Improvements on contours based segmentation for DNA microarray image processing

Yang Li\textsuperscript{a}, Andrei Păun\textsuperscript{b,c,*}, Mihaela Păun\textsuperscript{c,d}

\textsuperscript{a} College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272, USA
\textsuperscript{b} ICB/Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest 010014, Romania
\textsuperscript{c} Bioinformatics Department, National Institute of Research and Development for Biological Sciences, Splaiul Independenței, Nr. 296, Sector 6, Bucharest, Romania
\textsuperscript{d} Faculty of Administration and Business, University of Bucharest, Bucharest 010014, Romania

\textbf{A R T I C L E   I N F O}

Article history:
Received 8 November 2016
Received in revised form 28 March 2017
Accepted 17 April 2017
Available online xxxx

Keywords:
Microarray
Gene levels
Segmentation methods
Segmentation based on contours

\textbf{A B S T R A C T}

In this paper we present an improvement of the Segment Based Contours (SBC) method by implementing a higher order of finite difference schemes in the partial differential equation used in our mathematical model. Two methods are presented: one is a 4th order method and the other a 8th order method. The 4th order method could be applied to segment both the cDNA microarray images and the Affymetrix GeneChips, while the 8th order method could only be applied to processing the cDNA microarray images, due to the limitation of the current image resolution. Additionally, we provide both the mathematical derivations for the partial differential equations (their 4th or 8th order approximations) as well as the validation through simulations of the microarray images by using real images as seeds for the Nykter’s 2006 methodology. We conclude by showing that both the 4th order method as well as the 8th order one are superior to the SBC and the widely used GOAC method implemented in the Affymetrix standard processing package for microarrays.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Segmentation is a fundamental step in the microarray image analysis and the segmentation accuracy has a significant influence on the subsequent gene expression generation [1, 3], more accurate and efficient segmentation algorithms are being pursued all the time. In this paper we modify the SBC method introduced in [23] and improved in [27] as a refinement of the ACWE method [6], which has been shown to be more accurate in terms of segmentation than the GCOS software: the GOAC method.

The refinements in this paper represent two new segmentation techniques/refinements to the SBC: a 4th order and a 8th order method that approximate the curves of the spot segmentation. The mathematical derivations show that both the 4th order method and the 8th order method are better approximating the C–V model [6] than the SBC method does, which means they will offer better segmentation results as compared to the SBC method. Besides the mathematical proof, we do the practical experiments to double check the conclusion drawn from the mathematical derivation. The 4th and the 8th order methods are used to segment the microarray images and the intensities obtained will be compared to the results generated by the SBC method. The obtained results are also compared to those generated by the Globally Optimal...
Geodesic Active Contours (GOGAC) method from the GeneChip Operating System (GCOS) software, for a broader evaluation and comparison.

To give the ground true values of intensities as the standard for different segmentation methods comparison, a microarray image simulator is introduced to generate the simulated images used in the experiments. The simulated microarray images have all the characteristics that real microarray images have, but more importantly, the true intensity values of each probe spot in the image are now known for the simulated picture. We then segment the same picture with various methods obtaining intensity values that can now be compared to the true intensity values. Therefore, one could evaluate the performance of each segmentation method and decide whether it is more accurate than other methods by comparing its results in intensity values and investigating whether these are closer to the true values than any other segmentation method.

This work lies at the intersection of Mathematics, Computer Science and Biotechnology: it is using tools from Mathematics such as a partial differential equation framework used to detect shapes and segment the results of DNA arrays experiments stemming from Biology and Biotechnology. Furthermore, Computer Science automated methods are used for simulating thousands of replicates of the DNA array pictures and then we use Statistics tools to estimate the capabilities of the new segmentation tools proposed. Many of these areas were active areas of research of the regretted Solomon Marcus who published in areas covering Mathematical Analysis [16,24,17,18] as well as Natural Computing [10–15] or Theoretical Computer Science [4,5,19,20].

To obtain the performance of the gene expression, one needs to analyze the DNA microarray image. There are three primary steps for DNA microarray image analysis: Addressing, Segmentation, and Information extraction. The current investigation is focused on the segmentation step, which is to identify the foreground of each DNA spot in the image from the background of each spot. To be more specific, by segmentation we intend to detect the boundary of the bright part of each spot. The GeneChip Operating System (GCOS) by Affymetrix, Inc. which implements the Global Optimal Geodesic Active Contours (GOGAC) method [2] is the most used microarray image segmentation method at the moment. One could argue that one could improve it and, thus, the results of the microarray experiments.

The main idea of this paper was to develop more accurate microarray image segmentation methods than the SBC method [23] based on the current resolution of the microarray images. To achieve this goal, the following steps are performed:

1. (1) Simulate cDNA microarray images and Affymetrix GeneChip images, which are used to evaluate our improved segmentation methods.
2. (2) Develop the numerical algorithms of our improved segmentation method and deduce the truncation errors for the numerical approximation.
3. (3) Apply our two improved DNA microarray image segmentation methods to the simulated images, and then compare the performance of our method to the SBC, the GCOS, and the GOGAC methods.

These are the three major subsequent sections of the paper.

2. Simulating microarrays

In what follows, the reasoning behind simulating the realistic microarrays that was performed for this work is given: find a way to improve the methodology that is estimating the expression levels of the genes in a cell. Unfortunately, there is no a priori knowledge about the “perfect values” of the gene expression levels in that cell. Several technologies that were used have incorporated manners of quality assurance on the chips, but these are rather qualitative. Our intention is to go in the direction of quantitative microarrays (which is not achieved at the moment), hence the need to know all the correct values for the entire chip. In this way we will be able to better evaluate any new method of segmentation of these pictures. The paper presents the steps to simulate such images and a statistical approach to evaluate and compare the segmentation methods used. As a small overview of the process: images with the same statistical properties as the cDNA and Affymetrix GeneChip images starting with real images are simulated and segmented with the currently used methodology (GOGAC from GCOS) and then, using those values for each spot up to 10,000 replicates (simulated images) are generated and statistically analyzed to determine the better method. By simulating them one knows the “perfect values” for all the spots and genes investigated by that type of microarray. Armed with this information we then proceed to segmenting the images with the currently used segmentation methodology (again GOGAC from GCOS) but also with the SBC method, as well as with the two new refinements defined in Section 3. The results are presented in Section 4 showing that in some cases our methods outperform the current methods.

The DNA microarray image simulator, proposed by [25], is used to validate different kinds of data analysis algorithms. It can simulate both spotted two-channel and oligonucleotide one-channel microarrays, by using the true intensity values of each probe spot as an input. This simulator contains all the steps that affect the quality of real microarray data. For example, these steps include specific error models applicable to biological measurement technology issues, and simulating the microarray slide manufacturing and hybridization errors. With this in mind, the simulated data has realistic biological and statistical characteristics. Therefore, we use this simulator to simulate the microarray images used in our experiment. The aforementioned method contains six main modules: data input, slide manufacturing, biological noise, slide hybridization, slide scanning, and image reading. By setting various module parameters, the simulation process will provide three different quality images, which are high, normal, and bad. High and normal settings are used in this study.

Please cite this article in press as: Y. Li et al., Improvements on contours based segmentation for DNA microarray image processing, Theoret. Comput. Sci. (2017), http://dx.doi.org/10.1016/j.tcs.2017.04.013
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات