报业务中的主要监视对象及指标

监视对象		主要指标
基础平台	服务器	CPU 负载、内存使用率、磁盘使用率
	存储设备	I/O 读写速率、空间使用率
	网络	网络连通性、速率
系统软件	数据库	集群节点状态、连接数、并发数、请求数、耗时最长 SQL
	消息中间件	集群节点状态、队列积压状态(队列名、积压数)
应用软件	模式解码、降尺度算法等	进程是否存在或僵死、进程占用 CPU 比率
服务信息	用户交互(网站/客户端)	用户日访问量、总访问量、按 IP 或分地市访问量
	业务数据调用	热点数据分布、数据调用次数、调用量、调用速度

表 2 对业务监视对象及其主要指标进行了汇总。根据监视指标类数据具有高吞吐量写入、低延迟读取、高频页面查询刷新、按时间粒度汇聚等存取特征, 选用

Cassandra 数据库对指标信息进行存储^[12], 将每一条记录转换为键值对, 数据结构设计如表 3 所示。

表 3 监视指标类数据 Cassandra 表结构设计与示例

Cassandra 术语		对应描述	数据类型	示例
ColumnFamily		基础平台监视指标	字符串	Platform
RowKey		监视对象	字符串	Server
Value	Columns1	Column name	CPU 负载	CPU Load
		Column value	负载值	浮点型 85.4%
	Columns2	Column name	内存	字符串 Memory
		Column value	使用率	浮点型 75.8%
	Columns3	Column name	磁盘	字符串 Storage
		Column value	使用率	浮点型 60.3%

对于每个 Columns 都有一个时间戳 `timestamp`, 即某时刻点记录的监视值。通过这种方式将指标数据进行持久化, 具有很强的灵活性和扩展性, 对于不同的监视对象 RowKey 都可以通过不同时刻 `timestamp` 的指标 columns 记录其监视值 value, 并且指标数量 column 是可以进行灵活增删, 使得不同监视对象的监视指标互不影响。

2.2 监视信息采集技术

针对不同的监视对象及特点, 采用 C/S 架构, 监视客户端“推”或采集服务端“拉”两种方式实现监视信息的主动推送或被动采集。涉及采集技术包括以下四种。

(1) 标准接口。对于数据库和消息中间件等规范化的系统软件, 一般都提供了类似 JDBC 连接或消息队列监听等标准化的服务接口, 采集服务端直接调用客户端提供的接口抽取信息, 然后二次加工后即可采集到相应指标的监视信息。

(2) 采集代理。针对基础硬件平台监视信息的采集, 使用封装 `psutil` 工具的客户代理实现。 `psutil` 是

基于 Python 语言开发的跨平台库, 主要应用于系统监控, 分析、限制系统资源及进程管理等。通过 `psutil` 内置的丰富的函数, 能够快速获取系统运行的进程和系统利用率(包括 CPU、内存、磁盘、网络)等指标的信息。

(3) 日志解析。针对应用和服务信息的监视信息采集, 一般采用日志解析方式。其中, 日志类型包括结构化存储的日志数据表、非结构化存储的标准格式或自定义格式日志文件两大类。结构化日志数据一般可直连源数据表, 并使用 SQL 语句进行简单处理后即可实现解析, 而非结构化日志则采用 Logstash 技术进行采集与处理。Logstash 是一款轻量级的日志搜集处理框架, 可以方便地把分散的、多样化的日志搜集起来, 并进行自定义的转换处理, 然后传输到指定的目标位置^[13]。Logstash 包括 Inputs(日志收集)、Filters(日志过滤器)、Outputs(日志转存)三个阶段, 形成了一个类似管道的数据流, 其中 Filter 是 Logstash 管道中间处理的核心组件, 包含对多种可插拔的日志解析插件的管理。智能网格预报监控平台主要采用了 Filter 中的


```

"number_of_shards": 3,
"number_of_replicas": 2  },
"mappings": {
  "_DI": {
    #定义 DI 索引的内部结构
    "properties": {
      #定义 DI 的基本属性,包括编码和名称
      "di_codes": {
        {"type": "string"}, {"di_name": {
          {"type": "string"} },
      #定义 DI 的具体字段
      "fields": {
        "receiver": "string",
        "sender": "string",
        "data_type": "string",
        "source_counts": "integer",
        "product_counts": "integer",
        "process_state": "boolean",
        "record_time": "date"
      }
    }
  }
}

```

其中,DI 的具体字段包括接收系统、发送系统、数据编码、源数据记录数、产品记录数、加工状态、运行时间等。针对建好的 DI 索引,插入一条文档的方法示例如下:

```

PUTMonitor_log/_DI/1
{
  "di_codes": "FORECAST.DCOEF.DI",
  "di_name": "交叉取优算法处理详细信息",
  "fields" {
    "receiver": "降尺度系统",
    "sender": "背景场系统",
    "data_type": "F.8299.0001.S001",
    "source_counts": "384",
    "product_counts": "16",
    "process_state": "true",
    "record_time": "2020-10-28 02:35:12"
  }
}

```

EI 文档设计与 DI 类似,区别在 fields 中增加了告警级别和响应状态,不再赘述。

2.5 数据可视化技术

数据可视化的主要目的是通过统计图形、信息图形和图表等方式清晰有效地传达信息^[17]。在智能网格预报监控平台中的数据可视化包括对气象业务数据、监视指标、日志事件的可视化。其中,气象业务数据具有多类型、高维度、弱模式等特点,可在后台定时绘图,前端直接加载轮询图像产品即可。而监视指标、日志事件等需从时间、空间等维度进行统计整合,并将分析结果以表格及绘图呈现。该文采用 Matplotlib 和

D3.js 进行数据可视化的实现。

Matplotlib 是采用 Python 语言开发的 2D 绘图库,便于将数据进行图形化,且提供多样化的格式输出。Matplotlib 提供一套与 Matlab 语法相近的 API,既可用于跨平台的交互式绘图,也可作为服务端绘图组件,设置固定参数后,定时从业务数据库中读出数据并生成相应的图形产品^[18-19]。D3.js 全称是 Data-Driven Documents(数据驱动文档),是基于数据来操作文档的 JavaScript 库,用于在浏览器前端中创建交互式可视化。D3.js 采用声明性编程,支持大型数据集,为线性、分层、网络和地理数据布置视觉元素,并将数据关联到 html 页面中的元素或元素组。采用 Matplotlib 结合 D3.js 的解决方案,满足对二维气象数据及监视指标数据常见的时序图、折线图、直方图、饼图、散点图、仪表盘、流程图、热力图等,在前端直接绘制和渲染。针对三维标量场气象数据绘制等值线/面、色斑图、落区图、切片图等,采用 Matplotlib 库提供的插值算法进行网格化,并与 Basemap 相结合,实现气象 GIS 图像的绘制,也可叠加至 Web GIS 地图上展示^[20]。针对气象多维数据,采用几何投影、图标、像素等多种方法,对其进行映射、投影和变换,实现对雷达、模式资料的降维展示^[21]。

3 结束语

智能网格天气预报监控平台已在陕西省气象部门进行了技术实现及推广应用。该平台不仅对智能网格预报数据环境中涉及到的地面、高空、雷达、模式等十余类基础气象数据,以及网格预报业务中生成的 3X3 公里客观背景场、网格实况产品、网格订正预报、站点预报、线路预报、公共服务产品的接收、制作和存储状况进行监视,而且对业务中用到气象私有云中的 30 余台虚拟服务器资源的实时运行状态、网络拓扑及流量、应用软件运行、数据调用和服务信息实现了一体化监控。

平台上线后,业务人员对智能网格预报业务的整体运行状况有了更为全面的把控。当异常发生后,平台能通过多种渠道快速地向值班人员发出告警,大大提高了下一步运行控制和故障处理的提前量,缓解了值班压力,有效保障了核心预报业务的质量,有助于提供及时、高效的气象服务。目前,该平台在陕西省气象部门运行稳定,已成为全省智能网格预报业务开展的关键支撑系统。

参考文献:

- [1] 胡争光,薛峰,金荣花,等.智能网格预报应用分析平台设计与实现[J].气象,2020,46(10):1340-1350.

- [2] 沈文海,何文春,孙超.从两个典型应用看气象信息业务的数据工作[J].中国信息化,2017(9):70-76.
- [3] KRISHNA K V, BALLISH B, STOUTDT J. Real time data monitoring at NCEP[EB/OL]. 2006-02-01. <https://ams.confex.com/ams/Annual2006/webprogram/Paper103227>.
- [4] DAHOU M, ISAKSEN L. Towards a unified observations monitoring system. [EB/OL]. 2009-11-02. <https://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/2009-meteorological-operational-systems>.
- [5] 熊安元,赵芳,王颖,等.全国综合气象信息共享系统的设计与实现[J].应用气象学报,2015,26(4):500-512.
- [6] 孟永红.综合气象观测系统运行监控平台(ASOM)设计[J].科学技术创新,2018(16):81-82.
- [7] 李峰,秦世广,周薇,等.综合气象观测运行监控业务及系统升级设计[J].气象科技,2014,42(4):539-544.
- [8] 莫云音,吴盛洪,陈亮,等.气象信息发布监控系统应用研究与实现[J].计算机技术与发展,2019,29(8):147-151.
- [9] 徐浩彬,赵亮,徐碧裕,等.基于云平台的气象数据监控服务系统实现[J].陕西气象,2020(2):59-63.
- [10] 孙超,肖文名,陈永涛,等.气象综合业务实时监控系统的的设计[J].气象科技进展,2018,8(1):153-157.
- [11] 何林,薛春芳,邓凤东,等.陕西省精细化气象格点预报数据环境的设计与实现[J].陕西气象,2017(4):17-21.
- [12] 王若瞳,黄向东,张博,等.海量气象数据实时解析与存储系统的设计与实现[J].计算机工程与科学,2015,37(11):2045-2054.
- [13] 鲜征征,叶嘉祥.一种改进的ELK日志采集与分析系统[J].软件导刊,2019,18(8):105-110.
- [14] 李涛,刘震,李亚玲,等.省级气象资料传输监控平台的设计及实现[J].计算机与现代化,2019(8):117-120.
- [15] 王慕华,唐卫,丰德恩,等.基于消息驱动的气象图形产品加工系统[J].气象,2018,44(5):692-698.
- [16] 何林,邓凤东,曹波,等.基于ElasticSearch的气象大数据原型系统[J].中低纬山地气象,2019,43(1):59-63.
- [17] 徐达,孙超,刘乖乖,等.大屏可视化技术在气象领域的应用[J].科技传播,2019,11(10):133-135.
- [18] 宋永生,黄蓉美,王军.基于Python的数据分析与可视化平台研究[J].现代信息科技,2019,3(21):7-9.
- [19] 王亚东. Python在气象数据可视化中的应用[C]//第34届中国气象学会年会论文集. 郑州:中国气象学会,2017.
- [20] 李新庆,陈海波,杨有林,等.宁夏综合气象信息共享与管理系统设计研究[J].计算机技术与发展,2019,29(5):135-141.
- [21] 王轩,路明月,郑建琴,等.雷达数据的多维可视化[J].图学学报,2017,38(4):497-503.
- +++++
- (上接第196页)
- [10] 程裕强,刘爱民,周国军.二维码生成算法研究[J].玉林师范学院学报,2014,35(5):188-122.
- [11] 骆超,靳晟,张太红.赛马赛事Android应用设计与开发[J].计算机技术与发展,2019,29(9):169-174.
- [12] 李佳怡,刘东旭.基于Kinect体感识别的自平衡车交互系统[J].实验室研究与探索,2019,38(7):77-79.
- [13] 杨思瑶,宁宝宽,陆海燕,等.基于Unity3d的家装设计展示系统[J].建筑设计管理,2018,35(2):63-66.
- [14] 王兴鲁.基于Unity3D与Kinect的体感交互技术应用研究[D].兰州:兰州交通大学,2017.
- [15] 任重庚,沈捷,王莉,等.基于Kinect骨骼数据的手势识别[J].计算机工程与设计,2019,40(5):1440-1444.
- [16] 骆金维,曾德生,张劲波,等.基于Kinect体感技术开发幼儿认知系统技术可行性研究[J].智能计算机与应用,2016,6(5):109-111.
- [17] 张金成.基于Kinect的健身游戏的设计与实现[D].武汉:华中师范大学,2017.
- [18] 孙浩然.西柏坡智慧旅游服务平台设计与实现[D].石家庄:河北经贸大学,2016.