Computers in Biology and Medicine 51 (2014) 93–103

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Segmentation of colon tissue sample images using multiple graphics accelerators

Sándor Szénási *

John von Neumann Faculty of Informatics, Óbuda University, 96/B Bécsi út, H-1034 Budapest, Hungary

ARTICLE INFO

Article history:
Received 17 February 2014
Accepted 2 May 2014

Keywords:
Medical image segmentation
Cell nuclei detection
Data parallel algorithm
Distributed algorithm
GPGPU
CUDA

ABSTRACT

Nowadays, processing medical images is increasingly done through using digital imagery and custom software solutions. The distributed algorithm presented in this paper is used to detect special tissue parts, the nuclei on haematoxylin and eosin stained colon tissue sample images. The main aim of this work is the development of a new data-parallel region growing algorithm that can be implemented even in an environment using multiple video accelerators. This new method has three levels of parallelism: (a) the parallel region growing itself, (b) starting more region growing in the device, and (c) using more than one accelerator. We use the split-and-merge technique based on our already existing data-parallel cell nuclei segmentation algorithm extended with a fast, backtracking-based, non-overlapping cell filter method. This extension does not cause significant degradation of the accuracy; the results are practically the same as those of the original sequential region growing method. However, as expected, using more devices usually means that less time is needed to process the tissue image; in the case of the configuration of one central processing unit and two graphics cards, the average speed-up is about 4–6 ×. The implemented algorithm has the additional advantage of efficiently processing very large images with high memory requirements.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Nowadays, digital microscopes are becoming increasingly popular among pathologists. The processing of microscopic tissue images and the segmentation of tissue components are now done through digital imagery and special immunodiagnostic software products [1]. These are fast and accurate products and can serve several additional functions, like remote access, archiving [2,3], searching and tagging [4], semi-automatic diagnostics [5–7], registration [8], computer-aided tissue engineering [9], etc. This kind of processing offers a very promising way of using different segmentation techniques with the images received; this way, the different components of the tissues can be separated effectively. Appropriately, precise recognition of the tissue components would provide a safe background for automated status analysis of the examined patients, or at least promote the work of pathologists with this pre-processing.

Our work focuses on the segmentation of images containing haematoxylin and eosin (HE) stained colon tissue samples. There are several procedures to identify the main structures in these images and many are based on a reliable cell nuclei detection method. There are several image processing algorithms for this purpose [10–13], but some factors could increase the challenge. The size of the images can easily reach 100 MB; therefore, the image processing speed plays an important factor.

In this paper, after the presentation of the technical background (related work, evaluation method, etc.), we propose a new cell nuclei segmentation algorithm implemented in a heterogeneous environment. This method uses all the available GPUs of the system for the most computationally intensive tasks (data-parallel cell nuclei segmentation), and all the available CPU cores for the less computationally intensive additional tasks (splitting and merging images, and controlling the GPUs).

2. Evaluation of cell nuclei detection methods

2.1. Accuracy of nuclei segmentation algorithms

For comparison, we have to evaluate the accuracy of the different algorithms. We have 39 colon tissue sample images manually annotated by qualified pathologists (we will refer to these as the Gold Standard slides), therefore we can compare the outputs of the algorithms to these results. There are several available evaluation methods for this purpose, but most of them

* Tel.: +36 16665579.
E-mail address: szenasi.sandor@nik.uni-obuda.hu

http://dx.doi.org/10.1016/j.compbiomed.2014.05.002
0010-4825/© 2014 Elsevier Ltd. All rights reserved.

are not suitable for this task, therefore we designed a new methodology. We have to know the exact position and shape of cell nuclei for further diagnosis purposes, therefore the basic object-level comparison methods are not applicable (for example, just compare the number of cell nuclei, etc.); we need a pixel-level comparison method. The widely used confusion matrix gives very clear and easily understandable results, based on the below equation:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where

- TP: Number of true-positive pixels (the pixel is correctly classified as part of a nucleus in both the reference result set and in the test result set).
- TN: Number of true-negative pixels (the pixel is correctly classified as not part of a cell nucleus).
- FP: Number of false-positive pixels (in the test result, the pixel is classified as part of a nucleus, but in the reference result it is not).
- FN: Number of false-negative pixels (the pixel is incorrectly classified as not part of a cell nucleus).

However, this pixel-level evaluation itself will not give us perfect results, because during the segmentation our task is not only to determine whether a pixel belongs to a nucleus or not, we have to identify the closed nuclei objects themselves. For example, in the case of false-negative hits, the pixel-level evaluation cannot indicate how many nuclei the algorithm misses (for diagnostic purposes, it really matters whether we miss only one big nucleus, or a lot of small nuclei). Another problem can be when there are several small nuclei in the reference slide, but the algorithm identifies them as one large nucleus. In this case, the pixel-level comparison indicates relatively small errors; however, this can be very important information for detecting malicious cells.

Our specialized measurement number is not based only on the pixel-by-pixel comparison; instead it starts by matching the cell nuclei together in the reference and the test results. One cell nucleus from the reference result set can only have one matching cell nucleus in the test result set and vice versa. After matching the cell nuclei, we can compare the paired elements using the confusion matrix. There are some other improvements: for example, we use some weighting in the case of false-positive and false-negative pixels based on the distance from the nearest valid pixel, which is important for the appropriate results near the borders of the nuclei.

The implementation of this evaluation method raises several problems. The pairing of the test-reference nuclei is a very resource-consuming step (in the case of several overlapping nuclei, the number of valid pairings can be billions); therefore we use a backtracking-based method to find the optimal result [14]. In this paper, we will use this evaluation method for every task where we need to check the accuracy of the nuclei detection algorithm (evaluation of algorithms, testing, parameter optimizing, etc.).

2.2. Comparison of nuclei detection methods

The main purpose of these algorithms is the same: we have to select the pixels of the sample, which could belong to any nucleus. The first thought would be to use the colours of the pixels for this separation, but in practice, this causes many difficulties. In the case of a specific image, we can achieve good results because we can easily teach the program whether a given colour represents a nucleus pixel or not. However, our experiences show that the colours of the images pre-processed by different labs are significantly different. This problem can be solved with some profile files (one profile for each lab), since we can transform all images into a standardized colour space. However, in practice, it turned out that there are significant differences between results from the same laboratories as well. Even if the same tools and materials are used, a different amount of stain and processing time can cause different colours (in some cases, the nuclei are very strong dark areas, but in the case of some other images, these are significantly less contrasting).

There are various automatic threshold based techniques to solve this problem. Several papers deal with segmentations using the K-means procedure [15], which produces very quickly and with impressive results. The main limitation of this method is the insufficient accuracy [16]. Further options are the texture based methods [17] and colour clustering [18]. It is easy to achieve the quick results initially with moderate accuracy, but further development is generally impossible. Nevertheless, it is worth considering these techniques as they are quite flexible in regard to various staining conditions. Therefore, these procedures can be used for fast pre-processing.

Region growing is a more sophisticated technique [19]. This is because we can exactly define and fine-tune the iteration steps by choosing an arbitrary fitness function and stopping condition. Both of these may consider the colour of the pixels, the environmental conditions, the size of the increased region, their position, etc. Another important advantage of the region growing approach is that it provides information not only about the individual pixels (whether a given pixel belongs to a cell nucleus or not), but it gives detailed information about the whole cell nuclei objects (the result of the region growing is a list of cell nuclei). This information is essential by itself for the diagnosis (number of nuclei, density of nuclei, etc.), and it is useful for the further segmentation of the image (glands, surface epithelium, etc.).

However, region growing has some disadvantages as well. First, the biggest problem is that this method is rather slow. The process is slow to the extent that practical use seems almost impossible, because the segmentation of large images (8192×8192 pixels size or even greater) containing a moderate number of nuclei may require up to one hour to complete. However, because the process offers good accuracy, it is definitely worth dealing with this drawback, though, we have tried to speed up the process as much as possible (without loss of accuracy). For the implementation, we use the graphics hardware, because it is used in similar projects with good results [20,21].

3. Cell detection with data parallel region growing

3.1. Parallel region growing

The implemented region growing algorithm iterates the following three steps until one of the stopping conditions is met. Due to space limitations, this paper contains only a brief description of parallel region growing. Detailed introductions can be found in [22].

1. It checks the four possible directions in which the contour can be expanded. In case of the first iteration, this means the four neighbours of the starting point (seed point), in the latter iterations the pixels around the lastly accepted contour point (see below). We can check all directions at the same time; therefore, four threads examine the different neighbours, whether they are suitable for further expansion or not.
2. In the next step, all contour points are evaluated to decide the direction in which the region should be expanded. The algorithm evaluates a fitness function for every point. Unfortunately, some parameters of this fitness function change at the insertion of every new point (centre of the region, average intensity of the

Download English Version:

https://daneshyari.com/en/article/6921606

Download Persian Version:

https://daneshyari.com/article/6921606

Daneshyari.com