Online Tools for Processing and Classifying Multi and Hyperspectral Images: Overview and Opportunities

Yuri K. Fernandes and Ronaldo M. da Costa

February 24, 2021

Online Tools for Processing and Classifying Multi and Hyperspectral Images: overview and opportunities

1st Yuri Kuivjogi Fernandes
Instituto de Informática
Universidade Federal de Goiás (UFG)
Goiânia, Brasil
yurikfernandes@gmail.com

2nd Ronaldo Martins da Costa
Instituto de Informática
Universidade Federal de Goiás (UFG)
Goiânia, Brasil
ronaldocosta@inf.ufg.br

***Abstract*—This work presents a brief review of publications related to web tools focused on processing and classification of spectral images. These are mainly used in the large area of Geographic Information System (GIS) and also Re-mote Sensing (SR). Among the main applications we can mention: classification of hyperspectral images, use of time series for classification and monitoring, detection of anomalies, classification and interpretation of scenes, detection of objects, among others. The research method will be briefly described, as well as the main characteristics of the most relevant articles. The main objective of this article is to define the state of the art and possible research possibilities, as which encompass different configurations of architectures, metrics of performance evaluation and application of Artificial Intelligence (AI) in the classification methods.**

***Index Terms*—WebGIS, remote sensing, classification, processing**

I. Introduction

With the advancement of data processing capacity and the large amount of images available for free, SR applications have been transferred from personal and business computers to the internet environment. Such online tools that deal with data with spatial reference have been called WebGIS.

Initially, the programs developed to deal with GIS, and more specifically with SR, were extremely expensive and, therefore, little accessible. In addition, only specialized people with high-performance computers would be able to work in this area. Therefore, only large organizations were able to benefit from SR applications [1].

Systems embedded on the internet have many advantages over those based on personal or business computers. Those can be accessed from anywhere, they do not depend on the user's processing capacity, they tend to be less polluted with many options for the user, they do not require installation or technical knowledge, among others. With so many advantages, SR applications have become increasingly accessible to non-specialized audiences.

However, despite the great growth in tools for visualizing and monitoring areas via SR using online services, there are still few contributions made regarding the classification of images. According to [2], most articles related to image classification in the Internet environment are unsupervised, since the interaction with the user is less complex.

Within this context, this work aims to present the most current scientific contributions made in the area of image processing and classification using online services. Enabling, through this, to understand the state of the art and identify main opportunities for the next lines of research.

II. Bibliographic review

A. GIS

There are many definitions for GIS in the literature, some of which will be cited and discussed below. According to [3] "GIS is a decision support system involving the integration of spatially referenced data in a problem solving environment." Also, according to [4] GIS is defined as "a tool that stores geographic data, retrieves and combines this data to create new representations of geographic space, provides tools for spatial analysis and performs simulations to help expert users organize their work in many areas. " More recently [5] referred to GIS as "a computer based system to aid in the collection, maintenance, storage, analysis, output and distribution of spatial data and information." It is realized that GIS has always been seen as a tool to aid decision based on spatial data. However, it is observed that more recent articles include in their description the capacity to produce new georeferenced data.

GIS has proven to be a tool with great potential, mainly with the appearance of areas such as BigData. In one of his studies [6] stated in 2003 that it was estimated that 80% of the data had spatial aspect information. However, according to [1] initially GIS was very expensive and also monolithic, that is, an unique and rigid system.

In 1993 there was the first initiative in relation to WebGIS with [7]. He was a pioneer in introduce a tool that display spatial information on an internet site in an interactive way, in addition to being able to visualize and produce information from static files.

In 1994 the Open Geospatial Consortium (OGC) was founded, a worldwide community focused on setting standards for the development and implementation of geospatial content and services, GIS data processing, location services, sensors and data models [8]. From then on, many new tools were made available and pre-established development standards were followed to promote better interaction between online services.

[9] was one of the first to discuss the emergence of WebGIS. In 1998, he presented the factors that made possible the development of this area, since the creation of internet browsers and the HTML and their versions, even solutions for user interfaces such as Java, ActiveX and Scripts.

Another important aspect regarding WebGIS is the configuration of its architecture. There are several ways to define the architecture. [10] named and described three main types of architectures: client-server architecture, service-oriented architecture (SOA) and cloud computing.

B. Remote Sensing

Likewise, there are many definitions for SR, of which we will present some below. According to [11] "remote sensing has been variously defined but basically it is the art or science of telling something about an object without touching it." As the author himself states, this is still a very generic definition.

Due to the large number of definitions, many of which are not widely accepted, [12] compared the history of definitions presented, concluding that a good definition must address the main characteristics: contactless acquisition, use of the electromagnetic spectrum property (REM), use of instruments, fixed or mobile platforms, with treatment of collected data and use of interpretation techniques.

As a more recent definition, we can quote [13], "remote sensing is defined, for our purposes, as the measurement of object properties on the earth's surface using data acquired from aircraft and satellites."

III. Work Development

This work follows a specific research method called Systematic Review, consolidated in the area of Software Engineering by [14]. This method has been shown to be efficient in terms of building a reference base focused on a research topic. In summary, the steps involved are: planning, conducting and reporting.

To carry out this research the tool Parsifal[1] was used. It was developed specifically for this purpose, following protocols established by [14].

A. Planning

During the planning phase, articles classified as relevant to the area were used to carry out exploratory research. This step enabled the definition of several important research items such as: general objective, research questions, keywords, criteria for selecting articles and format of extraction data.

During the reading of these, the following questions were formulated that led the systematic review: **Q1**: what tools have been used for manipulating spatial data in a web environment? **Q2**: Which of these tools use artificial intelligence methods? **Q3**: which databases are used for testing? **Q4**: What metrics are used for performance analysis? **Q5**: what architectures are used? **Q6**: What are the challenges and opportunities in the area?

The literature review was carried out in the databases ACM Digital Library[2], IEEE Digital Library[3], Science Direct[4], Scopus[5] and Springer[6]. Only the period from 2015 to 2021 was addressed and publications of book chapters, incomplete publications and articles in other languages than English were also excluded.

Each research base has defined its own specific syntax for performing searches, logical operators can be used and different types of filters can be applied. The search string used, with minor changes depending on the database, was: "remote sensing" and ("user interface" or "recognition" or "open-source").

The selection criteria for articles used were:

- Approval: Presentation of a web tool for manipulating georeference data.
- Disapproval: **R1**: Articles of the abstract type; **R2**: articles outside the remote sensing area; **R3**: articles that do not have a web tool; **R4**: articles that do not use multi and hyperspectral images.

B. Conduction

In this step, the articles are selected and listed in order of relevance. Based on the criteria presented above, the title and summary fields of each article were read in order to select and classify them. In total, 502 articles were obtained and of these, 47 were selected. Figure 1 shows the distribution of articles in each search base.

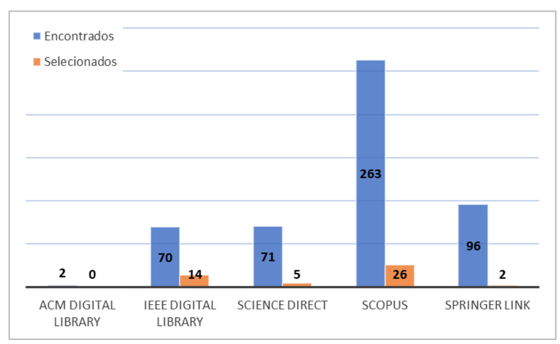

Fig. 1. List of articles obtained and selected in each search base

During the selection, 27 duplicate articles were found using the Parsifal tool, which were excluded. After the selection stage, all articles were again revisited in order to list them for relevance according to the following quality criteria: **1)** Does it have artificial intelligence methods? **2)** Does it describes the architecture? **3)** Does it describe the implementation in detail?

[1]Available at: https://parsif.al/

[2]Available at: https://dl.acm.org/

[3]Available at: https: //ieeexplore.ieee.org/

[4]Available at: https://www.sciencedirect.com/

[5]Available at: https: // www .scopus.com /

[6]Available at: https://link.springer.com/

4) Does it presents programming languages and tools used in the implementation? **5)** Does it have evaluation metrics? The step of ranking the articles is important because it increases the efficiency in the information extraction process. As this is a brief systematic review, only the 4 most relevant articles were considered to carry out the discussions.

IV. GENERAL COMMENTS

In summary, [15] presented a tool for viewing, processing and monitoring high resolution images. The GaoFen-1 and GaoFen-2 satellites were used to provide hyperspectral images. Since high resolution images have hundreds of spectral bands, resolutions in the order of a few meters and dimensions of kilometers, this type of data is commonly included in the BigData area, as quoted by [15].

A few years earlier, [2] developed an online system for supervised classification of Remote Sensing images. In addition, a detailed discussion is presented of which tools and approaches are most appropriate to implement each part of the tool, as well as the most used evaluation metrics for presentation and comparison of results.

Another contribution in this same area was made by [16], who developed an integrated development environment called ”JupyTEP IDE”. This tool can be used for processing Earth Observation (EO) data and is based on the Jupyter notebook environment. This work focuses on making a complete tool available to the user. It has the ability to integrate with several existing web services, execute several common algorithms for processing spatial data and customize the work environment through the inclusion of extensions.

In [17] is presented a platform for viewing and managing maps that address the topic of coverage and land use. The objective of the work is to facilitate the visualization of maps produced by different initiatives in order to enable the comparison between them. The tool encompasses several different map servers allowing for continental scale visualization in the Europe region.

A. Architectures

A feature that should be highlighted in these articles is the use of different types of architectures. As [10] stated, the architecture to be implemented will depend on the type of problem to be solved. Below, the architectures employed in each article are briefly presented.

- [15]:
  - Data processing (offline): Global Mapper software for manipulating and improving images and PostGIS as a spatial data storage service.
  - Infrastructure: workstations and servers for storing remote sensing heavy files, Geoserver as a basemap server, PostgreSQL to manipulate the system database, and Jetty as a web server.
  - Front-end display: based on the Openlayers API, WMS (OGC Map Service) as a dynamic map service, HTML and CSS to create structures, and JavaScript to build the visualization system.
- [2]:
  - Client: data visualization, sample collector for classification, classification tree builder and attribute tree builder.
  - Servlet: module for interaction between web services, capable of transferring messages between services and recording all the actions performed, enabling the correction and reproduction of processes.
  - Interface Service: several online services, such as WMS, for access and standardization of georeferenced data.
  - Service Function: module for data processing. Among them we can mention: segmentation, processing and classification of images, publication in WMS, merging of attributes, data transformation, evaluation of results, among others.
- [16]
  - Client: management of environmental tools, navigation for loading data, map browser through the map service (WMS).
  - Server: GitHub or OAuth with authentication from the client, virtual machines for data processing and management.
  - Data Repository: data storage service, user data storage, external data provider.
- [17]
  - Client: interaction made in the browser itself, including user interface implemented in HTML, JavaScript and CSS3.
  - Mapping library: module based on the Leaflet library for building maps, implemented in the JavaScript language responsible for processing and presenting maps and vectors.
  - Parallel component server: based on Geoserver to provide dynamic maps to the tool.
  - Data storage: GeoTIFF to store matrix data (images) and PostGIS for vector data (geometric shapes).

For data visualization tools, there is no need for large information processing, so there is the possibility of implementing an architecture called ”thin client architecture” [10]. This type of architecture does not include an intermediate service, that is, the user inputs on the interface are forwarded directly to a GIS server that will process the data and return the results. The biggest advantage of this system is its lightness and simplicity of implementation, besides being a cheap option. However, it is recommended only for services that will require little processing as this will take place on the GIS server, if very complex instructions are passed, it can compromise the system's response time.

Another aspect that can be taken into account is that despite having complex architectures, the recent artifice of cloud computing was not found in any of these articles. According to [18] the processing of large volumes of SR data requires unprecedented hardware configurations, which generally exceed the capabilities of conventional computers. One of the great advantages of this innovation is the customization of virtual hardware enabling the user to have access to large computing infrastructures through the cloud. The biggest disadvantage is the complexity of integrating many different services.

B. Artificial Intelligence Methods

Of the articles cited, only [2] implementation of artificial intelligence methods. The author decided to use the supervised classification method called K nearest neighbor (KNN). This is one of the most simplistic methods when it comes to supervised classification. In summary, the KNN routine is based on the statement that the closest data is more similar when compared to more distant data. In general, instructions are executed, in a cyclic way, to classify each data in classes stipulated by the user. After each classification, some parameters are modified and the data is reclassified. The method ends when it reaches the stop criterion, that is, when no data changes classes during the next executions.

In [16] no artificial intelligence method is mentioned, however, as the system was mostly implemented in Python and it has great customization capacity through extension additions, it appears that there will be no difficulties in implementing AI methods.

In a revision work [18] he highlighted the gain in importance of machine processing techniques in classification problems in the SR. By reading this, it is noticed that the methods of employees are from commercial initiatives or offline open source software.

C. Evaluation metrics

Both in the articles selected for this work, as in other articles read for theoretical foundation and exploratory analysis, it was noticed that there is a very small number of studies that present methods for evaluating tools. Of the four articles previously discussed, only [2] assesses the implemented tool. The evaluation metric used was processing time between the author's solution, an offline system and "eCognition"[7]. Other metrics can be used, mainly in classification problems such as accuracy, Kappa coefficient, F1 measure, among others, as mentioned by the author.

V. Conclusions

It is well known that there is a great growth in online services in the SR area. Such process of migration of systems to the internet has been observed in all areas. The rapid progress in the data processing capacity stands out as the cause of this process. For example, the advent of cloud processing and the large number of online services optimized and standardized to be used together.

Users who were previously tied to the capacity of their personal computers can now enjoy high-performance computing infrastructures. In addition, web services also have the ability to be applied within multiple platforms, such as servers, computers and smaller single board systems or mobile devices.

[7]Commercial supervised classification software.

Another migration factor is the immense amounts of hyper and multi-spectral images generated. One of the characteristics of the most current satellites is the smaller revisiting interval (temporal resolution), that is, the images that were previously captured every 15 days today are generated every day. As an example, we have the Sentinel-2 satellite that provides images of daily from all over the globe.

By reading the articles it is clear that there is a great deal of space to be explored regarding the WebGIS tools. There are many possibilities for services available and the number of applications is immeasurable. However, there is still a small amount of work that proposes to do it completely.

In general, the tools have many limitations in view of the great advances witnessed in the era of large databases. There are a huge number of AI methods available, which can be used far exceeding the results generated by conventional methods.

Another characteristic observed is the lack of comparison with related works. This may be due to the recent emergence of the area, just a few decades, or the great diversity of applications. However metrics such as processing time and accuracy, mean square errors, among others, can be compared with other conventional solutions within the work itself. The existence of image databases created specifically for performance tests may be another possibility for analysis and comparison of results in order to prove the scientific contribution of the works.

References

[1] M. F. Goodchild, Geographical information science. International journal of geographical information systems, v. 6, n. 1, p. 31-45, 1992.

[2] Z. SUN, H. Fang, L. Di, P. Yue, X. Tan & Y. Bai. Developing a web-based system for supervised classification of remote sensing images. GeoInformatica, v. 20, n. 4, p. 629-649, 2016.

[3] D. J. Cowen. GIS versus CAD versus DBMS: What area the differences. Introductory Readings in Geographic Information Systems, p. 52-61.

[4] P. Rigaux, M. Scholl & A. Voisard. Spatial databases: with application to GIS. Elsevier, 2001.

[5] P. Bolstad. GIS Fundamentals: A First text on Geographic Information Systems . St. White Bear. 2012.

[6] B. Klinkenberg. The true cost of spatial data in Canada. Canadian Geographer/Le Géographe canadien, v. 47, n. 1, p. 37-49, 2003.

[7] S. Putz. Interactive information services using World-Wide Web hypertext. Computer Networks and ISDN Systems, v. 27, n. 2, p. 273-280, 1994.

[8] E. Van Rees. Open geospatial consortium (OGC). GeoInformatics, v. 16, n. 8, p. 28, 2013.

[9] A. Hardie. The development and present state of Web-GIS. Cartography, v. 27, n. 2, p. 11-26, 1998.

[10] S. Agrawal & R. D. Gupta. Web GIS and its architecture: a review. Arabian Journal of Geosciences, v. 10, n. 23, p. 1-13, 2017.

[11] W. A. Fischer, W. R. Hemphill & A. Kover. Progress in remote sensing (1972–1976). Photogrammetria, v. 32, n. 2, p. 33-72, 1976.

[12] J. Fussell, D. Rundquist & J. A. Harrington. On defining remote sensing. Photogrammetric Engineering and Remote Sensing, v. 52, n. 9, p. 1507-1511, 1986.

[13] R. A. Schowengerdt. Remote sensing: models and methods for image processing. Elsevier, 2006.

[14] B. Kitchenham. Procedures for performing systematic reviews. Keele, UK, Keele University, v. 33, n. 2004, p. 1-26, 2004.

[15] H. A. N. Peng, L. I. Yan, W. A. N. G. Ying, L. I. Linqi, M. A. Haotian & W. A. N. G. Jinkuan. Designing, Analysis and Implementation of Open-source Visualization System for High-Definition Remote Sensing Data Processing. In: 2020 Chinese Control And Decision Conference (CCDC). IEEE, 2020. p. 5011-5015.

[16] J. Rapinski, M. Bednarczyk, D. Zinkiewicz. JupyTEP IDE as an online tool for earth observation data processing. Remote Sensing, v. 11, n. 17, p. 1973, 2019.

[17] M. A. Brovelli, F. C. Fahl, M. Minghini & M. E. Molinari. Land user and land cover maps of Europe: a WebGIS platform. 2016.
[18] L. Wang, Y. Ma, J. Yan, V. Chang & A. Y. Zomaya. pipsCloud: High performance cloud computing for remote sensing big data management and processing. Future Generation Computer Systems, v. 78, p. 353-368, 2018.
[19] A. E. Maxwell, T. A. Warner & F. Fang. Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, v. 39, n. 9, p. 2784-2817, 2018.