Anna PŁAWIAK-MOWNA, Małgorzata MAZURKIEWICZ

University of Zielona Góra

doi:10.15199/48.2018.01.20

On Effectiveness of Human Cell Nuclei Detection Depending on Digital Image Color Representation

***Abstract**. The paper presents results of research on effectiveness of automated detection of human body cells nuclei depending on the digital image color representation used. The problem importance is presented, data representation and processing problems are discussed. The standardized machine vision-based nuclei detection procedure is proposed. Nuclei detection effectiveness measurement algorithm is presented and results are discussed. The conclusion is drawn and future work areas are indicated.*

***Streszczenie.** W artykule przedstawiono wyniki badań skuteczności zautomatyzowanego wykrywania jąder komórkowych, w zależności od zastosowanej reprezentacji koloru przetwarzanego obrazu. Przedstawiono problemy związane z przetwarzaniem cyfrowych obrazów medycznych. Zaproponowano ujednoliconą procedurę komputerowego przetwarzania obrazu. Przedstawiono algorytm pomiaru skuteczności wykrywania jąder komórkowych w zależności od zastosowanej przestrzeni barw. Omówiono wyniki, sformułowano wnioski i wskazano przyszłe obszary badań. (**Analiza efektywności detekcji jąder komórkowych w zależności od sposobu reprezentacji danych w przestrzeni barw**).*

Keywords: image processing, image segmentation, object detection, image color analysis.
Słowa kluczowe: przetwarzanie i segmentacja obrazu, detekcja obiektów, analiza kolorów.

Introduction

Certain medical diagnosis methods widely use analysis of human body cell nuclei. One of the important areas is oncology, where the samples of tissues are used to decide about disease classification, possible treatment and expected results. The samples to be analyzed are subject to number of chemical and biological coloring processing procedures to enhance required cell and/or nuclei elements, such as membranes, nuclei and nucleoli, chromatin, etc. Diagnosis effectiveness depends strongly on numerous factors, such as repetitiveness of coloring process, experience of medical technicians and doctors, color schemes applied, and even hardware in use.

Since many years there are developed algorithms and devices to support medical image analysis and new solutions are still needed, to handle new types of images available and to produce better results. Nevertheless, there are significant problems in applying automated vision mechanisms, emerging from lack of standards in tissue samples coloring procedures, great variety of scanning devices and different color representations used in analysis process.

Nuclei-based medical diagnostic

Quick and accurate diagnosis is a key in effective treatment. Pathologists pay a key role in making decisions related to treatment [1]. The number of cancer incidences is significantly increasing [2], so development of improved methods of assisted analysis and interpretation of the test material becomes a priority. Another very important issue is the time of making decisions, which in many cases is critical to effectiveness of therapy and machine-vision based methods of image analysis can efficiently support the medical personnel.

Quantitative analysis of cytological images can be effectively used to support pathologists in classifying the analyzed material as lesion, illness or lack of thereof. Moreover, such an analysis can help to assess the disease progression. Currently used computer-based diagnostic systems directly affect the performance and accuracy of the work of pathologists in classifying diseases of cancer.

Despite the fact that methods and techniques to support diagnostic decisions has been developed since many years, there is observed need of development of new methods, to cope with new type of data, enhance effectiveness, improve operation time and precision, etc., especially in area of cancer tissue classification [3, 4, 5].

A particularly important element in interpretation of biomedical images in histology and cytology is detection and measuring the cell nuclei. Automation of this process requires an effective mechanisms of nuclei separation and evaluation of particular morphometric parameters, as in most cases the nuclei are superimposed on each other, which makes it a complex research problem itself [6, 7, 8].

Methods of automated image segmentation are subject to extensive development [9, 10, 11, 12] and new techniques are proposed, to fully explore data available and to meet the growing requirements of diagnosis support. Among other methods, there are applied machine learning algorithms, using so-called classifiers and producing important information about disease development and treatment effectiveness. Therefore, proper selection of classifiers, to capture deviations in the construction of the cell nucleus is important element of the developed image processing methods.

Data representation in cell and nuclei medical analysis

Color displays, to introduce in a viewer a sense of color, for a single picture element (pixel) use three sub-pixels, of red, green and blue colors, of which any color can be reproduced.

Basic color coding scheme, the RGB representation, is used widely, as it contains image color components in a form that is directly displayable on any color display: computer, portable device, medical equipment, etc. RGB format supplies the data in the form directly usable to present the color image.

Despite popularity and simplicity, the RGB format did not prove to be very useful in image analysis, e.g. segmentation or feature extraction. There are many other color representations used in medical imaging and analysis.

Digital images used in cell and nuclei medical diagnostics are processed and analyzed in a variety of ways to improve cell and nuclei detection and to enhance required nuclei features used e.g. for disease classification or therapy efficiency estimation. The features in interest can be of morphometric nature, such as nuclei size, elliptic parameters, irregularity measures such as ratio of the circumference to the field or convex deficiency, or of statistic nature, such as number of nuclei of particular type, number of nucleoli, etc.

Review of research materials [3 - 21] indicates there is no standard procedure or technique of acquisition of digital medical images. This is one of major problems in comparing effectiveness of particular methods of computer-aided diagnosis, as the acquisition conditions can vary strongly.

The aim of computer-aided diagnosis is to improve the nuclei detection process, and there were proposed different color representations [13, 14]. Image segmentation in cytological images were proposed to be carried on the data represented in the following formats: lung cancer: RGB, HSV, L*a*b, LUV, xyY [15, 16], breast/nipple RGB, HSV, LUV, YIQ, L*a*b, CIElab, YCbCr [7, 13, 14], prostate Lab [17], neuroblastomy RGB [17], brain cancer Luv, L*a*b [4], meningiomy CIElab [19] and lymphoblastic leukemia RGB, HSV [17].

Data processing problems

Computer-aided diagnostics can be applied in a very wide areas of medicine. In particular areas of research the body cells samples are subject to specific staining (coloring) processes, aimed to enhance required sample elements. Obtained results are scanned to produce digital images. The basic format of color digital data storage is some variant of Red-Green-Blue representation, used e.g. by Tagged Image Format (TIF). Although such a representation is very popular for its simplicity and easiness of processing, it is of some important disadvantages.

As the data are stored in very plain format, every pixel must be represented by three numbers, corresponding to Red, Green and Blue color component values. Therefore every single point must be represented by a triplet of integer numbers, and depending on the integer representation in computer memory, it occupies 24, 48 or 96 bits. For a single full HD image of 1920 by 1080 pixels the amount of memory required is over 49, 99 or 199 million bits, respectively. Medical scanning devices produce diagnostic images of huge resolution, far above simple HD, e.g. 200 000 x 80 000 pixels in multilayer TIFF structures so-called Virtual Slides. In result, the data consume considerable amount of storage space. Although, nowadays this problem is of less importance than decades ago, it is still a significant one.

The other problem is that diagnostically important information or features can manifest in a variety ways in scanned images, depending on the human body part in interest, sampling device type and characteristics, biological and/or chemical processing applied, disease type, etc.

Despite disadvantages mentioned, the RGB representation is used widely, as it contains image color representation in the form that is directly displayable on any color display.

One of the key issues in computer-aided diagnostics is selection of an appropriate color image representation to obtain required features of image contents, mainly human body cells and their elements, such as cell and nuclear membranes, nucleus and nucleolus, chromatin, etc.

The following chapter describes briefly characteristics of the most popular color representations.

Research materials and procedures

The research visual material was selected from a set of images supplied by pathologists. Images were acquired with use of Olympus VS120 Virtual Microscopy Slide Scanning System, producing virtual slides of size approx. 200 000 x 80 000 pixels, with magnification of 40 times. Breast tumor tissue samples were taken by biopsy, stained with hematoxylin and eosin and classified by pathologist as malignant tumors, benign or fibroadenoma.

As the source image is of huge size and can contain thousands of cells, for diagnostics there are selected some portions of image, with material that gives required information about optical characteristics of the cell nuclei. Such areas are called Regions Of Interest - ROI. Doctors make their decision about disease classification on basis of analysis of such ROIs.

For the research there were taken 60 ROIs, of which 41 were classified as malignant, 9 as fibroadenoma and 10 as benign. As the doctor makes his decision on basis of visual analyze of the visible cells, for research purposes there was taken only the cell nuclei that are clearly distinguishable and with all boundaries visible. Such a set of images with manually segmented nuclei areas was used as a reference for verification of results of examined machine-based image segmentation methods and color schemes. Verification procedure makes comparison of reference image produced by human versus machine-segmented one and there are counted nuclei that were properly detected. Percentage of detected nuclei is taken as the effectiveness factor.

Unified method of detecting the cell nuclei with use of machine vision algorithms

The aim of this research was to compare effectiveness of image segmentation and detecting the cell nuclei in dependence of color representation scheme used. There was designed an unified procedure of applying the machine vision segmentation techniques. The procedure is composed of six main elements: image acquisition, pre-processing (data normalization, etc.), machine-based segmentation for particular color representation, object identification, identifying detected nuclei locations and comparison of the locations with the reference image.

In detail, the procedure comprises following elements:

1) Opening the image in RGB format.
2) Storing copy of the image in grayscale form.
3) Converting image to particular color coding scheme.
4) Extracting particular color scheme component.
5) Obtaining black-and-white representation of analyzed color scheme component, thresholded with Matlab *graythresh* function that uses Otsu algorithm to determine the optimal segmentation threshold; grayscale copy of image is used in *graythresh* function.
6) Creating morphological structuring element of disk shape, as filtering window.
7) Performing operation of morphological image opening, which is achieved by imposition of expansion on the result of the erosion of the original of analyzed color component.
8) Performing operation of erosion on opened image color component.
9) Morphological reconstruction of image component, performed by repeated dilatations.
10) Performing operation of morphological image closing, achieved by imposing result of erosion on result of dilatation of the original of analyzed color component; the operation removes holes and concavities of size below the structural element.
11) Dilatation of closed image.
12) Morphological reconstruction of dilated image.
13) Transforming image into complement form.
14) Edge detection, in order to obtain better separation of disc-shaped nuclei.
15) Creating morphological structuring element of square shape, as filtering window.
16) Image segmentation, consisting of the operations of dilatation, filling, opening, creating a matrix with Euclidean distance transform of the binary image.
17) Watershed segmentation based on extended-minima transform to obtain a mask matrix with watershed

starting points, and morphological reconstruction so it only has regional minima wherever mask is nonzero.

18) Filtering the original color component image with use of watershed results, the result is an image with well-separated nuclei.
19) Labeling individual areas in image (detected nuclei) with consecutive numbers and plotting the boundaries.

Results

Nuclei detection procedure was carried on over all 60 test images, with the following color components tested:

- RGB format: all R and G and B components, RG components, RB components, GB components, and single R, G and B components, straight and converted to gray.
- HSV format: all HSV components, HS components, HV components, SV components, and single H, S and V components.
- Lab format: all Lab components, La components, Lb components, ab components, single L, a and b components and negatives of single L, a and b components.
- XYZ format: all XYZ components, single X, Y and Z components and negatives of single X, Y and Z components.
- YCbCr: all YCbCr components, single Y, Cb and Cr components and negatives of single Y, Cb and Cr components.

The results of applying the unified nuclei detection procedure (number of properly detected nuclei) were verified with results of human-based detection.

It is important to note that a pathologist does not take into account the all cell nuclei present in the material. There are cases that the nuclei are no good for diagnostics, e g. they are not clearly visible, they are overlapped or not whole nuclei is present in the analyzed image. In result the machine vision algorithms may produce false detections, both positive and negative. To verify effectiveness of particular methods it was necessary not only to count the number of detected nuclei but also verify localization of particular nuclei versus images produced by human detection, as the machine-based algorithms can detect objects that either are not nuclei at all, or such nuclei that are not diagnostic, e.g. overlapping, or visible only partially.

Fig. 1. presents graphic chart of results of machine-based nuclei detection for each color representation analyzed.

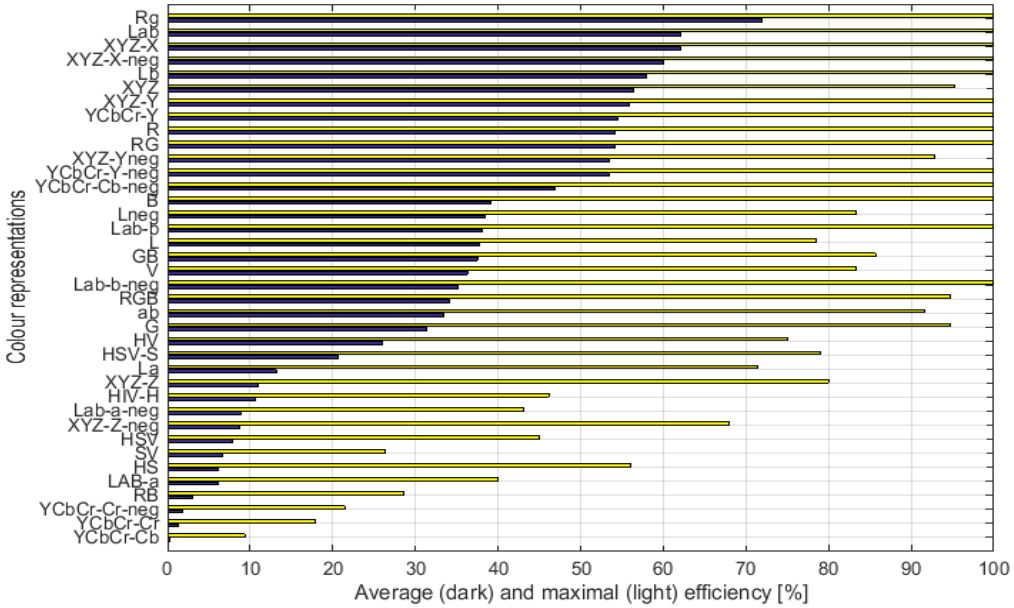

Fig.1. Average and maximal nuclei detection efficiency for particular color representations

Table 1. Efficiency of nuclei detection (average and maximal) for selected color representations

Color representation	Average efficiency [%]	Maximal efficiency [%]
Rg	71.98	100.00
Lab	62.17	100.00
XYZ-X	62.15	100.00
XYZ-X-neg	60.06	100.00
Lb	58.01	100.00
XYZ	56.47	95.23
XYZ-Y	55.81	100.00
YCbCr-Y	54.41	100.00
R	54.11	100.00
RG	54.11	100.00

Table 1 and Fig. 2. presents results of nuclei detection efficiency for ten color representations with the best performance. The most effective was representation of Red component of RGB color scheme, converted to grayscale (Rg). Average efficiency over the whole set of test images was nearly 72 percent and maximal efficiency reached 100 percent for this representation. The following three representations are: Lab, with average efficiency of over 62 and maximal efficiency of 100 percent, X component of XYZ representation, with average efficiency of very similar 62 percent and maximal efficiency of 100 percent and negative of X component of XYZ representation with average efficiency over 60 percent and maximal efficiency of 100 percent

Fig. 3. presents a ROI example showing comparison of results of nuclei detection by unified procedure based on three the best color representations: Rg, Lab and XYZ-X, compared to the results of manual nuclei segmentation by a pathologist. White circles indicate nuclei selected for analysis by human, nuclei indicated by number 1 were

automatically detected by standard procedure based on Rg color representation, number 2 indicates nuclei detected in Lab representation and number 3 designate X component representation.

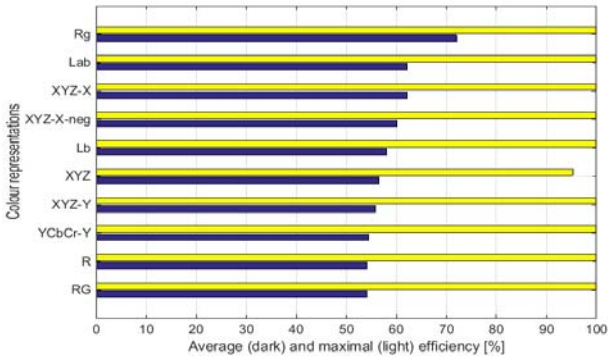

Fig.2. Average and maximal nuclei detection efficiency for the best ten color representations

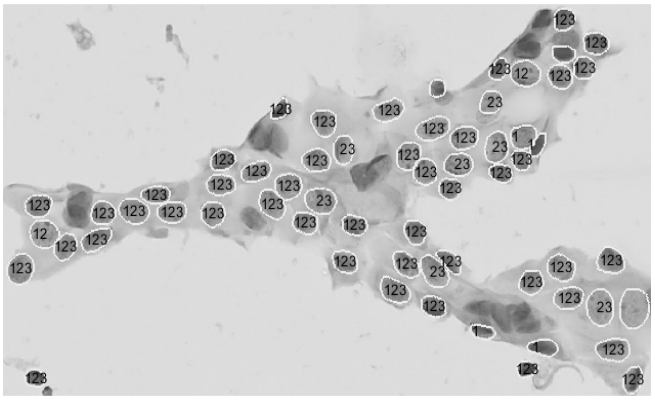

Fig.3. A fragment of ROI example showing results of human nuclei detection (white circles) and automated procedure based on Rg color representation (nuclei marked with "1") , Lab representation (marked with "2") and XYZ-X representation (marked with "3")

Conclusion and future work

In this paper there are presented results of nuclei detection obtained by standardized machine vision-based nuclei detection procedure in different color representations. Obtained results indicate that there is strong dependence in effectiveness of machine-based image segmentation results and color representation used. Therefore it is important to match properly the type of analyzed material and color scheme used, as it can affect strongly the automated detection results. Development of some hybrid mixture of color schemes would allow to improve nuclei detection results, which will be subject of future research.

***Authors**: dr inż. Anna Pławiak-Mowna, Uniwersytet Zielonogórski, Wydział Informatyki, Elektrotechniki i Automatyki, Instytut Sterowania i Systemów Informatycznych, ul. Licealna 9, 65-417 Zielona Góra, E-mail:a.mowna@issi.uz.zgora.pl; dr inż. Małgorzata Mazurkiewicz, Uniwersytet Zielonogórski, Wydział Informatyki, Elektrotechniki i Automatyki, Instytut Sterowania i Systemów Informatycznych, ul. Licealna 9, 65-417 Zielona Góra, E-mail:m.mazurkiewicz@issi.uz.zgora.pl*.

REFERENCES

[1] May M., A better lens on disease: computerized pathology slides may help doctors make faster and more accurate diagnoses, *Sci. Am.*, 302 (2010), 74-77

[2] Didkowska J. et al., Prediction of cancer incidence and mortality in Poland up to the year 2025. Department of Epidemiology and Cancer Prevention. Warsaw. Poland. http://onkologia.org.pl/wp-content/uploads/Prognozy_2025.pdf

[3] Kowal M. et al., Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images, *Computers in Biology and Medicine*, 43 (2013), 1563-1572

[4] Kong J. et al., Towards Building Computerized Image Analysis Framework for Nucleus Discrimination in Microscopy Images of Diffuse Glioma, Conference Proceedings 2011, 6605-6608

[5] Micsik T. et al., Experiences with an International Digital Slide Based Telepathology System for Routine Sign-out between Sweden and Hungary, AIMS Medical Science, 2 (2015), No. 2, 79-89

[6] Ali S., Madabhushi A., An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery, *IEEE Transactions On Medical Imaging*, 31 (2012), No. 7, 1448-1460

[7] Qi X. et al., Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set, *IEEE Trans. Biomed. Eng.*, 59 (2012), No.3, 754-765

[8] Plissiti M.E., Nikou Ch., Overlapping Cell Nuclei Segmentation Using a Spatially Adaptive Active Physical Model, *IEEE Transactions on Image Processing*, 21 (2012), No. 11, 4568-4580

[9] Kowal M., Filipczuk P., Nuclei segmentation for computer-aided diagnosis of breast cancer, *International Journal of Applied Mathematics and Computer Science*, 24 (2014), No. 1, 19-31

[10] Xu J. et al., Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images, *IEEE Trans. Med. Imaging*, 35 (2015), No. 1, 119-130

[11] Sirinukunwattana K. et al., Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, *IEEE Transactions on Medical Imaging*, 35 (2016), No. 5, 1196-1206

[12] Kumar R. et al. (2015, Jul.). Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features. Journal of Medical Engineering. http://dx.doi.org/10.1155/2015/457906

[13] Zarella M. D. et al. (2015), An optimized color transformation for the analysis of digital images of hematoxylin & eosin stained slides. Journal of Pathology Informatics. http://www.jpathinformatics.org/text.asp?2015/6/1/33/158910

[14] Jitaree S. et al., Cell type classifiers for breast cancer microscopic images based on fractal dimension texture analysis of image color layers, *Scanning*, 37 (2015), No. 2, 145-151

[15] Mohamed S. et al. Cancerous nuclei detection on digitized pathological lung color images. *Journal of Biomedical Informatics*. 35 (2002), No. 2, 92-98. http://dx.doi.org/10.1016/S1532-0464(02)00501-4

[16] Sajith Kecheril S. et al., Segmentation of lung glandular cells using multiple color spaces, *International Journal of Computer Science*, *Engineering and Applications*, 2 (2012), No.3, 147-158

[17] Amin M.M et al., Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K-Means Clustering and Support Vector Machine Classifier, *Journal of Medical Signals and Sensors*, 5 (2015), No. 1, 49-58

[18] Gurcan M. N. et al., Image analysis for neuroblastoma classification: Segmentation of cell nuclei, in Proc. IEEE 28th Annu. Int. Conf. Eng. Med. Biol. Soc., 2006, 4844-4847

[19] Anari V. et al., Computer-aided detection of proliferative cells and mitosis index in immunohistichemically images of meningioma, in 6th Iranian Conference on Machine Vision and Image Processing, 2010,1-5

[20] Irshad H. et al., Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential, *IEEE Rev. Biomed. Eng.*, 7 (2014), 97-114

[21] Basavanhally A. N. et al., Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology, *IEEE Trans. Biomed. Eng.*, 57 (2010), No. 3, 642-653