Rio´s International Journal on Sciences of Industrial and Systems Engineering and Management
Editor-in-Chief: Dr. Heitor Luiz Murat de Meirelles Quintella, Universidade do Estado do Rio de Janeiro

IMAGE RECOGNITION APPLIED TO CYTOPATHOLOGY IN THE NECK OF THE WOMB

Claudio Azevedo Passos, Emmanuel Lopes Passos
ICA: Computational Intelligence Laboratory, Department of Electrical Engineering, PUC-Rio,
R. Marques de S. Vicente 225, Gávea, Rio de Janeiro, Brazil
cpassos@ugf.br, emmanuel@ele.puc-rio.br

Keywords: Cytopathology in the Neck of the Womb, Image Recognition.

Abstract: This article describes a contribution to the modeling of an intelligent system for processing and analysis of digital images, capable of automatically detecting lesions on cells in the neck of the womb to provide support to the Cytopathologist. In order to do that, this modeling shall use Image Processing techniques and Artificial Neural Networks. Image Processing techniques shall be used to improve the images analyzed and to extract characteristics presented by them. Such characteristics shall serve as data input for the Neural Network, which shall have as its goal to classify the objects shown in the images. Thus, the modeled system for detection of sick cells, together with the Cytopathologist, is intended to decrease the number of false-positive results and increase exams productivity.

1 INTRODUCTION

Due to the vast technological advances in recent times, it is noticeable the ever increasing use of computers in the tasks of extracting information from images, to assist in solving several practical problems, such as automatic character recognition, biometrical systems (fingerprints, face, iris, etc) automatic discovery of fissures in materials, white cells count in blood samples, calculating blood volume in the heart chambers, counting microorganisms in organic samples, among others.

Image treatment resources with the aid of a computer enable us to obtain more precise information on the anatomical structure and pathology, assisting doctors to diagnose diseases that until then could not have been detected through traditional procedures (Dev, 1999) (Rhodes, 1996).

Nowadays, image-processing techniques are being widely used in the areas of Medicine and biology. Below, there are some examples of their use:

- Automatic recognition and counting particles of matter described by means of shape and size (cells, bacteria, viruses, powders);
- Analysis of chromosomes images;
- detecting tumors in X-Ray images;
- characterizing tissues using ultra-sounds;
- interpreting electrocardiograms;
- medical diagnosis

The proposed system intends to assist the Cytopathologist in identifying sick cells. For that purpose, we developed a particular modeling that uses filters to improve images, enhancing the nucleuses and afterwards cell segmentation, because, it is through morphology, texture and the nucleus/cytoplasm relation that the Cytopathologist can identify lesions in cells**.**

Currently, Cytopathology laboratories are facing a large demand for exams, causing a saturation of the resident Cytotechnician and Cytopathologist. Because of the overloading of such professionals, there is a considerable amount of false-negative results.

Due to the great variety in womb pathology, this work shall be limited to the scope of recognizing the presence or not of a lesion cells located at the neck of the womb.

2. CYTOPATHOLOGY

Cytopathology is the study of cell alterations in cases of disease, to wit, nothing more than examining the detailed structure of a small amount of cells, through smear techniques, and coming to a correct pre-operative diagnosis (Fonseca, 1975). However, the cell's structure extreme complexity justifies the imposing of exhaustive descriptions (Gompel, 1997). Where the Cytopathologist must have a certain basic morphology knowledge, to recognize, understand and interpret the cell anomalies observed during the practice of cytological method.

2.1 Collection Techniques and Specimen Preparation

In order for the material to be considered proper for examination, several types of indicators are taken into account, such as: the presence of cells (squamous and or glandular or metaplastic), preservation, coloring and mounting (collection and fixation of the material).

It is important to highlight that such technical problems in collecting material are the cause of a significant percentage of "false-negative" results, being considered as a problem at the time of collecting the material: inappropriate place for collection and insufficient amount of specimen, poor distribution of the specimen obtained, and fixation defects, consequently causing desiccation of the material to be studied.

As it was already seen, cytological specimens may be obtained through two different collection methods: *conventional cytology* and *liquid based cytology* (which was used to obtain material for this project).

Liquid based cytology came up to meet the demands of a computerized balloting, such as the one in this project. In order to enable computer reading the slides, there was the need for a product which presented the least possible number of artifacts and superposed cells.

This collection method means a great step forward in cytological products, since there is evidence that only part of the material collected through the *conventional method* is transferred onto the slide, the rest remains adhered to the brush/spatula and is discarded (Burghardt, 1998). Besides that, it has been proven that liquid based cytology has advantages compared to *conventional cervical* cytology, such as: characteristics of cyto-morphological clearness (to facilitate interpreting), better cell placement (homogenous distribution), smaller analysis area. There is also an obvious decrease in the number of red blood cells, inflammatory exudate and mucus. Another advantage is in the preparation of additional slides for the same case without the need to recall the patient, and the use of residual material for molecular tests with the purpose of identifying etiological agents such as HPV (Burghardt, 1998).

3. RECOGNITION OF IMAGES

As the careful interpreting of microscopic images demands time, and it is expensive to train cyto-technicians to process a large number of images, this work intends to assist in the interpreting of such computer images. Such system may be particularly used for triage purposes, as for instance, in identifying cells with lesions. Considering that in triage mode the probability of a true positive is relatively low and that the use of manual reading, besides being tedious, is very time consuming, the automatic image analysis system may signal abnormal or questionable cells for posterior interpretation by the cyto-technician or cytopathologist.

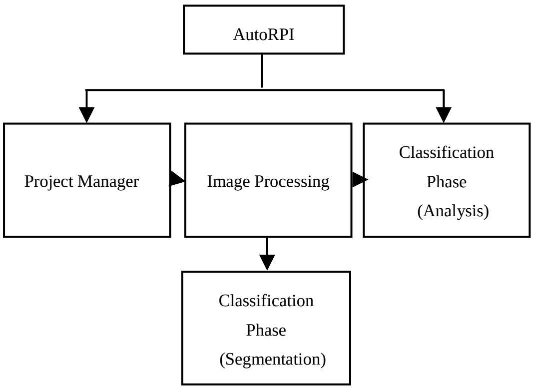
4. TOOLS FOR BUILDING THE MODEL

The tool developed here (AutoRPI[1]) allows for different combinations of artificial intelligence techniques integration, statistics classifiers and image processing for one same

[1] Friendly interface graphic environment for building solutions in pattern recognition of images found in the technical report of Claudio Azevedo Passos doctor's degree thesis, in COPPE Sistemas 2007

or different kinds of problems. AutoRPI presents, in its composition, three large function modules as shown by picture 1, described as follows:

- Project Management: The functions in this module have the purpose of assisting the user in tasks referring to the project's configuration.
- Image Processing: In this module, the functions are in charge of improvement, segmentation and extraction of image descriptors.
- Classification: Such group of functions is in charge of classifying and identifying the images analyzed. To achieve that, we have implemented Bayesian, neural and k-nearest-neighbor classifiers. Such classifiers may also be used as assistants[2] in image treatment.



Picture 1 – AutoRPI Module Diagram.

The successful use of those modules functions and consequently the generated project's efficiency are directly connected to the intelligence of the existing agents in the proposed model.

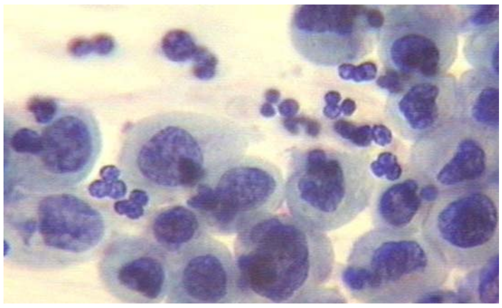
[2] In such cases, classifiers are used to assist in parameters for brightness and glare adjustment, for instance, or in classifying objects in the segmentation phase.

5. METHODOLOGY

In order to develop this project we used the fundamental stages of image processing which are: image acquisition, pre-processing, segmentation, description and recognition. The tool used in the stages of pre-processing and segmentation was AutoRPI. Following is a detailed description of all stages and how those stages were used during the proposed project.

5.1 Acquisition

To prepare a model, we used 60 samples of images acquired through a picture camera attached to a microscope. Said acquisition is preceded by a set of cytological stages consisting of techniques in collection, fixation and coloring the material. Picture 2 presents an image of cytopathology in the neck of the womb.



Picture 2 – Example the image of cytopathology in the Neck of the Womb

5.2 Pre-Analysis

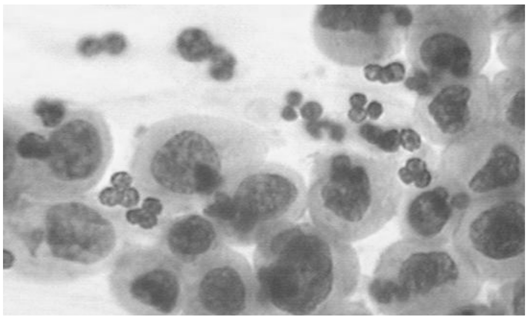
During the conduction of researches, some analysis in the image bank were performed, with the purpose of verifying several problems that could possibly hinder the pattern recognition process. Among them, we can mention:

- Difference between microscopic approximation rates;
- High level of cell superposing;

5.3 Pre-Processing

The purpose of this stage is to improve the quality of images and enhance the nuclei it contains, because it is the nucleus information that shall serve as input for the neural networks.

During the execution of specific pre-processing for identifying nuclei, the first step to be taken is the image's chromatic conversion. To wit, as the image captured by the microscope is colorful RGB 256, and the algorithms employed have no use for such information, this stage of the process consists in converting said image into 256 shades of grey. That can be observed on picture 3.



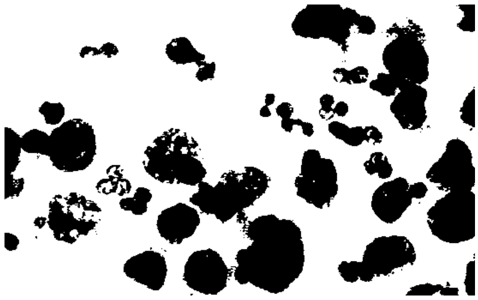
Picture 3 – Image resulting from the converting process into 256 of grey.

5.4 Segmentation

This stage is responsible for dividing an input image into its objects or constituting parts, to wit, segmentation consists of identifying and extracting homogeneous structures present in a scene.

Continuing the pre-processing for nuclei identification, the next step consists of the binarization of the resulting image, to wit, transforming the shades-of-gray images into a binary image, verifying the intensity values of *pixels* for the gray-shades images in order to decide if they received the black or white value.

As some of the nuclei will not be utilized, due to their being compromised by image residues which are not part of any nucleus, there may still be cases where the nuclei are superposed, rendering their utilization impossible, for in such cases they invaded the established diameter for one cell. That can be observed on picture 4.



Picture 4 – Nuclei resulting from the Segmentation process.

5.5 Post-Processing

Once the image is binarized, the next step is to tag the nuclei, to wit, placing an identification for each cell shown on the image. After placing the tags on the image, all objects on the image (superposed cells, leukocytes and cell fragments) which have a diameter out of the established pattern for a cell must be discarded, and it is at this point that calculations are made over each cell.

5.6 Descriptors

This process, also designated as characteristics selection, intends to extract image characteristics that result in some quantitative information of interest, or that are fundamental for discrimination (Gonzalez, 2001).

The descriptor used in the project was the quantization, after the nuclei has been tagged, those which are fit for research shall be identified, each fit nucleus shall have its range of pixels divided into groups of eight (where those groups shall be used in the Neural Network input layer), each group of 8 shall receive 32 different pixels values. The chart below explains how that division was made:

Chart 1. Neural network input attributes

1st Group	group pixels which vary their value from 0 up to 31
2nd Group	group pixels which vary their value from 32 up to 63
3rd Group	group pixels which vary their value from 64 up to 95
4th Group	group pixels which vary their value from 96 up to 127
5th Group	group pixels which vary their value from 128 up to 159
6th Group	group pixels which vary their

	value from 160 up to 191
7th Group	group pixels which vary their value from 192 up to 223
8th Group	group pixels which vary their value from 224 up to 255

Such descriptors shall serve as input for the Neural Network that shall have as its goal the sick cells identification process, in this project, we shall have two distinctive classes, with their respective characteristics, and they shall be named Sick cell class and Normal cell class.

5.7 Classifiers by Neural Network

The concept of Artificial Neural Network (ANN) has been successfully applied into the area of pattern recognition in digital images, by its skill in learning from examples. Therefore, the proposed project shall be focused on its use as a classifier. The knowledge acquired from the training set is extracted and stored in the neurons' connections (synapses), during the learning phase.

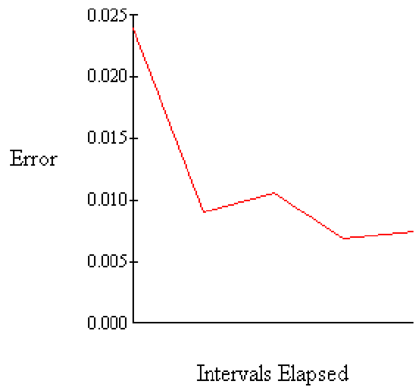
The architecture chosen was "Networks Fed Directly with Multiple Layers", where said network has nine (eight referring to the groups and one referring to the total of pixels in the nucleus) neurons on the first layer (input vector), 13 neurons on the second layer (first layer hidden) and 1 neuron on the third layer (output vector). To normalize the data, a division was made for the total value of pixels in the nucleus.

The criterion for stopping defined at the training stage was by minimum error average, configured for 0.0001 and the learning rate used was 0.1 and momentum 0.1.

6. RESULTS

The project's modeling was applied to a sample of 60 images.

Picture 5 shows the error rates presented on the neural network's tests.



Picture 5 – Errors graphic extracted from NeuroShell® 2 release 3.0.

Of the 60 specimens analyzed, 48 were for training the network and 12 for tests, the rate of correct interpretation was at 88 %.

7. OTHER RESEARCHES

In Brazil, works on recognition of cells in the neck of the womb with lesions are yet scarce, due to the still modest use of liquid based cytology. Some studies are already being carried out in the field of automatic cell recognition, such as project for "Computational Vision in the Diagnosis of Leukemia" developed by UNISantos teacher Daniela Mayumi Ushizima. Thiago Figueiro and others have presented a method for automatic detection of red blood cells in medical images using a correlation of images and a flood map. Vinicius Vilac, A. Reis and Silvio Jamil F. Guimarães presented a paper on Quantification of Cancer Cells in Microscopic Images by means of Digital Image Processing.

8. CONCLUSION

Image processing algorithms and the neural networks architecture chosen in the project's modeling have shown efficiency in performance and results.

Results may improve with the acquiring of new images arising out of the liquid based cytology process, a primary requirement for computerized balloting and automatic identification of cells with lesions or not.

9. REFERENCES

Burghardt, E.; Pickel, H.; Girardi, F.. 1998. Colposcopy Cervical Pathology. Stuttgart, Thieme, 323 p.

Dev. P. Imaging and Visualization in Medical Education. 1999. IEEE Computer Graphics and Applications. 19(3):21-31, Maio/Junho.

Fonseca, Dr Nisio Marcondes. 1975. Atlas de Citopatologia Ginecológica. Rio de Janeiro, ed. Atheneu, Instituto de Ginecologia. UFRJ. 228p.

Gompel, Claude e Leopold G. Koss. Citologia Ginecológica, e suas bases anatomoclínicas. Tradução de Ignacy Wulkan. Bela Vista (SP), ed. Manole, 1997. 200p. Original Francês.

Gonzalez, Rafael C. e Woods, Richard E. 2001. Digital Image Processing, 2º ed, Addison Wesley Pub.

Rhodes M.L., D.D. Robertson. Applicattions in Surgery and Therapy. 1996. IEEE Computer Graphics & Applications.

.