PAPER • OPEN ACCESS

Processing and analysis of botanical micro-and macroobjects using computer vision technologies

To cite this article: A V Vaganov *et al* 2021 *J. Phys.: Conf. Ser.* **2142** 012003

View the article online for updates and enhancements.

Processing and analysis of botanical micro- and macroobjects using computer vision technologies

A V Vaganov[1], O S Krotova[2] and L A Khvorova[2]

[1]South-Siberian Botanical Garden, Altai State University, 61 Lenin ave., Barnaul, 656049, Russia
[2]Altai State University, 61 Lenin ave., Barnaul, 656049, Russia

E-mail: vaganov_vav@mail.ru, kr.olga0910@gmail.com, KhvorovaLA@gmail.com

Abstract. This article discusses the problems of automation of processing and analysis of botanical micro- and macroobjects. The problem is a comprehensive interdisciplinary research aimed at recognizing and searching for botanical micro- and macroobjects using computer vision technologies. Computer vision is an area of artificial intelligence that deals with the analysis and extraction of information from images (in this case, the fern dispute). The purpose of this study is to solve a number of applied and fundamental problems in the biosystematics of botanical objects and the study of microevolutionary processes. The results of the research are stored in a specially developed database "Spore morphology of ferns of the Pteridaceae family". The database implements a set of queries relevant to biometrics tasks. The database is based on images and biometric data on the morphology of fern spores obtained in the form of measurements of systematically important signs of spores, as well as data on the shape and surface of the studied objects. The diagnostic characters of taxa of the Pteridaceae family have been standardized for solving the research problems. In the course of the study, fundamentally new knowledge was obtained that will supplement the natural classification of the Pteridaceae family, as well as test hypotheses about the relationship of a number of taxa represented on the continents of Eurasia and Africa, for an objective reconstruction of events of past geological periods. Processing and analysis of botanical micro- and macro-objects and the development of specialized software are carried out in the Python programming language using an integrated development environment for Python - PyCharm. When developing the graphical interface of the program, we used the sets of Python libraries PyQt5 and a visual tool for designing and creating graphical user interfaces (GUI) from Qt components - Qt Designer.

1. Introduction

Interdisciplinary research related to the complex solution of automation of processing, recognition and search of botanical micro- and macroobjects using computer vision technologies is aimed at solving the following applied and fundamental problems in the biosystematics of botanical objects and the study of microevolutionary processes:

1. Development of software for processing, analysis and semi-automatic measurements of biological micro- and macroobjects images.

2. Development of a database of images and measurements of botanical objects images by scanning electron microscopy for subsequent comparative analysis.

3. Development of a digital model and design of the process of recognition of botanical micro- and macroobjects using computer vision technologies, methods of image mining, machine learning and artificial intelligence.

4. Obtaining fundamentally new knowledge that will complement the natural classification (using the example of the Pteridaceae family), put forward hypotheses about the connection of a number of taxa presented on the continents of Eurasia and Africa.

This article discusses the development of a tool for performing measurements on images obtained by using microscopes, on images of virtual herbariums, entomological collections or photographs of micro - and macro-objects of the biological world made in a natural environment. Biologists face this problem during data collection and when studying botanical and zoological objects.

Biologists and ecologists who specialize in collecting biometric data with subsequent statistical processing, when studying biological objects, have a need for a universal (for macro - and micro-world objects) tool for collecting and analyzing data in the electronic form in a semi-automatic mode (combined with manual work).

Currently, there are the following programs that have similar functionality: PhotoM, Altami Studio and Image Sound. Program PhotoM [1] is intended for cytophotometry and performs calculation of optical density of photographs. You can download black and white images in the .bmp and .jpg format and change the zoom. The optical density can be calculated taking into account the background both from the average (in the selected area) and from a separate photo. In addition, it is possible to subtract the dark field of the video camera. In addition to calculating optical density, inverting, increasing contrast and smoothing the image, generating a binary image, determining the distance between objects and the area in the photo are provided. A calibration mode is provided to convert all coordinates to metric units.

Altami Studio [2] is designed for measurement, research and image processing. The features of Altami Studio include: taking measurements and processing images in real time, the ability to capture frames from the camera and control its settings, measuring the distance between two points, calculating the area and perimeter of a regular figure or object of arbitrary shape in real units of measurement. You can select a line style and color, background color, label type, and edit text in captions. All measurements carried out are displayed in the measurement table, basic statistics (average, maximum, minimum values) are calculated from them, and a histogram of the distribution by the selected value is generated.

Image Proc is also designed to process and analyze streaming and static digital images.

It should be noted that, as a rule, programs for measuring biological objects go with scientific equipment in the kit. At the same time, these programs have a number of restrictions and inconveniences that cannot be eliminated without the developer. In addition, biologists and environmentalists of the Institute of Biology and Biotechnology, employees of the South Siberian Botanical Garden, have a need to purchase their own software product that meets certain requirements, with convenient functionality and interface. The development of its own specialized software is also due to the ability to solve reverse problems in the biosystematics of botanical objects – this is the recognition of botanical micro- and macroobjects (using the example of fern spores) in order to identify and classify them by machine learning methods.

A more species-specific one than the appearance of the plant, subject to changes under the influence of environmental factors, is the sculpture of fern spores. It is a stable morphological feature. Thus, the study of the morphology of the spores Pityrogramma calomelanos (L.) Link (Pteridaceae) [3] showed that the sculpture of the spores is conservative in nature regardless of the continent and environmental conditions of growth. Another study showed a strong correlation between the nature of the spores and the size of their genome [4]. Therefore, the overall size and ultrastructural features of fern spores have greater evolutionary stability than external morphological features, and can be used to determine related relationships between taxa along with phylogenetic data.

2. Research Information Support

The object of this research is spores of ferns images of the Pteridaceae family – one of the most diverse and complex, taxonomic, fern families. To date, a significant backlog has been carried out based on three subfamilies of the Pteridaceae family: original images on the morphology of spores obtained by scanning electron microscopy have been accumulated; comparative analysis of features has been carried out to solve issues of systematics of the family. It is planned to investigate the morphology of spores on more than 200 representatives of ferns of the Pteridaceae family from Eurasia and Africa.

A fundamental aspect of this research is the acquisition of fundamentally new, conceptual information obtained by the method of mining the images of fern spores, which allow us to form a more comprehensive idea of the biosystematics of the Pteridaceae family.

The applied aspect of the research is the development of software, algorithmic and methodological support for intelligent image analysis on a training sample of fern spores, which can be used to solve automation problems when recognizing images of botanical objects. In order to improve the quality of identification and the speed of searching for the required object, proposals can be formed for the field of agrobio-industry, forensic science, paleontology and archeology: quality control of grain crops, biological protection of agricultural plants, biological and soil examinations by spore-pollen analysis, reconstruction of vegetation cover and historical events of the past.

Various resources were used to inform the research: scientific articles [5–8], global databases [9, 10] and personal data archives [3, 11]. Samples of the original data are given in figure 1.
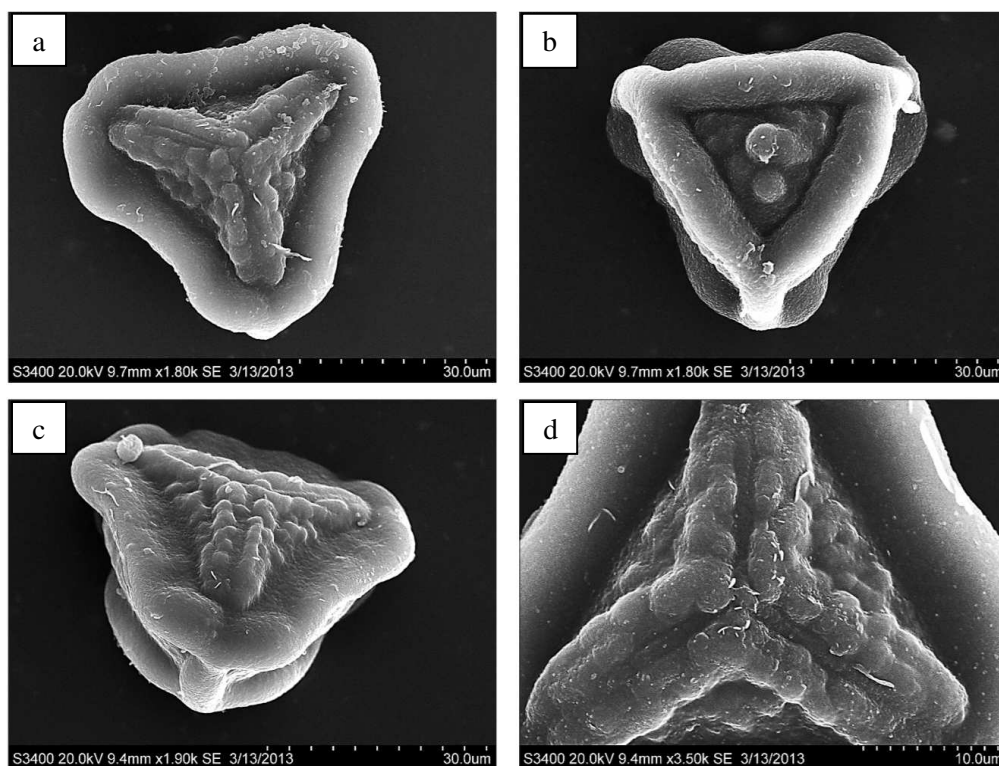
Figure 1. Electronic spores micrographs: a – proximal side of spores; b – distal side of spores; c – dispute in the proximal-equatorial position; d is a fragment of the proximal surface of the spores [10].

3. Development of Software for Image Analysis

The programming language in which the software will be developed deserves a lot of attention, since work with it will take place from the beginning of the program design to the very end. A review and

analysis of the literature showed that the choice of a programming language significantly affects the productivity of programmers and the quality of the code they create.

The development of the IT-model FAST (Functional Automated System Tool), software and algorithmic support is carried out in the Python programming language using the integrated PyCharm environment. The graphical interface of the program is developed using the Python library set PyQt5 and the GUI design and creation tool – Qt Designer.

Python is a high-level programming language focused on improving developer productivity and code readability. The Python core syntax is minimalistic. At the same time, the standard library includes a large amount of useful functions. Python supports structured, object-oriented, imperative functional, and aspect-oriented programming. The main architectural features are dynamic typing, automatic memory management, full introspection, exception handling mechanism, support for multithreaded computing, high-level data structures. It supports splitting programs into modules, which, in turn, can be combined into packages. The integrated development environment for Python PyCharm provides tools for code analysis, a graphical debugger, a tool for running unit tests and supports web development on Django.

Here are the main features of the FAST software module that meet customer specifications, as well as the additional features of the program [5]:

• ability to work in the program with images of biological objects from the micro level (micrometers, 10-6) to the macro level (centimeters);

• export measurement results to the table after manual measurements (Excel MO Office);

• supported image format – jpg, png, tiff, bmp;

• recording of fractional values of measurements – through «,»;

• image scaling;

• edit and delete entries in calibration and measurement tables;

• hot keys;

• change the thickness and color of the line.

The developed program takes measurements from images of objects of the micro- and macrocosm, structures and analyzes the data obtained, which made it possible to expand knowledge in the area under study. The main part of the main program window is occupied by the area for displaying the loaded image (figure 2). One photo can be uploaded at a time. On the left under this area there is information about the length of the current measurement (the default value is "0"). On the right, information about the applied calibration is shown (the default is "no"). Below is information about the current scale of the image and buttons for changing it. Nearby there are menus for choosing the color and thickness of the line (by default, the line color is "White"; the line width is "1"). Below there are the control buttons. They provide access to all the functions of the program.
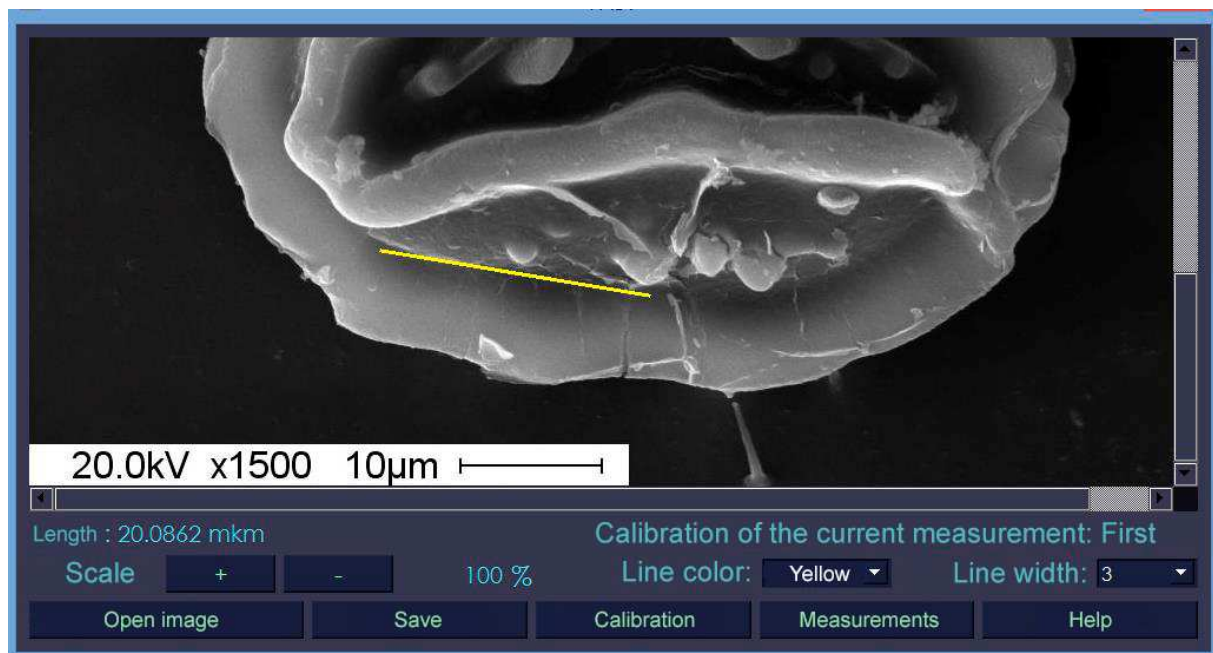
Figure 2. Sample measurement with specified calibration.

When the user changes the original size of the application window, all objects in this window are automatically adapted to the new window size.

Control panel buttons (from left to right):

• "Open image" - when you left-click on the button, the file selection dialog is displayed.

• "Save" - when you left-click on the button, a window for entering data on the performed measurement is displayed.

• "Calibration" - when you left-click on the button, a window with a list of calibration records is displayed.

• "Measurements" - when you left-click on the button, a window with a list of all saved measurements is displayed.

• "Help" - when you left-click on the button, a window with instructions for working in the program is displayed.

Images are downloaded through a file dialog with the user. Measurements are carried out by drawing a segment on two points. The measured distance is displayed below the image area on the left (default is in pixels (px)).

Control buttons are located at the bottom of figure 2. They provide access to all features of the program. If you change the original dimensions of the application window, all objects in the application window are automatically adjusted to the new window size. Calibration is provided for measuring objects. Each image has individual calibration parameters (figure 2).

Clicking the "Save" button on the main window saves the image measurement. The window is with fields for data entry (figure 3). All data is entered manually by the user, except for length and units (this data is set automatically by the program after the measurement is made).

The coordinates obtained from the spores' image can be translated into metric units (microns). To do this, you need to know the calibration factor, which depends on the lens of the photometer, the parameters of the video camera and the setting of the optical system.

Figure 3. Measurement Save Window Interface.

To calibrate the system, it is necessary to obtain photographs of the object on all used lenses and with the same resolution (image size) with which further work will be carried out. After that, it is necessary to measure the distance between the extreme risks of the subject of the photograph. The line connecting the markers must be perpendicular to the micrometer object lines. The distance between markers is automatically recorded in the "Length (pixels)" column. After clicking the "Calibration" button, in the window that opens, we select the "Add" button. In the appeared window for adding a new calibration, we enter information about the calibration: "Calibration name", "Length (meters)" and "Number of divisions". For Metric Units, we select "mkm", "mm", "cm" from the drop-down list. When you click "Save", the record is added to the calibration table and you are informed that the result was saved successfully. You can view the saved records in the «Measurements» section.

4. Conclusions

To store the results of the studies, the database "Morphology of spores of ferns of the Pteridaceae family" was developed with an implemented set of biometrics-relevant queries. The database is based on images of spores and biometric information on the morphology of fern spores, obtained in the form of measurements of diagnostic signs of spores, as well as data on the shape and surface of the examined objects. During the study, the diagnostic features of Pteridaceae family taxa were standardized. In the course of working with images, important evolutionary signs of the morphology of spores of taxa of the family from remote territories of the Old World were revealed. The acquired fundamentally new knowledge will supplement the natural classification of the Pteridaceae family, as well as put forward hypotheses about the connection of a number of taxa presented on the continents of Eurasia and Africa in order to objectively reconstruct the events of past geological eras. In addition, the developed software and methodological software for intelligent analysis of images of fern spores will be used to solve automation problems when recognizing images of botanical objects.

The developed software is actively used in the educational process and scientific research of the Institute of Biology and Biotechnology (Altai State University, Russia, Barnaul), Perm State University (Perm, Russia) and Centre for Agricultural Research (Institute for Soil Sciences, Budapest, Hungary). Using the FAST software module, the genus Schizotergitius Verhoeff, 1930, subfamily Lithobiinae, is briefly redescribed, rediagnosed, and shown to comprise only two species: Schizotergitius longiventris Verhoeff, 1930 (the type species from Tajikistan) and S. altajicus Loksa, 1978 (Mongolia). Based on type and fresh material of both sexes, S. altajicus is redescribed, and its distribution is mapped. An identification key to all eight genera of Lithobiidae occurring in Central Asia is presented [12].

References

[1]   http://www.t_lambda.chat.ru

[2]   http://www.altamisoft.ru

[3]   Vaganov A V, Gureyeva I I, Shmakov A I, Kuznetsov A A, Romanets R S and König V A 2017 *Turczaninowia* **20** 95–102

[4]   Barrington D S, Patel N R and Southgate M W 2020 *Applications in Plant Sciences* **8** e11339

[5]   Vaganov A V, Fast O V and Khvorova L A 2020 *Mathematics – Altai Krai: a collection of works of the All-Russian conference on mathematics with international participation* 152–154 (in Russian)

[6]   Irfan M, Jan G, Jan F G, Murad W 2021 *Microscopy Research and Technique* **84**(6) https://doi.org/10.1002/jemt.23835

[7]   Yu J, Wang Q X, Bao W M 2001 *Acta Phytotaxonomica Sinica* **39** 224–233

[8]   Chao Y-S and Huang Y-M 2018 *PLoS ONE* **13** e0207712

[9]   http://gis-lab.info/

[10]   https://www.google.com/search?rlz=1C1GCEU_ruRU820RU821&sxsrf=ALeKk02BUhLOz ur7CrxRNZ6kX6Le5R7Slw:1622783968864&source=univ&tbm=isch&q=%D0%B8%D0%B7 %D0%BE%D0%B1%D1%80%D0%B0%D0%B6%D0%B5%D0%BD%D0%B8%D1%8F+% D1%81%D0%BF%D0%BE%D1%80+Pteridaceae&sa=X&ved=2ahUKEwjinYuwnf3wAhVvh osKHV6NChcQjJkEegQIAhAB&biw=1536&bih=664

[11]   Vaganov A V, Shalimov A P and Shaulo D N 2014 *Plant World of Asian Russia* **2** 29–36 (in Russian)

[12]   Dyachkov Yu V, Farzalieva G Sh, Danyi L 2021 *Russian Etmological Journal* **30**(3) 361–371