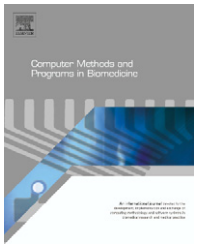
ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Towards high performance cell segmentation in multispectral fine needle aspiration cytology of thyroid lesions

Edgar Gabriel[a,*], *Vishwanath Venkatesan*[b,1], *Shishir Shah*[b,2]

[a] *University of Houston, Department of Computer Science, 524 Philip G. Hoffman Hall, Houston, TX 77204-3010, USA*
[b] *University of Houston, Department of Computer Science, 501 Philip G. Hoffman Hall, Houston, TX 77204-3010, USA*

ARTICLE INFO

Article history:
Received 19 February 2009
Received in revised form
16 April 2009
Accepted 2 July 2009

Keywords:
Fine needle aspiration cytology
Image analysis
Parallel computing
PC clusters
Multi-core processors

ABSTRACT

Thyroid nodule is a common cancer of the thyroid gland that affects up to 20% of the world population and approximately 50% of 60-year-old persons. Early detection and screening of the disease, especially analysis by fine needle aspiration cytology (FNAC), has led to improved diagnosis and management of the disease. Simultaneously, advances in imaging technology has enabled the rapid digitization of large volumes of FNAC specimen leading to increased interest in computer assisted diagnosis (CAD). This has led to development of a variety of algorithms for automated analysis of FNAC images, but due to the large scale memory and computing resource requirements, has had limited success in clinical use. In this paper, we present our experiences with two parallel versions of a code used for texture-based segmentation of thyroid FNAC images, a critical first step in realizing a fully automated CAD solution. An MPI version of the code is developed to exploit distributed memory compute resources such as PC clusters. An OpenMP version is developed for the currently emerging multi-core CPU architectures, which allow for parallel execution on every desktop system. Experiments are performed with image sizes ranging from 1024 × 1024 pixels up to 12288 × 12288 pixels with 21 spectral channels. Both versions are evaluated for performance and scalability.

1. Introduction

Cancer continues to remain a major health problem in the United States, with one of two men and one of three women developing cancer in their lifetime. Among various cancers, thyroid nodule is a common cancer of the thyroid gland. It has been estimated that up to 20% of the world population and approximately 50% of 60-year-old persons have palpable thyroid nodule or nodules [1]. The clinical spectrum ranges from the incidental, asymptotic, small, solitary nodule, in which the exclusion of cancer is a major concern, to the large, partly intrathoracic nodule that causes pressure symptoms, for which treatment is warranted regardless of cause. In spite of the growing incidences of thyroid lesions, the rate of thyroidectomies is on the decline. Early detection of the disease has been partly responsible for improved outcomes. Among screening and detection procedures, fine needle aspiration (FNA) is believed to be a safe, inexpensive, and minimally invasive procedure to diagnose tumors [2]. For cytological eval-

* *Corresponding author*. Tel.: +1 713 743 3857; fax: +1 713 743 3335.
E-mail addresses: gabriel@cs.uh.edu (E. Gabriel), venkates@cs.uh.edu (V. Venkatesan), shah@cs.uh.edu (S. Shah).
[1] Tel.: +1 713 743 3857; fax: +1 713 743 3335.
[2] Tel.: +1 713 743 3360; fax: +1 713 743 3335.
0169-2607/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.cmpb.2009.07.008

uation of FNA samples, smears are appropriately prepared and stained. Typically, the stain changes the color of the cells and tissue so that examination of the smear under standard microscopes with moderate magnification (20–40 ×) is sufficient for clinical evaluation.

With the advances in imaging technology, there is considerable interest in automated analysis of FNA cell smears that could help to reduce the time required for manual screening and increase the detection rate of abnormalities [3]. Several commercial products such as ScanScope from Aperio Technologies, DX-40 from DMetrix, Inc., and iScan from BioImagene, Inc. have been developed to automate the process of digitizing microscope slides. They provide high throughput capabilities to digitize cell smears, resulting in a stitched image of scan areas of the order of 1.5 cm × 1.5 cm in less than 5 min. This provides a single image per smear that can be as large as 60,000 × 60,000 pixels under 40 × magnification (resolution of 0.25 μm/pixel). More recently, multispectral microscopes capable of acquiring spectral images under transmitted illumination have also been used to digitize and analyze cell smears [4,5]. Spectral imaging allows for the simultaneously measurement of spectral and spatial information of a sample such that the measurement of the spectral response at any pixel of a two-dimensional image is possible. A spectral image consists of a series of images and each image is acquired under a narrow band wavelength of light. Studies have shown that biological tissue exhibits unique spectra in transmission. By exploring the spectral differences in tissue pathology, many chemical and physical characteristics not revealed under traditional imaging systems can be realized and used to improve the analysis efforts. Several efforts have already resulted in algorithms for cell segmentation, morphometric and karyometric feature analysis, as well as computer assisted diagnosis (CAD), with cell segmentation being the most challenging step for automated systems. However, to our knowledge, most of these efforts have been relatively limited in size due to the large data size and the computational bottlenecks. It is not uncommon to acquire anywhere from 5 to 31 spectral channels for each sample. Considering an average size of the smear to be 1.0 cm × 1.0 cm, the image cube to be analyzed would be approximately 8 GB to 50 GB in size. This creates difficulties in analyzing the entire data set on a standard desktop.

In this paper, we present our experiences with two parallel versions of a code used for texture-based image segmentation. Our main interest here is not to define the best segmentation algorithm, but to define a set of processes that would be realistically required in a typical CAD system. Specifically, we use Gabor filters for texture measurement and combine it with absorption computed from the spectral image stack to generate a feature vector for each pixel. k-Means clustering is used to group pixels into different classes resulting in the segmentation of thyroid cells. An MPI version of the code has been developed to exploit distributed memory compute resources such as PC clusters. An OpenMP version has been developed for the currently emerging multi-core CPU architectures, which allow for parallel execution on every desktop system. The rest of the paper is organized as follows: Section 2 provides a brief overview of multispectral microscopy, the digitization of thyroid lesion cell smears and gives an overview about the most relevant related work in this area. The texture-based approach for segmentation of thyroid cells is presented in Section 3. Sections 3.1 and 3.2 present the parallelization strategy developed for this application. The experiments performed for large scale analysis of entire scans of smear samples and according results are presented in Section 4. Finally, the paper is concluded in Section 5.

2. Background

The core element of any spectral imaging system is the spectral dispersion component that separates light into its spectral components, and is coupled to a two-dimensional (2D) optical detector such as a CCD camera, or to an array of photomultiplier tubes (PMT). In our system, we use a quarter-meter class, Czerny-Turner type monochromator that provides a tunable light emission spectrum at 10 nm resolution. We utilize a wavelength range from 400 to 700 nm for this study. The monochromator is connected to an Olympus BX51 upright optical microscope such that the light output from the monochromator feeds in to the transmitted light path of the microscope. This allows for the use of conventional optical microscopy to acquire brightfield images at desired wavelengths (transmitted light). An Olympus UPlanApo 40X NA 0.9 was used for imaging. The Photometrics SenSys ™ CCD camera having 768 × 512 pixels (9μ m × 9μm) at 8-bit digitization is used which provides for high resolution low light image acquisition. Fig. 1 (left) shows the system that has been assembled. To image each sample, the illumination from the monochromator was adjusted by achieving Köhler illumination for uniform excitation of the specimen. The condenser, aperture diaphragm, and the field stop were kept constant during measurements. Focusing was performed at the central wavelength of 550 nm to minimize the chromatic aberration at all wavelengths. The system was calibrated as per the method proposed in [6]. Using a stepper controlled microscope stage, multiple images were acquired to cover the extent of the smear on the microscope slide. Resulting images were stitched to generate a composite mosaic.

A multispectral image allows the possibility to locate, discriminate, measure, and enumerate many entities within a specimen by detecting subtle differences among their individual spectral signatures [7]. Clearly, different stained cells will be spectrally distinct. However, spectral information in any cell can come from such optical processes as reflection and scattering. As long as the phenomenology is based on reproducible physical reality, classification of spectrally distinct species can be of great utility. Fig. 1 (right) shows a subset of the spectral image (400 nm, 500 nm, 600 nm, and 700 nm) of a Papanicolaou stained cytological specimen. As seen the light absorption across various cellular constituents vary as a function of wavelength. This forms the spectral signature for each cellular entity.

To understand the spectral characteristics of biological samples, gray level image intensities may be used to determine the proportion of light transmitted by each cell across the exciting spectra. The transmission factor, T, is defined as:

$$T = I_t/I_i \quad (1)$$