Novel Approach for Generalizing the Process Chain of Optical Satellite Images Based on Knowledge Capitalisation

Razafinimaro, A.,\textsuperscript{1,2,3*} Hajalalaina, A. R.,\textsuperscript{1,2,3} Rakotonirainy, H.,\textsuperscript{1,2,3} Delaitre, E.,\textsuperscript{4} Zafimarina, R.\textsuperscript{5} and Libourel, T.\textsuperscript{6}
\textsuperscript{1}Department of computer science, Ecole de Management et d’Innovation Technologique (EMIT), University of Fianarantsoa, Madagascar, E-mail: tokyarisetra@gmail.com
\textsuperscript{2}Laboratoire d’Informatique et des Mathématiques Appliquées au Développement (LIMAD), Fianarantsoa, Madagascar
\textsuperscript{3}Informatique- Géomatique et Mathématiques Appliquées (IGMA), Fianarantsoa, Madagascar
\textsuperscript{4}Research Institute for Development, Montpellier, France
\textsuperscript{5}Higher Polytechnic School of Antananarivo, University of Antananarivo, Madagascar
\textsuperscript{6}LIRMM, University of Montpellier, Montpellier, France
\textsuperscript{*Corresponding Author}

DOI: https://doi.org/10.52939/ijg.v18i6.2463

Abstract

Optical remote sensing from satellites such as SPOT, Landsat, Sentinel2, Terra Modis, offers exceptional solutions for monitoring the earth’s surface. Indeed, these satellites deliver multi-source, multi-date images. These are Bigdata, voluminous, heterogeneous and their treatment are not only complex but also require a lot of time. Studies have been carried out on the formalization of experimental protocols in satellite image processing. Those propose automated processing chains. But, these solutions are specific and their reuse are not possible for other cases. Our reflection allows us to propose a new approach for generalizing the optical satellite image processing chains. This approach respects standards in knowledge capitalization. The objective is to ensure the interoperability to the use of optical satellite images. Metamodels for the generalization of processing chains, a method of execution, generalization rules and implementation prototypes have been developed. These results are obtained through theoretical analyses and also experiments on the processing of multidate, multisource optical satellite images applied in multisectoral domains. Thus, this research solves today’s interrogability problem in the processing of satellite images optical.

Keywords: generalization, interoperability, knowledge capitalisation, optical satellite images, process chain

1. Introduction

Life and environmental sciences (biology, risks, remote sensing, etc.) have accumulated observation data over many years and developed a wide variety of processing methods. For the last two decades, remote sensing has been strongly mobilized to monitor changes in land cover caused by natural or unnatural phenomena [1] [2] and [3]. In particular, optical satellites such as Sentinel-2, SPOT, Landsat, Terra Modis, have widely opened the way to research for scientists since these satellites provide multi-temporal series of images [4] and [5]. These satellite images contain pertinent informations about the processes that affect the earth’s surface, that’s why their use provides a great deal of knowledge in various research areas. The conservation of natural resources such as natural forest, silting, land use are examples that require their uses. Then, Satellites images processing are not only complex but also require a lot of times because the images delivered by these satellites are noisy, voluminous and heterogeneous. Moreover, analyses often need multi-source and multi-temporal satellite images in order to monitor changes over time. Consequently, Those generate increasing demands of formalization experimental protocols for the process satellite image. This allows to obtain processing chains and to make easy the exploitation data whatever the computer environment used.
In addition, it allows the capitalization of experiences for reuse in future research. In the formalization field, we have done works on the study of forest dynamics such as Hajalalaina et al., [6], Razafinimaro et al., [7] with the goal of making experimental protocols shareable, reusable and capitalizable across distributed systems.

If it is undeniable that the formalization of processing chains makes easy the process satellite images, through process automation. However, the formalization produces a particular solution specific to a single case of satellite image corresponding to the study realised. Consequently, this is not enough to answer the question of interoperability of processing chains, to ensure their reuse on different cases of study and future images for phenomena analysis. Hence, the research question is: How can we generalize cases of optical satellite images processing chain to make them interoperable and reusable? This in order to improve the processing of optical satellite images applied to the monitoring of change earth surface. Moreover, this interoperability is one of today’s challenges in the field of geomatics [8]. Thus, we propose an approach to generalize optical satellite image processing chains with varieties of satellite images from different sensors. So, our method is based on the creation of metamodels and generalizations rules. This approach respects standards in knowledge capitalization. These metamodel will use the satellite image processing generic model proposed by Hajalalaina et al., [9]. Also, we use the execution principle of the abstraction and concretization of processing chains of Lin [10]. In addition, the reliability of the result will be ensured by the machine learning algorithms [11] [12] and [13]. These are efficiency in terms of time optimization and accuracy of the results than the traditional techniques [14] and [15].

To present this research, we propose the following structure. In the second section, we will overview the related work on the issue of formalization of satellite image processing chains such as knowledge capitalization and experiments on satellite image processing. In the third section, we will discuss the characteristics of the formalization approach that we want to propose. In the fourth section we will propose the abstraction process chain method, we talk about our proposition for concretising and implementing the abstract process chain and we will discuss the results obtained by taking the characteristics of the approach. Finally, we will conclude the work carried out as well as propose the possible perspectives of this research.

2. Related Works
In this section, we will briefly discuss related work in the field of formalization of satellite image processing chains, experimental work and the state of the art in satellite image processing as well as the generalization in question. Indeed, these works concern the proposal that we want to put in place in terms of capitalization of knowledge and experience particularly on optical satellite images.

Within the framework of the formalization of a chain of treatments, the work focuses on the implementation of the organizations for the data, the treatments and the human resources [10]. Such an organization provides the automated chain by taking into the account characteristic of the data; the nature of each treatment (input parameter, the type of return), the sources of data as and the objective fixed during the treatments. Of course, it recommends the collaboration between domain experts and computer scientists. Therefore, this research work is part of the capitalization of knowledge and experience. In addition, methods can be applied in terms of formalization but the most common used are based on ontologies [16] and considering the concept of working context [10].

The latter brings more on the formalization of satellite image processing chain because of the simultaneous use of two formalisms: SWM and UML [17]. To this effect, the obtained models are legible and easy to understand. The scientific works of Hajalalaina et al., [6] [9] and Razafinimaro et al., [7] include some case studies on the formalization of satellite image processing chain SPOT and Landsat employing the concept of the work context. Hence, the automations worked well through of the concretization of processing chains and their implementations on the Orfeo ToolBox library and the Zoo-Project WPS web service. However, they have dealt with only a few sensors’ cases of these satellites. In addition, each sensor has its own characteristic. The consideration of the other cases deserves further research in order to make their satellite images usable.

The term generalization is often used in the world scientific research. Generalization is a cognitive process that consists of abstracting a set of concepts or objects by neglecting the details so that they can be considered in a comparable way. It can be obtained through various forms of analysis. In general, it includes experimental work on the study of various cases [18], general theoretical studies [19] and [20] in order to propose models that take into account the different cases.
Generalization work can be initiated from the case studies to find their common characteristics. That is to say, it is the experimental work that allows the verification of hypotheses and from which the synthesis is deduced like a solution.

The generalization can also be carried out using a theoretical study such as the state of the art, in order to analyze the various underlying cases. Intuitively, the confrontation of the characteristics also produces to the synthesis of compromises between the cases. In most cases, the product of this generalization work in computer science is restituted in the form of a diagram or a model illustrating the knowledge in a general way. The work of Hajalalaina et al., [6] proposed a generic model of satellite image processing. Similarly, The paper of [21] performed a classification of concepts related to change detection algorithms while Boussaid et al., [22] developed an ontology for satellite image classification. These various approaches are the manufacturing of a generalization of processing chain’s necessary parts, the fact is that these works are part of the particular studies used to highlight the common characteristics of satellite images.

3. Descriptions of Generalization

3.1 Characteristics

First, we consider a new approach to the generalization of satellite image processing chains for optical remote sensing, and for which we take into account four very important points. First, interoperability is the main objective of this research and will ensure the reusability of the processing chains to be designed. Then, the genericity takes into consideration the common characteristics of all the optical satellite images in order to propose the generic model. Then, this generalization will also verify the reliability of the results by proposing algorithms suitable for the input images to be able to give a higher accuracy and a lower processing time. Finally, this generalization ensures the capitalization of the initial knowledge and those resulting from experiments on the processing of multivariate and multisource optical satellite images.

3.2 Choice of Methods

As we mentioned in the introduction section, this generalization work enhances the proposal made by Hajalalaina et al.,[6] in terms of methodology on multitemporal and multisource satellite image processing. Indeed, he proposed a generic model containing three phases of satellite image processing following the analyses on the earth observation, namely pre processing, intermediate processing and final processing. It seems that this methodology is commonly used in multitemporal and multisource satellite image processing. Many research works have confirmed the efficiency of this method in applications on the monitoring of evolutions generated by a natural or not phenomenon on the earth surface. Let us quote the example of the evolution of the forest cover which requires this method to realize the change in a time interval [23]. Figure 1 illustrates us this generic model for the treatment of satellite images of observation.

From this generic model (Figure 2), we consider the design of a metamodel that addresses not only interoperability across optical satellite image types (Sentinel-2, SPOT, Landsat and Terra Modis), but also the extraction of knowledge on the accuracy of the results and the optimization of processing to produce reliable and efficient results. It should be noted that we take into account the ISO19115 and ISO19119 standards which are the two frameworks allowing the description of resource metadata such as processing. The proposal of these metamodels must respect them so that it can be capitalized and reused. In addition, the UML formalism is to be used during modeling. It is not only practical for the presentation of knowledge but also recommended by the notion of knowledge capitalization. To do this, this modeling requires prior very thorough investigations. These include experiments on satellite image processing techniques, theoretical analyses that will lead to generalization. Similarly, we must take into account the existing satellite image processing tools and the implementation of each treatment in order to propose generalized concepts.

Figure 1: Generic model [6]
Moreover, its use will have to be conditioned by the type of satellite image to be used, the established experiments and the criteria taken into account such as the suppliers of satellite images, the level of correction, the algorithms adapted to the images etc. The latter define the generalization rules in the form of conditions in order to propose a generalized and reliable processing chain for each case exposed. Also, adopting the concepts proposed by Lin [10] on the existence of abstraction stage and concretization of processing chain, which would make it easy for users to implement this generalization in which they will fill in only some information about the input images. Then, the Generalization simulates them in order to establish the abstraction, and generates the abstracted processing chain in the form of an easily exploitable file such as XML to ensure the following steps as well as its reuse in the next processing. This is recommended by the ISO19139 standard. Then, it proceeds to the concretization, the implementation of the chain of treatments by using the programming language (R, Python, Matlab,...), the libraries, the requested parameters. Finally, it launches the treatments. The implementation method is illustrated in Figure 2.

4. Results

In this part, some results of our research will be presented successively. It includes meta-model proposals for optical satellite image processing that it is presented in Figure 3, generalization rules showed by Figure 4 and the mode of access to services.

4.1 Abstraction the Process Chains for the Generalization

First, preprocessing is a primordial operation phase in an analysis of multisource and multispectral optical satellite images. Indeed, it is composed of steps depending on factors related to the images needed, such as the level of correction for example. Therefore, we propose a concept of pre-processing based on three categories:

- Visual analysis: it deals with the spectral and/or spatial aspects of the image. Thus, visual enhancements are generally adopted because of bad appearances of the image. Let’s take for example the presence of clouds in an image which consequently requires filtering.
- Area delimitation: it depends on the size of the study area in two cases. Firstly, if the study area is included in a tile, then cutting is necessary. On the other hand, tile mosaicking should be used in the case of a very large area.
- Corrections: According to the confrontation of theories on the processing of optical satellite images, some corrections can be applied and their using depend on several parameters. In general, we should take into account the level of correction acquired by the image and their providing, the radiometric or geomatic problems of the image to be used and the analysis to be conducted such as a multi-temporal analysis. These can be resolved by radiometric or geometric correction.
Then, the intermediate phase consists in classifying the optical satellite image resulting from the preprocessing phase. In this context, we consider the use of techniques proposed by artificial intelligence, more precisely machine learning, in view of their reliability not only in terms of accuracy of results but also in the optimization of processing. Moreover, these efficiencies are confirmed by research works such as Aaron et al., [14] and Phan and Martin [15], in which they stated that these algorithms are very effective compared to traditional techniques. These are: supervised, unsupervised and semisupervised learning. In addition, the index calculation is also a promising category in this phase for the simple reason that it can make the classification in question easy. Thus, our proposal on the concept of metamodel in this phase should be done around these techniques. Finally, The phase of final processing gives the final result of the treatments such as, for example, the change map. Its implementation could include a cleaning (optional) as well as the detection of change by specifying that the inputs must consist of at least two classified images of an area but at two different dates.

4.2 Generalization Rules
The generalization rule is the first responsible that controls the use of our metamodel. In fact, it is formed by algorithms using more condition. That have the objective of building the method adapted to the input images. On the one hand, this last one is obtained during the analyses of the technical cards of images but also the experiments carried out. On the other hand, it works from the parameters corresponding to the images in entry. The Lin [10] Simple Workflow Model formalism is used to present the mechanism of these generalization rules.

4.3 Proposal Concretization and Implementation of Generalization
The concretization and the execution of this generalization require main components:

- First, the concretization and implementation of processing chains rely on skills particularly in Computer Science. Indeed, these recommend knowledge to the tools and libraries of treatment more precisely on the treatments of satellite images. Then, the work focuses on the reading of abstract chains and then convert them into concrete chains well implemented. The products are then processing strings ready to be executed. For this purpose, we propose a way of concretization that is presented in Figure 5. It combines an extraction of the attributes of the processing strings and transforming them into queries, a database containing information about the processing tools and libraries, and then a storage medium (table) on the fly.

- The planning of the execution of the services constitutes the launching functionalities of the processing chain. Indeed, its realization is conditioned by the presence of three very important points. It includes the concrete chains, the execution planning system and the service of treatments. Their correspondences absolutely ensure the launching of each treatment and produce the final result. Figure 5 shows the relationship between these points.

- The processing department provides the functionalities capable of performing the operations. Indeed, it can be composed of files, modules, functions. These do not have to have a structure of chains. However, the planning of execution orders them as the concrete chains of treatments. In addition, the processing department must be made up of multi-language modules. That is, it must accept any language of implementation of the modules (R, Python, Java, Matlab, C++, etc.). In this case, the already developed modules can be used. On the other hand, we also suggest the use of standards on access to these modules. For example, we can use Java and R interfaces such as rJava, JRI. In web service mode, we can implement the WPS or Web Processing Service.

- The Generalization is designed for the needs of users who focus on optical satellite image processing. It includes people working extensively in the forestry field. Certainly, the type of processing automation is included. However, user interaction is essential to get it started. Consequently, we have distinguished two interactions, namely the interaction in the phase of abstraction of the processing chains and the interaction in the phase of concretization and implementation of the chains. Figure 6 represents the interface for the abstraction and the concretisation phase in our prototype.
Figure 3: Metamodel for processing chain of optical satellite images
Figure 4: Some generalization rules formalized in SWM
Figure 5: Concretization and implementation of chains (a) and Execution planning of processing chains (b)

Figure 6: Interface for the abstraction and the concretisation phase
4.4 Discussions

Genericity: When we generalize a system, it seems that we have to understand the common characteristics of the objects or also their genericity. Indeed, the modeling of concepts using the UML diagram is similar to the formation of a decision tree. The latter is formed by nodes and leaves, on which a node is considered as a category of treatments, while the leaves correspond to the techniques that can perform the treatments. During the design of our metamodel, we have chosen not only the most used techniques on the processing of optical satellite images but also the techniques confirmed as reliable by experimental works and scientific articles. This is the reason why the other techniques are not represented on our metamodel. Similarly, we have not mentioned deep learning which is considerably independent since it has almost techniques for each phase. We could thus consider a generalization dedicated typically to their case. Moreover, the algorithms themselves are not illustrated on this metamodel because of their variations. Intuitively, these variations make the organization of this metamodel and its readability very difficult.

Reliability: After modeling the concepts of these three proposed phases, we found that the pre-processing phase focuses heavily on improving the data sources needed for the studies. Indeed, this step is essential since the reliability of the results of the following steps depends a lot on this source. Then, the intermediate phase deals a lot with the reliability of the results in terms of accuracy and the optimization of the processing results due to the use of promising machine learning techniques. Then, the final processing phase engages in the final production of the result in the presence of their techniques. All these steps are therefore very important in a study to be conducted especially in the case of multidate and multisource analysis. Moreover, the effectiveness of the generalization rules in the form of conditions intervenes on the one hand in the implementation of the established metamodel but on the other hand in the reliability of the abstraction of the processing chain so that the proposed generalization is reliable during the processing. In addition, the use of hybrid techniques also brings new developments on the mode of satellite image processing, namely a higher accuracy and reduced processing time.

Capitalization: The generalization that we have proposed gives rise closely to the capitalization of knowledge and experience. First, the capitalization of knowledge allowed us to build the metamodel for the generalization of optical satellite image processing. Then, it helped us to design the generalization rules especially in the preprocessing phase. In addition, the capitalization of the experiments ensured the reliability of the results. For example, in the intermediate phase, thanks to our experiences on the classification of optical satellite images, particularly on land use, using machine learning, we were able to establish the generalization rules defining the algorithms to be used according to the spatial resolution, the number of special bands of the image to be classified. In any case, these two types of capitalization have formed a single efficient system for this generalization of optical satellite image processing chain.

Interoperability: This generalization is mainly designed to answer the question of the reusability of the treatments on Sentinel 2, Landsat, SPOT and Terra Modis optical satellite images, multi-dates and multi-sources. Since, we have taken into consideration its persistence regardless of the combination in between them. Therefore, interoperability is obviously assured on these images. Moreover, this characteristic exceeds the limit of a processing chain formalization, which proposes a specific solution only for the cases of study conducted. That is, each case should have an appropriate formalization. On the one hand, this situation recommends more time studies of the experimental protocols. And on the other hand, its implementation remains a very expensive task. Thus, the Generalization approach brings more to the optical satellite image processing mode thanks to this interoperability.

Implementation: the implemented processing chain allows users to execute the source code whatever the chosen environment. Then, the automation is included not only with the aim of making easy and fast the chained treatments but especially to decrease the intervention of the specialists in the field. In addition, it should be noted that Python, R, Matlab are very practical languages in the framework of the implementation of satellite image processing. Consequently, it is preferable to propose these options in the Generalization. Moreover, this code generation aims at sharing experiences through distributed systems such as grid computing, cloud
computing, cluster, web service, etc. Since the obtained strings are already implemented in a programming language, it is only necessary to comply with the standards of these sharing means such as the OGC standards on WPS.

5. Conclusion
In conclusion, satellite image processing is essential for monitoring our planet earth. In this paper, we have focused on the problem of multi-temporal and multi-source optical satellite image processing which is heavy, time consuming, difficult because of the volumes and heterogeneity of the required data. We have proposed a new approach to the generalization of processing chains adapted to optical satellite images Landsat, Sentinel2, SPOT and Terra Modis. This approach aims to establish, in addition to interoperability, reliability in terms of accuracy of results and optimization using machine learning, the capitalizability of knowledge and experience as well as the genericity of satellite image processing chains. Then, the generalization metamodel, rules of generalization have been obtained following a huge theoretical analysis and big experiments on optical satellite image processing. Thus, we can say that this new Generalization approach is able to overcome the limitation of specific formalization work at the level of interoperability.

Some perspectives have been considered through this research work. First, the analysis of the concepts of radar satellite images is a logical continuation of this research in order to design their generalization. Second, due to the more intelligent property that the Multi-Agent System [24] has, it seems that its application in this generalization is considerably possible to reinforce the interoperability. At the moment, the use of deep learning is carried out on voluminous data as satellite image processing. Moreover, it proposes architectures that can be adapted to each processing phase that we have modeled. A generalization based on deep learning is thus considerably interesting for satellite image processing. Moreover, the integration of Model Driven Engineering would allow to create a dedicated language facilitating the automatic generation of the satellite image processing chain. Also, the generalization using parallel and distributed architecture like grid computing [25] [26], cloud computing [27], cluster [28] would be interesting in the exploitation of large data.

References


