Управление разработкой программных комплексов и вычислительными процессами. Интеллектуальные информационные системы

UDC 004
https://doi.org/10.36906/AP-2020/25

TECHNOLOGY FOR DEVELOPING A PROTOTYPE OF AN INFORMATION SYSTEM FOR MONITORING REMOTE SENSING DATA FOR THE ARCTIC REGION

Bizyukin M.
Admiral Makarov State University
of Maritime and Inland Shipping
Saint-Petersburg, Russia
Abrahamyan G.
Doctor of Pedagogy, Professor
Admiral Makarov State University of Maritime and Inland Shipping,
Herzen State Pedagogical University of Russia
Saint-Petersburg, Russia

Abstract. The work is aimed at developing a model of an information system for the analysis and monitoring of remote sensing data by the example of processing hyper- and multispectral satellite images, which are widely used to analyze the state of static and dynamic objects in the Arctic region of the Russian Federation. For automatic analysis and decryption of Arctic data in the development of the model, methods of high-performance computing, radiometric calibration, filtering and clustering of images, as well as intelligent data processing methods using deep learning convolutional neural networks were used. Object-oriented design and united modeling language notation were used to develop the model. A data-level model, a conceptual model of the structure of system modules, including a resource storage center, a resource and results management center, and a presentation-level interface have been developed. To develop a diagram of the use cases of the information system, the structure of actors, use cases and their interrelations were identified. The logical model of the information system was created based on a class diagram consisting of the Resource and Results Manager Center, Intellectual Information System, Functional Neural Modules packages. The practical significance of the study is due to the fact that the results obtained will allow the development of a prototype of an information system that can be used for effective monitoring of “useful data" of the Arctic region of the Russian Federation, as well as to automate the processes of analysis, updating, storage and processing of data from objects in various areas of the Arctic infrastructure.

Keywords: technology for developing, a prototype of an information system, monitoring remote, sensing data, arctic region.

Hyperspectral satellite data are widely used in the analysis of atmospheric and ionospheric processes, the analysis of climatic changes, as well as in various monitoring systems (the state and dynamics of the Earth's forest resources, vegetation and ice covers, the state of soils, dumps of industrial products, fire regimes of forest ecosystems, marine hydrophysical processes, etc.) [1, p. 124826]. At the same time, it becomes necessary to develop qualitative methods for processing hyperspectral satellite images with high spatial resolution. An important place in such processing methods is occupied by the so-called atmospheric correction - the elimination of the effects on the results of image detection of effects arising from multiple scattering of light in the earth's atmosphere and its absorption by atmospheric gases and aerosol. The processing of satellite images in the Arctic region makes it possible to identify many important characteristics of various earth surfaces, such as: the properties of the mineral composition of open rocks, the prevailing species composition of the vegetation cover (and the degree of its disturbance), the composition, condition and moisture of the soil, the location of territories with forest fires, burns, illegal logging, condition and thickness of ice massifs. In addition, the processing of satellite Arctic data makes it possible to determine the gas composition and the degree of aerosol pollution of the atmosphere, create maps of the temperature of the air, the earth's surface and the surface of water bodies, and track the movements of ships in the Arctic region of the Russian Federation.

Difficulties in solving the problems of analyzing hyperspectral Arctic data are associated with three circumstances:

1) high dimensionality of the measured data to be processed,

2) a large number of spectral channels,

3) complex (highly variable) behavior of spectral curves [2, p. 45; 3, p. 220].

It should be noted that modern trends in the development of computational algorithms and approaches to the organization of information systems for processing Arctic Big Data allow solving a number of important problems associated with the difficulties of reorganizing the structural approach of the computing process within the framework of the digitalization trends of the Arctic infrastructure, taking into account Decree No. 164 "On the Fundamentals of State Policy of the Russian Federation in the Arctic for the Period up to 2035", adopted on March 5, 2020.

Analysis of the current state of research in the field of Earth remote sensing (ERS) data processing in the Arctic region shows that the development of ERS data processing technologies is moving towards the development of higher-level, intelligent methods, offering the end user solutions that allow, to a large extent, to automate the Arctic data when solving applied problems. Recently, researchers and developers are increasingly resorting to high-performance computing technologies, due to the growing need to extract “useful” Arctic data from large data structures and data sets, the size of which is growing every year [4, p.102189; 5, p. 381].

In this regard, researchers and software developers are currently looking for ways to improve the efficiency of remote sensing data processing [5, p. 381]. The methods and algorithms developed by the authors are based on the most modern and relevant solutions and make it possible to transfer the mathematical methods and algorithms incorporated in the procedures for processing remote sensing data to the paradigm of high-performance computing, for this they are used as parallel programming technologies for multicore and / or multiprocessor (cluster) systems, and distributed computing, cloud and General-purpose computing on graphics processing units (GPGPU) technologies.

Modern trends in the processing of Arctic Earth remote sensing data require the development of new effective integrated solutions, including the design, development and use of intelligent information systems. Designing and developing effective methods and computational architectures that transform large volumes of Arctic Earth remote sensing data into information required by customers is of great importance. The growth in the volume of Arctic remote sensing data is constantly growing all over the world, while Russian and international organizations and users require more and more efficient systems for the exchange of these data and resources. Active research is being carried out in the field of application of methods and systems for high-performance distributed computing of Arctic data [6, p. 100236].

The authors propose an integrated approach to the development of an information system model, which will solve a number of problems arising in connection with the processes of digitalization of the Arctic infrastructure. The use of modern technologies for processing, storing and managing big data allows not only to reduce the time spent on solving applied problems associated with processing large amounts of data, but also opens up the possibility of using more resource-intensive algorithms, which ultimately improves the quality of algorithmic processing of these objects.

Materials and methods

For processing large arrays of hyperspectral satellite, data, along with a developed arsenal of traditional computational methods, neural network intelligent approaches using convolutional neural networks (CNN) and deep learning (DL) are increasingly being used. Analysis of hyperspectral satellite data by deep convolutional networks is an important new subarea of neural network approaches [6, p.100236].

One of the key features of hyperspectral images is their increased graininess due to the spectrum noise. During visual interpretation of hyperspectral images, the human brain identifies the outlines of individual objects, despite their internal brightness heterogeneity, in contrast to algorithmic methods of automatic processing. Of great importance for thematic analysis of hyperspectral data is their preliminary processing, since the original images are difficult to interpret. To obtain an image suitable for automatic analysis and decryption, it is necessary to carry out procedures, among which the most important are: 1) radiometric calibration (transition from brightness values to physical characteristics of backscattering - specific effective scattering area (SESA), 2) filtering, 3) clustering. [7, p. 111919]

The use of deep learning conventional networks (DCNN) in hyperspectral remote sensing problems allows extracting important additional information of a global nature from Arctic data. The additional use of multilayer auto-encoder networks allows data compression in combination with the inclusion of joint spatial-spectral information in images, and also allows obtaining a more accurate solution of inverse problems of remote sensing using hyperspectral satellite Arctic data, solving problems of classification of hyperspectral images, problems of pattern recognition and detection objects in images. Deep learning theory supplements conventional machine learning methods with special algorithms that provide the ability to analyze input information at multiple levels of representation. Thus, deep learning contributes to a more comprehensive study of objects in the Arctic region, revealing hidden relationships and correlations of features between objects. This turns out to be especially important in situations when, at the initial stages of processing Arctic data, sufficient adequate information levels of data presentation were not found.

Thus, the application of DCNN methods in the processing of large hyperspectral data is one of the promising research methods in the field of remote sensing of the surface of the Arctic region. At the same time, the use of DCNN methods allows solving the problems associated with the analysis of large high-resolution hyperspectral images.

In parallel with the technologies of intelligent neural network processing of remote sensing data, modern methods of designing the architecture of systems focused on working with a large amount of data - global remote sensing databases — are actively developing. These methods make it possible to distribute the computational load, reduce the processing time of the task, and organize storage systems for Arctic data.

At present, distributed information and computing system based on the Apache Hadoop project is used to process Earth remote sensing (ERS) data using the example of hyper- and multispectral satellite images in the research process. A distinctive feature of the system is the high speed of operation of standard algorithms for processing remote sensing data, integrated into the environment of massively parallel execution of program code relative to its execution within one machine. The efficiency of using the Apache Hadoop platform is based on the implementation of the MapReduce distribution and reduction algorithm, which underlies Google's approach to organizing queries on distributed datasets. Apache Hadoop allows you to expand your compute cluster from one to a thousand machines, each offering local computing power. The software provides high resiliency by detecting and handling failures at the application layer, bypassing the hardware layer, resulting in high service availability on the cluster's master node, where each node may be down due to a hardware failure. According to the authors, distributed processing of spectral data in Apache Hadoop systems makes it possible to improve performance by changing the cluster parameters without making changes to the mechanisms of software algorithms.

Results

In the process of designing an information system model, the main requirements for its architecture were identified, including: a) the presence of a full-fledged client-server architecture, b) the presence of a distributed computing architecture, which allows increasing the efficiency of calculations, c) the use of modern intelligent algorithms for data processing and obtaining “useful” information, d) the presence of an archived distributed system for storing data and results [8, p. 1730; 9, p. 106275].

The data-level model proposed in the article ensures the maintenance of distributed storage systems and is focused on the dynamic formation of complex virtual information products at the time of their request based on stored processes and the response of the system to the moment the processes appear in the system. The specification and features of the model allow it to be highly optimized and perform satisfactorily when dealing with large amounts of data.

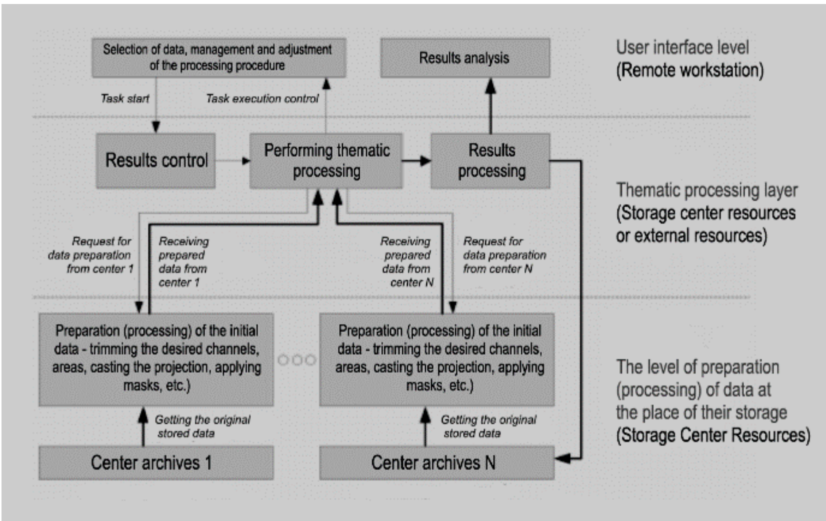

Fig. 1. Data level model

Figure 1 shows a data model of a distributed information system, represented by ultra-large global archives of remote sensing data.

To create a conceptual model of the information system, the methodology of object-oriented analysis and design (OOAD) was used, the functional modeling of the United Modeling Language (UML), on the basis of which six objects of IP - actors were identified:

1. The Information system (IS) administrator who will interact with the IS, have access to data editing, the ability to receive statistical reports, the ability to view and edit customer data;

2. A client who will interact with the functional modules of the system directly — use the services of the satellite imagery monitoring service - download and export data, receive analytical information, create personal classification profiles;

3. IIS (intelligent information system) for monitoring Arctic remote sensing data. The system is the foundation of the entire functional-process part and includes all classes, objects and other subsystems necessary for the successful operation of the software. In its subsystem, there are 3 process modules: an image preprocessing module - performs preliminary filtering of hyperspectral images, converts images into a convenient format for working with an image recognition system, clusters the image into smaller sectors; processing module - implements the work of the system of recognition and classification of objects in images, the basis of which is a neural network. Analytical module - implements the diagnostic system, as well as detecting anomalies. Performs the function of providing post-process data in the form of reports or graphical diagrams. In turn, it was decided to create a resource and results management center. This separate component is responsible for managing and sharing resources and tasks between distributed High-Performance (HP) systems, as well as processing results at the topic processing level and displaying at the user interface level;

4. Resource storage center — storage and access to data, preparation of data for processing;

5. Resource and Results Management Center — the main controller that manages the processes and controls all data flows in the system. Controls the execution of tasks and the organization of parallel computing;

6. The interface of the presentation level — is responsible for displaying and updating information at the presentation level, data exchange between the user and the system.

In the process of UML modeling, use cases were developed for each actor:

Client: 1) Downloading data for system processing, in particular — satellite imagery files in various formats; 2) Obtaining data processing information by the system — in the order of visual diagrams and graphs, displaying class zones on an interactive map, in the snapshot view mode; 3) Personal configuration of the data processor - filter of parameters and classes of object recognition; 4) Data export in various formats; 5) Keeping records and generating statistical documents; 6) Data selection; 7) Management and customization of processing procedures; 8) Analysis of results

Administrator: 1) Direct administration of the system: management of client profiles of the system, differentiation of rights, management of tariff plans of system users; 2) Providing technical support; 3) Ability to obtain broader data statistics based on user requests and data used to process specific requests from all users. Tracking global trends and the formation of dependencies and patterns.

IIS: 1) Pre-processing module: 1.1) Reduction of the level of principal noise based on the method of principal components; 1.2) Correction of geometric, radiometric and atmospheric distortions; 1.3) Formation and construction of an ensemble of hierarchical grid clustered groups for the segmentation of multispectral images; 1.4) Image compression to improve system efficiency (using auto encoders, Field-Programmable Gate Array (FPGA) and GPGPU); 1.5) Processing, identification and determination of the coordinates of objects in the image; 1.6) Decomposition of the image into frequency components using the method of two-dimensional Hilbert-Huang transform for the extraction of texture features; 1.7) Improving the quality of clustering through the use of agglomerative dendogramming methods to increase the resilience of clusters to grid changes;

1.8) Spectral separation, matched filtering (will help to improve the recognition accuracy of the final object by increasing the contrast against the background, which has a structured or stochastic nature); 1.9) Enhancing the contrast of the differences between the background and the target object through the main components; 1.10) Application of the automatic segmentation method based on a combination of graph theory and fractal net evolution approach (FNEA).

Processing module: 1) Selection and recognition of various geospatial objects and regions, based on the search for image areas that correspond to the spatial characteristics of the target fields, taking into account their main sign properties and parameters, using fuzzy algorithms of convolutional or combined neural networks. Possibly pixel by pixel; 2) Division of input data (remote sensing images) into classification clusters that satisfy the characteristics of the original search agent problem; 3) Morphological processing of clusters, merging and deleting nonexistent fragments from the point of view of the problem under consideration; 4) Selection of objects based on the use of the scale-space filtering method; 5) Description of the shape, position, texture and spectral features of the identified objects. The inference of the algorithm is implemented in the form of logical inference, with the help of which the search for a combination of features that satisfy the task is carried out. The output is a set of solutions with the value of the weight function exceeding the fitness threshold for this type of problem.

Analytical module: 1) Generation of reports and statistical summaries; 2) Forecasting changes in the analyzed region; 3) Highlighting anomalies and critical situations

Resource Storage Center: 1) Distribution and storage of resources in archiving centers; 2) Preparing data for processing; 3) Responsibility for issues of fault tolerance and data recovery; 4) Maintaining an archive of data and results.

Resource and Results Management Center: 1) Distribution and parallelization of tasks and processes between the computing components of the system; 2) Processing the results for display at the user interface level; 3) Monitoring the execution of tasks and their launch.

Presentation Layer Interface: 1) Display and update of information for the user; 2) Sending exported data to the client; 3) Transfer of changes in objects to the controller: 3.1) Sending downloaded data; 3.2) Send filter settings; 3.3) Submitting processing settings.

Together with the conceptual model, a UML use case diagram was developed that reflects the static representation of use cases and the actors — the actors. The diagram shows an example of the concept of information system behavior. Figure 2 is a use-case diagram that describes the concept of information system behavior.

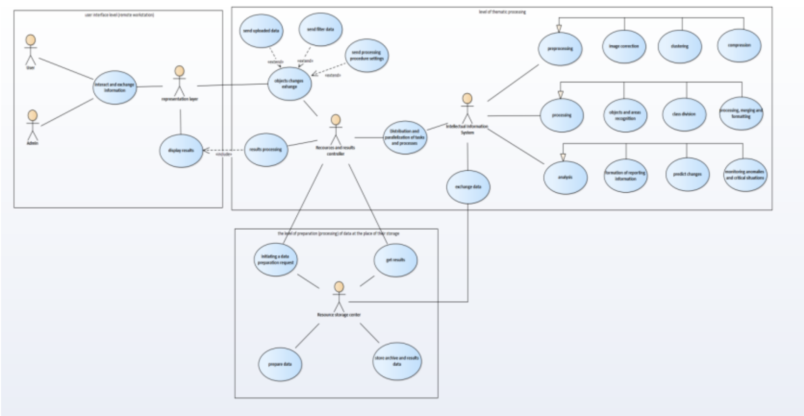

Fig. 2. Use case diagram

The diagram shows three border areas:

1. User Interface Level.

2. Level of Thematic Processing.

3. Level of Data Processing at Storage, which represents a specific logical level of the system and are responsible for the stage of the system presentation.

The central place in the OOAD methodology is the development of a logical model of an information system in the form of a class diagram, which allows you to logically represent the IS model and describe the internal structure of the IS.

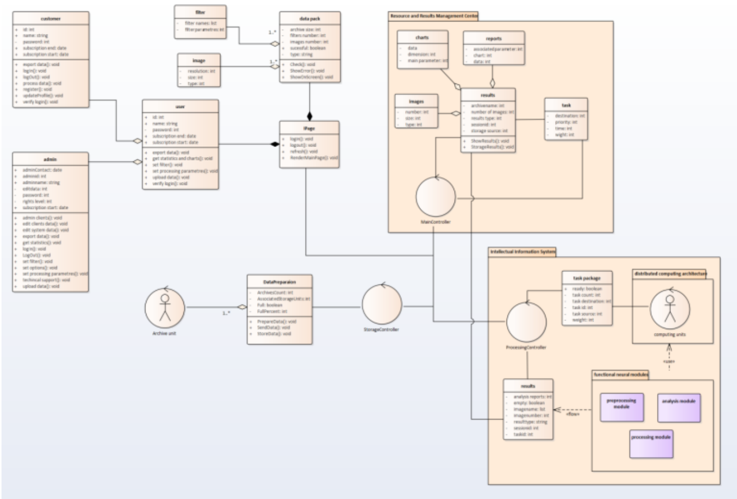

Fig. 3. Class diagram

The diagram (Figure 3) shows the hierarchy of information system classes, operations and attributes, and the relationships between them. The main contiguous classes were the “customer” and “admin” classes. These classes directly interact with the system through the system presentation layer — a web service that processes and displays information sent by the system and back. During the design, the Model-View-Controller (MVC) system-modeling pattern was used, since it significantly simplifies the work on large projects. When using MVC, it is much easier to maintain and test the code; it provides the convenience of displaying various representations of elements and various IC devices.

The ResouceandResultsManagerCenter package of the IS model contains a set of classes that allow you to implement the functions of control, monitoring the system and managing its processes. This package controls the distribution of tasks and processes between distributed computing resources, controls the path of data through the system and processes the results. The Main Controller class represents the package subsystem controller, which is connected by an association link with other IS levels.

The IntellectualInformationSystem package of the IS model contains a list of classes required for processing remote sensing data. The Processing Controller class represents the package control controller, which is responsible for processing and transmitting “useful” Arctic data. The DisturbedComputingAchitecture package is a set of architectures used in high performance computing for Arctic data. Actor ComputingUnits, included in the IntellectualInformationSystem package, is a set of key computers or their systems used to process Arctic data. The Functional Neural Modules package includes three main IS interface modules: 1) Preprocessing Module, 2) Processing Module, 3) Analysis Module. These interfaces are directly responsible for various stages of processing and extraction of “useful” Arctic information.

The discussion of the results

Because of the work, a study was carried out, and a model of an information system for monitoring Arctic data was created, the subject area was studied, and a brief overview of modern technologies and solutions was provided:

1. intelligent processing of Arctic data.
2. high performance computing.
3. analysis and processing of hyperspectral data.

The research carried out and the results obtained are of practical importance for the development of a prototype of an information system, which can be used for effective monitoring of "useful" data from the Arctic region of the Russian Federation. The proposed functional model can be used for monitoring and analyzing remote sensing data. The use of the information system will automate the processes of analysis, updating, storage and processing of objects in the Arctic zones. [11, p. 111457].

Conclusions

The article analyzes the existing models of remote and traditional systems for monitoring, processing and storing remote sensing data in the Arctic region. The advantages and disadvantages of existing tools and methods for monitoring hyperspectral data are analyzed. It is shown that the information system model can be represented in the categories of classes and actors, in the conceptual design format based on the OOAD and UML methodology. It is proposed to use a client-server distributed architecture for the implementation and deployment of an information system. The logical structure of the system has been substantiated, including: 1. IS Administrator, 2. Client, 3. IIS, 4. Resource storage center, 5. Resource and results management center 6. Presentation level interface, including module interfaces, classes and dependencies.

For the implementation and implementation of the information system for monitoring remote sensing data of the Arctic region in practice, its optimization and further development, it is necessary to clarify the technical means of implementation, a more extensive study of the network components of the system, as well as algorithms and methods for implementing processing and analytics modules. The proposed solutions will make it possible to develop new approaches and introduce modern technologies and high-performance computing tools to ensure the effective application of technologies for processing Arctic remote sensing data in various areas of the Arctic infrastructure.

In conclusion, I note that this work was carried out with the support of the FSI (contract 179GUTSES8-D3 / 54994 of 12.24.2019)

References

1. Chawla I., Karthikeyan L., Mishra A. K. A review of remote sensing applications for water security: Quantity, quality, and extremes // Journal of Hydrology. 2020. Vol. 585. P. 124826. https://doi.org/10.1016/j.jhydrol.2020.124826

2. Zhao J., Cao Z., Zhou M. SAR image denoising based on wavelet-fractal analysis // Journal of Systems Engineering and Electronics. 2007. Vol. 18. №1. P. 45-48. https://doi.org/10.1016/S1004-4132(07)60048-6

3. Gao F., You J., Wang J., Sun J., Yang E., Zhou H. A novel target detection method for SAR images based on shadow proposal and saliency analysis // Neurocomputing. 2017. Vol. 267. P. 220-231. https://doi.org/10.1016/j.neucom.2017.06.004

4. Nigro L. Parallel Theatre: An actor framework in Java for high performance computing // Simulation Modelling Practice and Theory. 2021. Vol. 106. P. 102189. https://doi.org/10.1016/j.simpat.2020.102189

5. Chen S. et al. How big data and high-performance computing drive brain science // Genomics, proteomics & bioinformatics. 2019. Vol. 17. №4. P. 381-392. https://doi.org/10.1016/j.gpb.2019.09.003

6. Wang S., Wang Q., Zhao J. Multitask learning deep neural networks to combine revealed and stated preference data // Journal of choice modelling. 2020. Vol. 37. P. 100236. https://doi.org/10.1016/j.jocm.2020.100236

7. Lee S. et al. Machine learning approaches to retrieve pan-Arctic melt ponds from visible satellite imagery // Remote Sensing of Environment. 2020. Vol. 247. P. 111919. https://doi.org/10.1016/j.rse.2020.111919

8. Hug C. et al. A method to build information systems engineering process metamodels // Journal of Systems and Software. 2009. Vol. 82. №10. P. 1730-1742. https://doi.org/10.1016/j.jss.2009.05.020

9. Ozkaya M., Erata F. A survey on the practical use of UML for different software architecture viewpoints // Information and Software Technology. 2020. Vol. 121. P. 106275. https://doi.org/10.1016/j.infsof.2020.106275

10. Lu H. C., Hwang F. J., Huang Y. H. Parallel and distributed architecture of genetic algorithm on Apache Hadoop and Spark // Applied Soft Computing. 2020. Vol. 95. P. 106497. https://doi.org/10.1016/j.asoc.2020.106497

11. Wang C. et al. Classification of the global Sentinel-1 SAR vignettes for ocean surface process studies // Remote Sensing of Environment. 2019. Vol. 234. P. 111457. https://doi.org/10.1016/j.rse.2019.111457