

doi: 10.3969/j.issn.1000-8349.2024.03.10

图像数据自动化处理与光变曲线抽取程序

邹乾一^{1,2}, 王炜¹, 郑捷¹, 毛益明^{1,3}

(1. 中国科学院 国家天文台 光学天文重点实验室, 北京 100101; 2. 中国科学院大学 天文与空间科学学院, 北京 100049; 3. 西藏大学 理学院, 拉萨 850000)

摘要: 对变星、太阳系外行星等瞬变源的巡天监测产生了大量的图像数据。快速处理这些数据, 生成光变曲线, 探测瞬变源对时域天文学的发展非常重要。以中国科学院国家天文台兴隆观测站拥有的一批米级光学望远镜为适用对象, 开发了一套图像数据自动化处理与光变曲线抽取程序 PPCLP, 它能够批处理原始图像并输出归一化较差测光光变曲线。程序拥有模块化、通用化、易使用化的特点, 适用于多种操作系统, 且使用了改进的较差测光方法, 能获得较高的较差测光精度。除适用于赤道式望远镜外, 也可用于处理由地平式望远镜获得的未经像场改正的数据。文章首先详细介绍了程序的架构、特点、使用方法; 然后以变星 V670 And 的数据为例, 展示了程序的处理结果; 最后讨论了其局限性与后续发展方向。

关键词: 光变曲线; 自动处理程序; 变星; 太阳系外行星; 测光数据处理

中图分类号: P111.2 **文献标识码:** A

1 引言

随着光学天文观测方法、观测仪器和数据处理技术的不断发展, 时域天文学已经成为当代天文学的一个重要组成部分^[1]。时域天文学的一个发展趋势是在多波段或多信使方面以尽可能短的时频对全天各类天体进行监测, 力争更全面地理解已知的瞬变源, 发现更多新天体、新现象, 为建立或完善相关理论或模型提供丰富的观测数据支撑。

在开展变星^[2]、太阳系外行星^[3]等时域目标的研究时, 测光光变曲线是目前为止最重要的观测数据之一。获取光变曲线需要利用望远镜对目标天区或天体开展长时间连续曝光或者多次曝光, 并从所获得的原始图像数据中开展测光分析, 获得多波段的高质量、高精度的测光数据与光变曲线。在 CCD 大量普及、大视场巡天越来越多的今天, 这样的观测数据非常丰富。为了有效地利用这些数据, 天文学家们基于不同的计算机性能、操作系统环

收稿日期: 2023-12-04; 修回日期: 2024-03-11

资助项目: 国家自然科学基金 (11988101, 42075123)

通讯作者: 王炜, wangw@nao.cas.cn

境、编程语言, 开发出多种多样的测光程序与软件包, 其中最著名的是 Butcher 和 Stevens 于 20 世纪 80 年代初编写的 IRAF^[4]。IRAF 精度高、功能多, 安装方便, 是天文学界最普遍使用的数据处理程序。不过, 由于其编写时间较早, 编程语言古老, 已逐渐落后于时代, 现已停止维护。

近年来, 随着国内外巡天计划的开展, 涌现出了许多图像处理与光变曲线抽取程序^[1, 6-8]。但这些程序都是围绕某架望远镜, 如兴隆站 1.26 m 近红外望远镜, 或某一类型望远镜, 如赤道式望远镜, 其生成的数据也可能格式不同, 一定程度上存在着通用性欠佳的问题, 不利于开展基于不同望远镜数据的科学研究。

本文的目的是在一定程度上与一定范围内改善上述问题, 提供一种适用于国家天文台兴隆观测站多台米级望远镜^[9]的光学波段天文图像数据处理程序, 实现光学上的多波段测光数据自动化快速图像处理与光变曲线抽取。用户仅需提供符合要求的原始图像观测数据, 并将其与代码放入指定目录、指定少数运行参数, 该程序便可实现对观测数据类型的识别、预处理、找源、测光、星点对齐、参考星选取及较差测光等步骤, 并进行光变曲线图像绘制。本程序命名为 Photometry Process and Light-Curve Pipeline, 简称 PPLCP, 已发布于软件项目托管平台 Github[®]。通过程序运行, 可以实现对原始图像数据的较差测光处理, 获得相对于绝对定标而言精度更高^[14-17]的目标光变曲线, 辅助制定下一步的观测计划。本程序也可监测望远镜视场位移, 为望远镜伺服机构维护提供帮助。

2 程序架构

2.1 软件环境

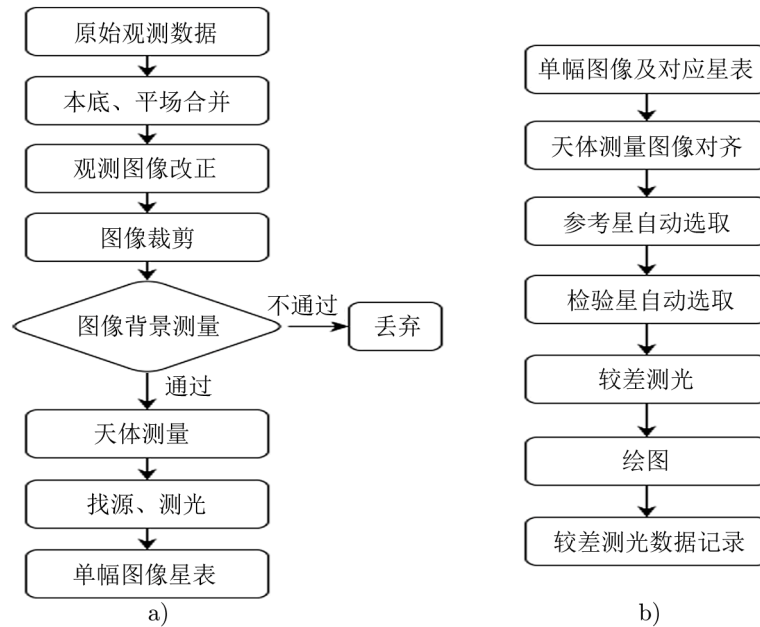
PPLCP 采用 Python3 语言编写, 调用天文软件包 astropy^[10]、astroquery^[11] 和 photutils^[12] 等。同时, 在数据处理中也使用了 Python 的一些通用软件包, 诸如: numpy、scipy、warning、matplotlib、os、glob 等。在天体测量部分, 本程序使用了 astroquery 包中提供的 Astrometry.Net^[13] 的 API 接口进行操作。由于此 API 需联网使用, 本程序天测模块也仅在联网时可用。

2.2 程序流程与总体架构

PPLCP 采用模块化结构, 由地址生成、图像改正、图像裁剪、图像背景评估、找星和测光、图像对齐、参考星和检验星选取、较差测光以及绘图模块共同组成。除地址生成预处理模块外, 程序整体运行流程如图 1 所示。按运行方式, 程序可以分成单幅图像处理与多幅图像处理两种模式: 在单幅图像处理模式下, 原始观测数据经本底平场合并、观测图像改正、图像边缘裁剪、天体测量坐标转化、找源和测光等步骤, 输出包含赤经、赤纬天测信息的经预处理的科学图像与星表; 在多幅图像处理模式下, 单幅图像及其星表经天体测量图像对齐、参考星检验星选取、交叉测光、绘图等步骤, 输出目标星的较差测光结果。各模块之

[®]PPLCP 代码网址为: <https://github.com/KeeliZou/An-Optical-Photometry-Data-Automatic-Processing-and-Light-curve-Extraction-Pipeline>

间的耦合主要通过模块间产出中间数据、用户指定参数实现。本程序允许用户按照需求或观测数据特性替换或弃用部分模块, 如图像背景测量模块、天体测量模块、绘图与较差测光模块等。



注: a) 单幅图像处理模式; b) 多幅图像处理模式。

图 1 程序整体流程

2.3 PPLCP 程序整体优势

PPLCP 在以下几方面具有一定优势。

(1) 自动化程度高。除模块调用选择和 8 个必要的运行参数外, 用户仅需将原始观测文件与代码放入对应目录, 程序即可自动识别观测数据类型并进行自动化处理; 对于不符合程序要求的文件, 仅需修改文件名或加入个别 FITS 头部关键词即可运行。

(2) 多波段同时处理。经用户设置波段信息后, 程序可同时进行多波段数据处理, 无须反复运行程序。

(3) 可处理带有场旋的图像数据。有少数地平式望远镜没有配置消旋器, 它们拍摄的图像存在场旋, 一般程序并不能直接处理。本程序可直接处理含有场旋的图像数据, 实现图像对齐并按天球坐标信息找星。

(4) 程序轻量化。通过将天体测量工作交与 API 处理, 在获得较好的天体测量精度前提下, 免去了用户下载庞大天测匹配库的存储要求和计算需求。

(5) 模块化与替代性强。用户可自行选择模块使用, 也可根据自身需求使用局部程序或增加自定义模块。

(6) 自动挑选参考星与检验星。通过找星与图像对齐模块, 程序能够自动识别给定测光

数据中有效数据点多的源,并挑选仪器星等比较稳定的源作为候选参考星与检验星,并在后续算法中通过多次迭代,自动挑选出各波段的最佳参考星与检验星。

(7) 较差测光方法更新。本程序适配于使用类似 CCD、CMOS 终端仪器的望远镜。对于瞬变源的研究,较差测光数据比较重要,能获得毫星等量级的测光精度^[14]。与普通的较差测光方法相比,PPLCP 将 10 颗参考星以其标准差倒数为权重拟合为 1 颗虚拟参考星的方法,减少红噪声对较差测光的影响,提高测光精度^[16, 17]。同时,程序还会在各波段挑选一颗与目标星星等差小于 1 mag 的星作为检验星,用以衡量星等临近情况下光变曲线中的白噪声水平。

(8) 辅助监测望远镜视场位移。本程序能够自动计算并记录每幅图像中心天球坐标位置,以图像方式监测观测过程中望远镜视场的位移。

(9) 对操作系统类型要求低。本程序为纯 Python 语言编制,除需联网调用 API 外,对操作系统类型无过多要求,可在 Linux、macOS、Windows 系统下运行。

3 程序关键步骤及实施方案

3.1 文件类型识别

望远镜产生的观测数据可大致分成定标数据和科学数据,自动化识别文件类型并分类可减轻用户使用负担。不同望远镜生成数据文件的命名规则不同,FITS 文件头部有效关键词命名方式也会有区别,这给自动化识别和调用增加了难度。

PPLCP 以文件名关键词匹配为主,头文件关键词检测为辅,完成文件识别。初步识别之后,还可通过数据的统计信息,比如中值、最大值等来确定数据类型。兴隆观测站米级望远镜产生的观测数据名通常含有部分识别度较高的名词,如本底文件一般包含 bias,而数据文件则包含目标名。通过适当的关键词与通配符组合便能识别不同类型的数据文件。

3.2 图像背景评估处理

在庞大的观测数据中,有少数数据因自然或人为因素导致质量不佳。此类数据在进行后续天体测量、找源、测光等步骤中通常会产生异常,影响程序运行和整体数据精度。

PPLCP 在必要的预处理步骤外,加入了图像背景测量模块。旨在通过对各图像进行背景估算,剔除某些因上云、起雾、月光、天文晨光、杂散光污染、读出噪声过高等因素导致的背景过高图像。通过在图像背景测量模块中将其剔除,既节约了后续运行时间,又提高了整体数据精度。

3.3 找源与测光

对于科学图像数据,通过合适的方法找源,测量相对流量或星等,将图像转化为包含时间信息的精确测光星表,是较差测光工作中最重要的工作之一。

在读取包含天测信息的图像后,PPLCP 使用 photutils 包中的找星函数进行找源与测光。找星模块已根据兴隆观测站常见天气质量及米级望远镜工况,对星点椭率、锐度、最小间隔距离、检测星点阈值、初始半高全宽等具体参数进行了设置;并以第一次找星结果的半

高全宽为参考值进行二次找星, 进一步剔除面源与异常值。考虑到观测站存在少见的极端气象变化, 本模块为初始半高全宽参数提供接口, 以方便处理观测条件较差时的数据。

3.4 图像对齐

将不同星表中散乱的星点转化为目标天体的有效连续测光数据, 是 PPLCP 中的难点之一。此过程中参考星矩阵的选取、星点判别方法及判别阈值的选择将直接影响产出数据有效性与精度。

本程序中, 图像对齐实际是星点坐标位置的对齐。不同于为兴隆站赤道式望远镜设计的“光变曲线自动抽取程序”(以下简称 QLCP) 中使用的“平移法”, PPLCP 利用图像头部中包含的天测信息, 在找源和测光模块将星点平面坐标信息转化为天球坐标信息, 使用天球坐标信息进行星点对齐。此模块以基准星表为参考, 将同一波段下其他星表中距离基准星表某一颗星一定距离内(此距离一般定为天体测量误差的 1.5 倍左右)的点源认定为同一颗星。之后, 可以根据输出的目标源坐标位置与输入的坐标位置的关系来判断图像是否对齐并进行调整。由于同一天区的多次科学曝光不一定包含同样的星表, 在天气有变化的情况下, 甚至不同曝光里被识别点源的数目有明显变化。用于对齐的基准星表通常星数量刚好大于均值, 以避免因星矩阵过大而增加计算时间、耗费过多资源。在完成所有星点对齐操作后, 还将剔除所有周围 2 倍半高全宽范围内存在疑似亮斑的源, 防止人造卫星、背景天体污染星光, 降低较差测光精度。

3.5 参考星、检验星自动选择

开展较差测光需要在视场内指定若干非变源作为参考星。假设它们的亮度不会随时间变化, 可以用于校正地球大气、仪器设备所产生的变化, 从而校正目标星的光变曲线。这一过程一般由数据处理人员手动完成。事实上, 仅凭肉眼识别并不能很好地挑选出各波段参考星与检验星。同时, 作为一个数据自动处理软件, 程序应尽量减少需用户指定的参数, 减轻用户使用难度及工作量。

因此, PPLCP 部分使用了 QLCP^[1] 中的自动选择算法, 可以从连续观测的多幅图像中自动找到亮度稳定的亮星作为参考星。首先, 利用对齐后的测光星表, 计算每一幅图像与基准图像之间的零点差, 即每幅图像的仪器星等和基准图像的仪器星等的差值。具体来说, 在每一张星表中选择信噪比较高的星, 剔除异常点之后计算星等均值, 这一张星表的星等均值与基准图像星等均值的差别即是这一副图与基准图的零点差。随后, 对每一张星表应用此零点差, 将仪器星等改正至基准图像的水平。最后, 筛选出较亮的星并计算这些高信噪比星改正后星等的标准差。标准差越小, 说明星越稳定。因此, 标准差最小的 10 颗星被选为参考星。检验星的选择类似于参考星, 仅需将信噪比条件更改为星等和颜色最接近目标源的即可。

3.6 较差测光

使用传统的较差测光方法时, 即使参考星数量充足, 仍不可忽视红噪声较大的参考星对较差测光精度带来的影响。因此, 本文的较差测光过程部分参考了 Everett 和 Howell^[14]、Wang 等人^[15]、石亚卿等人^[16, 17]的较差测光方法。本程序在进行较差测光时, 针对每一颗

参考星, 先使用 Sigma-clipping 方法, 以 2σ 为标准迭代 3 次剔除异常点; 接着将每颗星的光变曲线归一化到中值; 然后, 以每颗参考星光变曲线的标准差的倒数为权重, 对多条光变曲线取平均, 得到一条虚拟参考星光变曲线; 最后, 将这颗虚拟参考星的光变曲线再次归一化到中值。这样, 在进行较差测光时, 我们就可以扣除大部分红噪声, 减小参考星的弥散程度。在观测气象多变或较差的情况下, 此较差测光方法也能将参考星光变曲线的弥散度约束在 10^{-3} 数量级。除对参考星做此操作外, 程序也会对检验星使用 Sigma-clipping 方法, 以 2σ 为标准迭代 3 次剔除异常点, 并将检验星光变曲线归一化至中值后, 再进行较差测光。

4 程序性能测试

4.1 程序调用

PPLCP 程序已发布于 Github 平台, 它包含 7 个代码文件, 以 main.py 为主控代码。下载程序后, 创建 code 文件夹与 raw_data 文件夹, 分别将代码与原始观测数据置入, 并在 main.py 中设置必要运行参数即可运行。

4.2 使用示例

以兴隆观测站 60 cm 望远镜于 2021 年 10 月 16 日对 δ Scuti 型变星 V670 And 的观测为例。此次观测持续了约 4.25 h, 数据包含 B、V、R、I 四个波段, 共获得 552 幅图像, 此外还有 10 幅本底及每波段各 5 幅平场。

在 main.py 文件中对必要运行参数按表 1 中数值进行设置。由于此观测数据由更换 CMOS 前的 60 cm 望远镜产出, 因此调用 initial 包中的 address.cope.60_old 进行文件类型识别。接着, 依次调用 bfc_1、bfc_2、bkg、star.choose、plot 包分别进行图像改正、图像裁剪、图像背景评估、找星与测光、图像对齐、参考星、检验星选取、较差测光及绘图。单幅图像处理过程一般在 1~2 min 内完成, 总程序运行时间由科学曝光的数量和望远镜视场偏移度决定。最终结果详见下一节。

表 1 程序必要运行参数

参数名称	参数含义	示例用值
band_info	波段信息	['B','V','R','I']
target_name	目标名	V670 And
target_coor	目标天球坐标	23 30 37.2536 +46 24 04.2822
edge_cutx	X 轴图像裁剪量	20
edge_cuty	Y 轴图像裁剪量	20
ast_sigma	天体测量 API 星点检测 sigma 倍数	50
fwhm_initial	找星函数半高全宽初始值	5
star_num	图像星点阈值	20

4.3 输出结果说明

PPLCP 程序运行中主要产出文件如表 2 所示 (以 B 波段图像为例), 此表中包含本程序所有中间产出及最终结果。

表 2 程序主要产出文件

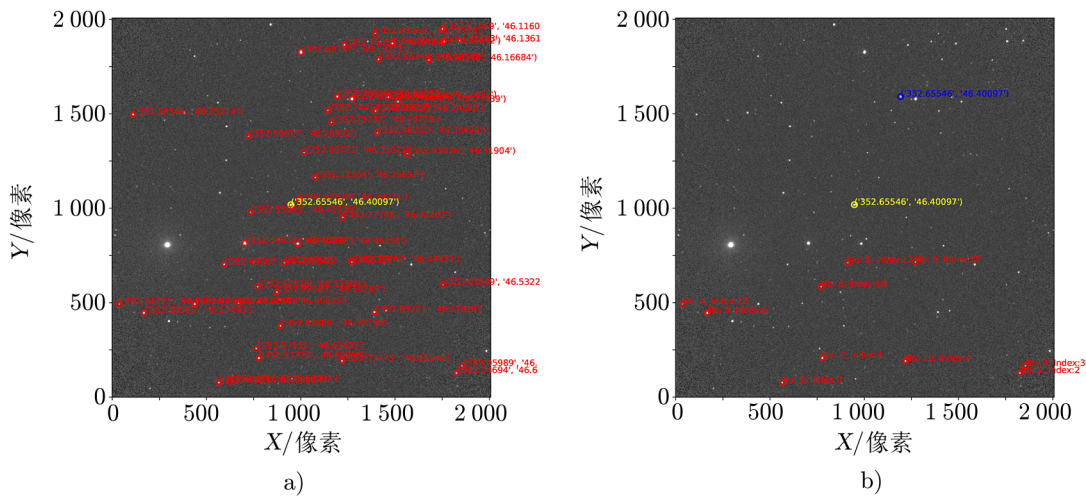
文件名称	文件内容	格式
superbias.fits	合并后的本底	简单图像 FITS 文件
superflat_B.fits	合并后的平场	简单图像 FITS 文件
Cor_V670And_B_001.fits	本底、平场改正后的观测数据	简单图像 FITS 文件
Cut_Cor_V670And_B_001.fits	裁剪和处理后的科学图像	简单图像 FITS 文件
Cut_Cor_V670And_B_001-ast.fits	带天测信息的处理后科学图像	简单图像 FITS 文件
Cut_Cor_V670And_B_001-ast-cat.fits	科学图像星表	二进制表格 FITS 文件
B_band_confirmed_star_001.fits	候选参考星光变数据	二进制表格 FITS 文件
B_band_target_star.fits	目标源光变数据	二进制表格 FITS 文件
B band differential photometry result.fits	归一化较差测光数据	二进制表格 FITS 文件
B band DEC movement.png	视场赤纬方向移动	PNG 图像文件
B band RA movement.png	视场赤经方向移动	PNG 图像文件
B band reference stars candidates.png	候选参考星、目标星证认图	PNG 图像文件
B band reference stars(AutoChoose).png	参考星、检验星和目标星证认图	PNG 图像文件
B band lightcurve.png	目标星、检验星较差测光图	PNG 图像文件

4.4 程序输出

以 B 波段为例, PPLCP 程序产出表 2 所示的二进制表格文件和 PNG 图像文件, 核心文件是星表数据和归一化较差测光数据。图 2a) 标注了所有候选参考星的天球坐标 (红色所示) 及目标星的天球坐标 (黄色所示), 图 2b) 标注了经自动选择算法选出的参考星及其索引 (红色所示)、检验星及其索引 (蓝色所示) 和目标星及其索引。图 3 为目标星与检验星的光变曲线。由检验星得出的 B 波段 RMS 在 0.0077 左右。图 4 为观测期间 60 cm 望远镜的视场中心沿赤经和赤纬方向的位移。不难看出, 当目标高度角降低时, 望远镜在赤经方向跟踪并不稳定。此图有助于后续望远镜伺服机构改进工作。本次观测中气象条件一般, 曾多次上云, 且星点位移较大, 因此光变曲线信噪比一般。但检验星 RMS 仍符合前文所述光变曲线弥散量级, 与 QLCP^[1]、广视角行星搜寻项目^[18] pipeline 的较差测光精度处于同一数量级。

4.5 存在的问题

本程序仍存在可扩充、可改进之处。首先, 利用 Astrometry.Net 的 API 开展天体测量需要上传数据、服务器响应和计算以及下载计算结果等多个步骤, 单张图像的天测需要大约 20 s (具体时间取决于服务器响应速度), 因此处理大量数据时总体耗时较长。后续我们将尝试将单张程序处理模块运行于服务器端, 在单张观测图像产生时即进行天体测量、找源、测光等工作, 或对天体测量、找源、测光步骤进行并行化操作, 这将极大缩短程序整体运行时间。



注: a) 候选参考星、目标星证认图; b) 参考星、检验星和目标星证认图。

图 2 参考星、检验星与目标星证认

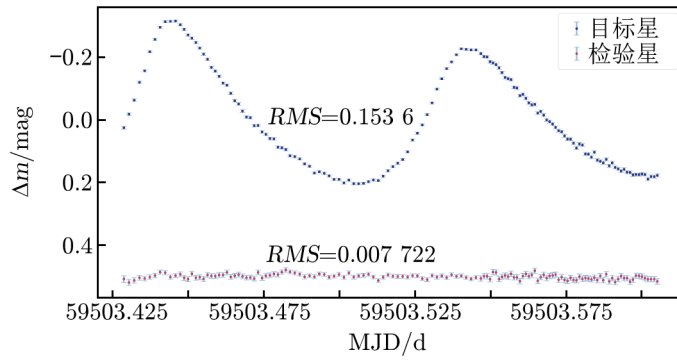
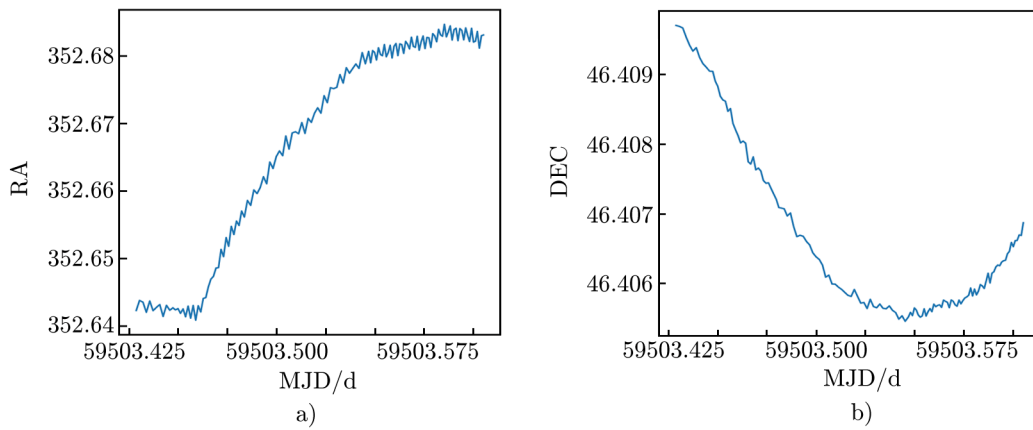


图 3 目标星、检验星较差测光



注: a) 望远镜视场中心沿赤经方向移动; b) 望远镜视场中心沿赤纬方向移动。

图 4 望远镜视场中心位移

此外, 由于 PPLCP 当前主要处理对象为各类变星及太阳系外行星测光, 此两种情况一般对视星等无特别要求, 且较差测光精度通常高于绝对流量定标, 因此, 本程序现阶段不包含与绝对流量定标相关的操作。

最后, 在对 85 cm 望远镜拍摄的科学图像进行处理时, 本程序需预先在 FITS 文件头部中加入望远镜指向信息。

5 总结与展望

本文介绍了我们开发的光学测光数据自动化处理与光变曲线抽取程序 PPLCP, 主要包含程序结构、使用方法与处理结果等内容。该程序可以在极大程度上简化光学多波段实测数据处理的复杂度, 并且具备了一定的模块化特点。但目前版本仍存在一定的局限性和不足, 有较大的改进空间, 我们将在后续使用中继续改进。

致谢

感谢中国科学院国家天文台兴隆 60、80、85 厘米望远镜全体工作人员的支持。本工作得到国家自然科学基金 (11988101, 42075123) 的支持以及中国科学院光学天文重点实验室开放课题资助。感谢包璐璐、高兴等人试用程序时提出的意见与建议。

参考文献:

- [1] 郑捷, 江林巧. 天文研究与技术, 2023, 20(01): 83
- [2] 沈强生. 云南天文台台刊, 1979(01): 45
- [3] 王佳琪, 王汇娟, 王炜, 等. 天文学进展, 2021, 39(01): 1
- [4] Butcher H, Stevens R. Kitt Peak National Observatory Newsletter, 1981, 16: 6
- [5] 钟文杰. 硕士论文, 广州: 广州大学, 2021: 27
- [6] 范舟, 赵刚, 王炜, 等. 天文学进展, 2018, 36(02): 101
- [7] 郑捷, 赵刚, 王炜, 等. 天文研究与技术, 2019, 16(01): 93
- [8] David Motl. C-Munipack. <http://c-munipack.sourceforge.net/>, 2010
- [9] 国家天文台兴隆观测站. <http://www.xinglong-naoc.cn/>, 2023
- [10] Astropy Collaboration, Price-Whelan A M, Sipőcz B M, et al. AJ, 2018, 156(3): 123
- [11] Ginsburg A, Sipőcz B M, Brasseur C E, et al. AJ, 2019, 157(3): 98
- [12] Bradley L, Sipőcz B, Robitaille T, et al. astropy 和 photutils 程序包: Zenodo, 2020
- [13] Lang D, Hogg D W, Mierle K, et al. AJ, 2010, 139(5): 1782
- [14] Everett M E, Howell S B. PASP, 2001, 113(789): 1428
- [15] Wang W, van Boekel R, Madhusudhan N, et al. AJ, 2013, 770(1): 70
- [16] 石亚卿, 王炜, 赵景昆. 中国科学院大学学报, 2023, 40(6): 761
- [17] Shi Y, Wang W, Zhao G, et al. MNRAS, 2023, 522(1): 1491
- [18] Pollacco D L, Skillen I, Cameron A C, et al. PASP, 2006, 118(848): 1407

A Pipeline for Image Data Automatic Processing and Light-curve Extraction

ZOU Qian-yi^{1,2}, WANG Wei¹, ZHENG Jie¹, MAO Yi-ming^{1,3}

(1. Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; 2. School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China; 3. College of Science, Tibet University, Lhasa, 850000, China)

Abstract: The monitoring of transient sources such as variable stars and extrasolar planets yields a significant volume of imaging data. Timely reduction of the data, producing light curves and triggering transient event are pivotal for advancing research in this field. Focusing on a group of meter-class optical telescopes located at the Xinglong Observatory of the National Astronomical Observatories of Chinese Academy of Science, an automatic pipeline to reduce time-series raw images and generating normalized differential photometric light curves named PPCLP is built and introduced in this paper. The program is modular, universal, and easy to use, making it adaptable across various operating systems. Employing an improved differential photometric method, PPCLP achieves a differential photometric accuracy of 10^{-3} . In addition to its compatibility with equatorial telescopes, PPCLP seamlessly processes field-uncorrected images acquired from altazimuth telescopes. The components, features, and usage of the program are described in detail in this paper. A brief introduction on how to use the program is presented using the data obtained from the 60 CM Telescope at the Xinglong Observatory for the variable star V670 And. Lastly, this paper discusses the limitations of the program and outlines potential avenues for future development.

Key words: light-curve; automatic pipeline; variable star; extrasolar planets; photometric data process