



ELSEVIER

journal homepage: [www.intl.elsevierhealth.com/journals/cmpb](http://www.intl.elsevierhealth.com/journals/cmpb)

# Automatic cervical cell segmentation and classification in Pap smears

Thanatip Chankong<sup>a</sup>, Nipon Theera-Umpon (Senior Member, IEEE)<sup>a,b,\*</sup>,  
Sansanee Auephanwiriyaikul (Senior Member, IEEE)<sup>c,b</sup>

<sup>a</sup> Department of Electrical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>b</sup> Biomedical Engineering Center, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>c</sup> Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

## ARTICLE INFO

### Article history:

Received 12 June 2013

Received in revised form

26 October 2013

Accepted 18 December 2013

### Keywords:

Cervical cell classification

Cervical cell segmentation

Pap smear

ThinPrep

Cervical cancer screening

## ABSTRACT

Cervical cancer is one of the leading causes of cancer death in females worldwide. The disease can be cured if the patient is diagnosed in the pre-cancerous lesion stage or earlier. A common physical examination technique widely used in the screening is Papanicolaou test or Pap test. In this research, a method for automatic cervical cancer cell segmentation and classification is proposed. A single-cell image is segmented into nucleus, cytoplasm, and background, using the fuzzy C-means (FCM) clustering technique. Four cell classes in the ERUDIT and LCH datasets, i.e., normal, low grade squamous intraepithelial lesion (LSIL), high grade squamous intraepithelial lesion (HSIL), and squamous cell carcinoma (SCC), are considered. The 2-class problem can be achieved by grouping the last 3 classes as one abnormal class. Whereas, the Herlev dataset consists of 7 cell classes, i.e., superficial squamous, intermediate squamous, columnar, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. These 7 classes can also be grouped to form a 2-class problem. These 3 datasets were tested on 5 classifiers including Bayesian classifier, linear discriminant analysis (LDA), K-nearest neighbor (KNN), artificial neural networks (ANN), and support vector machine (SVM). For the ERUDIT dataset, ANN with 5 nucleus-based features yielded the accuracies of 96.20% and 97.83% on the 4-class and 2-class problems, respectively. For the Herlev dataset, ANN with 9 cell-based features yielded the accuracies of 93.78% and 99.27% for the 7-class and 2-class problems, respectively. For the LCH dataset, ANN with 9 cell-based features yielded the accuracies of 95.00% and 97.00% for the 4-class and 2-class problems, respectively. The segmentation and classification performances of the proposed method were compared with that of the hard C-means clustering and watershed technique. The results show that the proposed automatic approach yields very good performance and is better than its counterparts.

© 2013 Elsevier Ireland Ltd. All rights reserved.

\* Corresponding author at: Biomedical Engineering Center, Chiang Mai University, Chiang Mai 50200, Thailand. Tel.: +66 5394 2083; fax: +66 5322 1485.

E-mail addresses: [tchankong@gmail.com](mailto:tchankong@gmail.com) (T. Chankong), [nipon@ieee.org](mailto:nipon@ieee.org), [nipon.tor@gmail.com](mailto:nipon.tor@gmail.com), [DrNipon@chiangmai.ac.th](mailto:DrNipon@chiangmai.ac.th) (N. Theera-Umpon), [sansanee@ieee.org](mailto:sansanee@ieee.org) (S. Auephanwiriyaikul).

0169-2607/\$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.cmpb.2013.12.012>

## 1. Introduction

Cervical cancer is the fourth leading cause of cancer death in females worldwide [1]. The prognosis for cervical cancer depends on the stage of the cancer at the time of detection. The disease can be cured if diagnosed in the pre-cancerous lesion stage or earlier. Papanicolaou test or Pap test is a physical examination technique widely used to prevent cervical cancer by finding cells that have the potential to turn cancerous. It was estimated that, in the year 2006, systematic screening can reduce mortality rates from cervical cancer by 70% or more [2].

In Thailand, cervical cancer is the second most common cancer among women [3], with a high mortality rate of nearly 300,000 per year. The screening program administered by the Ministry of Public Health and the National Health Security Office suggested that women aged 35–60 years should undergo Pap smear examination every five years [4]. With an overall population number of 67 million in Thailand [5], the number of samples to be examined per year is large comparing to the number of cytologists who process the screening. The study of automatic cervical cell classification has been done for over 40 years to cope with this labor-intensive task of manual screening and also to reduce the error of screening result. A number of commercial automated screening systems have been approved by the FDA for quality control with examples including PAPNET (Neuromedical Systems Inc.), FocalPoint Slide Profiler™ (formerly AUTOPAP; BD TriPath), ThinPrep Pap Test, ThinPrep Imaging System (Hologic Inc.) and Imager™ (Cytyc). Several research works have shown that these automated systems indeed improve the accuracy of the screening result and reduce the false-negative rate [6–8]. However, cost effectiveness is a major drawback of these systems with the cost of PAPNET test far exceeds that of manual screening.

Uncertainty in diagnostic capability was also reported [9]. It is therefore suggested that the automated system should be used as an aiding tool in conjunction with the expert's opinion rather than relying on the system as a primary screening and diagnosing tool [10,11]. Over the past few years, trend of research in automated cervical cancer screening has shifted from cytology screening to histology image [12,13] and colposcopic image [14,15]. Histology image is not only used in cervical cancer screening, but it is used in the other kinds of cancer screening also [16–18]. However, cytology screening is still a default screening method in most countries due to its relatively low cost and its effectiveness in cervical cancer prevention if the screening is routinely performed.

The screening process normally starts with gathering cervical cell samples from the uterine cervix and mounting it on a glass slide. The collected sample is visually inspected under a microscope to identify the target cell or classify each cell into categories. The basic characteristics used to classify the stage of cells are mainly the characteristics of cell nuclei and cytoplasm such as shape, size, texture, ratio of nucleus and cytoplasm. From image processing point of view, the first step in extracting information from cell components is to correctly identify a region of each component (nucleus, cytoplasm, and non-cell components) by segmentation process. There are several research works on nucleus segmentation

[19–22]. However, when one would like to classify each cervical cell into categories with only nucleus information, it might not yield a good performance. Hence, segmenting whole cell is more desirable [23–26]. However, there is no classification result reported in these works. After the segmentation step, each cell is then classified using specific classifiers based on the extracted features from cell components as mentioned earlier or by using filters to discriminate classes without feature extraction process [27]. However, the classification performance in Ref. [27] is not quite high.

In this research, we propose a method for automatic segmentation and classification of cervical cell images. In the segmentation process, we use a patch-based fuzzy C-means (FCM) clustering technique. A cell image is segmented by using the over-segment FCM technique into nucleus, cytoplasm, and background. For comparison, we use the hard C-means clustering technique and the watershed segmentation technique in the segmentation step as well. The segmented image is then used to extract related features to be the input to classifiers. Five classifiers including Bayesian classifier, linear discriminant analysis, K-nearest neighbor, artificial neural networks, and support vector machine, are considered. The usefulness of features based on nucleus and entire cell is also investigated.

This paper is organized as follows: the following section describes basic knowledge of the segmentation technique, mathematical morphologies, feature extraction, classifiers, followed by the details of cell datasets used in this research. Section 3 provides the experimental results and discussion. The conclusion is drawn in Section 4.

## 2. Materials and methods

### 2.1. Cervical cell segmentation using patch-based fuzzy C-means clustering

We invented the patch-based fuzzy C-means (FCM) clustering method to segment nuclei and cytoplasm of white blood cells [28]. It was later applied to segment nuclei of cervical cells from the conventional Pap test [29]. The FCM is good for clustering data with uncertainty. We, therefore, chose the FCM to cluster uncertain cell image intensity data. In this research, the segmentation method is tested with cervical cells from, not only the conventional Pap test, but also the ThinPrep® Pap test. The segmentation of cytoplasm is also performed. The segmentation algorithm is as follows:

*Convert color input image to grayscale*

*Apply median filter*

*Apply FCM algorithm*

*Sort FCM centers in ascending order*

*For each patch corresponding to sorted FCM centers (from dark to bright)*

*If ((FCM center of patch)  $\leq T_N$ ), then label patch as nucleus*

*Else if ( $T_N < (\text{FCM center of patch}) \leq T_C$ ), then label patch as cytoplasm*

*Else label patch as background*

*End (For each patch)*

Download English Version:

<https://daneshyari.com/en/article/468470>

Download Persian Version:

<https://daneshyari.com/article/468470>

[Daneshyari.com](https://daneshyari.com)