

Matlab 在天文底片数字化中的应用

傅承启, 赵建海

(中国科学院上海天文台, 上海 200030)

摘要: 采用 Matlab 语言开发了天文底片数字化处理软件 SDAPM。利用该软件可以处理商用平板扫描仪输出的照相底片扫描文件。由于 Matlab 语言强大的图形处理功能, 使处理程序比用 FORTRAN 编制的处理程序大为简化。除了介绍程序的主要流程以外, 还对该软件用到的一些 Matlab 特殊函数的具体用法 (特别是曲面拟合) 和如何解决内存不足的问题做了简要介绍。

关键词: 天文底片; 数字化; Matlab; 曲面拟合

中图分类号: TP319, P123.1⁺⁴

1 引言

照相底片是 CCD 被广泛应用之前天文观测最主要的探测器件, 全世界约有 300 万张天文照相底片, 其中我国有 3 万余张。这些底片记录了 100 余年来天体及其活动的信息, 因此, 底片数字化对永久保存和充分利用这些宝贵的天文资料极其重要^[1, 2]。

传统上, 人们采用 PDS (自动坐标量度仪) 进行天文底片的测量。PDS 十分昂贵, 而且拥有 PDS 设备的天文台屈指可数, 测量精密又要求它不能随意搬动。此外, PDS 的测量效率很低, 一张底片往往需要测量 5~6 个小时甚至 1 天, 这使它在天文底片数字化中的应用受到限制。因此, 各国天文台纷纷研制快速测量方法来实现底片的数字化, 利用高端商用平板扫描仪对底片扫描则是其中一条有效而简单的途径^[1-4]。

市场上有多种平板扫描仪, 选择哪一种扫描仪, 测试它的性能, 研究它的扫描精度及需要作的改正, 是采用这种方法实现天文底片数字化首要的也是最基本的任务。为了处理扫描仪数字化了的天文图像, 我们利用具有强大图像处理功能的 Matlab 编制了一个专用软件——天文底片数字化软件 SDAPM (Software for Digitalization for Astrographic Plates with Matlab), 以便对各种扫描仪进行比较和选择。本文介绍该软件的流程, 并结合流程介绍 Matlab 某些特殊函数以及应用经验, 这些函数或未见载于有关教程, 或在天文底片数字化上特别有用。

2 SDAPM 的处理流程

流程参见图 1。输入信息包括扫描仪型号、扫描分辨率、底片的拍摄年份、底片光心赤经赤纬和底片扫描图像文件名, 以及该天区的星表和初始证认文本文件。其中初始证认文本文件需事先建立, 它包含扫描星的序列号以及所对应的 Tycho 星表序号, 该文件至少需要包含 4 颗以上的恒星, 以便求解 6 个底片常数。

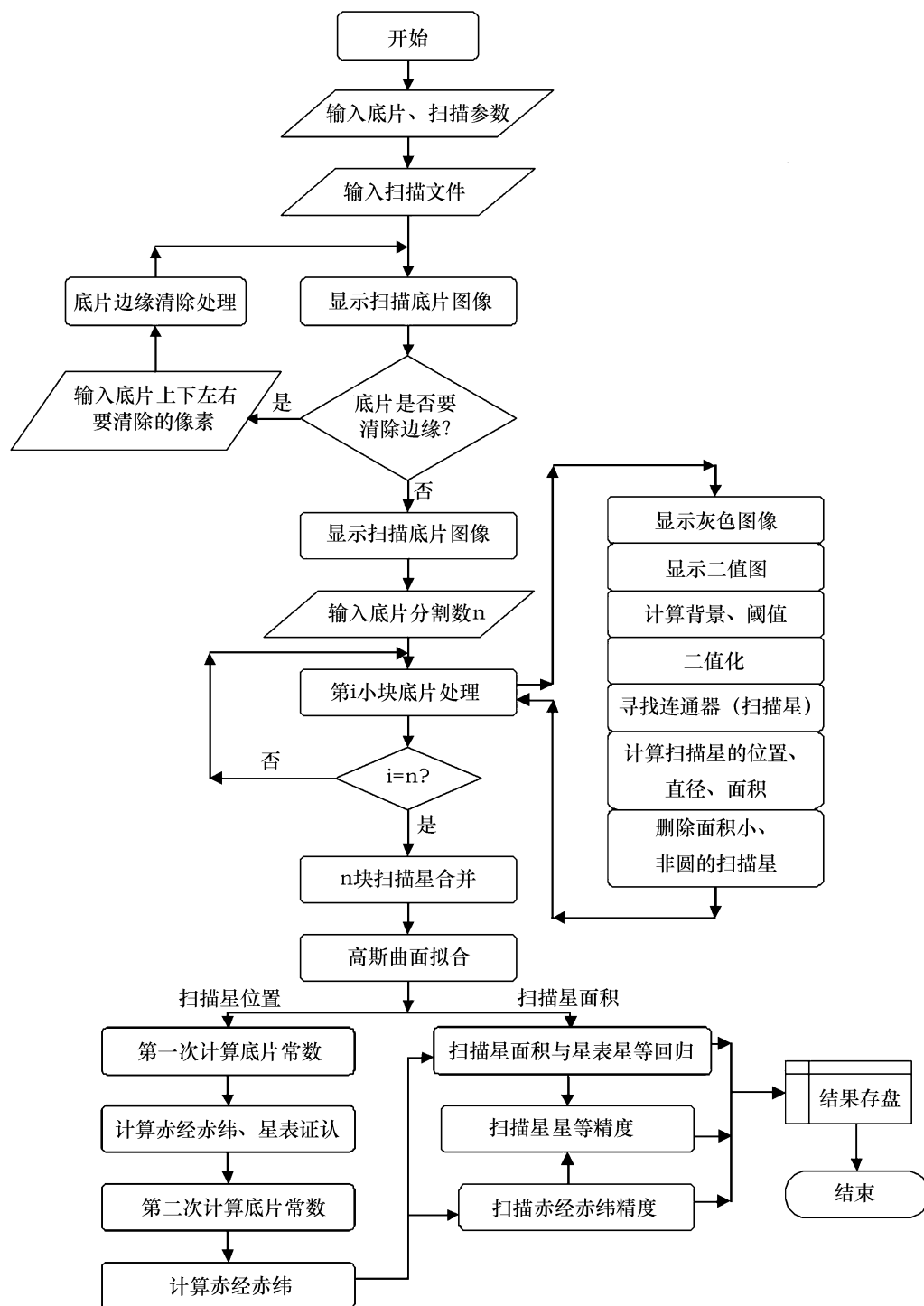


图 1 照相底片数字化程序流程图

输入的底片扫描图像文件需要进行初处理, 主要去除当年观测者在底片上书写拍摄情况而造成的污染以及扫描产生的透明白边。通过目测, 判断并输入底片被污染的边缘宽度 (以扫描像素为单位) 加以清除, 清除可反复进行, 直至没有污染边缘为止。

干净的底片可以开始处理, 但是, 底片扫描文件相当大, 2000 dpi 分辨率扫描一张 15 cm×20 cm 的小底片, 文件达数百 MB。由于 Matlab 存储变量采用连续内存区的方法, 因此电脑常常出现内存不足情况。为此, 将 2000 dpi 以上分辨率的扫描图像分割成 n 块, n 由目测判定后从键盘输入, 然后对每一小块图像分别处理。

每小块图像文件的处理方法都相同, 最后给出这块图像上所有扫描星的图像及参数 (位置、面积、直径、周长和偏心率)。处理的第一步是图像二值化。Matlab 图像二值化的命令为函数 `im2bw(filename, level)`, 其中不同扫描的阈值 `level` 可能不同。为此, 我们用图像上天空背景平均值及其均方差的 3 倍之和作为二值化阈值。为了排除星像对平均天空背景的影响, 阈值通过循环计算得出。每次循环剔除部分超过阈值的星像像素, 直至所有星像被剔除为止, 并得出最后的阈值。二值化后寻找扫描星的工作非常简单, 可以采用 `bwlabel(bw, n)` 函数寻找图像上的 n 连通器。不过, 图像上的连通器不完全是恒星, 更多的是底片上孤立的污点, 如划痕、字迹、霉斑或宇宙线作用痕迹。前几项一般没有固定形状, 不呈圆形, 而宇宙线痕迹面积很小。因此, 可以根据形状和大小来剔除污点。计算连通器特性的函数很多, 功能最全的当属 `regionprops` 函数, 它能轻易地求得每个连通器的几何中心位置、面积、有效直径、周长和偏心率等, 从而剔除面积太小和不圆的连通器, 留下恒星, 这里称之为扫描星, 以区别星表星。剔除污点后连通器的数量可减少约 70%, 具体数量取决于剔除的条件。

由于扫描星的数量已大为减少, 于是可将 n 小块图像合并成一个完整的底片图, 小图像上的扫描星合并成一个矩阵变量进行统一处理。这个扫描星矩阵包含扫描星编号、几何中心位置、面积、有效直径和偏心率。利用几何中心位置可以计算底片常数以及扫描星的赤经赤纬, 当然更好的是通过高斯曲面拟合进一步计算更精确的扫描星中心位置。

高斯曲面拟合采用 `fit` 函数, 这是 Matlab 2009a 版本开始新增加的, 而此前的版本没有专用的曲面拟合函数。拟合对每颗扫描星分别进行: 以扫描星几何中心位置为中心从原始图像中萃取其周围一小片天区, 然后用高斯曲面进行拟合。曲面拟合包含天空背景、星像幅度、中心位置和曲面宽度等五个参数。利用中心位置可以计算底片常数, 利用星像幅度和曲面宽度可以计算扫描星星等。

底片常数的计算分三步进行。首先, 人工识别 4 颗以上对应星表星的扫描星, 这是扫描星的初始识别。因此, 在程序开始运行前需要建立一个初始识别文本文件, 它包含 4 颗以上扫描星的编号和与其对应的星表星编号。程序根据初始识别文本文件将计算出 6 个初始底片常数。第二步, 根据初始底片常数以及底片的光心位置, 计算全部扫描星的赤经赤纬, 然后与输入的星表进行比对, 识别出所有的扫描星。第三步, 从识别出的扫描星中自动选出位置偏差小的、且不太亮也不太暗的扫描星 (太亮的星底片密度饱和, 太暗的星信噪比太低, 均使位置确定误差变大), 用它们进行底片常数的第二次计算和赤经赤纬计算, 并计算全部证认扫描星的赤经赤纬偏差。同时, 计算每颗已识别的扫描星的高斯积分面积, 并与星表星等进行回归, 得到计算面积与星等的标定关系, 据此关系, 可以得出其他所有未识别扫描星的星等。

最后, 全部计算结果与图像进行存盘, 并结束计算。

作为实验,我们用两种商用扫描仪——EPSON PERFECTION V200 及中晶 1000XL 进行了试扫描。扫描分辨率为 1000 dpi 及 2000 dpi, 字节为 16 bit 及 8 bit 两种。扫描的样品底片是 1987 年 2 月 4 日拍摄的 M79 底片, 大小 12 cm×15 cm。在 2000 dpi 扫描分辨率时扫描图像分割成 4 小块进行处理。图 2 绘出该 M79 天区恒星的证认结果。

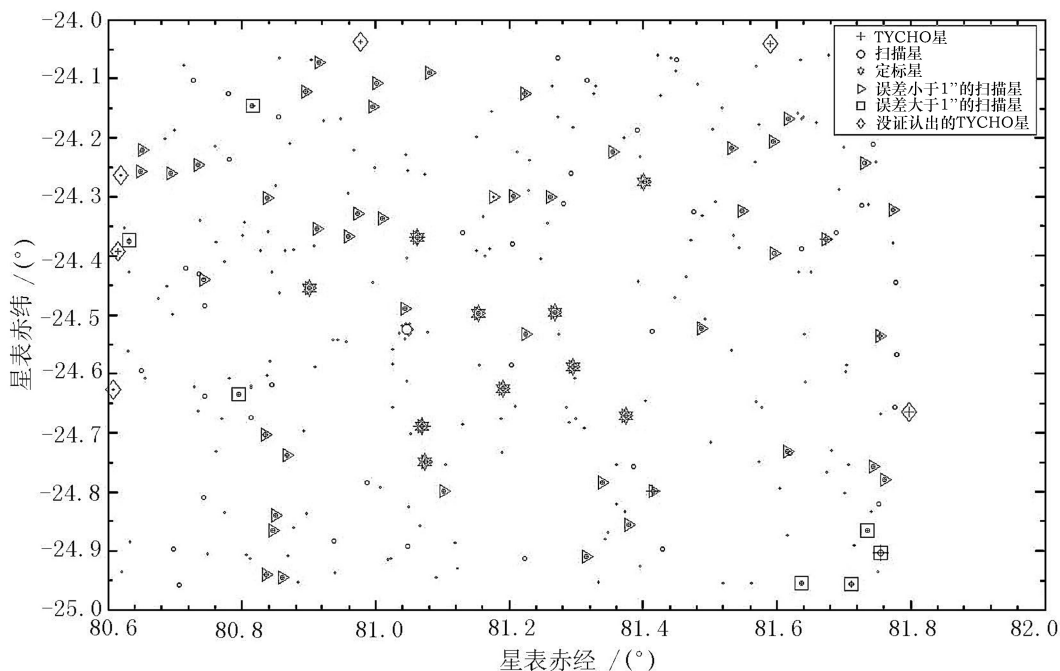


图 2 1987 年 2 月拍摄的 M79 星团底片经 EPSON 扫描仪扫描并经 SDAPM 程序处理后得到的底片图像。该星场有 62 颗 TYC2 星, 其中 56 颗 (三角和方块) 得到证认, 6 颗未证认的都位于底片边缘 (金刚符号)。所有证认扫描星与 TYC2 星表赤经赤纬差的均方根 $\Delta\alpha$ 与 $\Delta\delta$ 分别为 $\pm 0.403''$ 和 $\pm 0.288''$ 。

3 某些特殊 Matlab 函数

Matlab 软件处理数字化底片时,用到了一些特殊的函数,这些函数在有关教科书上的介绍十分简单,甚至完全没有。下面简要介绍这些函数的特殊应用。

3.1 图像文件的读入

商用平板扫描仪扫描图像通常为 TIF 或 BIMP 格式, Matlab 软件中可以采用 imread 函数读入。但天文图像常采用 fits 格式, 并有现成的软件 SAO Image ds9 可以直接查看 fits 格式的图像文件。因此, 为了用 ds9 软件对扫描星进行初始识别, 我们将 TIF 或 BIMP 格式的图像全部转换为 fits 格式, 然后用读取 fits 格式的函数 fitsread 读入数据, 语句为 data=fitsread(filename)。注意读出的数字就是底片的密度值 (即图片的强度值), 无需根据 fits 格式头部文件给出的零点 (BZERO) 和比例尺 (BSCALE) 进行换算, 可直接用 SAO Image ds9 软件校对。

fits 格式头部文件可采用函数 fitsinfo(filename) 读取, 该函数返回 fits 格式文件的结构。

3.2 图像的二值化

图像二值化并计算图像上各连通器几何特征的算法, 用早期的 Matlab 软件实现比较

烦琐, 现在的 Matlab 已有能轻松完成这些计算的函数。下面用我们编写的一小段程序加以说明:

```

bw=im2bw(scandata, threshold); % 扫描图像 scandata 二值化
figure(1); clf (1);
resizestar=imresize(bw, m); % 改变图像显示大小, 由 m 控制
imshow(resizestar); % 显示二进图
[Leb, n]=bwlabel(bw, 4); % 4 连通器标记、计数
states=regionprops(Leb, 'Centroid', 'Area', 'Eccentricity', 'EquivDiameter', ...
    'Perimeter'); % 计算连通器的几何特征
idx=find([states.Area]>25&[states.Area]<100000&[states.Eccentricity]<0.8 ...
    &[states.Perimeter]<(1.6*pi*[states.EquivDiameter])); % 对符合恒星像特
    征的连通器加标记
bw=ismember(Leb, idx); % 保留有标记的连通器
[Leb, n]=bwlabel(bw, 4); % 对连通器重新编号
resizestar=imresize(bw, m); % 改变图像的显示大小, m 为缩放比例
imshow(resizestar); % 显示剔除非恒星像以后的二进图

```

上述程序中, 二值化函数 `im2bw` 中的参数 `threshold` 为二值化阈值, 取为天空背景与 3 倍背景起伏之和。函数 `imresize(bw, m)` 只是为改变图像显示大小所设, 由数值 `m` 控制, $m < 1$ 为缩小。由于照相底片很大, 扫描分辨率很高, 因此命令窗口会出现显示图像太大的警告信息。增加该语句后, 不再出现此警告信息。

函数 `bwlabel(bw, 4)` 是采用 4 连通器方法对连通器进行标记、计数, 输出标记矩阵 `Leb` 和连通器数量 `n`。然后, 可以运用另一个非常有用的函数 `regionprops` 对连通器进行几何特征计算。这里我们仅仅计算几何中心、面积、偏心率、直径和周长 (上述程序的第 6 行), 实际上这个函数还可计算许多别的几何特征, 如包含连通器的矩形顶角坐标 (`boundingBox`)、长短轴长度 (`MajorAxisLength`、`MinorAxisLength`) 等。这个函数使图像处理程序大为简化。根据我们的计算, 这个函数给出的几何中心位置已相当准确, 与 TYC2 星表表载赤经赤纬相比, 差值的均方根已达到 $\pm 0.4''$, 可以用它们作为高斯曲面拟合时的初始位置, 而连通器的面积、偏心率、直径和周长则被用于判别恒星像的属性。

为了剔除非恒星像, 程序用了 3 句语句——`find`、`ismember` 以及 `bwlabel` 语句。它们分别对符合恒星像属性的连通器进行标记、保留和重新编号, 从而达到剔除的目的。

3.3 曲面拟合

通常采用质心或高斯曲面中心作为底片星像中心, 我们采用后者。从 2009a 版本开始, Matlab 可用 `fit` 函数进行曲面拟合, 该函数也可用于二维拟合。拟合对每颗扫描星分别进行, 否则会产生内存不足问题。以扫描星几何中心为中心抠取一块包含星像和天空背景的 k 行 \times k 列大小的图像进行处理, 下面是相应的程序:

```

u=[1:k]; % 给出网格化坐标
u=meshgrid(u);
v=u';

```

```

ft=fitttype('b+a*exp((-u-u0).^2-(v-v0).^2)/(2*sigma^2)'), ...
    'independent', 'u', 'v', 'dependent', 'z'); % 定义拟合曲面公式
opts.Lower=[0, 0, 0, u0low, u0low]; % 拟合参数的下限
opts.Upper=[a0, b0, di*1.1, u0up, u0up]; % 拟合参数的上限
opts.StartPoint=[z0, b0, diam/2, k0, k0]; % 拟合参数的初始值
opts.DiffMinChange=1e-9; % 拟合参数的变化步长
opts.TolFun=1e-8; % 拟合收敛时要求的函数符合精度
opts.TolX=1e-8; % 拟合收敛时要求的系数精度
[cfun, gof, output]=fit([u(:), v(:)], z(:), ft, opts); % 高斯拟合

```

从程序可见, fit 函数用于三维曲面拟合时包含 4 部分语句:

- (1) 网格化的二维坐标: 程序前 3 句语句定义的 u 和 v , 它们是 $k \times k$ 矩阵。
- (2) 拟合函数的定义: 由第 4 句 fitttype 函数定义。它的一般用法为:

$$\text{fitttype}(\text{expr}, \text{param1}, \text{value1}, \text{param2}, \text{value2}, \dots),$$

其中 expr 为拟合模型的表达式, 由用户提供。param-value 为选项参数对, 主要的选项有自变量、应变量、待定系数等。如以上设定 u, v 为自变量, z 为应变量。需要留意, fit 函数默认的自变量是 x , 应变量是 y , 其他未加设定的变量都被默认为模型的系数。因此, 曲面拟合时不要用 x, y 作自变量, 否则将出现错误。另外, 自变量 u, v 分别对应图像数据矩阵的列与行, 而不是行与列。

(3) 拟合参数选项: 由第 5-10 句所设定, 分别给出拟合的下限、上限、初始值、系数变化步长、拟合收敛时要求的函数精度和系数精度。系数的初始值十分重要, 不正确甚至偏差太大的初始值可能获得不同的收敛结果。另外, 好的初始值对快速收敛十分有益。

(4) 曲面拟合函数: 程序最后一句的 fit 函数。该函数用 ft 定义的模型拟合 u, v, z 数据, 给出模型的待定系数, 其中 z 即为图像上的扫描测量的密度值, u, v 为二维坐标 (图像的列与行)。

4 内存不足问题

Matlab 在天文照相底片数字化工作中的最大问题是内存不足, 常常出现 “out of memory” 的提示信息, 并停止运算。这个问题的产生固然与 Matlab 内存管理方法——需要连续内存区存放变量有关, 更重要是底片的信息量太大, 一张 $15 \text{ cm} \times 12 \text{ cm}$ 的小底片, 用 1000 dpi 分辨率、双字节扫描, 文件大小约 60 MB, 2000 dpi 分辨率的文件达到 240 MB。对于内存 2 GB、虚拟内存 3 GB 的 32 位 windows 操作系统电脑, 2000 dpi 就会出现内存不足的错误。输入 feature memstats 命令可检查内存情况。

解决内存不足的问题, 需要从多方面改进, 例如优化程序, 尽量少设变量, 清除和保存暂时不用的变量, 用 pack 命令, 增加虚拟内存等措施, 对于大文件, 读入并提取所需信息后应立即清除, 需要时再读入。即使如此, 处理 2000 dpi 图像时依然会出现内存不足问题, 因此不得不采用前面提到的分割底片处理办法, 有时还需要关闭 Matlab 释放内存后重新计算。当然, 最彻底的办法是采用 64 位操作系统。

SDAPM 程序现已用于天文照相底片的数字化处理工作,但要详细地分析扫描仪的扫描误差及其改正尚需对程序作进一步改进。另外, Matlab 许多功能强大的图像处理函数,还有待我们在实际工作中不断加以开发与应用。

参考文献:

- [1] 金文敬, 唐正宏, 王叔和等. 天文进展, 2007, 25(1): 1
- [2] Griffin E. In: van der Hucht K A ed. Reports on Astronomy 2006–2009, Trasnction IAU, Volume XXVIII
- [3] Vicente B, Abad C, Garzon F. A&A, 2007, 471: 1077
- [4] Nesci R, Barbieri C, Bucciarelli B *et al.* Men. S.A.It. Suppl. 2003, 3: 364

Application of Matlab in Digitization of Astrographic Plates

FU Cheng-qi, ZHAO Jian-hai

(*Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030*)

Abstract: A software for digitization of astronomical plates with Matlab is introduced in this paper. The software is used to process the digital files of these plates scanned by commercial scanners, and give positions and magnitudes of stars on the plates. The graphic processing functions of Matlab make the program much more simplified than that compiled with FORTRAN. This paper introduces the flowchart of the program, some special Matlab functions used by the software, especially the surface fitting function, and how to solve the problem on “out of memory”.

Key words: astrograph; digitization; Matlab; surface fitting