RevMexAA (Serie de Conferencias), **35**, 261–262 (2009)

MASSIVE DATA REDUCTION FOR MASSIVE SURVEYS: SPM AND NDWFS

D. Herrera[1]

RESUMEN

Presento algunos métodos para la reducción efectiva y automática de datos en dos grandes sondeos: el proyecto Southern Proper Motion (SPM) y el NOAO Deep Wide Field Survey (NDWFS). La reducción de datos en ambos casos incluye una interacción un tanto intrincada pero exitosa entre programas, lenguajes de programación y scripts, resultando en un uso mejor y más inteligente del tiempo del astrónomo durante el procesamiento de datos, igualmente a una consistencia más confiable de los resultados. Este automatizado del proceso de reducción se puede aplicar fácilmente a cualquier estudio con grandes detectores y grandes cantidades de datos.

ABSTRACT

I show some methods for effective and automatic reduction of data in two large surveys: the Southern Proper Motion project (SPM), and the NOAO Deep Wide Field Survey (NDWFS). The data reduction in both cases involves a somewhat intrincate but successful interaction among programs, languages and scripts, leading to a better, smarter use of the astronomer time during the reduction process, and to a more reliable consistency of the results. This pipelining of the reduction process can be easily applied to any survey with large detectors and large amounts of data.

Key Words: methods: data analysis — surveys

1. INTRODUCTION

In large surveys, data processing has become a difficult and lengthy task. Here I show two kinds of projects and reduction processes: one taking a large number of images with small detectors, and another taking huge images with large detectors. I show part of the reduction process, the tools used, some examples, how much it helps, and before the conclusions I'll show some examples of simple scripting.

2. SOUTHERN PROPER MOTION SURVEY

This program intends to get proper motions of aprox. 1 million objects in the Southern Hemisphere (for details see van Altena 1999). The total raw data is around 700 GB, between 4 and 5 GB per night. The computer usually spends between 2 and 10 minutes and 150 MB per pointing. The software we counted on were IRAF, Fortran, Linux, Perl, Visual (vi), SExtractor, and Awk. We assembled these in the right way in a huge script to obtain the reduced images.

To check how much we were improving the science by the reduction, we did the following: (1) We chose several fields that have two observations in different nights, then for every field; (2) we processed both images independently; (3) found the astrometric solution for both couple of images: the raw pair and the processed pair, one as reference to the other; (4) we plot the error in the astrometric solution found in each pair, for all the fields, in different magnitude bins. The results showed that the reduced images led to smaller errors in the astrometric solutions, specially for fainter objects (R >18) which actually were not even found in the raw images.

In our reduced images from one of our detectors (the ALTA camera), there was an outlier group of pixels left that yielded to "bumps" in the pixel value distribution (PVD) (see the top panel of Figure 1). We realized that the PVD of the raw science and dark images were somewhat different among them, apparently due to small temperature changes in the CCDs during the night. To fix this we multiplied all the darks (beyond the second) by the difference between the first and second maximum in the PVD of the first dark. This procedure equalized all the PVD peaks to the first dark. Then, we made a combined dark for every science image, matching its PVD peaks. What we obtain is shown in the lower panel of Figure 1.

Our pipeline has been working OK so far. Very few things change from night to night.

[1]National Optical Astronomy Observatory, 950 N. Cherry Ave, Tucson, Arizona 85716, USA (herrera@noao.edu).

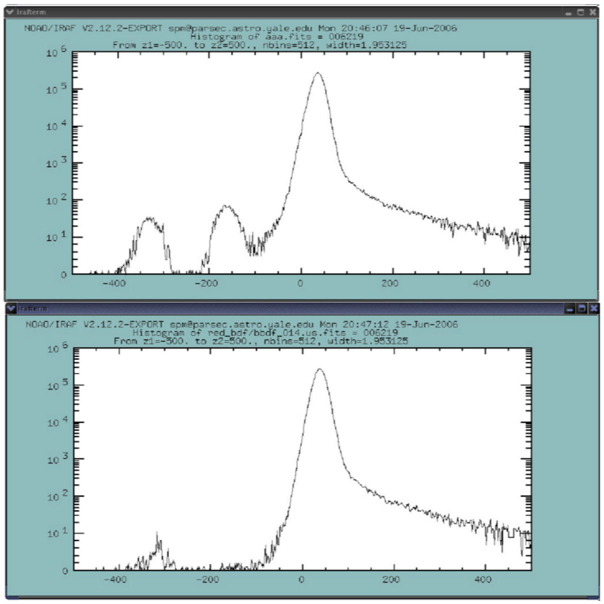

Fig. 1. The PVD in the ALTA camera in a regular science image before (top panel), showing the two bumps left, and after including the dark scaling process in the regular pipeline (lower panel).

3. THE NOAO DEEP WIDE FIELD SURVEY

This survey focuses in the large scale structure formation and evolution in the redshift range of 0.5 to 5, by observing fields repeatedly and stacking them to detect faint objects (see Jannuzi et al. 2000 for more details). The images are 8×8 Kpixels. Every raw image is 270 MB. The total raw data for the survey could easily reach some 500 GB.

We use the "NOAO Deep WIde Field Data Reduction Guide" (see Jannuzi et al. 2003) for guidance using IRAF and the external package "mscred". Our reduction pipeline is a longer process than that for SPM, with many more tasks to run and longer scripts, but at the moment we have been successful in making it all the way to the end with two scripts only.

4. REDUCTION HIGHLIGHTS

We developed pipelines that use mostly IRAF, but also include simple forms of tools like vi, Awk, shell scripting, Perl, for processes repeated for several data sets. Here is an example: I want to create a couple of lists, an input and an output list of all images in my current directory. The input list is created by a simple `ls` command:

```
!ls B_*.us.fits > objdark_bus.list
```

The content of this list will be:

```
B_113.us.fits
B_114.us.fits
B_Dark_001.us.fits
B_Dark_002.us.fits
```

Afterward I copy that file into another file which will become the output list:

```
!cp -f objdark_bus.list objdark_bbdf.list
```

Then I have a vi script with fixed commands to make the output list as we want it. This vi script is called `changes_objb.vi` and its content is:

```
:%s/B_/red_bdf\/bbdf_
:%s/red_bdf\/bbdf_D/bb_d
```

These are the vi commands we want to apply to our output file, and we do it in the following way:

```
!vi -s changes_objb.vi objdark_bbdf.list
```

In that case our output list file `objdark_bbdf.list` will look like this:

```
red_bbdf/bbdf_113.us.fits
red_bbdf/bbdf_114.us.fits
bb_dark_001.us.fits
bb_dark_002.us.fits
```

which is what we wanted.

In a similar way, Awk and Perl can be used in a simple way to create lists, scripts, etc.

5. CONCLUSIONS

Reduction can be challenging when new problems emerge, but automation can make the solution very easy to carry on. By automatizing the same tasks applied to many and big sets of data, we can make our time way more efficient. Common software tools can be used for this automation without complicated, long codes. This automation in the SPM Program produced a first paper (Casetti-Dinescu et al. 2007) about proper motion of globular clusters. Most of the procedures can be applied to sets of data from any telescope and detector, and additions or modifications for a particular data set can be inserted in the scripts easily.

REFERENCES

Casetti-Dinescu, D., Girard T. M., Herrera, D., van Altena, W. F., López, C. E., & Castillo, D. J. 2007, AJ, 134, 195

Jannuzi, B. T., Claver, J., & Valdes, F. 2003, The NOAO Deep Wide-Field Survey MOSAIC Data Reductions (http://www.noao.edu/noao/noaodeep/ReductionOpt/frames.html)

Jannuzi, B. T., Dey, A., Tiede, G. P., Brown, M. J. I., & NDWFS Team 2000, BAAS, 32, 1528

van Altena, W. F., Girard, T. M., Platais, I., Kozhurina-Platais, V., Ostheimer, J., Lopez, & C. E., Mendez, R. A. 1999, BAAS, 31, 1229