International Journal of Image Processing Techniques – IJIPT
Volume 1 : Issue 1
Publication Date : 09 January 2014

# Colour Histogram Based Colposcopy Cervical Image Classification

P.S.RamaPraba
Research scholar, Department of Electronics and Communication Engineering,
Sathyabama University, Chennai, India.
pazrama@yahoo.co.in

Dr.H. Ranganathan
Principal,
Rajiv Gandhi College of Engineering,
Chennai, India

***Abstract*—In this paper colposcopy cervical image classification based on colour histogram and K Nearest Neighbor (KNN) is presented. The classification is achieved by extracting colour histogram features from the cervix. To extract the colour histogram features, the colour space of the given image is converted from RGB to CIE colour space because of its perceptual uniformity. KNN classifier is used to classify the cervical images into normal and abnormal images. The performance with overall sensitivity of 94.71% and accuracy of 93.75 % is achieved using k-NN classifier. The performance is evaluated using 240 images collected from the hospital.**

***Keywords—Colposcopy, Histrogram,colorspace KNN classifier, cervicalcancer***

## I. INTRODUCTION

India has a population of 366.58 million women ages 15 years and older who are at risk of developing cervical cancer. Current estimates indicate that every year 134420 women are diagnosed with cervical cancer and 72825 die from the disease [1]. Cervical cancer ranks as the 1st most frequent cancer among women in India, and the 1st most frequent cancer among women between 15 and 44years of age. About 7.9% of women in the general population are estimated to harbour cervical Human Papilloma Virus (HPV) infection at a given time, and 82.5% of invasive cervical cancers are attributed to HPVs 16 or 18.Cancer of the cervix uteri is the second most common cancer among women worldwide, with an estimated 529,409 new cases and 274,883 deaths in 2008. About 86% of the cases occur in developing countries, representing 13% of female can-cers. Worldwide, mortality rates of cervical cancer are substantially lower than incidence with a ratio of mortality to incidence to 52% [2]. The majority of cases are squamous cell carcinoma and adenocarcinomas are less common.

Cancer of the cervix is primarily caused by HPV infec-tion. In developing countries like India HPV infection is high [2]. Cancer refers to a class of diseases in which a cell or a group of cells divide and replicate uncontrollably, intrude into adjacent cells and tissues (invasion) and ultimately spread to other parts of the body than the location at which they arose (metastasis) [3]. In cervical cancer, (cancer of the uterine cervix), cancer develops in the tissues of the cervix, which is a part of the female reproductive system. The cervix connects the upper body of the uterus to the vagina. The endocervix (the upper part which is close to the uterus) is covered by glandular cells, and the ectocervix (the lower part which is close to the vagina) is covered by squamous cells. The transformation zone refers to the place where these two regions of the cervix meet.

There are several types of cervical cancer, classified on the basis of where they develop in the cervix. Cancer that develops in the ectocervix is called squamous cell carcinoma, and around 80- 90% of cervical cancer cases (more than 90% in India) are of this type [1,3]. Cancer that develops in the endocervix is called adenocarcinoma. In addition, a small percentage of cervical cancer cases are mixed versions of the above two, and are called adenosquamous carcinomas or mixed carcinomas. There are also some very rare types of cervical cancer, such as small cell carcinoma, neuroendocrine carcinoma etc.

To diagnosis the cervical cancer, two prevention programs are available today. They are Papanicolaou (Pap) smear test and colposcopy [4]. Even though cervical cancer prevention programs such as the Pap smear have been effective in reducing the incidence and mortality of cervical cancer in developed countries, developing countries often lack the sophisticate laboratory equipment, highly trained personnel and financial resources necessary to implement these programs [5] Pap test can only screen for potential problems, not diagnose them. A Pap test is reported as normal (negative) when all the cells are of a healthy size and shape. An abnormal (positive) test is reported if any cells of different sizes or shapes are found. An abnormal Pap test does not mean the patient has cancer. Without a cost effective cervical cancer screening solution, cervical cancer has remained a leading cause of cancer-related death among women in developing countries. To address this problem, alternative cost effective cervical cancer screening methods have been investigated The consistent and accurate diagnoses provided by digital image analysis will allow less experienced physicians to provide a standard of care on par with expert colposcopists.

Colposcopic images are characterized by color, texture and relief information. Thus, their automatic analysis is difficult. However, the diagnosis of experts about some much debated images is often different, because of the very high specialization required. An integrated analysis tool for helping gynecologists to build their colposcopic diagnosis is proposed in [6]. Colposcopic image classification based on contour parameters using different artificial neural network and the

KNN classifier is proposed in [7]. A set of original spatial and frequency parameters is extracted from 283 samples to characterize the attribute of contour. The spatial parameter is the number of the region around the edges and the frequency parameters are amplitude of first peak, frequency of the end of first peak, area under first peak and area under other peaks. Then the Principal Component Analysis is performed to test the parameters. Segmentation and classification of cervix lesions by pattern and texture analysis is presented in [8].

Three different images are generally used one for the observation of the cervix without any preparation, a second one after application of an acetic acid solution (dissolution of mucous and accentuation of typical areas) and a final observation after application of a Lugol's solution (dark brown color stain of glycogen). The given images are characterized by many attributes such as color, texture or relief. Thus, their automatic analysis is difficult. So the methods in [9] is focused on a particular aspect of the whole challenge including enhancement of image quality, abnormal vascular patterns quantification by texture analysis , lesion area measurements with 3D correction.

In this paper, an automatic classification of cervical images based on colour histogram features and KNN is presented. The remainder of this paper is organized as follows: The proposed cervical image classification method and experimental results are described in sections 2 and 3. The conclusion based on the performance of the proposed methods is made in section 4.

## II. PROPSOED SYSTEM

The proposed system for the classification of cervical images is built based on colour histogram. In cervical image, almost 40% of the whole image consists of imaging device and the region outside cervix. In order to remove this irrelevant information Region of Interest (ROI) is extracted from the cervical image. Before feature extraction, Specular Reflection (SR) in the ROI is also removed. The approach used for ROI segmentation and SR removal is presented in our earlier work [10].

### *A. ROI Segmentation*

A colposcopy cervical image contains major cervix lesions, regions outside the cervix and parts of the imaging device. In this method, only the major cervix lesion is segmented for further processing. The major cervix lesion is a reddish, nearly circular section approximately centered in the image. This feature is used to identify the ROI region.

For ROI segmentation, first the given cervix image in RGB colour space is converted into Lab colour space due to the fact that Lab colour space is a good choice for representing the colour. The Euclidean distance of a pixel from the image center is extracted for all pixels and it is represented as Euclidean distance array d. The Gaussian Mixture Model (GMM) parameters μ and σ are calculated for the Euclidean distance array [10] and the colour channel a from the Lab colour space.

By using the GMM parameters, Euclidean distance array and the colour channel a from the Lab colour space are normalized and aggregated into a single array which is given to K-means clustering algorithm as an input. The value of K is set to 2. Among the 2 cluster, the cluster which has the smallest *d* and largest *a* is chosen as ROI after the cluster centroids are de-normalized. Finally, morphological opening is used to remove the small regions and fill the holes to get the required ROI image that contains only the cervix lesion.

### *B. Specular reflection removal*

Specular reflections (SR) appear as bright spots heavily saturated with white light. These occur due to the presence of moisture on the uneven cervix surface, which acts like mirrors reflecting the light from the illumination source.

The segmented cervix image in RGB colour space is converted into HSI colour space due to the fact that HSI colour space represents the colour similarity how the human eye senses colours. The HSI color model represents every color with three components, hue (H), saturation (S) and intensity (I). Specularities always have very intense brightness and low saturation values. Hence, I and S component in the HSI colour space is used to find the SR region. The initial SR regions are identified by applying thresholding technique on image pixels.

After thresholding, morphological dilation is performed on the thresholded image by using square structuring ele-ment of width 5 to get SR regions. Boundaries are extracted from the SR regions by using 8-connected neighborhood. Finally, the SR regions are smoothly interpolates inward from the pixel values on the boundaries by solving Laplace's equation.

### *C. Cervical image classification*

The proposed cervical image classification system has two stages, feature extraction and classification stage. In the feature extraction stage, the proposed colour histogram features are extracted from the training normal and abnormal cervix images and these features are given to the KNN classifier for the classification.

To extract the colour histogram features, the colour space of the given image is converted from RGB to CIE colour space [11]. Then, the colour histogram is constructed from CIE colour space and stored in the database called feature database. The conversion equations are given below

$$l^* = 116 .h(G / G_W) - 16 \quad (1)$$

$$a^* = 500 [h(R / R_W) - h(G / G_W)] \quad (2)$$

$$b^* = 200 [h(G / G_W) - h(B / B_W)] \quad (3)$$

$$h(q) = \{(q)^\wedge 1/3\} \qquad q > 0.008856$$

$$\{7.787q + 16/116\} \qquad q \le 0.008856 \quad (4)$$

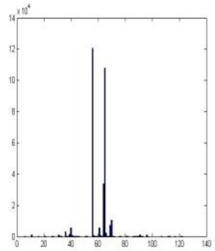
where $R_W$ , $G_W$ and $B_W$ are reference white tristimulus values, $l^*$ represents lightness of the colour, $a^*$ represents red minus green and $b^*$ represents green minus blue.

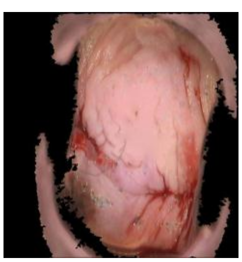
In the proposed system KNN classifier is used for the classification purpose. KNN is a method for classifying objects based on closest training examples in the feature space. KNN is a type of instance-based learning where the function is only approximated locally and all computation is deferred until classification. In KNN, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor. The neighbors are taken from a set of objects for which the correct classification is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The distance measure used in the proposed emotion recognition system is Euclidean distance. KNN classifier is trained by the feature database. For an unknown cervical image, the proposed colour histogram features are extracted and classified by using KNN classifier.

## III. EXPERIMENTAL RESULTS

The performance of the proposed method is tested on 240 colposcopy cervix images obtained from Government Kasturibaigandhi Hospital (KGH), Chennai, India. Figure 1 (a) shows sample normal image and the corresponding colour histogram is shown in Figure 1 (b). Similarly the sample abnormal image and the corresponding colour histogram are shown in Figure1 (c) and (d) respectively.


(a)


(b)


(c)

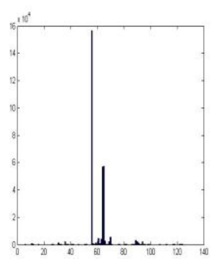

(d)

Figure 1 (a) Sample normal image (b) colour histogram of (a)

(c) Sample abnormal image (d) colour histogram of (c)

There are 70 normal images and 170 abnormal image are available in the database. For each category 2/3 of total images are used for training the KNN classifier. All the images in the database are used for testing the classifier using Euclidean distance measure. Table 1 shows the confusion matrix of the proposed cervical image classification system. Table 2 shows the performance of the proposed system in terms of specificity, sensitivity, classification accuracy, precision and recall rate in percentage

TABLE I. CONFUSION MATRIX OF THE PROPOSED CERVICAL IMAGE CLASSIFICATION SYSTEM

| **Test outcome** | **Condition** | |
|---|---|---|
| | ***Abnormal*** | ***Normal*** |
| Abnormal | 161 | 6 |
| Normal | 9 | 64 |

TABLE II. PERFORMANCE METRICS OBTAINED FROM THE CONFUSION MATRIX

| **Parameter** | **Performance (%)** |
|---|---|
| Sensitivity | 94.71 |
| Specificity | 91.43 |
| Accuracy | 93.75 |
| Precision | 96.41 |
| Recall | 87.67 |

## IV. CONCLUSION

In this paper, colour histogram features based colposcopy cervical image classification is proposed. KNN classifier is used for classification. The accuracy and sensitivity of the proposed classification system is 93.75% and 94.71%. The evaluation of the proposed algorithm is carried on totally 240 images and good results are achieved. The proposed system can only classifies the cervical images into normal or abnormal. In future, the work can be extended for the detection of lesion in the abnormal cervix by using colour histogram features.

## REFERENCES

[1] WHO/ICO Information Centre on HPV and Cervical Cancer

[2] IARC, GLOBOCAN 2008

[3] National Cancer Institute 2009

[4] "Pap Test and Colposcopy", by McKinley Health Center

[5] A. Batson, F. Meheus, and S. Brooke, "Chapter 26: Innovative financing mechanisms to accelerate the introduction of HPV vaccines in developing countries," Vaccine, vol. 24, pp. 219–225, 2006.

[6] Shiri Gordon and Hayit Greenspan, "Segmentation of Non-Convex Regions within Uterine Cervix Images", IEEE International Symposium on Biomedical Imaging, 2007, pp 312-315.

[7] Claude, I. Winzenrieth, "Contour features for colposcopic image classification by artificial neural networks", IEEE International Conference on Pattern recognition, 2002, pp 771-774.

[8] Bhakti Tulpule, Shuyu Yang, "Segmentation and Classification of Cervix Lesions by Pattern and Texture Analysis", IEEE International Conference on Fuzzy System, 2005, pp 173-176.

[9] J. Qiang, J. Engel and E. Craine, "Classifying Cervix Pattern With Texture Analysis,", Pattern Recognition, vol. 33, pp. 1561-1573,2000.

[10] P. S. RamaPraba and H. Ranganathan, "Automatic Lesion Detec-tion in Colposcopy Cervix Images Based on Statistical Features", Communications in Computer and Information Science, 1, Volume 270, Global Trends in Information Systems and Software Applications, Pages 424-430,2011.

[11] G.Sharma, M.J.Vrhel and H.J.Trussell, "Digital color Processing", IEEE Transaction on Image Processing, vol.6,no.7, pp.901 - 932, July 1997.