Automated Identification & Analysis of Cervical Cancer

SAMYA MUHURI, MONALISA BHATTACHARJEE
Department of Computer Science & Engineering RCC Institute of Information Technology, Kolkata, India
E-mail: smuhuri13@gmail.com, monalisa.rockzz@gmail.com

Abstract: Recent advances in imaging technology like USG, CT, MRIs, PET scans and other techniques have a huge impact on the diagnosis and treatment of disease. Medical imaging helps earlier detection of various malignant diseases such as breast cancer, lung cancer, cervical cancer etc. Cervical cancer is the second most common cancer among women in developing countries. Cervical benign lesions and invasive cancer exhibit some morphological features that can be detected during a conventional visual examination. If it can be detected earlier and treated properly then the patient can survive. Computational techniques in the processing of histopathological images allow us to assist the oncologist with a Computer Aided Diagnosis system. The overall objective of the work is to implement and evaluate a method for the analysis of cervical images. Epithelium that turns white after application of acetic acid is called acetowhite epithelium. Acetowhite epithelium is one of the major diagnostic features observed in detecting precursor lesions and cancerous region. Here we present a multi-step acetowhite region detection system to analyze lesions in uterine cervical images. First the Gabor filter is used to enhance the image. Second, the uterine cervix is segmented in different regions by using of Marker Control Watershed Segmentation algorithm. Third, acetowhite region are extracted through Support Vector Machine classifier. The above mentioned processes are tested on different uterine cervical images randomly retrieved from popular search engine and work has been implemented by MATLAB R2009b package.

Keywords: Cervical cancer; Acetowhite region; Computer-aided diagnosis (CAD); Feature extraction; Gabor filter; Support vector machine.

Introduction:
The automatic diagnosis of cancer using the various computational technique is the most important and major utilization of medical image processing and computer-aided diagnosis system. Nowadays, cervical cancer is a very common concern of the women of the developing countries. The cure of this cancer is related to early detection and proper treatment according to the malignancy level. An automated method to analyze the uterine cervical images and to extract the diagnostic features of the images can be helpful to detect the cancer. In this project, the histopathological images of squamous epithelium tissues of the cervix are used as preliminary data. During a medical test approximately 3%-5% acetic acid applied to the cervix [1]. After applying the acetic acid if the epithelium turns white then it is called acetowhite epithelium. The specialist in colposcopy evaluates the color and the density of the acetowhite reactions to assess the severity of possible lesions. Abnormal acetowhite epithelium varies from bright white to a grey white colour. Hence, acetowhite epithelium is one of the major diagnostic features in detecting cancer and pre-cancerous fields.

Automatic extraction of acetowhite epithelium from digital cervical images is a challenging task. All the published records of acetowhite segmentation are in a preliminary stage which has a scope for betterment [1]. Specular reflection (glare) and various illumination conditions have been bottlenecks in developing automated algorithms to extract acetowhite region of clinical significance. To the authors’ knowledge, only a very small number of prototype algorithms have been reported. Yang et al. [7] developed a sophisticated technique for precise detection of the acetowhite regions using K-means clustering and deterministic annealing technique. Gordon et al. [13] developed an unsupervised segmentation algorithm for three-tissue types in cervical imagery using a Gaussian mixture model. In another study [14], the acetowhite region was obtained from extracting the highest mean intensity cluster among the smooth regions. They also mention that due to illumination effects and large intra-patient variation, acetowhite lesions are wrongly detected. In addition, acetowhite lesions located in shaded areas of the image are not detected at all [8]. The work mentioned above is all based on cervigram images collected by the National Cancer Institute (NCI) [1] [2]. These studies serve as ground truth information for development of our algorithm.

Using the digitized colposcope data, we have developed an automated image analysis system to identify the acetowhite epithelium to match the expert theory. In order to extract the colour information of acetowhite properly, a multi-step procedure has been utilized. First, noise is removed by a low-pass filter. Second, the images have to be enhanced using any enhancing algorithm. After that, the anatomy of the cervix is analysed to identify different region of the cervix. So the cervix is segmented in different regions. Next multi-level acetowhite regions are identified and extracted using a classifier. The results demonstrate a significant correspondence with colposcopic and histopathologic annotations.

This paper consists of the following sections: section two deals with the basic preliminaries, section three


Basic Preliminaries

A. Image Enhancement

Image enhancement means to make the picture look good. It mainly depends on visual perception. Image enhancement is the improvement of digital image quality without knowledge about the source of degradation. If the source of degradation is known, it can be called as image restoration. Both are more or less identical process. Many different processes like elementary and heuristic methods are used to improve images in some sense. Apart from geometrical transformations some preliminary grey level adjustments may be indicated, to take into account imperfections in the acquisition system. Enhancement can be done pixel by pixel, calibrating with the output of an image with constant brightness. Frequently space-invariant grey value transformations are also done for contrast stretching, range compression, etc. Some examples of simple grey level transformations in this domain are shown in Fig.1.

![Image of transformations](image-url)

**Figure 1:** Image enhancement (a) Thresholding (b) Contrast stretching (c) Inversion

B. Gabor Filter

In image processing, a Gabor filter is a linear filter used for edge detection and image enhancement. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. It is a linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Its impulse response is defined by a sinusoidal wave multiplied by a Gaussian function. The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually.

Complex part:

\[ g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp \left( -\frac{x^2 + \gamma^2 y^2}{2\sigma^2} \right) \cos \left( 2\pi \frac{x}{\lambda} + \psi \right) \]

Real part:

\[ g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp \left( -\frac{x^2 + \gamma^2 y^2}{2\sigma^2} \right) \sin \left( 2\pi \frac{x}{\lambda} + \psi \right) \]

where \( x' = x \cos \theta + y \sin \theta \) and \( y' = -x \sin \theta + y \cos \theta \)

In this equation, \( \lambda \) represents the wavelength of the sinusoidal factor, \( \theta \) represents the orientation of the normal to the parallel stripes of a Gabor function, \( \psi \) is the phase offset, \( \sigma \) is the sigma or standard deviation of the Gaussian envelope and \( \gamma \) is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function.

C. Image Segmentation

In computing technology, image segmentation is the process of partitioning a digital image into multiple segments known as superpixels. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries like lines, curves, etc in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics such as colour, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. The result of image segmentation is a set of segments that collectively cover the entire image.

D. Marker-driven watershed segmentation technique

Marker-driven watershed segmentation technique extracts seeds that indicate the presence of objects or background at specific image locations [4]. Marker locations are then set to be regional minima within the topological surface and the watershed algorithm is applied. Separating touching objects in an image is one of the most difficult image processing operations, where the watershed transform is often applied to such problem. Marker-controlled watershed approach has two types: External associated with the background and Internal associated with the objects of interest. Image Segmentation using the watershed transforms works well if we can identify or “mark” foreground objects and background locations, to find “catchment basins” and “watershed ridge lines” in an image by treating it as a surface where light pixels are high and dark pixels are low [15].

E. Image Classification

Image classification analyzes the numerical properties of various image features and organizes data into categories. Classification algorithms typically employ...
two phases of processing: training and testing. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, i.e. training class, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features.

**F. Support Vector Machine classifier**

In machine learning a Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or other, making it a non-probabilistic binary linear classifier. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. Basically a support vector machine constructs a hyperplane or set of hyperplanes in an infinite-dimensional space, which can be used for classification.

**Proposed Methodology:**

Searching and detecting anatomic features are essential stages in the machine learning process in the medical field [9]. Anatomic features represent quality of the images and also the adequacy of the medical test. More importantly for the diagnosis of vital information of tissue, it can assist doctors by identifying the location of the diagnostic features like acetowhite epithelium. The original colposcopic image usually contains the edge of the speculum, the vaginal wall, and noise from the background. The cervix region needs to be extracted as the Region of Interest (ROI) for further processing. Our cervix region detection algorithm includes enhancement, segmentation and classification phase using Gabor filter, Marker Watershed method and SVM classifier respectively.

**G. Algorithmic steps**

- **Image Denoising:** The purpose of the stage is to make the image free from any unwanted noise. The wavelet method is used in the system to denoise the cervical cancer image.
- **Image Enhancement:** In this step the input image is enhanced by removing any corruption or interference. The Gabor filter method is used for the purpose.
- **Image Segmentation:** The image segmentation step involves divide and segmentation of the enhanced images. The used method is Marker-Controlled Watershed Segmentation approach.
- **Image classification & Feature Extraction:** To obtain the general features (acetowhite region) of the enhanced segmented image, classifier like support vector machine is used here.

**Procedure:**

Here we proposed a computer aided system based on cervical image analysis system for extracting acetowhite epithelium in cervical images. The three major processing steps are involved in this process. A pre-processing step includes denoising and enhancement of the cervical cancer images. At first the noise is removed from the image using a low-pass Gaussian filter. The noisy input image and denoised image is shown in Fig.3 (a) and Fig.3 (b) respectively. Then the denoised image is enhanced using Gabor filter. The enhanced output image is shown in Fig.4 (a) and (b).

**H. Block Diagram**

![Block diagram of the algorithm](image)

**Figure 2:** Block diagram of the algorithm

![Noisy Input Image](image)

(a)

![Denoised Image](image)

(b)

**Figure 3:** (a) Noisy Input Image (b) Denoised Image
The enhanced image is followed by a segmentation process. An anatomic feature extraction step analyzes different region of the cervix and extracts the cervix region. The segmentation process is done by Marker-driven watershed segmentation approach. The result of the segmentation is shown in the Fig.5(a),(b),(c), (d).

In the next step, acetowhite region is detected and characterized in different levels of acetowhite epithelium regions. The classification is done by linear Support Vector Machine classifier. Here we have taken some epithelium tissue image as training data and the kernel is working on Gaussian Radial Basis Function with a default scaling factor, sigma, of 1. The output of the classification is shown in the Fig.6. Acetowhite region is marked in green colour.

**Figure 4:** Enhanced Image using Gabor filter

**Figure 5:** (a) Modified Regional Maxima superimposed on input image (b) Reconstruction by thresholding (c) Markers and Object boundaries superimposed on the original image (d) Segmented image by watershed

**Figure 6:** Acetowhite region extraction using SVM classifier

**Conclusion:**
In this project we propose automated machine learning based cervical cancer detection method. The purpose of the work is to assist the physicians in clinical studies. In Future work, we can include improvement of the cervix region detection algorithm to remove the vaginal wall along the periphery of the cervix. A better cervix region will also remove false positives of the acetowhite region detection outputs. Future scope will also involve images taken at different time oscillation of the test. The system will also be combined with other diagnostic feature extraction algorithms to provide a better diagnostic method to the doctors.

**References:**


