Design of a modular feature extractor for hyperspectral images.

Paula Pacheco[1,2,3], Juan B. Cabral[1,2,3], Sebastián Heredia[1], Magdalena Borda[1,2,4,5], Fernando Gomez[4,5], Pablo Granitto[6]
paulapacheco@mi.unc.edu.ar

[1] Grupo de Innovación y Desarrollo Tecnológico, Gerencia de Vinculación Tecnológica, Centro Espacial Teófilo Tabanera, GVT-CONAE, Argentina
[2] CONICET - Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.
[3] Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba (FAMAF-UNC). Córdoba, Córdoba, Argentina
[4] Centro de Investigaciones en Ciencias de la Tierra (CICTERRA, CONICET-UNC), Córdoba, Argentina
[5] Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba (FCEFyN-UNC), Córdoba, Argentina
[6] Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas (CIFASIS, CONICET–UNR), Rosario, Santa Fé, Argentina.

Abstract. Hyperspectral image sensors capture surface reflectance across a range of wavelengths, which record detailed spectral information in terms of hundreds of bands. The classification of hyperspectral images has generated significant interest among researchers in the remote sensing community, as the large number of bands provides rich spectral information that can be used to classify objects, determine chemical components, or detect vegetation changes; all of which are useful in areas such as agriculture, geology, medicine, etc. However, due to dense sampling, some bands may contain redundant information; and sometimes, spectral information alone may not be sufficient to achieve the desired accuracy in results. Efforts have been made to describe all possible feature extraction mechanisms that add more information to multi-spectral images. Therefore, in this work, we present the design of a tool that implements a uniform and configurable interface for the extraction of spectral and spatial features from hyperspectral images.

Keywords: HSI, feature extraction, modular architecture.

Diseño de un extractor de características modular para imágenes hiperespectrales.

Abstract. Los sensores de imagen hiperespectral capturan la reflectancia superficial en un rango de longitudes de onda, las cuales registran la información espectral detallada en términos de cientos de bandas. La clasificación de imágenes hiperespectrales ha despertado un gran interés entre los investigadores de la comunidad de teledetección, ya que la gran cantidad de bandas proporciona una información espectral rica para ser utilizada para clasificar objetos, determinar componentes químicos o detectar cambios de vegetación; todo esto útil en áreas como la agricultura, la geología, la medicina, etc. Sin embargo, debido al muestreo denso, algunas bandas pueden contener información redundante; y a veces, la información espectral por sí sola puede no ser suficiente para obtener la precisión deseada de los resultados. Se han realizado esfuerzos para describir todos los posibles mecanismos de extracción de características que agreguen más información a las imágenes de múltiples espectros. Consiguiente, en este trabajo presentamos el diseño de una herramienta que implementa una interfaz uniforme y configurable para la extracción de características espectrales y espaciales de imágenes hiperespectrales.

Keywords: HSI, extracción de características, arquitectura modular.

1 Introduction

Hyperspectral images allow for capturing a wide and continuous range of the spectrum, which enables detailed analysis of their properties, considering that their spatial resolution is relatively low, which causes the mixing of materials in pixels and hinders precise characterization. Additionally, they present high correlation and redundancy between spectral bands Li et al., 2024. On the other hand, the availability of this type of images has increased considerably, along with the development of specialized software and increased processing capacity, which has facilitated their use on a larger scale.

These two factors do not allow solving the resolution, correlation, and redundancy problems that we mentioned in various ways, one of them being to extract information, characteristics, and features oriented to each application to achieve their effective use Yin et al., 2012. For this purpose, methods such as PCA, *wavelets*, and machine learning techniques, both traditional and deep, are applied through the use of architectures such as convolutional neural networks, autoencoders, and generative adversarial networks Kumar et al., 2020, as well as more recent approaches like Graph Neural Networks Zhao et al., 2025 and Transformer-based architectures Lu et al., 2023, specifically adapted to this type of data.

In this work, we present the design of HYPPO (Stylized from "Hyperspectral Processing"), a tool for extracting spectral and spatial features from hyperspectral images. HYPPO is based on the modular architecture of the feets feature extractor Cabral et al., 2018, which provides a uniform structure for data access, regardless of the requested features, and reduces processing times through internal computing parallelism compared to independent *scripts* that perform similar tasks. Initially, the objective is to implement the 32 features described in the article "*Feature extraction techniques for hyperspectral image classification: a comprehensive review"* Kumar et al., 2020. These features are organized into three main categories: *spectral methods* including Projection Pursuit (PP), Principal Component Analysis (PCA), Minimum Noise Fraction (MNF), and others such as Nonparametric Weighted Feature Extraction (NWFE) and deep learning approaches using autoencoders and CNNs; *spatial methods* such as Morphological Profiles (MP), Extended Morphological Profiles (EMP), and others like Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP); and *spectral-spatial methods* including 3D Morphological Profiles, Spectral-Spatial Autoencoders (SSAE), Deep Spectral-Spatial CNNs, among others. This comprehensive feature set covers the main categories of feature extraction techniques currently used in hyperspectral image analysis, leaving the possibility that we will incorporate new relevant functionalities in future versions.

2 The design

We can see in Figure 1 the general scheme of the proposed use for HYPPO, where the user interacts with the components: `IO` for input/output of data in common formats in which hyperspectral images are distributed, and `FeatureSpace` configures and provides the possibility to execute the set of available feature extractors. It serves as an intermediary between the user and the complexity of the extraction methods, operating them through its common interface.

HYPPO follows a modular architecture based on established software engineering principles to ensure maintainability, extensibility, and performance. Each feature extractor is implemented as an independent module conforming to a common interface. This enables the integration of new extractors without changes to the core system and ensures a consistent API for end users.

Feature extractors are grouped according to the type of information they process—spectral, spatial, or combined—and can be developed and maintained independently. The `FeatureSpace` orchestrates their execution, handles parameter configuration, and aggregates results, providing a flexible framework that promotes reusability and scalability.

On the other hand, Code 1 presents a sketch based on the functionality offered by feets Cabral et al., 2018 of how Hyppo could be used to orchestrate multiple feature extractors under a single feature space.

Given that HYPPO is currently under development and being implemented in `Python`, various local and distributed parallelization strategies are being evaluated. At the local level, the incorporation of techniques such as multiprocessing, thread-based parallelism, and the use of sub-interpreters is planned. Regarding distributed processing, the use of libraries such as *Dask* Rocklin et al., 2015 is contemplated, which would allow scaling the processing to large volumes of data in clusters or cloud environments.

It should be noted that this library is conceptually based on *feets*, a tool that was recently adapted in a final undergraduate project to operate with *Dask*, obtaining satisfactory results in the parallelization of tasks on light curves.

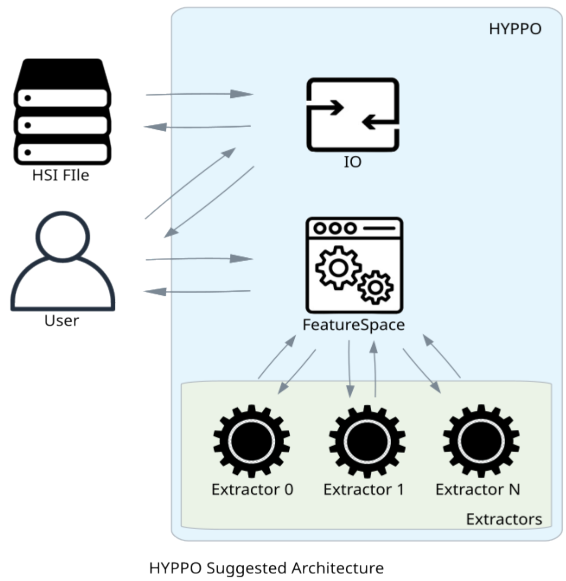

Fig. 1. Suggested architecture of the HYPPO system. The diagram illustrates the main components and their interactions, including data entry via the IO module, feature extraction through the FeatureSpace, and interactions with external services. Its modular design promotes scalability, maintainability, and interoperability across diverse applications.

```python
import hyppo  # Import the module
hsi = hyppo.io.load("path_a_archivo")  # Read the File
# configure the feature space
fs = hyppo.FeatureSpace.from_features(
    ["svm_rfe", "autoencoder", ...],
    c=.5,  # The 'c' parameter of the SVM
    autoencoder={...}  # some parameter for the autoencoder
)
the_features = fs.extract(hsi)  # extract the features
the_features.to_frame()  # convert the features to a DataFrame
```

Code 1: Example code showing the basic workflow of HYPPO: importing the library, loading a hyperspectral image, configuring a feature space with multiple extractors (where '...' indicates additional extractors that can be specified), setting parameters for the extractors (where '...' represents other possible parameter values), extracting features, and converting the results to a DataFrame.

3 Conclusions

We introduced the design of HYPPO, a modular tool for extracting spectral and spatial features from hyperspectral images. The unified and configurable interface aims to simplify, in its first version, the application of the 32 features compiled in Kumar et al., 2020, promoting integration into larger processing pipelines and addressing challenges commonly found in existing tools.

We also plan, in the subsequent releases, the addition of new techniques and performance optimization. These developments aim to enhance the tool's ability to efficiently handle diverse datasets and meet the specific needs of the hyperspectral imaging community.

We hope HYPPO will become a flexible and efficient resource for researchers and practitioners working with hyperspectral data.

Acknowledgments. This work was partially supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). P.P. and M.B. were supported by a fellowship from CONICET and CONAE.

This research employed the Cornell University xxx.arxiv.org repository, the Python programming language, the Numpy and Scipy libraries, and the other packages.

The grammar and spelling in this document have been checked and corrected using AI language assistants including ChatGPT and Claude. While these tools have been used to improve the clarity and correctness of the text, all content, ideas, and technical assertions remain those of the authors.

References

Cabral, J. B., Sánchez, B., Ramos, F., Gurovich, S., Granitto, P. M., & VanderPlas, J. (2018). From fats to feets: Further improvements to an astronomical feature extraction tool based on machine learning. *Astronomy and computing*, *25*, 213–220.

Kumar, B., Dikshit, O., Gupta, A., & Singh, M. K. (2020). Feature extraction for hyperspectral image classification: A review. *International Journal of Remote Sensing*, *41*(16), 6248–6287.

Li, N., Wang, Z., & Cheikh, F. A. (2024). Discriminating spectral–spatial feature extraction for hyperspectral image classification: A review. *Sensors*, *24*(10), 2987.

Lu, W., Wang, X., Sun, L., & Zheng, Y. (2023). Spectral–spatial feature extraction for hyperspectral image classification using enhanced transformer with large-kernel attention. *Remote Sensing*, *16*(1), 67.

Rocklin, M., et al. (2015). Dask: Parallel computation with blocked algorithms and task scheduling. *SciPy*, 126–132.

Yin, J., Wang, Y., & Hu, J. (2012). A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy. *IEEE Transactions on Industrial Informatics*, *8*(4), 935–943.

Zhao, X., Ma, J., Wang, L., Zhang, Z., Ding, Y., & Xiao, X. (2025). A review of hyperspectral image classification based on graph neural networks. *Artificial Intelligence Review*, *58*(6), 172.