2009 Ninth International Conference on Intelligent Systems Design and Applications

THREECOND: An Automated and Unsupervised Three Colour Fuzzy-Based Algorithm for Detecting Nuclei in Cervical Pap Smear Images

Fabio Vaschetto
DELEN – Dipartimento di Elettronica
Politecnico di Torino
Torino, Italy
S146042@studenti.polito.it

Eduard Montseny
Computer Engeneering and Industrial Systems Dept.
Universitat Politècnica de Catalunya
Barcelona, Spain.
eduard.montseny@upc.edu

Pilar Sobrevilla
Applied Mathematics II Department
Universitat Politècnica de Catalunya
Barcelona, Spain.
pilar.sobrevilla@upc.edu

Enrique Lerma
Pathology Department, Sant Pau Hospital
Universitat Autonoma de Barcelona
Barcelona, Spain.
elerma@hsp.santpau.es

***Abstract*—Visual examination and interpretation of microscopic images taken from the cervix are at the core for the detection and prevention of cervical cancer. However these visual processes are tedious and in many cases error-prone. This is why automated screening systems, interacting with the technologist, would be a tremendous improvement for reducing the likelihood of human errors.**

In this work we propose THREECOND, a three colour-based algorithm that integrates colour information, cyto-pathologists knowledge and fuzzy systems. This algorithm is designed to be integrated into the previously developed system [23], with the aim of improving its accuracy and efficiency for detecting and segmenting the nuclei of Pap smear images.

Keywords-color image analysis; cervical cancer screening; cytology; pap test; image understanding, fuzzy techniques

I. INTRODUCTION

Cervical cancer is the second most commonly diagnosed cancer in women worldwide; exceeded only by breast cancer [1]. The incidence and mortality related to cervical cancer can be reduced if this disease is detected at its precancerous state, known as Squamous Intraepitelial Lesion (SIL) [2-3].

The Pap test, or cervico-vaginal cytology [4], is globally the most used and suitable for screening of precursor lesions of cervical cancer [2], with a significant impact in reducing the incidence and mortality rates [5]. The test, that is simple and risk free, has two stages. At first stage a large number of cells obtained by scraping the cervical epithelium are smeared onto a slide which is then stained and fixed with lacquer to prevent the alteration of the cells. At second stage the cells are examined under a microscope by trained screening laboratory staff (cytologists or cytotechnicians), and the results are reported indicating whether the cells are normal or atypical.

Unfortunately, Pap smear test for cervical cancer detection suffers from subjective variability and no specificity [6]. The most controversial point of the cervico-vaginal cytology remains the persistence of false negatives (cases where pre-malignant or malignant cells are diagnosed as normal), which rate can vary from 5% to 55%. Approximately two thirds of false negatives are due to errors in the taking of samples, while the rest are caused by fault detection or interpretation in the laboratory [7].

The main factors that contribute to the high false negative rate are: the huge number of cells to be inspected in each sample (50,000–300,000 cells per slide), the tedium and fatigue associated with the current manual mode of performance, and finally the existence of the samples that are relatively difficult to process and may be classified differently by different experts.

Therefore, it exists a demand for a highly accurate, non-specialist and relatively cheap method to improve the sensitivity (correct classification percentage on the precancerous tissue samples) of the Pap smear.

Many health organizations such as College of American Pathology embrace an automated solution to cervical smear screening. An automated classification system could not only reduce the time required for sample classification, but also avoid misclassification of samples because of fatigue or other types of human error.

It should be borne in mind that the goal of automation-based cervical cytology is that the percentage of normal samples that are being classified as such must be as high as possible but getting an acceptable sensitivity degree. To achieve this goal it has to be considered that the basic concepts underlying the Pap test are: morphological characteristics of nuclei and cytoplasm, and nucleus/cytoplasm relation.

Moreover, an algorithm allowing automate the process should not just provide the automatic segmentation of the cells, but also integrate expert knowledge to design the proper rules for the classification of the cells, and consider the variability inherent to both the images and the rules that have to be considered for an accurate classification.

In this paper we present part of our ongoing work towards automation of cervical smear screening process. Specifically, based on the facts that in cytological studies nuclei are considered as the most informative regions, and that an accurate segmentation is needed for extracting meaningful cell features, we present a fuzzy-based automatic and unsupervised segmentation algorithm that allow extracting nuclei region from background in colour images.

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.212

II. BACKGROUN ON COMPUTERIZED SYSTEMS FOR ELIMINATING FALSE-NEGATIVE PAP SMEARS

The false-negative Pap smear is the major quality issue currently facing the practitioners of diagnostic cytology. As a report of no abnormal cells equates to a negative test, meaning that affected woman just need to follow-up in one year, a failure detecting cervical disease entails delay in diagnosis that can lead to progression of the disease and the need for more aggressive treatment.

Computer-assisted devices can reduce false negative Pap smear interpretations using computerized systems to assist the cyto-technologist in identifying Pap smear abnormalities and providing added value in their ability to consistently and objectively analyze all cells on slides without fatigue.

In the last years some computer vision based techniques have been introduced to improve screening and interpretation accuracy through improvements in sampling method [8-12]. These tools allow removing almost all the subjectivity of the process, which represents a qualitative leap in the ability of diagnosis and analysis.

Some tools to assist cyto-technologists have recently been introduced, these are: PAPNET, Autopap, and Thinprep.

AutoPap is a computerized scanning system for the primary screening of cervicovaginal smears. The system screens all slides and an initial diagnosis is made before the suspicious or abnormal slides are passed for review by the manual system. Slides diagnosed as normal by this technology are usually either archived or submitted to some quality control procedure such as rapid review.

The PAPNET is a re-screening system that combines the speed of low-level programming with the decision-making of neural networks. When using this system, once the entire Pap smear has been re-screened, the 128 "cell scenes" judged as most significant are selected for visual evaluation by trained technicians.

Many works have been presented trying to establish the real value of these methods. Some of them are in favor of its use because provide a good quality control, better detection of false positives, improving the time of screening, or increased sensitivity to diagnose injuries low-grade [13-16]. Others are against its use by not detect infections, reduce the identification of ASCUS adding cost, do not improve inter-observer variations, not increase the quality of manual screening, the greater influence of false-positive compared with false negatives, and low sensitivity for recognizing endocervical cells [17-22].

We have been involved in automating part of the screening process for several years. In a previous work [23] we developed an automated fuzzy-based cell screening detection system suitable for interacting with the human technologist. Taken into consideration the process followed by the technicians the process was carried out into two differentiated steps: (1) Detection and evaluation of the areas of interest (AOI), and (2) Nuclei detection.

The results of the tests carried out provided very accurate results in the case of AOI detection, and promising results in the case of nuclei detection. Although the number of false positives obtained using this system is low, the number of false negatives should be reduced.

In order to improve the accuracy and efficiency in segmentation and classification of our previous system, here we propose a new nucleus detection and segmentation system thought to be integrated in our previous algorithm for improving its efficiency.

III. PROPOSED APPROACH

Accurate cell nucleus segmentation is crucial for the development of automated cytological cancer diagnosis system, because nuclei features are heavily used in cell classification standards such as The Bethesda System (TBS) [24]. As most of the malignant characteristics of cells are contained in cell nucleus, the isolation of cell nucleus is an important part of segmentation for this kind of cell images.

Nuclei detection and segmentation is carried out analyzing the AOIs obtained using our previous algorithm [23]. To get these regions the monochrome image of the slide is scanned with a low magnification lens (2.5X objective) in order to cover the entire scene rapidly. To intelligently eliminating normal images (the *"hay"*) a degree of interest - μ_{AOI} -will be also associated to each region according with technicians' knowledge and information. So, normal images can be eliminated and suspicious images saved and evaluated for further processing.

A. Nuclei Detection and Segmentation

Once areas of interest are located, they are reviewed at high magnification (10X objective), emulating the human method of screening a slide. This process is carried out on colour images because they hold more information, what make easier both, nuclei detection and the analysis that will be performed in subsequent stages for identifying the kind of cells appearing within the image.

We consider color representation because prevents non-homogeneity problems due to illumination and shadows, allowing color recognition process to be independent of illumination [25]. Moreover, it has to be taken into consideration that within cervical smear images there is blood, which appears as red stains that in a monochrome image is converted into a dark grey-level very similar to the one of the cells nuclei, so obstructing their location. The process is based on fuzzy techniques, so that the natural variability of color data is well accommodated.

1) Colour Representation: Colour-order systems based on perceptual variables are more convenient for computer vision applications, as they are somehow correlated with human being's colour perception [25].

Among the perceptual colour models we have selected the Smith *HSI* colour model [26] because, besides its close relationship between chromaticity and how humans perceive colour, it decouples the intensity component (*I*) from the colour information (*Hue* -*H*-, and *Saturation* -*S*-).

Unlike our previous algorithm, wherein a palette of 92 colours defined by their *H* and *S* values and distributed over the *H–S* map was considered, for developing the new module of our algorithm we have considered only three colours.

This decision was taken after conducting a thorough and comprehensive analysis of more than ninety images; some of them are shown in figure 1. From this analysis, according with the information provided by the cytologists, the most relevant information contained within colour Pap smear imagery is provided by considering pink, blue, and transparent cells.

2) Nuclei Detection and segmentation: The final aim of this stage is twofold: to locate the cells feasible of being abnormal, and to account up the identified normal cells. Then, the potentially abnormal cells will be located to be analyzed and identified in a following stage.

To achieve this objective, the first step consists in precisely locating the cells and their boundaries to obtain a good description allowing their identification.

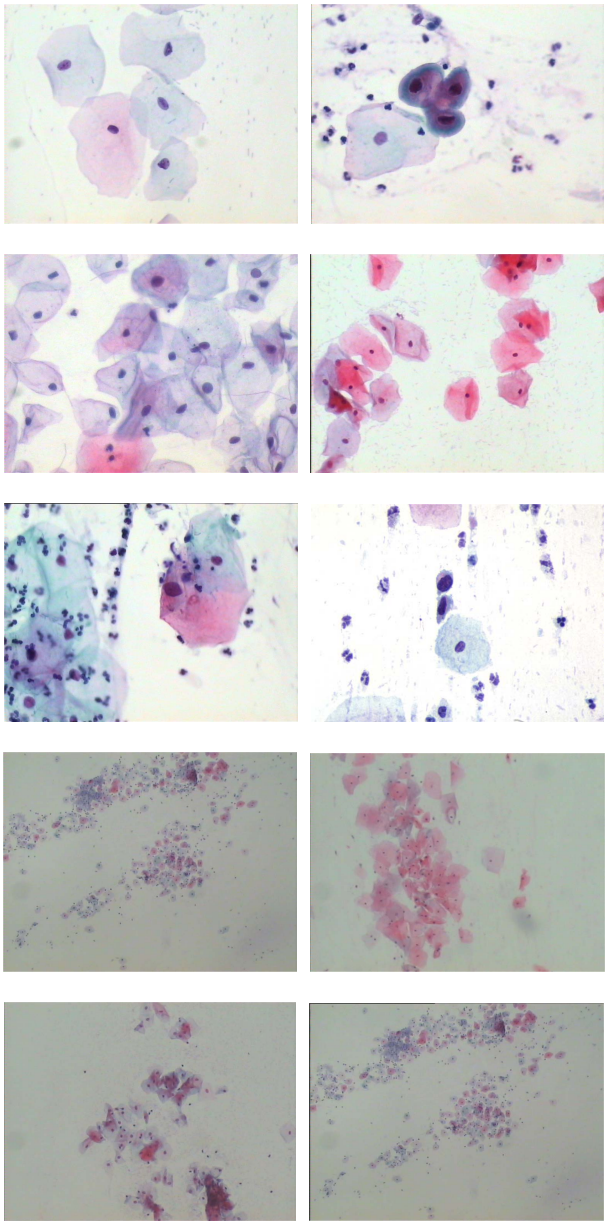

Figure 1. Exemples of real images.

In this work we propose THREECOND, a three color-based algorithm that integrates colour information, cyto-pathologists knowledge and fuzzy systems. This algorithm is designed to be integrated into the previously developed system [23], with the aim of improving its accuracy and efficiency for detecting and segmenting the nuclei of Pap smear images.

The main steps of our algorithm work as follows:

a) Getting the characteristics of pink, blue, and transparent cell nucleus: To do it we start from a training set of Pap smear colour images containing cells of the three colours. In order to get the characteristics that fit better the properties of the cells' nuclei, besides considering images wherein cells appears isolated, and overlapped, we have chosen images containing vaginal secretions, inflammatory cells, cell-clumbing, blood staining, and other cell sampling and preparation artifacts.

- After obtaining the *Hue*, *Saturation* and *Intensity* values from training images, these were analyzed, taken into consideration experts' knowledge looking for the characteristics of both, contour and inner part of the nucleus.
- From this analysis it was realized that the saturation component can not be considered because it is not stable for our purposes, i.e. does not provide neither relevant nor discriminant information.

b) Obtaining the membership functions of nucleus and edge of nuclei for pink, blue and transparent cells. From the above study, the functions for H and I for each case are given in Figure 2. Figures 2-(a) to 2-(c) are the membership functions for nuclei of pink, blue and transparent cells, respectively. Membership functions of pink, blue and transparent cells' edges are depicted at figures 2-(d) to 2-(f), respectively. It has to be pointed out that Hue membership functions takes care of its circularity property.

Then, for each pixel p_{ij} belonging to the area of interest, its membership degrees to nucleus -μ_{n*}- and edge -μ_{e*}- of pink, blue and transparent cells, we aggregate the corresponding *H* and *I* membership functions by the product.

So, for example, given a pixel p_{ij}, if $\mu_{nph}(p_{ij})$ and $\mu_{npi}(p_{ij})$ are their degrees to pink cell nucleus for *H* and *I* components, respectively, then the degree to which p_{ij} is *pink cell nucleus* is given by $\mu_{np}(p_{ij}) = \mu_{nph}(p_{ij}) \cdot \mu_{npi}(p_{ij})$.

c) Extracting the candidates for nucleus and edge of cell. As this step is carried out in the same way for the three types of cells, for simplicity we are going to explain the case of pink cells.

- *Candidates for nucleus*: Considering a 3x3 window that drags on the image, we look for clusters of pixels -*Cnp*- satisfying following conditions.
  - *i.* $\forall\, p_{ij} \in Cnp$: $\mu_{np}(p_{ij}) > 3/5$, and
  - *ii.* $2 \leq \#(Cnp) \leq 50$.

  All the obtained clusters will be considered as *candidates for pink cell nucleus*.

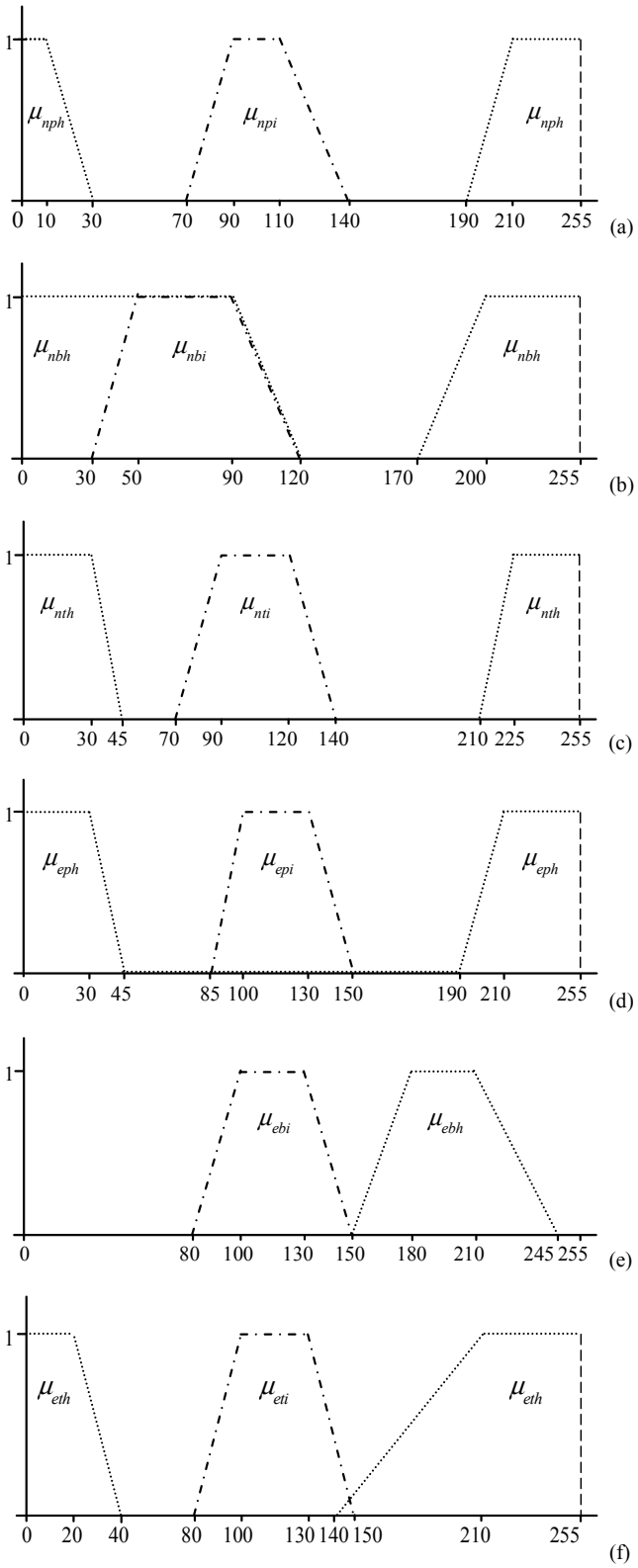

Figure 2. Hue and Intensity membership functions considered for nucleus and edge. Graphs a-c correspond to cell nucleus pink, blue and transparent, respectively. The functions for pink, blue and transparent cell edge, are depicted by graphs d to f, respectively.

- *Candidates for edge*: For each cluster *Cnp* obtained at previous step, its *candidate for edge* -*Cep*- will be constituted by all the pixels surrounding it.

d) Nuclei obtaining.. The purpose of this step is identifying, among the *candidates for nucleus*, those that can be considered as such. To do it, given a candidate, Cn_* ($*\in \{p, b, t\}$), we analyze the pixels belonging to its *candidate for edge* - Ce_* -, and we will say that Cn_* *is nucleus* (and we will denote N_*) if $\mu_{e*}(p_{ij}) > 3/5$ for at least half of the pixels of Ce_*.

IV. Results and Conclusions

The nuclei detection step has been tested on 15 cervical smear colour images from a database collected at the Hospital de Sant Pau de Barcelona, Spain. These images were obtained with a system formed by a JVC TK-C1381 video colour camera attached to a BH2 Olympus microscope and connected to a PC. In each slide, the nuclei detection was tested with a 10X objective.

The results of the tests carried out have been very satisfactory, providing very promising results. As a matter of example, figure 4 depicts the results obtained for the smear image of figure 3, which contains pink, blue and transparent cells. The first column of the figure depicts the candidates to nucleus, second column corresponds to candidates for edges, and third column shows the final results.

As can be observed, having a look at images appearing at first row of third column, in the case of blue cells (N_b) all nuclei have been detected, and there is only a possible false detection (marked with an arrow), which corresponds to the nucleus of a transparent cell that is partially hidden by a blue one. A similar result has been obtained in the case of pink cells (second row –N_p-), wherein our algorithm has detected a nucleus that corresponds to cell that is in transitional cellular state (marked with an arrow).

In the case of transparent cells (third row –N_p-), the algorithm has not detected one nucleus, that corresponds to the cell partially hidden by the blue one.

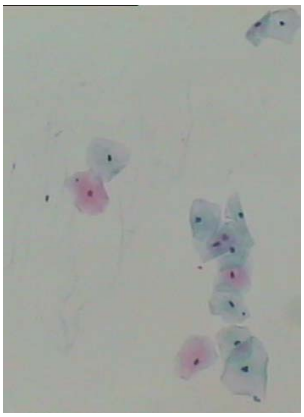

Figure 3. Original Image.

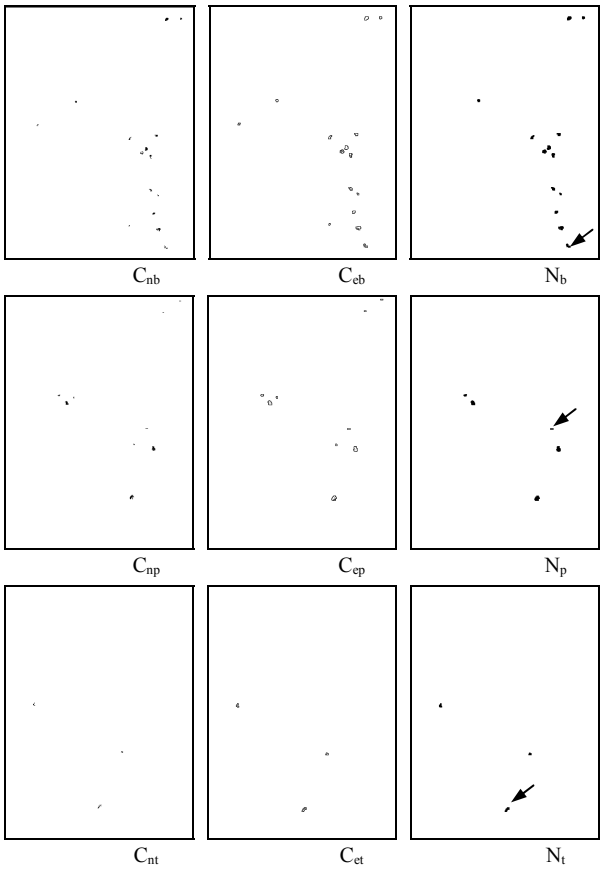

Figure 4. Results obtained for the image of figure 3. From top to button appear the results for blue, pink, and transparent cells, respectively. First column depicts the candidates to nucleus, second column corresponds to candidates for edges, and third column show the final results.

Moreover, there is a false detection corresponding to the nucleus of a pink cell (marked with an arrow).

The problems in the case of transparent cells, and possibly in the other cases are due to the algorithm is still in developing phase, and nuclei obtaining step has to be improved.

We have defined a quantitative measure that gives some insight into the effectiveness of the proposed algorithm. Our *Effectiveness measure*, η_* ($* = p, b, t$), is defined as follows:

$$\eta_* = \frac{\boldsymbol{CN}_*}{\boldsymbol{FD}_* + \boldsymbol{RN}_*} \qquad (1)$$

where $\boldsymbol{CN}_*$ is the number of detected nuclei that are of type $*$, $\boldsymbol{FD}_*$ is the number of false detections, and $\boldsymbol{RN}_*$ is the total number of type "$*$" nuclei that appear at the image. The results obtained for some of the tested images appear at rows two to seven of table 1. The last row of the table shows the average values obtained for the efficiency considering all the test images.

Given the way the effectiveness of the algorithm is assessed, in cases where the number of cells is very small, if there is some false detection and/or some nucleus is not detected, the resulting value decreases rapidly. This is the case of transparent nuclei for images 1 (that is the image of figure 3), 3, 4, and 6.

TABLE I. EFFECTIVENESS VALUES OBTAINED FOR SOME OF THE CONSIDERED TEST IMAGES

Image	***Blue***	***Pink***	***Transparent***
1	0.923	0.833	0.5
2	--	0.843	--
3	0.883	0.823	0.634
4	0.918	0.827	0.673
5	--	0.832	0.761
6	--	0.892	0.631
Total	**0.922**	**0.861**	**0.632**

In general, the efficiency values obtained for transparent cells are quite low, what is due to the absence of hue for these cells.

It has to be pointed out that although the results presented in our previous algorithm were quite good, the problems appeared with some of the images considered at present system.

The proposed algorithm for *Nuclei Detection* has proved a good performance and efficiency, specially considering that, in general, the number of false detections has been lower than the 0.5% for all the tested images. This has been in part due to the effectiveness of fuzzy techniques to deal with vagueness problem, but allowing to integrate experts knowledge.

At present we are working on improving the last step of the algorithm. In particular we are working to improve the way the candidates for cell contour are considered so as to take into account the problems that arise in the case of transparent cells.

ACKNOWLEDGMENT

This work has been partially supported by the Spanish CICYT Project TIN2007-68063.

REFERENCES

[1] D.M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global cancer statistics, 2002", CA Cancer J Clin, 55(2), 2005, pp.74–108.

[2] G.F. Sawaya, A.D. Brown, A.E. Washington, and A.M. Garber, "Clinical practice. Current approaches to cervical-cancer screening", N Eng J Med, 344, 2001, pp.1603-160.

[3] P. Sasieni, J. Adams, and J. Cuzick, "Benefit of cervical screening at different ages: evidence from the UK audit of screening histories", Br J Cancer, 89, 2003, pp.88–93.

[4] G.N. Papanicolaou, N.F. Traut, Diagnosis of uterine cancer by the vaginal smear, New York: Commonwealth Fund; 1943.

[5] S.M. Selvaggi, "Implications of low diagnostic reproductibilit of cervical and histologic diagnosis", Journal of the American Medical Association –JAMA-, Vol.285, 2001, pp.1506-1507.

[6] L. Denny, L. Kuhn, A. Pollack, H. Wainwright, and T. C. Wright, Jr., "Evaluation of alternative methods of cervical cancer screening for resource –poor settings," Cancer, Vol. 89, 2000, pp. 826-833.

[7] M.H. Stoler, and M. Schiffman, “Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytology and histologic interpretations. Realistic estimates from the USCUS-LSIL Triage Study”. Journal of the American Medical Association –JAMA-, 285, 2001, pp.1500-1505.

[8] J.C. Bezdek, and M.A. Sutton, Image Processing in Medicine, Applications of Fuzzy Systems, ed. H.J. Zimmerman, Kluwer, Norwell, MA, 1999.

[9] D.N.Chun and H.S.Yang, “Robust Image Segmentation Using Genetic Algorithm with Fuzzy Measure”, Pattern Recognition, vol. 29, no. 7, 1996, pp1195-1211.

[10] P. Kostic, B. Reljin, I. Milosavljevic, S. Cirkovic, I. Reljin, and I. Rakocevic, “CAMIA-Computer-Aided Medical Images Analyzer”, Proc. 10th Mediterranean Elechtrotechnical Conference (MELECON 2000), vol.2, 2000, pp.486 – 489.

[11] Li Zhong, and K. Najarian, “Automated classification of Pap smear tests using neural networks Neural Networks”, Proc. International Joint Conference on Neural Networks (IJCNN '01), vol. 4, July 2001, pp.2899 – 2901.

[12] V. Van Raad, “Design of Gabor wavelets for analysis of texture features in cervical images”. Proc. 25th Annual Int. Conf. of the IEEE Engin. in Med. and Biol. Soc. (IEEE EMBS), vol. 1, 2003. pp:806 – 809.

[13] C. Bergeron, “Quality control of cervical cytology in high-risk women. PAPNET system compared with manual rescreening”, Acta Cytol., 44, 2000; pp.151-155.

[14] M.A. Duggan, “PAPNET-assisted, primary screening ofcervico-vaginal smears”, Eur. J. Gynaec. Oncol., 21, 2000; pp.35-42.

[15] E. Lerma, Ll. Colomo, A. Carreras, E. Esteva, M. Quilez, and J. Prat, “Re-screening of atypical vaginal smears with PAPNET”, Cancer Cytopathology, 84, 1998; pp. 361-365.

[16] G. Ronco, et al., “Impact of the AutoPap (currently Focalpoint) primary screening system location guide use on interpretation time and diagnosis”, Cancer, 99(2), Apr 2003, pp.83-88.

[17] W. J. Meerding, et al., “Cost analysis of PAPNET-assisted vs. conventional Pap smear evaluation in primary screening of cervical smears” Acta Cytologica, vol. 45 num. 1, 2001.

[18] K. Losell, and A. Dejmek, “Comparison of papnet-assisted and manual screening ofcervical smears”, Diagn. Cytopathol., 21, 1999; pp.296-299.

[19] A. Cosentino, et al., “An Interlaboratory Study of the use of PAPNET in the quality control of cervico-vaginal cytology”, Pathologica, 91, 1999, pp.101-106.

[20] D.H. Grohs, “Impact of automated technology on the cervical cytologyc smear. A comparison of cost”, Acta Cytology, vol.42, 1998, pp. 165-170.

[21] B.H. Willis, P. Barton, P. Pearmain, S. Bryan, and C. Hyde, “Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK”, Health Technol Assess, 9(13), Mar. 2005, pp.1-207.

[22] M. Confortini, et al., “A feasibility study of the use of the AutoPap screening system as a primary screening and location-guided rescreening device”, Cancer, 99(3), Jun 2003, pp. 129-134.

[23] P. Sobrevilla, E. Montseny, and E. Lerma, “A Fuzzy-based Automated Cells Detection System for Color Pap Smear Tests –FACSDS–”, Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, Springer, Berlin / Heidelberg, 2008, pp. 657-674.

[24] R. J. Kurman, and D. Solomon, The Bethesda System for reporting cervical/vaginal cytologyc diagnosis. Springer-Verlag. N. York. 1994.

[25] S. Romani, E. Montseny, P. Sobrevilla, “An Approach to a Fuzzy Color Detection Method, Robust with Regard to Variable Illuminant Level Conditions”, Proc. 2002 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS-FLINT 2002), 2002, pp. 69-74.

[26] A. R. Smith, Color gamut transform pairs, IEEE Transactions on Computer Graphics, vol. 12, 1978, pp. 12-19.