A method of data structuring in the decision-making support system in oncological diagnostics of prostate diseases

S M Zaytsev¹, V G Nikitaev¹, A N Pronichev¹, E V Polyakov¹, N A Romanov¹, D Y Pushkar², E. A Prilepskaya², M V Kovilina², A V Govorov², A V Glotov², A O Vasilyev² and K V Kolontarev²

¹National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe highway 31, Moscow, 115409, Russia
²A.I. Yevdokimov Moscow State University of Medicine and Dentistry of the Ministry of Healthcare of the Russian Federation, Delegatskaya ulica 20/1, Moscow, 127473, Russia

E-mail: VGNikitayev@mephi.ru

Abstract. The study proposes the way of data organization to store the information obtained as the result of measurements of morphological structures of prostate gland preparations in the computer microscopy system. The proposed data structure provides fast access to object features in interactive analysis of preparation images during oncologic diagnostics of prostate gland diseases.

1. Introduction
The prostate gland cancer is a malignant neoplasm that grows in epithelium of alveolocellular elements of a prostate gland. It is one of the most widespread malignant new growths among men. More than 400000 occurrences are revealed every year in the world. On a number of countries it assumes the 2nd or the 3rd spot in the ranking after lung cancer and stomach cancer[1-2].

In Russia prostate gland cancer assumes the 7th – the 8th spot. It is the most often malignant tumor among men after 60. In Russia and the other countries of the Commonwealth of Independent States it is characterized with the late diagnostics, on the III-IV stages.

One of the criteria of the prostate gland cancer diagnostics is a Gleason grading system. Gleason score is based on the difference between cancer an normal cells in a prostate gland tissues.

For now the visual examination of microscopic images of histological specimen is used to reveal a disease. Adenocarcinoma is a kind of prostate gland cancer. In case of a late revealing it causes a death because of metastasizes, while on the early stages it has often no symptoms. The disease is detected by a pathologist using the visual research of microscopic images, this process is time-cost and subjective. The lack of skilled staff in oncology proves the urgency of the development of automated cancer diagnostic systems.

The intent of the research is a design of data structures and data access algorithms in the problem of the oncological diagnostics of prostate gland cancer. The morphology of tissue and cellular structures are studied in the investigation. The histological descriptions of preparations are given by specialists of the urology department of the A.I.Yevdokimov Moscow State University of Medicine and Dentistry of the Ministry of Healthcare of the Russian Federation[3-5].

2. Digital image processing
Object recognition is a common problem in most programs of automated diagnostics[6-9]. The object recognition criterion implies two conditions. Each pixel on the preparation image permits to determine, which object it belongs to. And for each object all pixels can be enumerated. To implement such a contract one should obviously store the results of recognition in special data structures that provide high access rate. Thus the decomposition of the object recognition problem into two subproblems can be made.

First, design and implementation of object isolation algorithms. This subproblem often needs an empiric solution. Second, design and implementation of data structures to store object information. This subproblem can be formalized. The study demonstrates its solution.

The input data for the problem is a so-called input matrix – a binary matrix (with one bit elements), which width and height are equal to those of a processed image. Elements corresponding to pixels of objects have a value of one elements corresponding to pixels of a background have a value of zero. Hence for a standard 24-bits image a matrix needs about 4 per cent of the image memory. For example for a 4K-image a corresponding matrix needs 8 Mbytes.

In the proposed solution the result of the recognition is stored in two interlinked structures: an object matrix and an object array. Matrix proportions are equal to recognized image proportions. Each matrix element one-to-one maps to an image pixel. The size of element is 6 bytes. 4 lower of them code a number of pixel-owning object and two higher code a relative position of a pixel in the object.

Because of the element size the maximum allowable number of objects is about 4 billion. This is more than can be located on the image with a resolution of 150000х150000. For standard 24-bit image a matrix needs about 200 per cent of the image memory.

A number of elements in object array equals to a number of objects plus one (this one corresponds to a background). An array element is a special data structure that describes an object and its position on the image. The size of an element depends on the number of used object features. An array element contains the coordinates of corners of a circumscribed rectangle, the coordinates of a center, the perimeter, the square, the number and the total square of spots etc. In addition an element contains a number of features, acknowledged informative in the problem of prostate gland cancer diagnostics: a form factor, absolute and relative inertia moments, maximum, minimum, mean and root-mean-square radiuses (distances between the center and the border), maximum, minimum, mean and root-mean-square Feret diameter. Each element can optionally contain a list of coordinates of border pixels and a list of coordinates of inner pixels. This in addition allows to increase object processing speed due to memory overspend.

In the problems of object recognition on the prostate gland images in cancer diagnostics the typical element size is about 1 Kbyte. Thereafter for 10000 objects an array spends about 10Mbytes.

The algorithm that fills stated combined structure of matrix and array has time complexity O(M), where M is a number of input matrix elements. While filling the structure the algorithm calculates most of the morphological parameters of objects. The algorithm memory complexity is O(1).

The designed data organization method allows for each pixel to determine the object number with time complexity O(1) using object matrix, and allows to get all information about an object with the same time complexity O(1) using object array.

All pixels of an object can be enumerated with complexity O(N), where N is a number of pixels in the object.

3. Conclusion
The designed data structure permits to perform all operations necessary in cancer diagnostics with linear complexity. The data structure and corresponding algorithms are implemented as a program library that is used if several diagnostic systems. Modern PCs need no more than dozens of milliseconds to make an object calculation for binary matrix with 2K resolution.

Acknowledgments
This work were supported by Russian Science Foundation (RSF) - project 14-15-01120
References

[1] Siegel R L, Miller K D and Jemal A 2016 *CA: A Cancer Journal for Clinicians* 66(1) 7
[2] Chissov V I and Davydov M I 2008 (in Russian) *Oncology* (Moscow: GEOTAR-medica publisher) p 815 (Original Russian title: *Onkologia. M.: GEOTAR-media*)
[3] Prilepskaya E A et al. 2016 (in Russian) *Archives of pathology* 78(1) p 51 (Original Russian title: *Arkhiv patologii*)
[4] Naik S et al. 2007 *The Second Int. workshop on Microscopic Image Analysis with Applications in Biology* (Piscataway: DIMACS) p 1
[5] Forsea A M et al. 2014 *British Journal of Dermatology* 171(1) 179
[6] Nikitaev V G 2015 *Measurement Techniques* 58(2) 215
[7] Nikitaev V G 2015 *Measurement Techniques* 58(6) 215
[8] Nikitaev V G 2015 *Measurement Techniques* 58(4) 68
[9] Davydov M I et al. 2015 *Bulletin of the Lebedev Physics Institute* 42(8) 237