EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2024)
M. Corsini, D. Ferdani, A. Kuijper, and H. Kutlu (Editors)

Short Paper

PlateUNLP: An Open-Source Solution for Automating the Digitization of Historical Astronomical Plates

S. A. Ponte Ahón[1] and Y. Aidelman[2,3] and J. M. Seery[1] and F. Quiroga[1,4] and F. Ronchetti[1,4] and W. Hasperué[1] and M. Iannuzzi[2] and R. Peralta[5] and M. López[5] and A. F. Bariviera[6] and L. Cidale[2,3] and R. Gamen[2,3]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, UNLP, Argentina
[2] Facultad de Ciencias Astronómicas y Geofísicas, UNLP, Argentina
[3] Instituto de Astrofísica de La Plata, CONICET-UNLP, Argentina
[4] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA), Argentina
[5] Museo de la Facultad de Ciencias Astronómicas y Geofísicas, UNLP, Argentina
[6] Universitat Rovira i Virgili (URV), España

Abstract

The Observatorio Astronómico de La Plata belonging to the Facultad de Ciencias Astronómicas y Geofísicas (FCAG) of the Universidad Nacional de La Plata (UNLP) is among the oldest astronomical institutions in South America. Its extensive collection includes thousands of spectroscopic and photographic glass plates, which were collected between the early 1900s and the 1980s by eminent Argentine astronomers. These plates represent a unique and valuable record that remains largely unpublished. In 2019, the Recuperation of Historical Observational Work (ReTrOH, for its initials in Spanish) project initiated a digitization programme to convert over $15,000$ *plates from the collection into digital format. Given that the manual processing of these plates is intricate, error-prone, and time-consuming, a multidisciplinary team of astronomers and computer scientists was assembled to create a software tool called* ***PlateUNLP****. This tool was specifically designed to streamline the digitization process, enhancing efficiency and reducing the potential for errors.* ***PlateUNLP*** *utilizes advanced signal processing and computer vision techniques to automatically identify and isolate each spectrum captured on the plates, attaching the relevant metadata and thus limiting the need for manual intervention.*

The system is released as open-source software, allowing for broad accessibility. While ***PlateUNLP*** *was tailored to meet the demands of the ReTrOH project, its adaptable design makes it suitable for digitizing many other plate collections around the world.*

CCS Concepts
• ***Applied computing*** → ***Astronomy;*** *Physics;*

1. Introduction

The *Observatorio Astronómico de La Plata* (OALP, the Astronomical Observatory of La Plata, Argentina), now part of the *FCAG* of the *UNLP* has an enormous collection of spectroscopic and photographic glass plates. These observations, which are unique, unparalleled and unpublished records in the world, were made in the 20th century, between the 1900s and the 1980s, by renowned Argentine astronomers.

This heritage includes spectroscopic observations recorded on glass plates (as shown in Figure 1 and section 2) of variable objects, now of great interest to modern astrophysics, as well as data on asteroids, comets, novas, proper motions, and notebooks with logs of observations and comments, which also include meteorological data.

Since 2019, FCAG began digitizing its vast collection of around

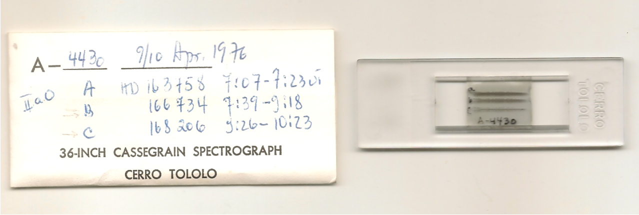


Figure 1: *Spectroscopic plate taken at Cerro Tololo Inter-American Observatory. Left: Envelope containing the plate with information about the observations. Right: Glass plate with three spectroscopic observations.*

DOI: 10.2312/gch.20241243

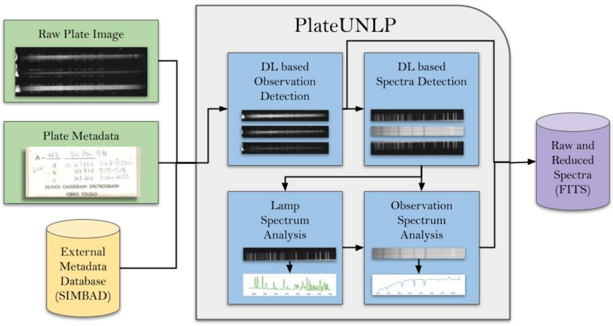


Figure 2: *Overview of the main tasks performed by the **PlateUNLP** software. The user inputs a scanned plate and a minimal set of metadata, that is complemented via external astronomical databases. The user then guides the software through the plate processing steps. The resulting FITS files store both the raw data and the results of the processing.*

15,000 spectroscopic plates through the ReTrOH project. So far, over 300 plates have been scanned using traditional digitization methods. This has proven to be a lengthy, error-prone process that requires expert intervention, often taking hours or days for a single plate and involving subjective decisions.

To address these challenges, Machine Learning, particularly Deep Learning models, offers a promising solution [C*18]. These methods can automate repetitive tasks in digitizing and processing spectroscopic data, providing a faster, more reliable, and scalable alternative to manual processing.

This work introduces **PlateUNLP**, an integrated system designed to digitize spectroscopic plates efficiently. The system automates metadata retrieval, spectrum segmentation, data reduction, and storage, using a user-friendly interface and public online databases (Figure 2). It also incorporates a Deep Learning model for spectrum detection, cropping and an automated calibration method. **PlateUNLP** is accessible through any web browser and is available under an open-source license, potentially adaptable for digitizing similar collections globally.

The most similar software to **PlateUNLP** is `PyPlate` [Tuv22], which assists in the digitization of photographic plates but does not support spectroscopic plate processing. With regard to the digitization of photographic plates used for spectroscopy, [WN17] and [Dav24] both evaluated the use of a commercial flatbed scanner for the digitization of spectroscopic plates and concluded that it is possible to eliminate and reduce the errors introduced by the scanner. However, only [Dav24] discusses the spectral extraction process. The tasks of the IRAF software were used to extract the spectrum and calibrate it in blue wavelength. Therefore, the methodologies used to carry out the digitization and reduction of these spectra are similar to those used by ReTrOH (Meilan, 2018 [Mei18] and Meilan et al., 2020 [MCA*20]). In other words, these processes are not yet automated.

2. Spectroscopic plates. The digitization and reduction process.

Photographic plates consisted of a glass base with a photosensitive emulsion attached to one side. The emulsion is capable of reacting to photons, which carry astrophysical information, via a photochemical effect. The intensity of the signal becomes visible in the darkening of the emulsion. Photographic plates could detect photons between the near ultraviolet (which were discovered with these detectors by Johann Wilhelm Ritter) and 8000 Å (the infrared limit of the visible range of the electromagnetic spectrum). These detectors were employed for both photometric and spectroscopic purposes.

In spectroscopy, a spectrum represents a picture in which the amount of energy received per wavelength is observed. The image can be obtained by shining light through a dispersing object, such as a prism or a diffraction grating. This process breaks down the white light into its constituent wavelengths. The stellar spectrum corresponds to the central white band in Figure 3.

The spectra of the acquired objects of study, designated as science spectra, are invariably accompanied by "comparison lamp spectra". A comparison lamp generates a spectrum from a light source composed of known gases, characterized by a specific set of emission lines with well-known wavelengths. Consequently, the wavelengths corresponding to the emission lines observed in its spectrum are also known. The lamp spectrum enables the construction of a function $f(x) = \lambda$, where x represents the position (in pixels) in the dispersion direction and λ is the corresponding wavelength. Therefore, the spectrum of a celestial body contains information about the source of the light, including its temperature and the chemical elements present in its photosphere.

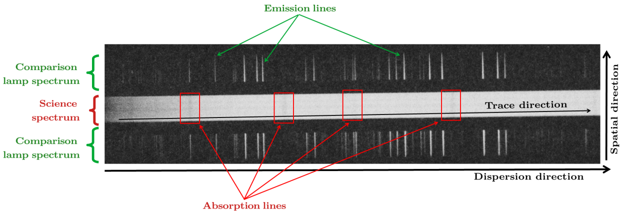


Figure 3: *A spectroscopic plate. The central white band corresponds to the science spectrum. The two bright vertical segments located at the top and bottom of the image represent the two comparison lamp spectra.*

The reduction process of the digitized spectra, when performed manually, is typically carried out using IRAF tasks, as explained in the work of Meilan (2018 [Mei18]). Afterwards, the wavelength calibration is done. Through this process, line identifications and other decisions are also performed manually.

3. PlateUNLP: A system for automatic processing of astronomical plates

PlateUNLP helps experts digitize plates efficiently and systematically. Given a scan, the system assists in segmenting the spectra and retrieving metadata (details in [PA23, PPA*23]). After processing, a FITS file is generated for each identified spectrum. The software includes a frontend built with Svelte, a Node.js backend for scan management, a YOLO-based spectrum detector, a Dwavelength calibration and extraction tool, and a Python service using Flask that connects detection with the web application.

Initially, the frontend retrieves a list of available scan files, allowing the user to select one. The system updates the stored information through the backend API as the user processes a file. This allows the user to pause and resume work on a file anytime.

Once the user selects a file to work with, the frontend retrieves its information, including the scan image and its technical details, which are displayed below the file selection menu. If the file has already been processed (partially or entirely), all available data is retrieved via the backend API. For files that have not been processed before, the frontend sends a request to the backend to use the spectrum detection service to identify the positions of each spectrum in the scan under analysis. More information about spectrum detection can be found in Sec. §4. An editing system for detections has been incorporated, allowing the user to modify the number and regions of detected spectra. Additionally, a filter system has been added to adjust brightness, contrast, and colour for better image manipulation. These changes do not affect the final cut of each object.

On the other hand, proper metadata management is crucial. The system requires certain standardized metadata in FITS files, even if specific data is missing. Metadata is divided into two types: plate-related (common to all spectra, e.g., plate number, observer) and object-related (e.g., star name, observation time). While manual entry is possible, it can be time-consuming. To streamline this, the system can automatically derive some metadata from the TIFF file, use default values for repeated fields, and retrieve additional information from databases like SIMBAD. Users can input minimal data, and the system will calculate the remaining required metadata.

4. Automated spectra detection

Once a spectroscopic plate is digitized, it is necessary to detect and crop the different spectra present on it, to export them individually. To streamline this process, we developed an automatic spectrum detector based on Machine Learning. To achieve this, we created a database and trained a convolutional model specific to this task.

Automating the detection of spectra presents several challenges despite the seemingly straightforward nature of manual identification. Multiple spectra on a single plate necessitate treating the task as a complex detection problem within an image. As a spectrum detection technique, we trained a YOLO model [RDGF15]. Figure 4 shows an example of detection on some plates from the test set. The model performs well in different scenarios, with varying lighting, noise, artefacts, and the number of observations. Additionally, deteriorated plates, which may be dirty or stained, further complicate the process (Figure 4 second case). Furthermore, the absence of comparison lamps or corresponding spectra requires discarding such cases, as this prevents proper analysis of the data.

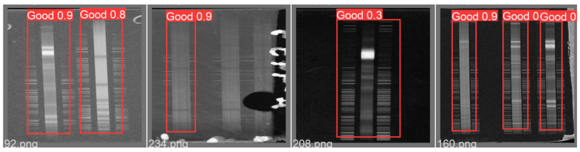


Figure 4: *Sample detections of the model in the evaluation set. In the second image, two observations were not detected due to being invalid for analysis because of noise.*

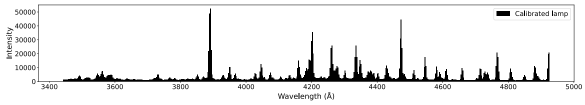


Figure 5: *Comparison lamp spectrum calibrated in wavelength (Å). On the X axis the wavelength number. On the Y axis, the intensity that was recorded in the lamp for each wavelength.*

5. Automated wavelength calibration

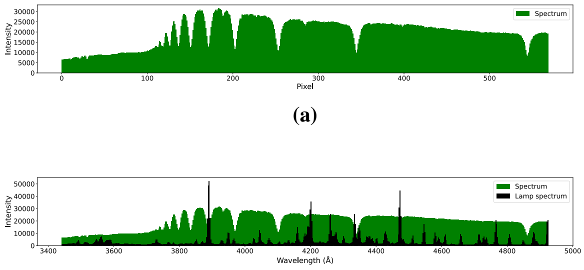
Wavelength calibration is a critical and challenging step in automating plate processing. After extracting the science spectrum, it is necessary to calibrate it in terms of its wavelength. Comparison lamp spectra are critical for this step.

By leveraging the known wavelengths of emission lines observed in comparison lamp spectra, we can derive a transformation between the position along the dispersion direction and the corresponding wavelength. Applying this transformation to the science spectrum allows us to obtain a wavelength-calibrated spectrum. Figure 5 illustrates a well-calibrated comparison lamp spectrum.

Accurate wavelength calibration requires reference spectra for the elements in comparison lamps (e.g., He, Ar, Ne, Fe). Users, typically aware of the elements in their lamps, can input this information into the software. The corresponding spectra can be obtained from databases like NIST, which provides peak wavelengths, or LIBS, which offers peak data and interpolated values based on user-specified resolution. After calibrating the comparison lamp spectra, the next step is the calibration of the science spectrum.

Figure 6a displays the graph of the extracted science spectrum function. With the calibrated lamp function obtained previously, the problem of wavelength calibration is nearly resolved. The wavelength of each position of the lamp emission line is known, and each lamp is aligned with its respective spectrum. By examining each horizontal pixel of both elements, a wavelength correspondence can be established between the calibrated comparison lamp and the spectrum to be calibrated. The calibrated science spectrum is the final product of this process. Figure 6b shows both the calibrated science spectrum and its corresponding calibrated lamp spectrum (showed in Figure 5).

As mentioned before, calibration involves aligning observed data with reference spectrum while addressing two challenges: non-uniform wavelength sampling along the X-axis and varying intensity ranges along the Y-axis. Manual calibration involves counting potential coincidences of emission lines for each chemical element, which can number in the dozens. The process requires exploring



(a)

(b)

Figure 6: *(a) Newly extracted science spectrum function. On the X axis, the number of horizontal pixels corresponding to the longest side of the spectrum, on the Y axis, the average intensity corresponding to each pixel. (b) Science spectrum function calibrated in wavelength (Å) based on a previously calibrated comparison lamp function (Figure 5). The X axis represents the wavelength, and the Y axis shows the intensity recorded in the lamp for each wavelength.*

Table 1: *Comparison of IoU of the best hyperparameters found for automated calibration based on DTW: Values show the mean and deviation IoU over 30 different samples. PR: peak detection in reference lamp, PL: peak detection in sample lamp, WN: window normalization, ZP: Zero padding to match resolutions between reference and sample spectra.*

PR	PL	WN	ZP	Windows	$IoU(\mu)$	$IoU(\sigma)$
✗	✗	✓	✓	1620	**0.93**	**0.01**
✗	✗	✗	✗	1039	0.84	0.03
✓	✗	✗	✗	**846**	0.80	0.02
✓	✗	✓	✓	846	0.79	0.02

many possible wavelengths and accounting for noise in the analyzed spectra. This task can be time-consuming and depends heavily on the astronomer's knowledge and ability to identify patterns.

For these reasons, to automate the calibration we use Dynamic Time Warping (DTW) based method, since DTW can adjust for X-axis irregularities and ensure accurate alignment. To handle Y-axis variability we include a normalization step.

Given a gold standard calibration of a lamp, typically performed manually by an expert, we can evaluate the quality of the estimated calibration using the "Intersection over Union" (IoU) metric. The $IoU = Intersection/Union$ is calculated by first computing the $Intersection$ and $Union$ of the wavelength range of both calibrations. If $IoU \approx 1$, the ranges coincide to a greater or total extent, therefore it is likely that the calibration being evaluated is good. If $IoU \approx 0$, the ranges have little or no overlap, implying that the calibration is incorrect.

As shown in Table 1, our method achieves a high degree of overlap in the 30 sample spectra we evaluated. While users may need to slightly adjust the final calibration, this automation can significantly speed up the processing of each plate.

6. Conclusions

In this work, we present innovative software to assist in digitizing, processing, and reducing spectra of stellar objects acquired on glass plates with different instruments and stored at the OALP. The software speeds up the digitization of large glass plate quantities, reduces error rates through signal processing algorithms and computer vision models, and minimizes user involvement. **PlateUNLP** effectively facilitates the rescue and preservation of historical observations of astronomical data and the storage of science spectra on appropriate (FITS) formats for future data analysis.

Future work will focus on validating the software and detection model by digitizing the entire collection of existing plates. This ongoing effort, part of the ReTroH project at the UNLP, is already facilitating the preservation of over 15,000 astronomical plates of significant historical and scientific value. We will also focus on enhancing the interpretability of the software to ensure users can understand the decisions it makes.

Acknowledgements

YJA thanks the financial support from CONICET (PIP 1337), and the Universidad Nacional de La Plata (PID EG001), Argentina.

This project has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Grant Agreement No. 823734.

References

[C*18] CHOLLET F., ET AL.: *Deep learning with Python*, vol. 361. Manning New York, 2018. 2

[Dav24] DAVIDGE T. J.: The Digitization of Photographic Spectra in the Dominion Astrophysical Observatory Plate Collection with Commercial Scanners: A Pilot Study. *Astronomical Journal 167*, 6 (June 2024), 249. `arXiv:2405.01456`, `doi:10.3847/1538-3881/ad38b7`. 2

[MCA*20] MEILÁN N., COLLAZO S., ALESSANDRONI M. R., LÓPEZ DURSO M., PERALTA R. A., AIDELMAN Y., CIDALE L. S., GAMEN R.: Proyecto de digitalización de placas espectrográficas del Observatorio de La Plata. *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina 61B* (July 2020), 251–253. 2

[Mei18] MEILÁN N. S.: *Recuperación del patrimonio observacional histórico*. PhD thesis, Universidad Nacional de La Plata, 2018. 2

[PA23] PONTE AHÓN S. A.: *Digitalización de placas astronómicas antiguas*. PhD thesis, Universidad Nacional de La Plata, 2023. 2

[PPA*23] PEREYRA N., PONTE AHON S., AIDELMAN Y. J., RONCHETTI F., QUIROGA F., GAMEN R., CIDALE L.: Software de procesamiento automático de placas espectrográficas. *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina 64* (Aug. 2023), 332–334. 2

[RDGF15] REDMON J., DIVVALA S. K., GIRSHICK R. B., FARHADI A.: You only look once: Unified, real-time object detection. *CoRR abs/1506.02640* (2015). URL: `http://arxiv.org/abs/1506.02640`, `arXiv:1506.02640`. 3

[Tuv22] TUVIKENE T.: Pyplate repository. `https://github.com/astrotuvi/pyplate`, 2022. Online repository. 2

[WN17] WYATT M., NAVE G.: Evaluation of resolution and periodic errors of a flatbed scanner used for digitizing spectroscopic photographic plates. *56*, 13 (Apr. 2017), 3744. `doi:10.1364/ao.56.003744`. 2