# Automatic segmentation of cervical cells in Pap smear images

Omelkhir Boughzala
Laboratory of Technology
and Medical imaging
(LTIM)
Monastir 5000,
Tunisia
University of Monastir
omelkhirboughzala @
gmail.com

Lamia Guesmi
Laboratory of
Technology
and Medical imaging
(LTIM)
Monastir 5000,
Tunisia
University of Monastir
lamia_guesmi0107@
yahoo.com

Asma Ben Abdallah
Laboratory of
Technology
and Medical imaging
(LTIM)
Monastir 5000,
Tunisia
University of Monastir
assoumaba@
yahoo.com

Mohamed Hédi Bedoui
Laboratory of
Technology
and Medical imaging
(LTIM)
Monastir 5000,
Tunisia
University of Monastir
MedHedi.Bedoui@
fmm.rnu.tn

## ABSTRACT

In the context of medical diagnosis by image analysis, segmentation is the most critical step in image processing. The problem of image segmentation has been studied for years and many methods have been suggested in the literature. However, there is not yet any automatic method able to correctly process any type of image. In this work, we present an automated method for cell segmentation in Pap smear images. The automatic analysis of Pap smear images is one of the most interesting fields in medical image processing. The object of this paper is to present the strategy of the first part of the system segmentation. It is based on a segmentation of color images tested with different classical color spaces, namely RGB, L*a*b, HSV, and YCbCr, to select the best color space using k-means clustering to separate groups of objects. The k means clustering treats each object as having a location in space. The method is aimed at developing an automated Pap smear analysis system which can help cytotechnologists reduce examination time in pap screening process.

## Keywords

Pap smear, medical image, processing, cervical cancer detection, cytologic screening, K-means clustring

## 1. INTRODUCTION

The Pap smear is a technique of cervical screening used to detect pre-cancerous changes of the uterine cervix. It is not designed to find cancer or any abnormalities of any organs. These changes are called cervical intraepithelial neoplasia (CIN) [O14]. They are obtained by opening the vaginal canal with a speculum and scraping the cervix with a wooden stick and a tiny brush, which are scratched on a glass slide in order to collect the cells. The collected cells are then examined by a cytotechnologist.

At present slide examination of cervical cytology image is performed manually. The cyto-technologist looks at a glass smear under the microscope and analyses the full image in order to determine the presence of disease. He / She is involved in the diagnosis of cancer, pre-cancerous lesion, benign tumors and infections from various body sites. Cytotechnologists work under the direction of a physician called a pathologist, a medical doctor who specializes in the study of diseases and determines the nature and cause of the problem. Most patients never meet the pathologist or the cytotechnologist who evaluates their samples, yet the treatment of their illness may depend on the work of the pathologist- cytotechnologist team. Screening consists in locating and visually assessing all the present cells on a slide. This mainly aimed at detecting abnormal or suspicious cells to reach a diagnosis. This is of a major interest for the cytotechnologist whose diagnosis will depend on the good recognition of the abnormal or suspicious cells during screening. This cytological manual screening is described as intense and complex. The results are based on the point of view of a human being. It is often tiring to screen for relatively small and abnormal cells, which is time-consuming and requires high concentration. [B03].For all these factors, some errors appeared and can cause false negatives [NAH97]. These errors are considered as being inseparable of the process of manual screening. A promising approach is to help the cyto-pathologist in his / her search for abnormal cells on a glass. An automatic system could contribute to the

detection of screening errors and thus allow a better reliability of the diagnosis. Such a system is programmed to verify the results of conventional screening and possible false negatives. This would prevent delays allow considerable time gain.

The automatic segmentation system of cervical cells in Pap smear images is the most crucial step for any system. Segmentation is aimed at detecting the cells along with their nuclei and cytoplasms (see figure1) [LEC+98]. The shape and report area of the cytoplasm and nucleus are two important factors in detecting pre-cancerous changes in the uterine cervix [CHL+13]. Due to the complexities of cells many studies have focused on segmentation of cytopathological images [PN12] [DUK08].In the literature, some methods used thresholding techniques [KSW07] [LN07]. Many methods of segmentation of the nucleus and the cytoplasm focused on edge detection [LCC09], CHL+13]. For example [RNRT12][PN11] adapted the mathematical morphology to segment both the nucleus and the cytoplasm of a single cell. Other works utilized the genetic algorithm[LH03] and the deformable template [GP00]. A watershed transform has been applied in [MKI]. Recently Sajeena T A and Jereesh A S proposed a framework for automatic analysis of single cellular pap smear slides[SJ15].Fuzzy C-means (FCM) clustering technique is proposed for single cell segmentation[KSN15].

In this paper, we propose a method of 2D color cytological image segmentation using the color information as a priori information. The full image is segmented into 3 regions: the nuclei, the cytoplasm and the background. When we consider color image segmentation, choosing a proper color space becomes the most important issue. In this work, a segmentation of color images is tested with different classical color spaces: RGB, HSV, L*a*b, and YCbCr, to select the best color space for the considered kind of images. The segmentation process is based on the K-means segmentation technique
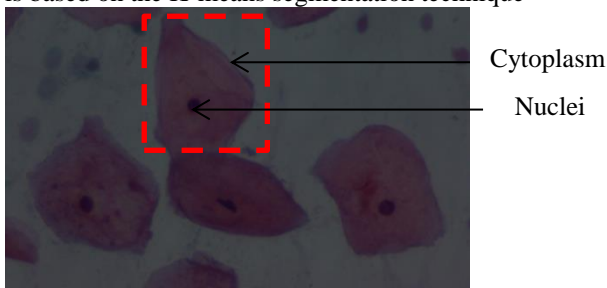


**Figure 1 Macroscopic image of cytology**

## 2. Method

The slides are examined by a microscope to which a color camera is fixed. The obtained images are color images of 1030X1300 pixels. It is necessary for us to isolate their cytoplasm and nuclei at the same time:

the cytoplasm to obtain information to characterize isolated cells and the nuclei to characterize the cell and estimate the wickedness. The cytoplasms and the nuclei will help the recognition of different cells. To carry out the segmentation step we should know the nature and the context of the images. Our color images present cells from the cytology of the cervical sample colored by the international coloration standard of Papanicolaou[GK96][MCL91]. Cells have a blue nuclei and a green cytoplasm with the exception of the red blood cells which are totally colored in red. The spatial configuration of the cells and their color are extremely variable. There can be isolated, attached but also heap cells which can overlap (nuclei or cytoplasm). The color of the nuclei can vary from very pale blue to very dark blue. This big variety in the spatial configuration and the color of cells raises problems of segmentation and requires a method of fine and strong segmentation at the same time

### 2.1. General presentation of the method

The strategy of segmentation which we organized is the segmentation of color images by the k-means method.
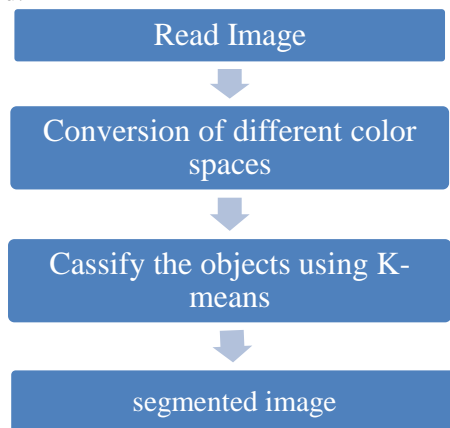


**Figure 2 the color image segmentation strategy Pap smear cytology**

### 2.2. Conversion in different spaces of colors

A color space combines numbers to the visible colors. Generally, it represents a color by a triplet of values. The visible colors can be seen as belonging to a three-dimensional space. There are many color spaces, each with its properties. The conversion of the image in different spaces of colors is the preliminary step to the achievement of the segmentation. Each color space has interesting properties and presents an interest for such an application. Generally the choice of the color space is done when the segmentation method is established. Several works were focused on comparing different color spaces [KEHT14] [LEC+98][LEC03]. Others chose to use multiple color spaces at the same time [MK12] [BA12] [M08].

### 2.2.1 RGB color space

Many color spaces are in use today. For pictures captured by digital cameras, the most popular one is the RGB model. This is an additive color model in which the colors red, green, and blue are combined together in different manners to reproduce a broad array of colors. This color space-based segmentation is not accurate for computer vision applications. The RGB color is device-dependent i.e the same signal or image can be viewed differently with different devices, Nunobiki and all reported the usefulness of RGB color specification in analyzing the variation of color properties for Papnicolaou-stained cervical smears. [NST+02].

### 2.2.2 HSV color space

The HSV color space (Hue, saturation, Value) define a model in terms of its components. The space has the ability to separate the intensity of the color information hue and saturation. For this reason it has been adopted for processing images having brightness variation characteristics. Many works use this space [OTDC02]. The conversion from RGB to HSV

$$H=\begin{cases} 0 & \text{if } \max(R,G,B) \\ 60*\dfrac{G\text{-}B}{\max(R,G,B)\text{-}\min(R,G,B)} & \text{if } R=\max(R,G,B) \text{ et} G{\geq}B \\ 60*\dfrac{G\text{-}B}{\max(R,G,B)\text{-}\min(R,G,B)}+360 & \text{if} R=\max(R,G,B) \text{ et} G{<}B \\ 60*\dfrac{R\text{-}G}{\max(R,G,B)\text{-}\min(R,G,B)}+240 & \text{if} B=\max(R,G,B) \\ 60*\dfrac{B\text{-}R}{\max(R,G,B)\text{-}\min(R,G,B)}+120 & \text{if} G=\max(R,G,B) \end{cases}$$

$$S=\begin{cases} \dfrac{\max(R,G,B)\text{-}\min(R,G,B)}{\max(R,G,B)} & \text{if } \max(R,G,B) \neq 0 \\ 0 & \text{if } \max(R,G,B) =0 \end{cases}$$

$$V = \max(R, G, B)$$

### 2.2.3 L*a*b color space

A L*a*b color space is a color opponent space with dimension L for lightness, the a* layer indicates where the color falls along the red green axis, and b* layer indicates where the color falls along the blue-yellow axis. The L*a*b* color space includes all perceivable colors, which means that its gamut exceeds those of the RGB and CMYK color models (for example, ProPhoto RGB includes about 90% all perceivable colors). The most important feature of this color space is that it ,is device-independent, that is to say provides us with the opportunity to communicate different colors across different devices [RKV12]. The solution to convert images from the RGB space to the L*a*b* color space is given by the following formula. The conversion from RGB to XYZ is:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 0.618 & 0.177 & 0.205 \\ 0.299 & 0.587 & 0.114 \\ 0 & 0.056 & 0.944 \end{bmatrix} * \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

The transformation from XYZ to Lab is performed with the following equations:

$$L^* = 116(Y/Y_n)^{\frac{1}{3}} - 16$$
$$a^* = 500 \left[ \left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \right] \quad (3)$$
$$b^* = 200 \left[ (Y/Y_n)^{\frac{1}{3}} - (Z/Z_n)^{\frac{1}{3}} \right]$$

### 2.2.4 YCbCr color space

The YCbCr color space is used in digital video image processing. It represents colors in terms of one luminance component (Y) and two chrominance components ( Cb and Cr). The Cb component is the difference between the blue component and a value, where Cr is the chrominance red component. In contrast to RGB, the YCbCr color space is luminance- independent, that is why it gives better performance.[BTP+11]The transformation used to convert from RGB to YCbCr color space is shown in equation (4):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} * \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$



**(a)**      **(b)**

**(c)**      **(d)**
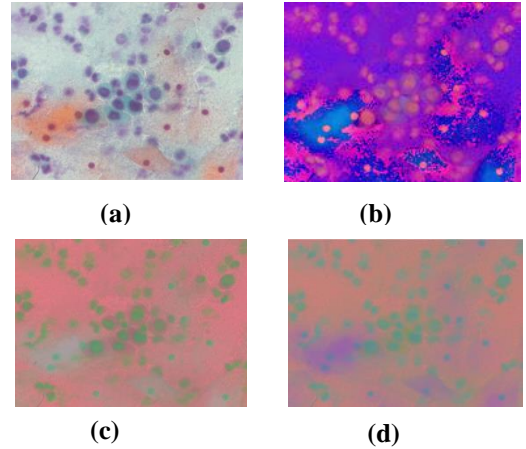
**Figure 3 (a) Image RGB (b) Image HSV (c) Image L*a*b (d) Image YCbCr**
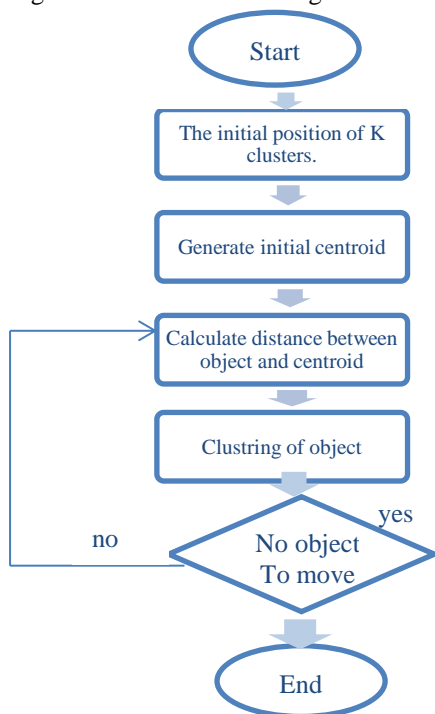
### 2.3 Image Segmentation Using K-means

Clustering is a process to separate groups of objects. The algorithm k-means is the algorithm of the most known and most used clustering, because of its simplicity of implementation [MK12]. K-means clustering treats each object as having a location in space. It allows to partition the data of an image K clusters. Contrary to other hierarchical methods, which create a structure in "tree of clusters "to describe the groupings, the k-means creates only a single level of clusters. The algorithm sends back a partition of the data, in which similar objects are

placed as close as possible to each other in the same cluster, and the different ones are placed as far as possible in another cluster. The k-means is an iterative algorithm which minimizes the sum of the distances between every object and the centroid of its cluster. The final result depends on the centroids' initial position. Therefore, the centroids must be placed as far as possible from each other so as to optimize the algorithm. K-means changes the objects of cluster until the sum cannot decrease anymore. The result is a set of clusters that are compact and clearly separated provided that the best K value of the number of clusters is chosen. The main stages of the algorithm k-means are (see figure 4) :
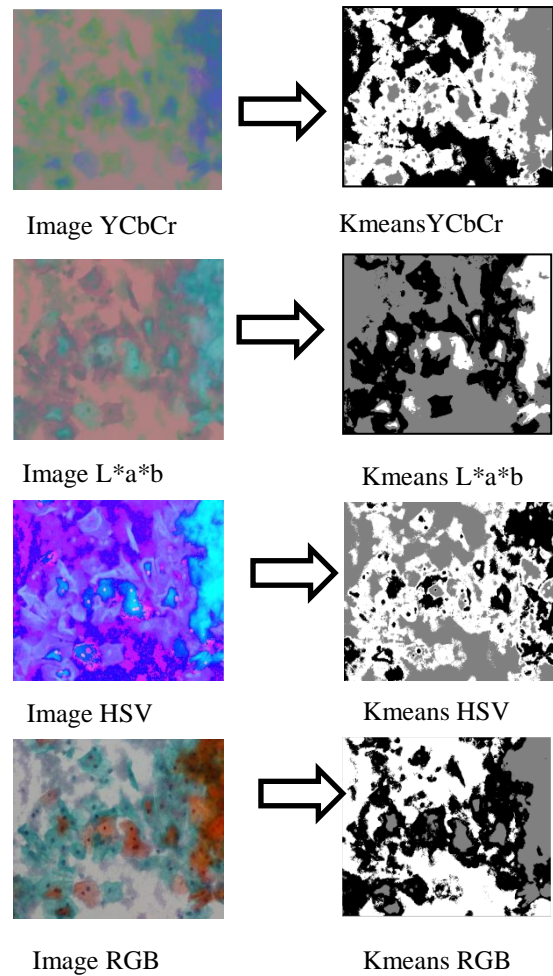
1. Random choice of the initial position of K clusters.
2. Allocate objects to a cluster following a criterion of minimization of the distances (generally according to a measure of Euclidian distance).
3. Once all the objects are placed, recalculate K - centroids.
4. Repeat stages 2 and 3 until no more reallocations are made.

We are going to look at each of the color spaces according to their influence on the algorithm of k-average and then we will present a method for choosing a color space.

Since the color information exists in the different color space our objects are pixels with channels values. Use k-means to place the objects into three clusters using the Euclidean distance metric. For every object in the input, k-means returns an index corresponding to a cluster. Label every pixel in the image with its cluster index figure 5.



**Figure 4 Steps of the k-means algorithm**



Image YCbCr      KmeansYCbCr

Image L*a*b      Kmeans L*a*b

Image HSV      Kmeans HSV

Image RGB      Kmeans RGB

**Figure 4 Example of Kmeans segmentation with different color space YCbCr,L*a*b, HSV and RGB.**

## 3. Results and discussion

The automatic K-means technique was experimentally tested using a dataset of 18 different images. It was applied with different color space models, including *RGB*, HSV, L*a*b and YCbCr. The features that identify each image pixel are only the values of its three components in the selected color space. We took the k values (number of clusters) as 3 for the K-means algorithm and the distance metric chosen is cosine. For every object in input, k means returns an index corresponding to a cluster. Label every pixel in the image with its cluster index. We can separate objects in an image by color, which will result in three images (color background, color cytoplasm and color nuclei) see figure 6.

For evaluation, a mathematic expression -described in (7) and (8)- is proposed to obtain the percentage of amount of cells that are in the input image. The percentage nuclei and cytoplasm segmentation are calculated as the fraction of pixels with segmentations of K-means technique divided by the total number of pixels in ground truth. The results

obtained are shown in tables 1 and 2. The color space L*a*b arises as giving the highest one (see Table 1). This color space is therefore more suitable for the segmentation of the nuclei. The color space L*a*b is done on representative images of the nuclei of the cells. The nuclei of the cells have an extremely variable color. We can see that the RGB space is the one that obtains better results for the segmentation of the cytoplasm table 2

$$nuclei\ \% = \frac{nuclei color pixels}{image pixels in ground truth} * 100 \ (7)$$
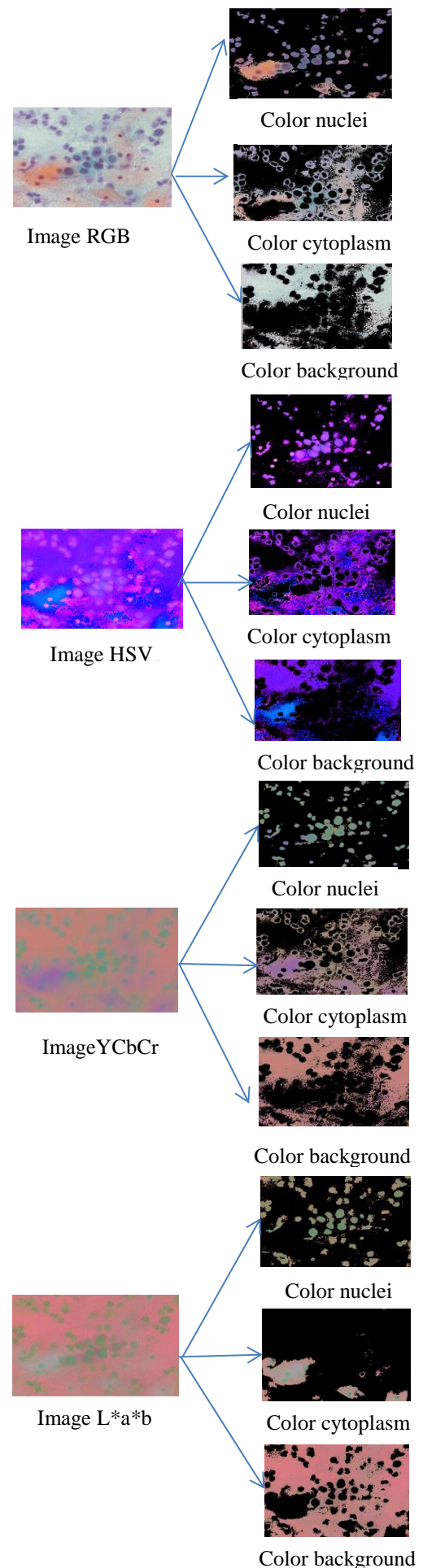
$$cytoplasm\ \% = \frac{cytoplasm\ color\ pixels}{image\ pixels\ in\ ground\ truth} * 100 \ (8)$$

**Table 1: Experimental nuclei segmentation results on different color spaces using K-means.**

| Image | cell percentage % | | | |
|---|---|---|---|---|
| | RGB | HSV | L*a*b | YCbCr |
| 1 | 98.43 | 34.43 | 99.06 | 53.94 |
| 2 | 40.63 | 36.27 | 36.21 | 53.79 |
| 3 | 79.70 | 68.31 | 65.25 | 97.20 |
| 4 | 14.72 | 14.50 | 14.38 | 12.17 |
| 5 | 78.92 | 61.10 | 79.80 | 26.63 |
| 6 | 52.59 | 45.39 | 56.38 | 98.93 |
| 7 | 93.39 | 47.51 | 52.56 | 70,59 |
| 8 | 43.67 | 99.40 | 40.04 | 43.51 |
| 9 | 80.16 | 37.40 | 53.64 | 85.93 |
| Avg | 64.69 | 49.36 | 82.88 | 60.29 |

**Table 2 Experimental cytoplasm segmentation results on different color spaces using K-means**

| Image | cell percentage % | | | |
|---|---|---|---|---|
| | RGB | HSV | L*a*b | YCbCr |
| 1 | 99.06 | 40.74 | 94.83 | 96.64 |
| 2 | 61.71 | 65.21 | 64.62 | 99.79 |
| 3 | 99.02 | 98.30 | 99.23 | 86.32 |
| 4 | 21.52 | 25.57 | 20.65 | 12.82 |
| 5 | 99.04 | 98.2 | 97.60 | 98.20 |
| 6 | 68.17 | 93.91 | 62.79 | 45.34 |
| 7 | 98.01 | 92.72 | 99.04 | 61.27 |
| 8 | 97.14 | 50.00 | 82.77 | 77.65 |
| 9 | 92.45 | 87.40 | 97.03 | 89.51 |
| Avg | 81.79 | 72.45 | 79.84 | 74.17 |



Image RGB

Color nuclei

Color cytoplasm

Color background

Image HSV

Color nuclei

Color cytoplasm

Color background

ImageYCbCr

Color nuclei

Color cytoplasm

Color background

Image L*a*b

Color nuclei

Color cytoplasm

Color background

**Figure 5 Image Obtained After K -Means Clustering**

## 4. Conclusion and Future work

the performance of the K-means clustering is evaluated using four different color spaces, RGB, HSV, L*a*b and YCbCr. The experimental results showed that the segmentation results depending on the L*a*b color space provided the best nucleus segmentation. The average rate of correct segmentation was 82.88% for nucleus segmentation in the RGB color space. This was the best result compared to other color spaces. The average rate of correct cytoplasm segmentation was 81.79%. The present study can be improved by enlarging the dataset and including different kinds of images.

As future work, we will incorporate other heuristics such as size to improve the first phase. We will also experiment with nucleus and cytoplasm classifications using different classifiers.

## 5. REFERENCES

[BA12] Patel Janak kumar Baldevbhai , R. S. Anand, *Color Image Segmentation for Medical Images using L*a*b* Color Space,* IOSR Journal of Electronics and Communication Engineering, 2278-2834 Volume 1, Issue 2, May-June 2012

[BTP+11]Jorge albertomarcialbasilio , gualbertoaguilartorres, gabrielsánchezpérez, l. Karina toscano medina , héctor m. Pérez meana, Explicit Image Detection using YCbCr Space Color Model as Skin *5th WSEAS international conference on Computer engineering and applications,* 978-960-474-270-7 2011

[B03] E. Bengtsson, Computerized cell image analysis: past, present, and future, in: Proceedings of 13th Scandinavian Conference on Image Analysis, 2003, pp. 395–407.

[CHL+13] Yung-Fu Chen, Po-Chi Huang, Ker-Cheng Lin, Hsuan-Hung Lin, Li-En Wang, Chung Chuan Cheng, Tsung-Po Chen, Yung-Kuan Chan, and John Y. Chiang. Semiautomatic Segmentation and Classificationof Pap Smear Cells. *IEEE journal of biomedical and health informatics, vol. 00, no. 00, 2013*

[DUK08] C. Duanggate , B. Uyyanonvara , and T. Koanantakul. A Review of Image Analysis and Pattern Classification Techniques for Automatic Pap Smear Screening Process. *International Conference on Embedded Systems and Intelligent Technology Bangkok, Thailand*2008

[GP00] A. Garrido*, N. PeHrez de la Blanca. Applying deformable templates for cell image segmentation.ELSEVIER Pattern RecognitionVolume 33, Issue 5, Pages 821–832 2000,

[GK96]Claude Gompel, LeopoldG.Koss, Cytologie Gynécologique et ses bases anatomo-clinique *Paris Pradel 1996*

[KSW07] Kwang-Baek Kim, Doo Heon Song, and Young Woon Woo. Nucleus Segmentation and Recognition of Uterine Cervical Pap-Smears.*Rough Sets, Fuzzy Sets, Data Mining and Granular Computing Springer Berlin Heidelberg.*pp 153-160 2007

[KEHT14] Dina Khattab, HalaMousherEbied, Ashraf Saad Hussein,andMohamedFahmy Tolba1, Color Image Segmentation Based on Different Color Space Models Using Automatic GrabCut, *Hindawi the Scientific World* Journal 2014,

[KSN15] S. Kaaviya, V. Saranyadevi ; M. Nirmala. PAP smear image analysis for cervical cancer detection,*Engineering and Technology (ICETECH), 2015 IEEE International Conference on1 – 4 2015*

[LCC09]Chuen-Horng Lin, Yung-Kuan Chan, Chun-Chieh Chen. Detection and Segmentation of Cervical Cell Cytoplast and Nucleus.*International Journal of Imaging Systems and TechnologyVolume 19, Issue 3, pages 260–270- 2009*

[LEC+98]Olivier Lezoray, AbderrahimElmoataz, Hubert Cardot, Gilles Gougeon, Michel Lecluse, et al..Segmentation of cytological image using color and mathematical morphology.*European conference on Stereology, 1998, Amsterdam, Netherlands.*10 pp, 1998.

[LN07]Zhong Li, KayvanNajarian. biomedical image segmentation based on shape stability. *Image Processing, 2007.ICIP 2007. IEEE International Conference on* VI - 281 - VI – 284-2007

[LH03]N. Lassouaoui, L. Hamami. genetic algorithms and multifractal segmentation of cervical cell images. *Signal Processing and Its Applications, 2003.Proceedings. Seventh International Symposium on1 - 4 vol.2- 2003*

[LEC03]O. Lezoray A. Elmoataz, and H. Cardot1. A color object recognition scheme: application to cellular sorting Machine Vision and Applications 14: 166–171 2003

[MCL91]Michel Mallet, Dominique Chiarasini, Sylvain Labbe, Cytologir Gynécologique Normale Et Pathologique, :*PICCIN 1991*

[MK12]C. MythiliV.Kavitha Color *Image Segmentation using ERKFCM,* International Journal of Computer Applications (0975 – 8887) Volume 41– No.20, March 2012

[M08]Mignotte, M., "*Segmentation by Fusion of Histogram Based -Means Clusters in Different Color Spaces,*" in *Image Processing, IEEE Transactions on* , vol.17, no.5, pp.780-787, May 2008

[MKI12]IzzatiMuhimmah, RahadianKurniawan, Indrayanti. Automated Cervical Cell Nuclei Segmentation Using Morphological Operation

and Watershed Transformation.*Computational Intelligence and Cybernetics (CyberneticsCom), 2012 IEEE International Conference on* 163 -167 2012

[NAH97] H.Z. Noorani, C. Arratoon, A. Hall, Assessment of Techniques for Cervical Cancer Screening, *Technical Report CCOHTA Report 1997: 2E, Canadian Coordinating Office for Health Technology Assessment* (May 1997)

[NST+02] O Nunobiki, M. sato E. Taniguchi, W. Tang, M. Nakamura,H. Utsunomiya, Y. Nakamura, I. Mori, and K. Kakudo, "Color image analysis of cervical neoplasia using RGB computer color specification,"Anal. Quant. Cytol. Histol., vol. 24, no. 5, pp. 289–294, 2002.[12]

[O14] Adekunle Oguntayo, Cervical Intraepithelial Neoplasia (CIN) (Squamous Dysplasia).*Intraepithelial NeoplasiaInTech* 2014

[OTDC02] F. Ortiz1 , F. Torres1 , E. De Juan2 and N. Cuenca3, Colour Mathematical Morphology For Neural Image Analysis, *Real-Time Imaging 8,* 455–465 2002

[PN13] Marina E. Plissiti,ChristophorosNikou. A Review of Automated Techniques for Cervical Cell Image Analysis and Classification. *Springer Science+Business Media Dordrecht* pp. 1-18, 2013.

[PN11] Marina E. Plissiti, ChristophorosNikou. Automated Detection of Cell Nuclei in Pap Smear

Images Using Morphological Reconstruction and Clustering.*IEEE transactions on information technology in biomedicine*, vol. 15, no. 2, 2011

[RKV12] Mr. Vivek Singh Rathore, Mr. MessalaSudhir Kumar , Mr. AshwiniVerma, Colour Based Image Segmentation Using L*A*B* Colour Space Based On Genetic Algorithm, International Journal of Emerging Technology and Advanced Engineering, 2250-2459, Volume 2, Issue 6, 2012

[RNRT12] Rahmadwati; Golshah Naghdy ; Montserrat Ros ; Catherine Todd. Computer aided decision support system for cervical cancer classification. *Proc. SPIE 8499, Applications of Digital Image Processing XXXV, 849919*pp. 1-13, 2012.

[SJ15]Sajeena T JereeshA SAAutomated Cervical Cancer Detection through RGVF segmentation and SVM Classification. IEEE*InternationalConference on Computing and Network Communications (CoCoNet)*663 − 669 2015