UDC 004.4

An algorithm on the preprocessing optical microscopy image

Olga Deeva[1], **Marko Ćosić**[2], **Inna Kolesnikova**[3], **Sanja Zoran Despotović**[4]

[1] *National Research Nuclear University MEPhI, Moscow, Kashirskoe shosse, 31, Russia*
[2] *Institute of Nuclear Sciences Vinča, National Institute of Republic of Serbia, University of Belgrade, Belgrade, Serbia*
[3] *Joint Institute for Nuclear Research, Dubna, Str. Joliot-Curie 6, Russia*
[4] *Institute of Histology and Embryology, The University of Belgrade, Belgrade, Serbia*

Email: deeva.olga.2020@gmail.com

The task of recognizing and classifying neurons is now relevant for histological image analysis. Nowadays, the cells are being counted manually, so image processing is error-prone and takes a lot of time. That is why it is important to create a tool to help classify cells according to some mathematical properties that faithfully reflect their histological classification. In our case, the classification was done according to the multiscale structural complexity. The first step in creating such a tool is developing a special method of preprocessing optical microscopy images. Experimental images possess two features that corrupt complexity calculations: a significant variation of the light intensity from the bright center to the dark edges of the image and a lack of clarity of small objects. An algorithm has been developed to solve both of these problems, which, in essence, reduces to the determination of the average light intensity and its subtraction from the initial image. This method enhances the visibility of smaller objects in the image, by making their edges sharper. Also, thanks to subtraction, the dark edges of the image, obtained due to the microscope, were lightened, which improves the visibility of cells located far from the center of the image. Both improvements enhance the distinguishability of the complexity distributions corresponding to different histological types, then those built using raw data. The developed algorithm should be useful, in general, for the preprocessing of histological microscopy images.

Key words and phrases: histology, brain cells, image preprocessing, YC_bC_r color space, multiscale structural complexity, image sharpness

1. Introduction

The identification and categorization of neurons has become a crucial task in the realm of histological image analysis. The analysis of the damaging effects of ionizing radiation on tissues and organs is also applicable to assess the risks associated with human spaceflight. Recent studies conducted in Russia and abroad indicate that exposure of the brain structures to the radiation of heavy charged particles (HCP) and high-energy protons of cosmic origin can lead to cognitive impairment. This, in turn, entails a partial or complete loss of operator functions by the spacecraft crew members. Therefore, an accurate assessment of biological effects caused by various types of cosmic radiation ubiquitous in interplanetary flights is critical.

Currently, histological analysis is done manually. Compared to all other image processing tasks, the classification requires, the largest amount of time, thus making it severely error-prone. It is crucial to make the classification as accurate as possible because the number and the distribution of cells determine the state of the mouse brain. To streamline this process, developing a tool that can automatically classify nontrivial patterns according to their mathematical properties, which additionally reflect their histological classification, is imperative.

Pattern formation ability is one of the distinctive features of complex systems, characterized by the nonlinear interaction between subsystems. Complex patterns can be self-similar, as in the case of fractals [2], self-organized, as in the case of crystals [3], or hierarchical, as in the case of patterns generated by biological systems [4]. The origin of biological complexity is still not adequately understood, even though it inspired the development of a new branch of mathematics known as catastrophe theory [5].

To classify such patterns various measures of complexity were proposed. Two major classes are measures based on the information entropy and Kolmogorov complexity, respectively. Recently, authors of Ref. [1] proposed a novel method for assigning complexity to the data represented by matrices called Multiscale Structural Complexity. Loosely speaking, structural complexity measures the amount of information about the object retained in the sequence of images of lower and lower resolution. It has been shown that changes in the complexity values accurately reflect phase transitions in the Ising model, bifurcations in magnetic labyrinths, and phases of the dissolving process [1].

Since the MSC measures the degree of similarity between images at different scales, it is reasonable to assume that it will also be useful for detecting any nontrivial pattern presented in microscopy images. However, in some cases, the complexity distributions obtained for each histological type did not differ enough to allow unambiguous interpretation. That is why the initial step in classification involves devising a specialized method for preprocessing microscopy images.

All preprocessing aims to minimize artifacts introduced by equipment and to improve the signal-to-noise ratio. Two primary preprocessing challenges faced and solved in this study are the darkening of image edges, which appeared due to the use of a microscope, and the low clarity of cell edges, measured by image sharpness.

2. Methods and materials

The test data used for the development of the algorithm was a set of 81 histological images of the mice hippocampus Ref. [6], acquired using an optical microscope (a BiOptic B-200 light microscope and Industrial Digital Camera 5.1MP 1/2.5 with color Aptina CMOS sensor). Software development was done in Matlab programming language Ref. [7].

The fixation of the sample was in formalin. The tissues were cleaned in xylene and filled with paraffin to strengthen the structure before cutting. Cutting on a microtome allowed us to obtain high-quality samples, which were then dried. Staining was carried out using hematoxylin and eosin, which made it possible to identify almost all cells and many noncellular structures. At the same time, the nuclei of cells are basophilic and are stained purple with hematoxylin, and the cytoplasm has oxyphilia and is stained pink with eosin.

A histologist inspected all images and classified neurons into the following five types: Type 1 normal and slightly altered neuron cells; Type 2 hyperchromic neuron cells; Type 3 degenerative neuron cells; Type 4 astrocytes; and Type 5 oligodendrocytes, microglia, and vascular cells. The number of cells corresponding to each type was found to be 1971, 119, 56, 561, and 1654, respectively. The total number of positively identified cells was 4361. Next, the average size of the square window containing the cell was determined for each cell type. The dimensions of the window side were 52, 52, 55, 15, and 17, respectively.

For each region of interest (a sub-image specified by a window having $n \times m$ pixels), we have created a sequence I_0, I_1,...I_N is an original subimage of maximum resolution) of images of lower and lower resolution by the down sampling procedure. The structural complexity at the given scale C_k ($k = 1, \dots, N-1$) is defined by the expression (1)

$$C_k = \sum_{j=0}^{n} \sum_{i=0}^{m} (I_{k+1}(i,j) - I_k(i,j))^2, \tag{1}$$

where indices i, j specify a pixel's vertical and horizontal position. Multiscale structure complexity is defined as a sum of partial complexity (2)

$$C = \sum_{k=1}^{N-1} C_k. \tag{2}$$

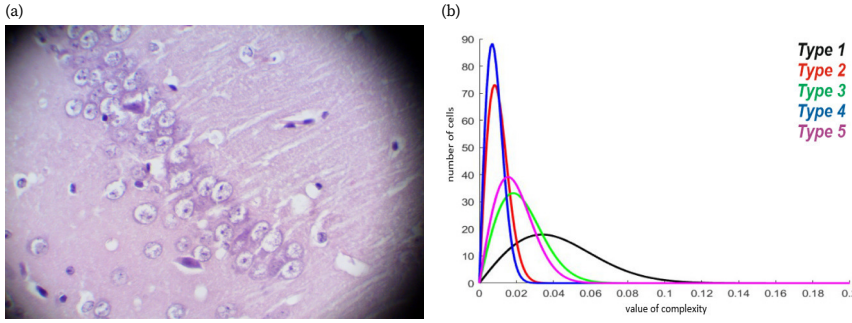

Figure 1. Raw data (a) image (b) fits of complexity distributions of 5 types.

For multi-channel images, the structural complexity was defined as an arithmetic average of the complexities of each channel (details of the calculations can be found in Ref. [1]).

The range of complexity values was from 0,0013 to 0,1416. For each type, we then constructed distributions of structural complexities. Distributions were represented by histograms with 15 bins of equal width, covering the interval [0, 0.15] in the complexity space, obtained by counting the number of cells having complexity values within the interval corresponding to each bin in the complexity space. In the end, the obtained histograms were fit by Rayleigh distribution given by expression (3)

$$f(C,\sigma) = \frac{C}{\sigma^2}\cdot e^{\frac{C^2}{2\sigma^2}}, \tag{3}$$

where σ is the scale parameter of the distribution.

In our analysis, we examined images in RGB and YC_bC_r color spaces. The RGB color space represents images as an m-by-n-by-3 numeric array whose elements specify the intensity values of the red, green, and blue color channels. Individual color components of YC_bC_r color space are luma Y, chroma C_b, and chroma C_r [8]. For transporting from RGB to YC_bC_r, we used formula (4) given in the Ref. [8],

$$\begin{aligned} C_b &= -0.169\cdot R - 0.331\cdot G + 0.5\cdot B + 128, \\ C_r &= 0.5\cdot R - 0.419\cdot G - 0.081\cdot B + 128, \\ Y &= 0.299\cdot R + 0.587\cdot G + 0.114\cdot B. \end{aligned} \tag{4}$$

Sharpness is the level of clarity of detail in a photo and is a valuable tool for emphasizing textures of subjects and subjects' details in an image. The sharpness of an image is described by two main factors: resolution and acutance. The IS parameter of an image detail s, is in Eq. (5) defined as the difference between the largest and smallest its light intensity values

$$S = \min(I) - \max(I). \tag{5}$$

3. Results

The image of row data and fits of normalized complexity histograms by a Rayleigh distribution is shown in Figs. 1a and 1b. Note that fits for type 2 and type 4 overlap significantly.

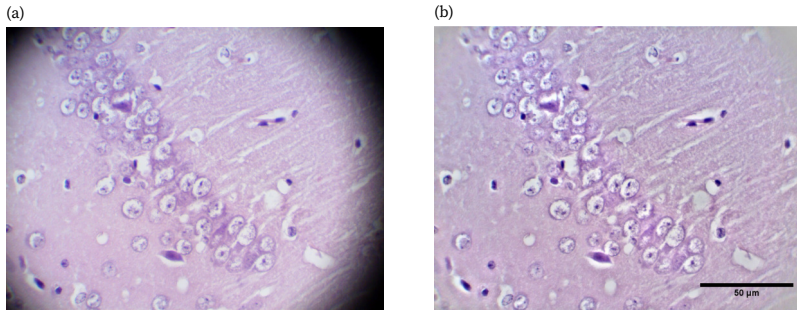

Figure 2. (a) Raw data, (b) filtered data

Similarly, distributions for type 3 and type 5 are also almost identical. Therefore, in many cases, the classification of the cells based solely on the complexity values is multivalued and thus ambiguous.

In our approach, the global IS, is defined by Eq. (6)

$$S = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\frac{I_{n+1,m} - I_{n,m}}{\Delta x} + \frac{I_{n,m+1} - I_{n,m}}{\Delta y}\right). \tag{6}$$

It actually measures an average gradient norm of the light intensity distribution I. As its representative, we have chosen the Y channel of the image's YC_bC_r color representation. To enhance the difference between pixels and increase IS, the values of image's Y channel were recalled to the interval from 0 to 255. Let Y_r represents the rescaled image, $m = \min(Y)$, and $M = \max(Y)$. The rescaled intensity is then given by the following relation

$$Y_r = \frac{255(Y - m)}{M - m}. \tag{7}$$

After rescaling, a low-pass filter was applied, implemented by the convolution method with the function of a two-dimensional rectangular window with a size of 52 pixels. The window size was determined iteratively, starting with the minimum window containing a neural cell, and was set to obtain the highest value of the IS for the final image. After that, the obtained values for Y_r of pixels were again rescaled to values from 0 to 255 using (7).

The last stage of preprocessing was the return from a single-channel image to a color one. It is an important thing to do, since color is a crucial feature when working with histological images of brain cells. There are cases when the final decision on whether a cell belongs to a particular type is made based on the color scheme of this cell. To ensure that colors are not affected by the processing, the obtained Y_r channel was combined with the C_b and C_r channels of the original image to produce a final color image.

Figs. 2a and 2b show the original and processed images side by side. First, as can be seen from the comparison, the black areas at the edges disappeared, which made it possible to see the cells that were previously in the dark. Note that due to the microscope, the area that was bright enough for analysis was only 67% of the image. The second result is that the cells became more clearly visible and had clearer edges than the initial data.

Fig. 3 shows new fits of rescaled complexity histograms by Rayleigh distribution for the processed data. The comparison of the fits shown in Fig. 1b and Fig. 3 provide the second

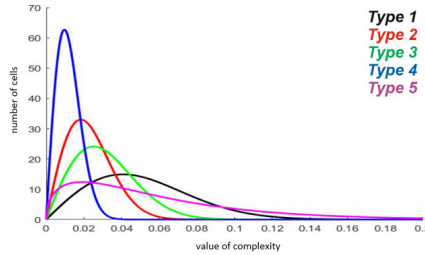

Figure 3. Filtered data. fits of complexity distributions of 5 types

proof of the algorithm's good performance. Obviously, the cell classification can now be done with significantly lower error, because the respective maxima of new fits differ more than those of the raw data.

This fact can be confirmed by quantifying how different the obtained curves are. Firstly, we defined the rectangle approximation for the distributions by the following relation (8)

$$f(C) = \max(f)\cdot\mathrm{rect}\left(\sigma_f\cdot\left(C - C_f\right)\right),$$
$$\mathrm{rect}(C) = \begin{cases} 1, & |C| \le \frac{1}{2}, \\ 0, & |C| > \frac{1}{2}, \end{cases} . \tag{8}$$

In our case the distinguishability was defined by relation (9)

$$D = 1 - \frac{\int_0^\infty g(x)\cdot f(x)dx}{\sqrt{\int_0^\infty g^2(x)dx\cdot\int_0^\infty f^2(x)dx}}, \tag{9}$$

where $g(x)$ and $f(x)$ are distribution functions for two different types approximated by rectangular functions.

The obtained values of the mutual-distinguishability parameters were summarized in Table 1). The distinguishability parameter improved in seven out of ten cases for process data. The improved ones are colored with green.

This algorithm has also been compared with an existing algorithm [9] to improve IS, that is used for superficial vein enhancement and blood vessel enhancement in retinal fundus images. As you can see, this algorithm was too aggressive, so even though the IS of the image in Fig. 4a exceeds the difference in Fig. 4b by more than 10 times (230 and 22, respectively) based on the presented images, it becomes clear that the algorithm is not applicable to the analysis of our data due to the following factors. Firstly, the black-and-white image cannot be analyzed in terms of tinctorial properties (it is impossible to determine hyperchromic and hypochromic neurons); secondly, the grain of the image does not allow us to establish where the nucleolus is located, and where other cellular elements; thirdly, clusters of glial cells or blood cells and blood vessels merge are also misleading, or even do not allow to determine the type of cell at all.

Table 1

Distinguishability in the case of: (a) the raw and (b) processed data.

(a)	Type 1	Type 2	Type 3	Type 4	Type 5
Type 1	0	0.811	0.323	0.878	0.445
Type 2	0.811	0	0.485	0.438	0.377
Type 3	0.323	0.485	0	0.608	0.074
Type 4	0.878	0.438	0.608	0	0.507
Type 5	0.445	0.377	0.074	0.507	0
(b)	Type 1	Type 2	Type 3	Type 4	Type 5
Type 1	0	0.489	0.328	0.888	0.143
Type 2	0.489	0	0.207	0.444	0.387
Type 3	0.328	0.207	0	0.62	0.224
Type 4	0.888	0.444	0.62	0	0.546
Type 5	0.143	0.387	0.224	0.546	0

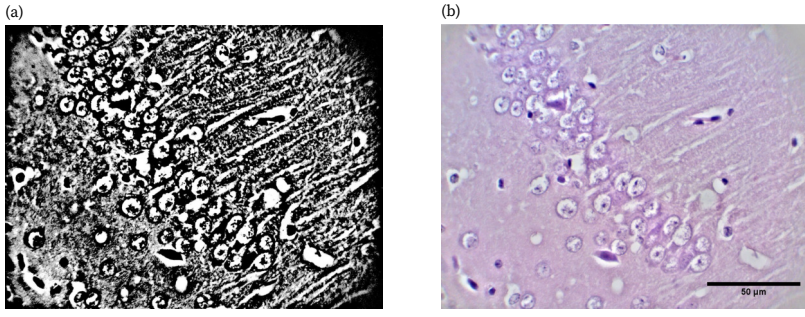

Figure 4. (a) An equivalent result obtained by application of the SOUCE algorithm, (b) The output of our image shapes improving algorithm.

4. Conclusion

A specialized method for preprocessing histological images of the brain has been developed. Thanks to this method, it was possible to improve the quality of images for analysis. The improvement includes: 1) getting rid of the darkening at the edges due to the microscope, which made it difficult to recognize the cells on them, 2) increasing the clarity of the cells, which made it possible to more accurately calculate the complexity for the cell.

Appendix: Algorithm

1. Transporting image from RGB to YC_bC_r;
2. splitting the image into channels Y, C_b, and C_r;

3. rescaling channel Y;
4. convolutional rescaled channel Y filtering using the 52-pixel rectangular window function;
5. subtraction of filtered channel Y from initial channel Y;
6. rescaling the subtraction;
7. combining channels: the one received at the point 6 and channels of color from the point 2 into one image.

References

1. A. A. Bagrov, I. A. Iakovlev, A. A. Iliasov, V. V. Mazurenko, Multiscale structural complexity of natural patterns, PANAS **117,** 48, 30241 (2020) doi: https://doi.org/10.1073/pnas.2004976117
2. B. Mandelbrot, The Fractal Geometry of Nature, (W. H. Freeman and Co., 1982)
3. P. Alberto, V. Jacques, Physics of Crystal Growth (Cambridge University Press, Cambridge, 2010)
4. A. S. Herbert, The Architecture of Complexity. Proc. Am. Philos. Soc. 106, 6,: 467 (1962)
5. R. Thom, Structural Stability and Morphogenesis (Benjamin, Reading, 1975).
6. Inna Kolesnikova, Natalya Budennaya, Yuriy Severiukhin, Maria Lalkovicova, https://ceur-ws.org/Vol-2743/7-16-paper-2.pdf Algorithmic Approach to the Recognition of Cells in the Sensorimotor Cortex from Microphotographs
7. Documentation for matlab
8. ITU-R BT.601-5 [1] and ITU-R BT.709-5 [2] standards of ITU (International Telecommunication Union)
9. Ravimal Bandara, A Fast, Simple and Powerful Contrast Enhancement Algorithm for Image Analysys

УДК 004.4

Алгоритм предварительной обработки оптического микроскопического изображения

О. Деева[1], М. Кошич[2], И. Колесникова[3], С. З. Деспотович[4]

[1] *Национальный исследовательский ядерный университет МИФИ, Москва, Каширское шоссе, 31, Россия*
[2] *Институт ядерных наук Винча, Национальный институт Республики Сербия, Белградский университет, Белград, Сербия*
[3] *Объединенный институт ядерных исследований, Дубна, ул. Жолио-Кюри, 6, Россия*
[4] *Институт гистологии и эмбриологии Белградского университета, Белград, Сербия*

Email: deeva.olga.2020@gmail.com

Задача распознавания и классификации нейронов в настоящее время актуальна для гистологического анализа изображений. В настоящее время подсчет клеток ведется вручную, поэтому обработка изображений сопряжена с ошибками и занимает много времени. Вот почему важно создать инструмент, помогающий классифицировать клетки в соответствии с некоторыми математическими свойствами, которые точно отражают их гистологическую классификацию. В нашем случае классификация была выполнена в соответствии с многомасштабной структурной сложностью.

Ключевые слова: гистология, клетки головного мозга, предварительная обработка изображений, цветовое пространство YC_bC_r, многомасштабная структурная сложность, четкость изображения