

TeraPro, a System for Processing Large Biomedical Images

K. N. Kozlov^a, P. Baumann^b, J. Waldmann^b, and M. G. Samsonova^a

^a *St. Petersburg State Polytechnic University, St. Petersburg, 195251 Russia*

^b *Jacobs University, 28759, Campus Ring, 12, Bremen, Germany*

e-mail: mackoel@gmail.com

Abstract—Image analysis plays an important role both in medical diagnostics and in biology. The main reasons that prevent the creation of objective and reliable methods of analysis of biomedical images are the high variability and heterogeneity of the biological material, distortion introduced by the experimental procedures, and the large size of the images. This paper presents preliminary results on creating a system called TeraPro, which combines a platform for image processing (ProStack) and a raster data storage system (rasdaman). This integrated system can be used in a cloud environment, providing access to the methods of visualization, analysis, and processing of a large amount of images through the Internet. Such an approach increases the speed and quality of image understanding and softens the limitations imposed by other systems. The system allows us to view uploaded images in the browser without having to install additional software on any device connected to the Internet, such as tablet computers and smartphones. This paper presents the preliminary results of processing biomedical images.

Keywords: image processing, medicine, biology.

DOI: 10.1134/S105466181304007X

Image analysis plays an important role in medicine and biology. In system biology its role is especially significant in obtaining quantitative data, describing the processes occurring in cells, from the concentrations of proteins and mRNA to the rates of biochemical reactions. In medicine, image analysis plays a critical role in the diagnosis and prediction of the outcome of diseases, in study of organs and tissues, in estimating the response of an organism to therapy, and in clinical testing of new drugs. Extracting features from biological images and clinical diagnostics are still largely based on a qualitative visual analysis of the images. It is evident that such analysis is subjective, depends on the qualifications of the expert, does not allow for standardization, and is labor-intensive and time-consuming. The wide variety of images and other types of data, which we will refer to as multidimensional raster data, serve as an additional complication. Image processing and analysis are very difficult tasks. Images consist of pixels, but for the expert, it is not the pixels but the objects composed of pixels that are of interest. Therefore, image analysis addresses the difficult task of transforming the data presented in the form of pixels into data in the form of objects. After successfully solving this problem, the properties of the objects can be quantified: morphometric measurements can be performed, and spectral and statistical characteristics can be estimated, as well as the degree of fluorescence.

The problem of creation of objective and reliable methods for the analysis of biomedical images is very relevant and is exacerbated by the lack of software designed for the needs of system biology and clinical diagnostics. The main reasons that prevent their creation are the high variability and heterogeneity of the biological material, distortion introduced by the experimental procedures and recording devices, insufficient development of the standards for description of biomedical images and data taken from them, etc. Another factor that complicates processing and analysis of biomedical images is their large size. Thus, for example, a slide scanner usually produces images of 2.7–10 GB and 30000 × 30000 pixels or more in size. This is approximately 100 times more than the size available for processing by specialized medical systems. Modern imaging systems use powerful technologies of processing, but in most cases the image size is limited by the size of the computer memory, for example, in eCognition, the processing of objects larger than 2 GB is difficult. Similar arguments are valid for packages of visualization and analysis, such as AVS/Express, IDL, and Envi. Since often an image serves for a doctor to view it and make out a diagnosis, visualization plays an important role. Sometimes, instruments perform “processing outside the channel” [1] for discharging part of the image in processing. This slightly reduces the limitations, but the algorithms used are not optimized for these operations, owing to which the rate of processing suffers. Further-

Received March 3, 2012

more, expansion of the volumes is possible, for example, by one order but not 3–6, which occurs when using multi-terabyte/multi-petabyte images.

Picture Archiving and Communication Systems (PACS) [2] are widely used for storing, managing, distribution, and display of medical digital images. Specialized software is available through companies with a global name, such as GE, Philips, and Siemens, and small ones, such as Mint Medical, while CMEDIC PACS provides a free solution with an open source code. The common weakness in the existing systems consists in that multidimensional images of large sizes are served by servers without high-level logic and complex processing is done in client programs. This leads to unnecessarily intensive exchange of data of large volumes and, consequently, reduced speed.

One approach to solving this problem is the creation of software that integrates the technologies of multi-dimensional databases and image processing. This paper presents the preliminary results on creating the TeraPro system, which combines a platform for image processing (ProStack) and a raster data storage system (rasdaman). ProStack, developed by experts for image processing, has proven its applicability in solving complex research and practical problems. The rasdaman raster data server allows us to store images of any size and dimension in a standard database and provides a raster query language that has proven its effectiveness in processing data of terabyte volume. The integrated system can be used in a cloud environment, providing access to the methods of visualization, analysis, and processing of large-sized images through the Internet. This approach increases the speed and quality of image understanding and softens the limitations imposed by other systems.

The interpenetration of various disciplines in the life sciences, without which achievements such as the decoding of the human genome would not have been possible, has led to the necessity of using ontologies. An ontology helps to set a controlled vocabulary, reducing the ambiguity in information processing and exchange. Ontologies are designed for a variety of disciplines, such as anatomy, biochemistry, medicine, human diseases, and phenotypes, for example, NCI Thesaurus, SNOMED CT, and OMIM. A wide variety of ontologies is available on the OBO (Open Biomedical Ontologies) Foundry. Owing to their different purposes and usage, they vary in complexity and size. Some consist of several hundred concepts, while others, such as GO (Gene Ontology), of several tens of thousands of concepts. Work is also under way to standardize the description of ways of obtaining images: the settings of the microscope and details of obtaining images (<http://www.berkeleybop.org/ontologies/owl/FBbi>), The Open Microscopy Environment consortium [3]. For the creation, storage, transmission, and visualization of medical images and documents of patients, the DICOM industrial standard has been developed [4]. In [5], an experimental version of the

ontology of image analysis developed in the OWL language has been discussed. As part of working on the ProStack package, the authors have developed an ontology of image processing methods, which covers all aspects of extraction of precise quantitative data from biomedical images and allows us to build complex multistep image processing procedures, which use the implementation of the necessary algorithms in various software packages.

MATERIALS AND METHODS

2.1. *Ontology of Image Processing Methods*

The word *ontology* has many definitions in the field of artificial intelligence. In particular, an *ontology* can be defined as a formal description of classes of image processing techniques, the properties and attributes of each class, and the restrictions on them. Classes can contain subclasses. Together with a multitude of class representatives, an ontology constitutes a knowledge base. The advantage of using an ontology is the systematic approach to the subject area.

Building an ontology should be started by highlighting the main concepts of the subject area, taking into account the objectives of the future use, i.e., the search for a desired method and realization by the properties of the input and output images. Figure 1 shows the first stage of the hierarchy of classes of image processing methods. The base class or the root element of the hierarchy is the “Method” class. The specific names of implementations of methods, the descendants of the base class, are stored in the respective properties of this class, which are inherited by all subclasses. Each package of image processing offers its own version via the menu system and, in addition, various embodiments can be found in the literature.

The most complete classification and description of the image processing methods is given in the book by Gonzalez and Woods [6]. The presentation of the classification methods through the menu system, for example, in a graphical user interface, is not an ontology, as it contains no description of the properties or attributes of the methods. In the ProStack ontology of image processing methods, a subdivision of all methods into ten classes is proposed: (1) “Combination”; (2) “Object Recognition”; (3) “Correction”; (4) “Segmentation”; (5) “Restoration”; (6) “Morphology”; (7) “Geometry”; (8) “Transform”; (9) “Arithmetic”; and (10) “Misc.,” which allows us to build a dialogue for selecting the appropriate method of solution of a particular problem. Each of these classes has its subclasses detailing operations on the images.

Consider the class “Segmentation.” It combines the methods that allow us to divide the entire image into zones according to some criterion, for example, by the presence of the sole object of interest. This class includes methods such as threshold filtering, edge

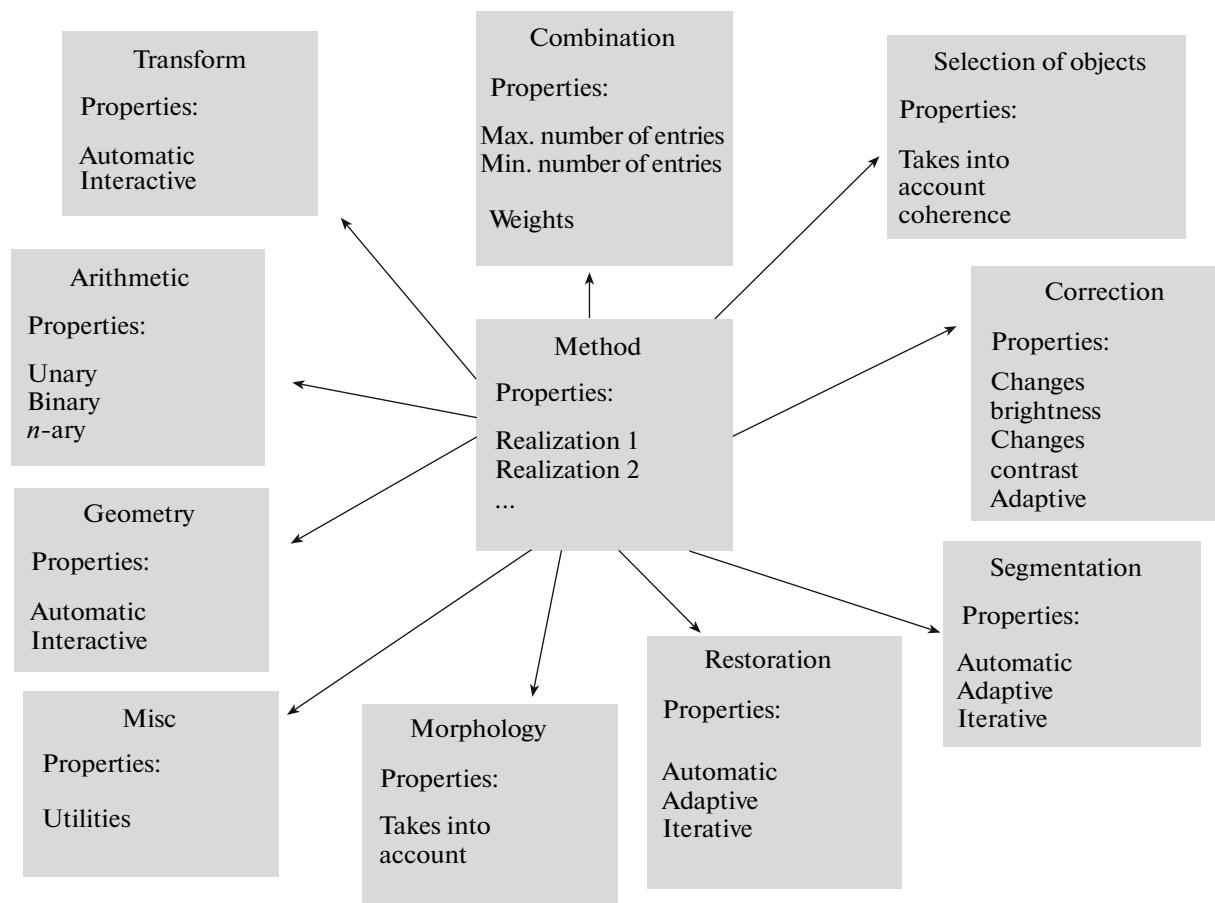


Fig. 1. Classes of methods that implement various operations and are descendants of the root element “Method.”

detection, and watershed transformation. The choice of a specific approach to segmentation in the mode of dialogue is possible with the use of properties. The properties of this class can take on the meanings TRUE or FALSE. A database with the use of DBMS SQLite version 3 has been developed for storing the ontology of the methods of image processing, which is one of the recognized leaders among embedded DBMS. Each class of the ontologies of image processing methods corresponds to a separate database table that lists the subclasses of this class with the keywords that characterize their functionality and references to respective tables. Thus, the hierarchical structure of the ontology is maintained by means of a DBMS.

A necessary condition for providing the user with a quick and easy search for the necessary method and implementation by keywords and properties of the input and output images consists in the presence of a correspondence table for the common terms of image processing and ontology concepts (“Dictionary”). In the developed database, the “Dictionary” is stored in the SUBSTITUTIONS table.

To store the information about the methods of image processing, the OPERATORS table is pro-

vided. Compliance of a specific method with each of the ontology classes is characterized by a single number from zero to one in the column of the table for this class. Ontology classes may overlap; that is, some methods or their implementations may relate to several classes. In addition, the table contains information on the correspondence of the methods not only to classes-leaves of the hierarchy tree, but also to ancestor classes. One method can have several implementations, which are interchangeable at the stage of realization. Each available implementation of the registered methods is assigned a whole number called the weight of implementation. Summoning of a method induces the ProStack runtime library to summon an implementation with the maximum weight. Owing to the fact that on different platforms, such as personal computers and servers, different implementations have different weights, the use of the mechanism of semi-automatic selection of the most appropriate implementation of the method requested by the user improves the efficiency of using computing resources.

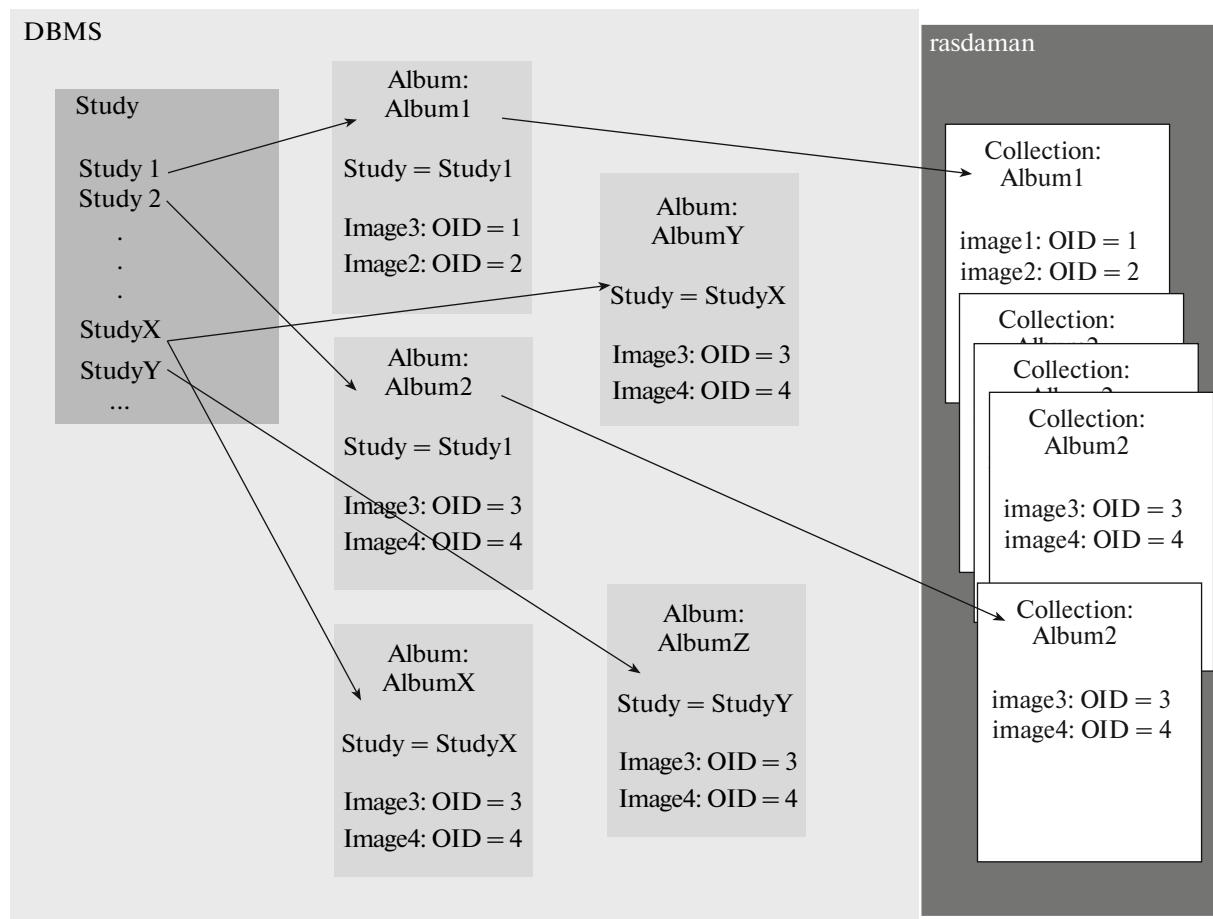


Fig. 2. Prototype of a base for storing metadata.

2.2. The TeraPro System Architecture

The TeraPro system [7] combines the ProStack platform for image processing, the rasdaman raster data storage system, and a database for storing image metadata. The main object of the metadata database is “Research.” For each trial, there can be several “Albums,” and for each one of them, several images. These conditions define the relationships within the base and make it possible to maintain its integrity through a DBMS. With the introduction of the product in the medical field, “Research” may be associated with a patient and “Albums” may be associated with sets of images obtained using specific equipment. With implementation in other fields of knowledge, the fields can be renamed. Each entity has a unique numeric identifier and a unique alphanumeric pseudonym. The “Research” table allows one to store identification data of objects, and “Albums” correspond to collections in rasdaman and refer to specific images in rasdaman through object identifiers (OID) (Fig. 2).

ProStack, developed by experts in data processing, has proven its applicability in complex research and practical problems. Development of the ProStack

software for processing images of various nature was started at St. Petersburg State Polytechnic University. In the ProStack software [8], methods for processing two- and three-dimensional images, such as removing a nonspecific signal, edge detection, identification and classification of objects, morphological methods, etc., have been developed. All the methods are implemented as separate operators that can be combined into image processing scenarios using the graphical user interface. Scripts can be used as new modules. The open architecture of the system allows the user to integrate new processing packages into it, both commercial and free. Some procedures of the OpenCV library can be used in the scenarios [9]. The rasdaman (“raster data manager”) server [10] is a software superstructure built on top of the DBMS (though often called a database itself), which allows one to store raster data of any volume and dimension in a standard database. It provides flexible access through the query language—an expansion of SQL—with declarative raster expressions that admit optimization. Considerable optimization of storing, such as storage of fragments (tiles), adaptive compression, and support archiving to tape, and optimization of performance,

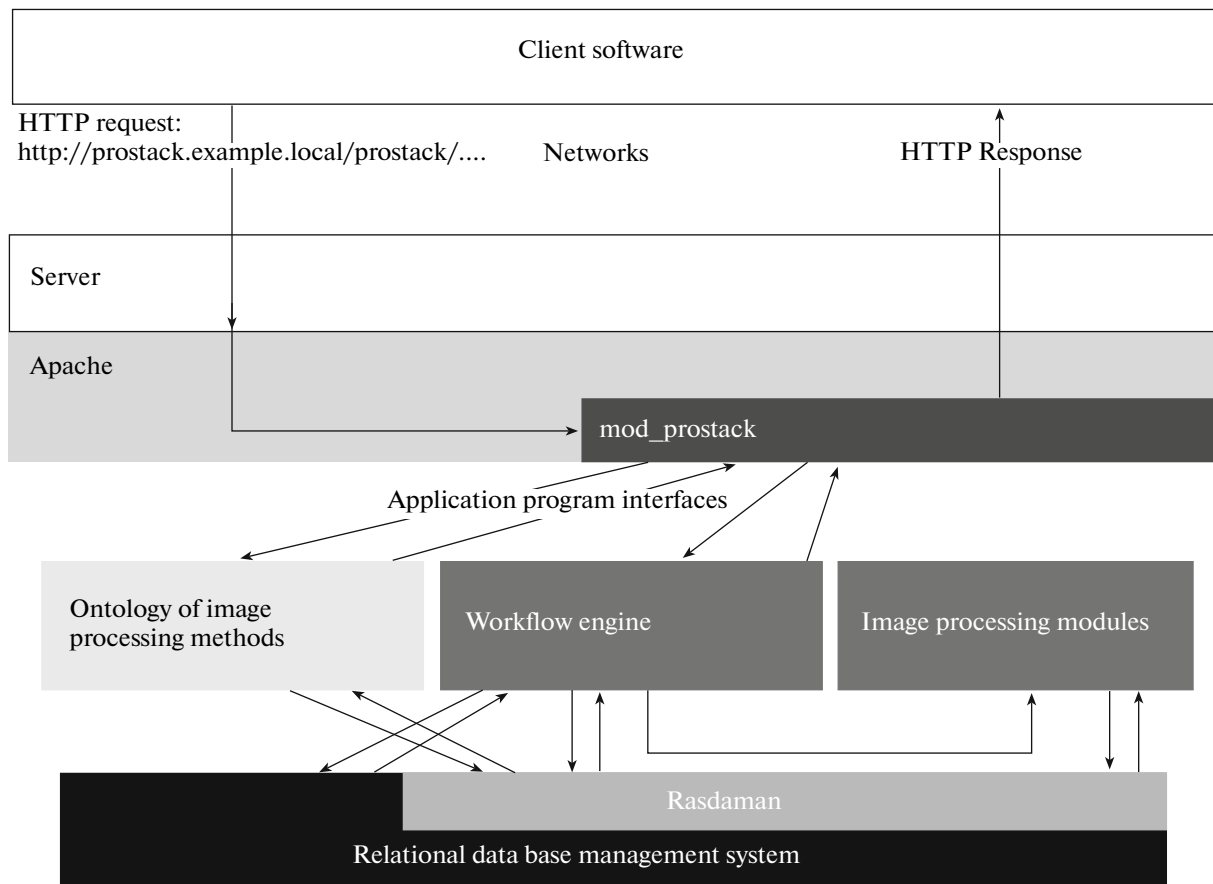


Fig. 3. Concept of software integration.

such as heuristic rewriting, dynamic preaggregation, and just in time compiling, provide access to multi-terabyte objects with a delay of less than a second (12 TB in 400 ms).

A plan of the interaction of TeraPro components is given in Fig. 3. The server component of the ProStack package is implemented as a standard dynamically loadable module to the Apache v.2 web server. The module allows one to perform the procedures, and scenarios are available in the ProStack package remotely. Management is carried out by means of the HTTP (s) protocol. The software uses GET, PUT, POST, and DELETE statements to provide the same or extended functionality compared with the version for the personal computer. The server module is not limited to the use of one specific client application. The interface of the desktop version of the ProStack package can be used for communication with the server, and, in addition, the possibility of using different servers in the same scenario is fulfilled.

2.3. Interaction of ProStack and Rasdaman

The interaction interface is based on using the rasql command-line interpreter of queries. Owing to the

ontology of image processing methods and the open architecture of the system, the ProStack knowledge base includes rasdaman operations that can reduce the dimension of the extracted data by selecting an area of interest or through arithmetic operations within rasdaman and extracting the resulting array. During the execution of the scenario, the operations implemented in rasdaman are not executed immediately, but accumulate in the buffer until the moment when their result is needed in the form of images for executing an operation from another package. The accumulated operations are converted into one query in the rasdaman language, which is executed by the interpreter. Increase in the interaction efficiency and reduction of delays is thus achieved through the conversion of several consecutive requests into one, the result of which has a smaller volume.

RESULTS

3.1. Working in the TeraPro System

The first pilot version of the TeraPro system is available on the Internet [7]. The demo version, admission to which is available on request, contains images of patterns of activity of segmentation genes of

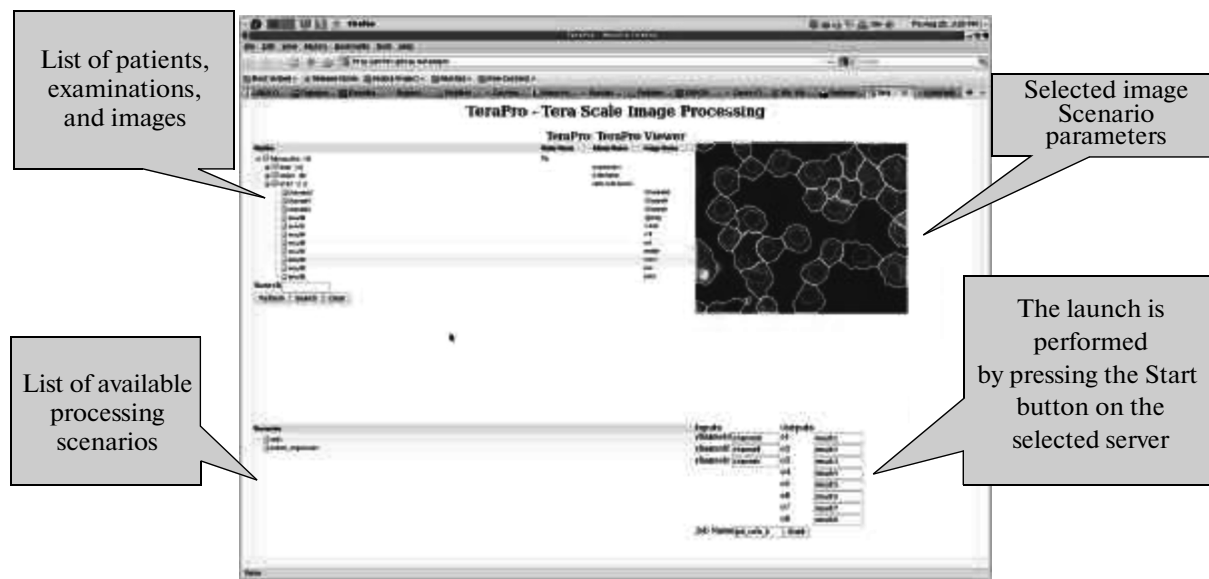


Fig. 4. User interface.

the *Drosophila* embryo, as well as images of cell cultures. Scenarios of processing of such images are registered in the system.

The system permits viewing the available images, as shown in Fig. 4. The image selected in the tree view in the top left corner appears in the center of the screen. A list of registered processing scenarios is on the bottom to the left. If the user selects one of the scenarios on the right bottom, a list of parameters for this scenario with input fields appears. The user can fill in all the fields and run the script by clicking on the “Start” button. The results of processing consist in the images being loaded into the system. The user can update the list of images by pressing “Refresh.”

The user creates and debugs scripts in the interactive mode in the graphic interface of the desktop version of ProStack. The presence of *rasdaman* on a PC is not required. The weight of implementation of methods from *rasdaman* on the TeraPro server is greater than the weight of other implementations, and therefore, when the script is run in the TeraPro system, ProStack operators are automatically and transparently replaced by *rasdaman* queries, which, in turn, are optimized through combining. As a result, at the stage of extraction of the data from *rasdaman*, the volume of data will be reduced, so that processing procedures obtain images of the minimum required size.

The graphical user interface of implementations in the packages of the same methods is identical. This approach simplifies the study of the system and increases the efficiency of work through the transition to implementations more suitable for the specific platform, which is more transparent for the user.

3.2. TeraPro: Visual Analysis of Large-sized Images

One of the complex issues of using advanced systems for processing biomedical images obtained on modern diagnostic equipment is using closed undocumented formats of data storage by manufacturers, conversion of which into common formats is possible only with the use of software of the same manufacturer. The OpenSlide project (<http://openslide.org/>) offers an alternative solution for the most common formats. In the TeraPro system, using the libraries of the OpenSlide project, a mechanism for downloading images taken on a slide scanner was developed. These images in a closed format have a typical size of 1.5 GB in compressed form, which corresponds to about 50 GB in the TIFF format. The software by the manufacturer allows the user to view images on the workstation running the MS Windows with sufficient system resources. After uploading to the *rasdaman* repository, the TeraPro system allows the user to view the images in the browser without installing additional software on any devices connected to the Internet, such as tablet computers and smartphones.

3.3. ProStack: Fundus of the Eye Image Processing

We have obtained preliminary results on the development of scenarios of automatic detection of objects in digital images of an eye's fundus obtained using a Fundus camera and those available on the Internet [11]. Analysis of such images is the main method of diagnosing glaucoma, a disease that leads to blindness and is diagnosed in 67–70 million people in the world. We have developed a scenario of segmentation of such images in order to detect automatically the area of the disk of the optic nerve, exudative manifestations, and

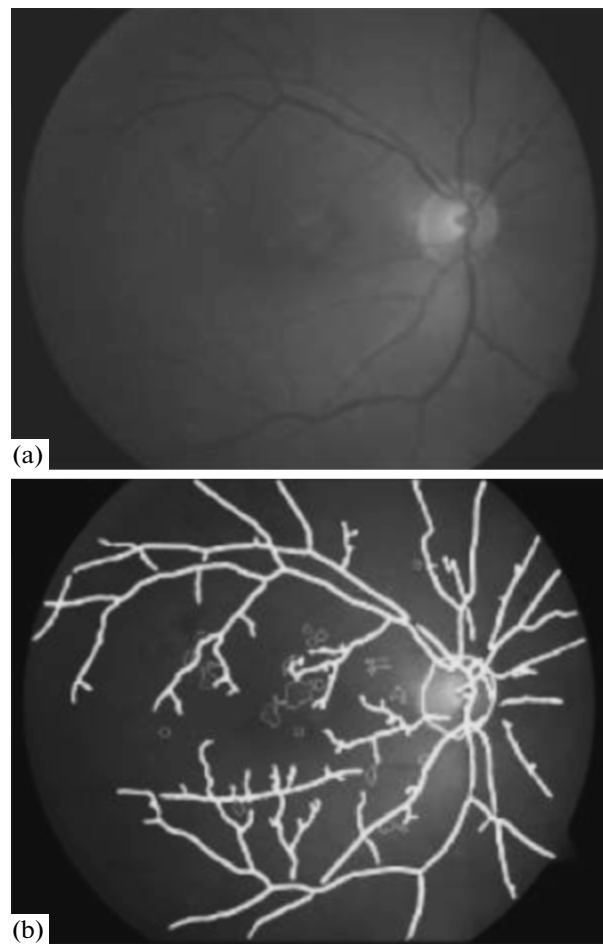


Fig. 5. (a) Typical image of the fundus of an eye. (b) Segmented blood vessels, optical disk, and exudative manifestations.

blood vessels. Automation of the registration of pathologies in images will help medical experts in the diagnosis of glaucoma due to acceleration and standardization of the process. The algorithm is based on [12] and uses morphological operations of expansion, contraction, and reconstruction with different structural elements. The modified procedure uses information from all color channels. Figure 5a shows the typical image, and Fig. 5b gives the result of processing.

3.4. *ProStack: Automation of Obtaining Quantitative Data on Gene Expression*

The quantitative information obtained using modern microscopes is necessary to understand the basic principles that govern the behavior of biological systems [13–18]. Using the methods of the ProStack package, a procedure for processing confocal images of the gene expression patterns was developed to extract quantitative data on gene expression from confocal images of expression patterns in the fruit fly *Drosophila* (Fig. 6a). The procedure consists of several steps, for each of which specialized data process-

ing scenarios [19] have been developed (Figs. 6b–6d). The procedure was successfully applied to obtain quantitative information on the cellular level in space and with a resolution of 6.5 min time. The data obtained allowed us to construct a time–space atlas of segmentation gene expression accessible in the FlyEx database (<http://urchin.spbcas.ru/flyex>) [20].

CONCLUSIONS

Problem-oriented image processing in the life sciences is currently accomplished by software adapted or developed by researchers themselves. Among the technological flaws of the available products, there exists a limit on the size of the data of the operational memory. In addition, the development of complex algorithms requires good skills in programming, while maintenance of information systems in working order demands great experience in administration.

The main purpose of creating the TeraPro system is to integrate rasdaman and ProStack into a functional, scalable, expandable, and reliable open system for processing raster data of large dimensions in the field

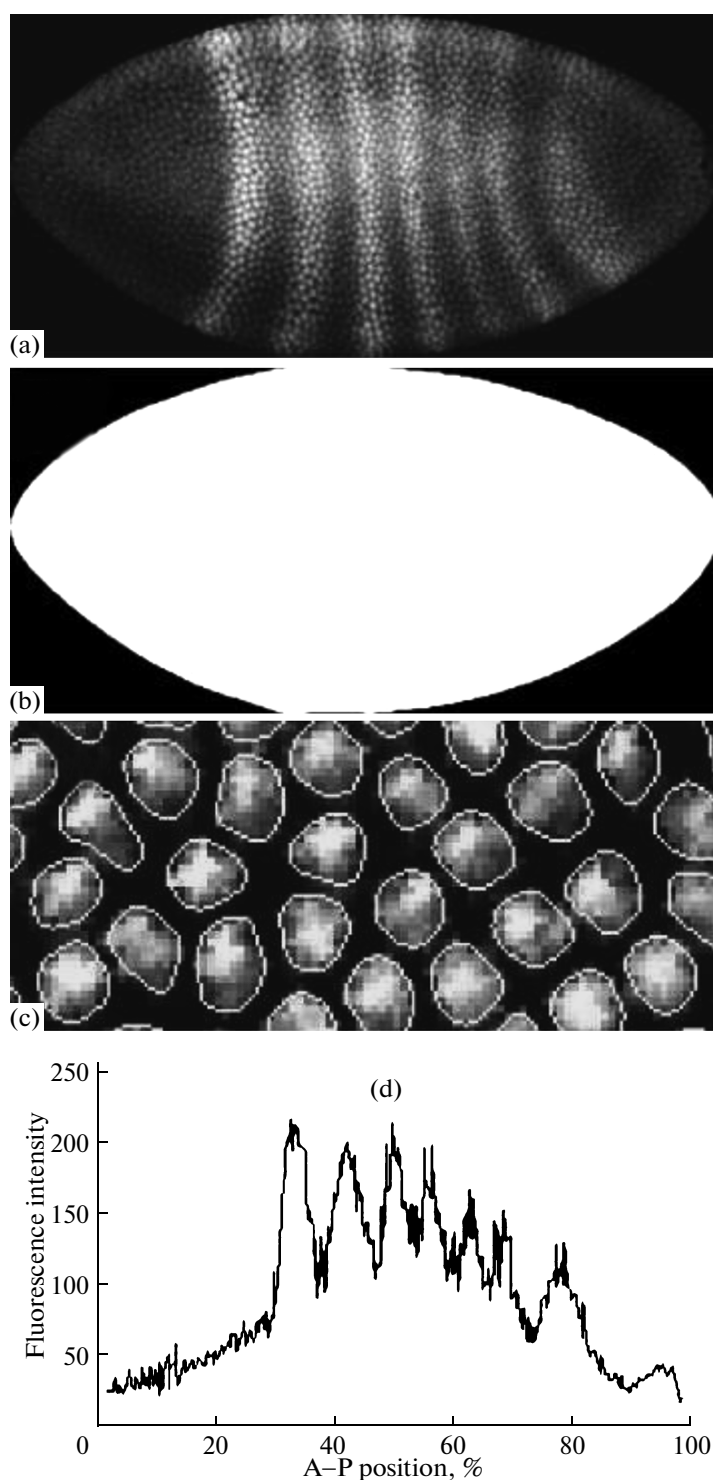


Fig. 6. (a) Extraction of quantitative information on gene expression in a *Drosophila* blastoderm embryo in the late cleavage cycle 14 from an image of the pattern of expression of the *even-skipped* (*eve*) gene. (b) Smooth mask. (c) Blown-up image of the boundaries of the nucleus superimposed on the original image of the expression pattern. (d) Quantitative information on the *eve* gene expression *eve*. The vertical axis shows the fluorescence intensity, and the horizontal axis shows the A–P positions along the anterior–posterior axis (0% corresponds the anterior pole). The graph shows the values only from the nuclei of the central 10%.

of medical imaging, biological raster data and data in the life sciences in general, and in the future, in the field of satellite images of the Earth.

Experience with ProStack in processing biological and medical images yielded successful preliminary results in using TeraPro for analyzing images of large volume.

ACKNOWLEDGMENTS

We thank A. Pisarev and E. Myasnikova for their helpful comments and suggestions. This work was supported by the Foundation for Assistance to Small Innovative Enterprises (FASIE) in Science and Technology, State contract no. 14069, and the Russian Foundation for Basic Research, project nos. 11-04-01162 and 10-01-00627.

REFERENCES

1. P. Widener, M. Barrick, J. Pullikottil, P. Bridges, and A. Maccabe, "A framework for out-of-band processing in large-scale data flows," in *Proc. 2007 Int. Conf. on Grid Computing (Grid 2007)* (Austin, 2007).
2. U. Bick and H. Lenzen, "PACS: the silent revolution," *Europ. Radiol.* **9** (6), 1152–1160 (1999).
3. I. Goldberg, C. Allan, J.-M. Burel, D. Creager, A. Falconi, H. Hochheiser, J. Johnston, J. Mellen, P. Sorger, and J. Swedlow, "The open microscopy environment (OME) data model and XML file: Open tools for informatics and quantitative analysis in biological imaging," *Genome Biol.* **6**, R47 (2005).
4. P. M. Kuzmak and R. E. Dayhoff, "Integration of imaging functionality into the healthcare enterprise using DICOM," *J. Digital Imag.* **11** (3), 67–70 (1998).
5. I. B. Gurevich, O. Salvetti, and Y. O. Trusova, "Fundamental concepts and elements of image analysis ontology," *Pattern Recogn. Image Anal.* **19** (4), 603–611 (2009).
6. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. (Prentice Hall, Upper Saddle River, N.J., 2002).
7. <http://urchin.spbcas.ru/trac/TeraPro>
8. K. N. Kozlov, A. S. Pisarev, and M. G. Samsonova, PC Software Certificate no. 2009611452 (March 16, 2009). <http://www.prostack.ru>
9. G. Bradski, "The OpenCV Library," Dr. Dobb's J. Software Tools (2000).
10. P. Baumann, "On the management of multidimensional discrete data," *VLDB J.* **4**, 401–444 (1994).
11. T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kalviainen, and J. Pietila, "DIARETDB1 diabetic retinopathy database and evaluation protocol," Tech. Rep. (2008). <http://www.it.lut.fi/project/imageret>
12. S. Ravishankar, A. Jain, and A. Mittal, in *Proc. 2009 Int. Conf. on Grid Computing (Grid 2009)*, *CVPR* (Banff, 2009), pp. 210–217.
13. J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz, "Dynamic control of positional information in the early *Drosophila embryo*," *Nature* **430**, 368–371 (2004).
14. J. Jaeger, M. Blagov, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, S. Surkova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz, "Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*," *Genetics* **167**, 1721–1737 (2004).
15. V. V. Gursky, J. Jaeger, K. N. Kozlov, J. Reinitz, and A. M. Samsonov, "Pattern formation and nuclear divisions are uncoupled in *Drosophila* segmentation: Comparison of spatially discrete and continuous models," *Phys. D* **197**, 286–302 (2004).
16. M. G. Samsonova, V. V. Gurskii, K. N. Kozlov, and A. M. Samsonov, "System approach for organisms researching," *Nauch.-Tekhn. Vedomosti St. Petersburg Gos. Tekhn. Univ.* **2**, 222–234 (2006).
17. Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A.-R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz, "Canalization of gene expression in the *Drosophila* blastoderm by dynamical attractors," *Plos Biol.* **7**, e1000049 (2009).
18. Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A. R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz, "Canalization of gene expression and domain shifts in the *Drosophila blastoderm* by dynamical attractors," *PLoS Comput. Biol.* **5**, e1000303 (2009).
19. S. Surkova, E. Myasnikova, H. Janssens, K. Kozlov, A. A. Samsonova, J. Reinitz, and M. Samsonova, "Pipeline for acquisition of quantitative data on segmentation gene expression from confocal images," *Fly (Austin)* **2**, 58–66 (2008).
20. A. Pisarev, E. Poustelnikova, M. Samsonova, and J. Reinitz, "FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution," *Nucl. Acids Res.* **37**, D560–D566 (2009).

Translated by N. Smolina



Peter Baumann is Professor of Computer Science at the University of Jacobs (Germany, Bremen), head of various national and international research and industrial projects since 1996, including in the 6th and 7th EU Framework Programmes. During the EU FP6 ESTEDI project, the group of Peter Baumann successfully collaborated with the group of Prof. Maria Georgievna Samsonova.

Peter Baumann is a world expert in the field of large-scale image services with more than 95 publications in reviewed journals, innovative work in the field of scientific databases, in which he explored various aspects of support for data arrays in DBMS, including mathematical modeling, development of a query language, architecture, optimization, implementation, and application in remote probing, the study of climate, oceanography, astrophysics, brain research, analysis gene activity, education, etc. He is the chief architect of rasdaman, which has been used for the past five years to support objects weighing dozens of terabytes.

Since 2002, Peter Baumann has been actively involved in standardization as an expert. He is a cochairman of the working groups on processing raster data in the Open Geospatial Consortium (OGC), which engages in standardization in the field of geo-services in collaboration with ISO, OASIS-Open, and W3C. He currently is the editor of nine specifications that are adopted as a standard (WCPS) or are in process of being approved in July 2010 (WCS 2.0).



Jost Waldmann works on a project at the Jacobs University under the supervision of Peter Baumann.



Konstantin N. Kozlov graduated from the Department of Applied Mathematics, Faculty of Physics and Mechanics, St. Petersburg State Polytechnic University in 2000. His experience of scientific work is more than 10 years. He has worked at the Department of Computational Biology, Center for Advanced Studies, since 2001. Dr. Kozlov is a specialist in applied mathematical methods and information technologies, and con-

ducts scientific research in the field of biomedical image processing and mathematical modeling of biological systems, is a coauthor of more than 35 works published in leading journals. He is member of international conferences, an invited speaker at the “Evolution, Systems Biology, and Supercomputing in Bioinformatics” International School Seminar at the International Conference on Bioinformatics of Genome Regulation and Structure (BGRS), 2006.



Mariya G. Samsonova is a Professor and Doctor of Biological Sciences. She is Head of the Department of Computational Biology at the Center for Advanced Studies. Prof. Mariya Samsonova has headed many Russian and international research projects. She has extensive experience in the creation of consortia, writing projects, and project management. Samsonova is the author of over 70 publications in reviewed journals, a member of the

Editorial Board of the Journal of Integrative Bioinformatics and Mathematical Biology and Bioinformatics, invited speaker at many international conferences on bioinformatics and system biology. She received an award for best paper at the ISMB99 conference. Samsonova is an expert in developing methods of extracting quantitative information on the activity of the genes from images, in database development, integration of distributed databases, programs, and services, as well as natural language interfaces. The group of Prof. Mariya Samsonova has developed innovative methods and software for system biology that are widely used by the scientific community. Among them are the ProStack image processing package, BREReA software for registration and removal of background signal from gene activity pictures, and the FlyEx database (<http://urchin.spbcas.ru/flyex/>).