GWAC工程背景简介及基本数据描述

徐 洋、吴 潮

文档导读：本文档主要提供对GWAC项目的技术和数据背景了解的辅助性文档，为大数据台开发的设计人员提供参考材料，而非设计需求约束文档。

1. 基本技术相关关键词
2. **GWAC系统基本组**
   1. 一个GWAC系统由40个GWAC相机组，每个相机的口径为18厘米，每个相机配备有4kx4k的CCD, 每4个GWAC相机由1个GWAC转台支撑。（按工程实现计划，先实现第一组20个相机，再实现第二组20个相机）
3. **唯一化星表（Uniq-Catalog）**
   1. **GWAC唯一化星表：** 对应GWAC观测到每颗星在天球上的位置属性，不包括时间的维度。
4. **唯一化星：唯一化星表中的每一颗星**
5. **天区（Sky-Region）：**
   1. 一个转台所对应的视场大小为S，将全天区按S进行分区编号，每个天区以skyid为标识
   2. 转台（4个望远镜）在观测时，在一段时间内会持续对一个天区进行观测。例如一个转台在20点到23点观测天区A，在24点到凌晨3点观测天区B等等。
6. **模板星表（Template-Catalog/Reference-Catalog）：**
   1. 每个相机对固定的skyid都有一个板模星表；
   2. 该星表会随时间更新，历史模板保留。数据量增加不大。
7. **模板星：**模板星表中的每一颗星
8. **观测星表：**
   1. 对望远镜观测所生成的新图进行点源提取后所形成的星表
   2. 需要对目标星表与模板星表进行交叉证认过程，使目标星表中的星与模板星表中的星相匹配
   3. 匹配成功之后，目标星即作为模板星的一条新记录进行存储
   4. 对未匹配的星需要做其他处理
9. **目标星：**观测星表中的每一颗星
10. **OT（optical transient）：**光学瞬变源目标星表匹配后，未匹配成功的星
11. **光变曲线**：唯一化星的所有星等观测记录随时间的变化
12. GWAC功能需求
    1. 数据量估算

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **望远镜** | **记录/20幅图** | **幅/观测夜** | **记录/观测夜** | **记录/年(300天)** |
| GWAC | 15W\*20=300W | 2000 | 60亿条 | 1.8亿亿条 |

* 1. 外部接口需求
     1. 数据入库

入库目标：目标图像点源提取后所生成的目标星表文件。现有星表属性请参考附录2。

数据入库程序需要完成星表文件的读取，交叉证认，匹配出已知星，找出未知星（OT），流量归一化，最终入库等操作。

现有入库流程请参考附录1。

* + 1. 星表查询需求

按以下查询功能提供页面查询、程序查询等接口需求

1. 按天球坐标（RA，DEC）查询（锥形检索）
2. 按类别查询
3. 按ID（名称）查询：按监控星ID查询出该星的所有观测记录（光变）
4. 按日期查询：查询该日期范围内监控星ID的所有观测记录
5. 按星等查询：
6. … （待定扩充）
   * 1. 数据输出接口

提供数据查询结果的下载功能。其他待定。

* 1. 功能需求

1. 光变曲线查询与显示；
2. 光变曲线分析工具。
   1. 其他需求（待讨论）
3. 星表的短期、中期、长期存储
4. 发现的科学目标切图、原始图像
5. 短期内原始图像的存储
6. 长期原始图像压缩
7. 是否需要通过Web页面的展示结果，追溯数据处理过程的详细信息，检查数据的有效性，需要追溯那些数据

附录

1. **现有星表入库程序（C语言版，基于Postgresql）流程简介（参考）**
2. **读星表文件**

目标星表文件主要以FITS二进制格式进行存储，当前项目中的FITS文件包含3个数据区（HDU）。第一个数据区没有存储数据，第二个数据区中存储原FITS图像文件的数据头信息，是一个一行一列的数据块。第三个数据区中存放星表数据，是一个N行M列的表格，N行代表有N颗星的信息，M行代表每颗星有M个属性。程序主要用到第二个和第三个数据区里的信息。

程序直接使用cfitsio库来对FITS文件进行读写操作。当用cfitsio库读出第二个数据区中的数据后，该数据实际是一个完整的FITS数据区头，不能直接进行操作，需要用到wcs库。

1. **交叉证认**

**交叉证认：**目标星表中的一颗星与模板星表中的每一颗星进行距离比较，当距离小于误差半径r，即认为两颗星是同一颗星。

**交叉证认有两种距离计算方式：**

1. 天球坐标（RA、DEC）

天球坐标通过大圆距离公式计算两颗星**(和(**之间的距离：

其中,为经度，,为纬度

1. 图像坐标（X、Y）

图像坐标通过笛卡尔坐标计算两颗星(和(之间的距离：

1. **星场判断**

待匹配的两个星表的星场之间可能只有一部分是相交的，因此目标星表中一部分没匹配成功的星并不是没有对应的匹配星，而是因为它的位置在目标星表星场之外，需要将星场之外的星去掉，找出真正的未匹配成功的星(OT)。如下图所示，两个星场只有部分重合。



目标星表星场与模板星表星场不完全重合

1. **流量归一化**

完成目标星表与模板星表的匹配之后，需要对目标星表的每个匹配上的结果进行流量归一化处理，具体过程如下：

1. 找出匹配结果中目标星表中的mage(请参考附录2.目标星表表结构)值小于0.05的所有星；
2. 对满足条件1的所有星计算：，其中，；
3. 求出Ratio的中值
4. 计算归一化因子；
5. 对目标星表中每个匹配上的星进行操作：magnorm = mag + magdiff。
6. **流量过滤**

完成目标星表与模板星表的匹配之后，需要对目标星表的每个匹配上的结果按流量进行过滤，具体过程如下：

1. 对目标星表中所有的匹配成功的星计算：，其中，；
2. 计算Ratio的标准差：  
   其中为均值；
3. 对目标星表中所有匹配成功的星进行过滤，过滤条件：  
   如果，则将第i颗星的信息写入到OT表中，其中为上小节中计算的Ratio数组的中值，K为可变参数，由用户输入。
4. **数据入库（PostgreSQL）**

目标星表与模板星表进行匹配及星场判断操作之后，目标星表中的星被分为3类：匹配成功的星S1，未匹配成功但在模板星场之内的星S2，未匹配成功但在模板星场之外的星S3。S1、S2、S3要分别进行存储。

1. **现有模板（目标）星表数据库表结构（参考）**

**星表文件表（**catfile**）表结构**

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 属性名 | 长度 | 备注 |
| catid | ID | 长整型 |  |
| catfile | 文件名 | 字符串 |  |
| airmass | 天文参数1 | 单精度浮点 |  |
| magdiff | 天文参数2 | 单精度浮点 |  |
| jd | 儒略日 | 单精度浮点 |  |

**模板（目标）星表（tamplate\_catalog）表结构：**

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 属性名 | 长度 | 备注 |
| tmpl\_id | ID | 长整型 |  |
| catid | 星表文件ID | 长整型 |  |
| ra | 赤经 | 单精度浮点 |  |
| dec | 赤纬 | 单精度浮点 |  |
| background | 背景亮度 | 单精度浮点 |  |
| classstar | 分类星 | 单精度浮点 |  |
| ellipticity | 椭率 | 单精度浮点 |  |
| flags |  | 单精度浮点 |  |
| mag | 星等 | 单精度浮点 |  |
| mage | 星等误差 | 单精度浮点 |  |
| magnorm | 归一化星等 | 单精度浮点 |  |
| fwhm | 半高全宽 | 单精度浮点 |  |
| magcalib | 矫正星等 | 单精度浮点 |  |
| magcalibe | 矫正星等误差 | 单精度浮点 |  |
| pixx | 图像X坐标 | 单精度浮点 |  |
| pixy | 图像Y坐标 | 单精度浮点 |  |
| thetaimage |  | 单精度浮点 |  |
| vignet |  | 单精度浮点 |  |
| pixx1 | 模板图像X坐标 | 单精度浮点 |  |
| pixy1 | 模板图像Y坐标 | 单精度浮点 |  |
| fluxratio |  | 单精度浮点 |  |