A SYSTEM FOR ON-LINE PROCESSING OF IMAGES OF GENE EXPRESSION PATTERNS

Blagov M.S.*, Poustelnikova E.G., Pisarev A.S., Myasnikova E.M., Samsonova M.G.

St.Petersburg State Polytechnical University, St.Petersburg, Russia
* Corresponding author: e-mail: blagov@spbcas.ru

Keywords: on-line processing, gene expression patterns, relational database, confocal scanning microscopy, image processing

Summary

Motivation: The development of software for quantification and analysis of gene expression in situ is an important task for bioinformatics.

Results: A software system to provide access and process images of gene expression patterns has been developed. The images can be stored in a relational database as BLOBs (Binary Large Objects) or in a file system as image files in any graphic format. The system is completely portable across different software/hardware platforms and supports both basic operations on images (scaling, cuts of rectangular area, filtering of fluorescence intensity, contrast enhancement, etc.) and subject domain oriented operations (generation of a multiple-stained image from several single-stained ones, masking by a nuclear mask, background removal, image registration). The system architecture provides for a user interface based on standard Web browsers and the HTTP protocol, and allows to use both FireWall and Proxy servers. The developed system is being successfully used for processing of images in the FlyEx database (Poustelnikova et al., 2004).


Introduction

The quantitative approach to the analysis of gene expression information allows to reveal fine details in gene regulation (Jaeger et al., 2004). Quantitative gene expression data can be extracted from confocal images of gene expression patterns as these images have very high quality and resolution (Sharpe, Hecksher-Sorensen, 2001). Unfortunately relative few methods are currently available for extracting quantitative data from such images (OME, Myasnikova et al., 2001) and linking this data to other biological information. Images are for the most part stored either in file systems or in relational databases, which do not have any built-in tools for image processing, while the image processing is performed by graphic packages installed on a local computer. This approach requires to download an image from a database or copy an image file to a local computer in order to process it, as well as to insert the processing results in the database or file system again. Thus the development of software for quantification and analysis of gene expression in situ is an important task for bioinformatics.

In this work we present a software system, which is designed for on-line processing of images stored in databases or graphic files. This system is portable across software/hardware platforms, supports image processing in the multiuser mode and publishing data in the Internet. It realizes both basic operations on images (scaling, cuts of rectangular area, filtering of fluorescence intensity, contrast enhancement, etc.) and subject domain oriented operations (generation of a multiple-stained image from several single-stained ones, masking by a nuclear mask, background removal, image registration). The architecture of the system provides for a user interface, which is based on standard Web browsers and the HTTP protocol, and allows to use both FireWall and Proxy servers.

The system designed is being applied to process the digital images of segmentation gene expression patterns. The determination of segments is the subject of intensive research over the last two
decades and the quantification and analysis of segmentation gene expression is an essential part of the work of many researchers (Houchmandzadeh et al., 2002; Wu et al., 2001; Stathopoulos, Levine, 2002).

**System and Methods**

**Methods for image processing**

The basic operations for image processing are implemented by use of the JMagick package, which represents a Java interface to the ImageMagick package. These packages are publicly available (Yeo, 2003; Still, 2003). The subject domain oriented methods for image processing were developed by the authors.

*Data normalization.* Quantitative gene expression data is rescaled in order to get rid of distortions caused by the presence of background signal. The method for removal of background signal is based on the observation that the level of a given gene expression in a null mutant embryo for that gene is well fit by a very broad two dimensional paraboloid. The background paraboloid is automatically determined from the areas of wild type embryos in which a given gene is not expressed and used to remove background from the entire embryo.

*Registration.* To eliminate small individual differences quantitative gene expression data is subjected to registration. Two registration methods were developed. Both methods are based on the extraction of ground control points (GCPs). For GCPs the extrema of the expression pattern of the *eve* gene are used. The affine coordinate transformation is applied to make the corresponding GCPs in different images coincide as closely as possible. In one registration method (the spline or SpA method) GCPs are extracted by a quadratic spline approximation, while in the other method (FRDWT or wavelet) the fast dyadic redundant wavelet transform is used (Myasnikova et al., 2001; Kozlov et al., 2002).

**System architecture**

Figure presents the architecture of the system

The architecture is three-tier. ImageServer (IS) is the middleware. IS is written in Java and represents an application server, which runs on the server under control of JVM (Java Virtual Machine). Use of the Java programming language provides platform independence of the system. IS is permanently waiting for client requests listening to an IP-port with a given number. Clients interact with IS via the HTTP protocol. IS realizes a subset of standard Web server functions, in particular, the GET and POST methods. Use of the HTTP protocol allows to use both FireWall and Proxy servers. To guaranty parallel and independent work of several clients a separate thread is created for each client’s request. The images (operands), operations and other settings are specified as the parameters of the HTTP request. Clients are usually Web browsers, which call IS by including the standard tags <image> into the body of a HTML page. The <image> tags have to contain the server URL and the parameters. IS can access images which are stored in a relational database as BLOBs, image files in popular graphic formats (JPEG, GIF, TIFF, BMP, PNG etc.), as well as the files in RAW format, which represents the byte array of intensities for each image pixel. By default, the target image has the JPEG format, however the output format can be specified explicitly using the corresponding parameter. To provide software independence IS interacts with a database via the JDBC protocol.
**Implementation and Discussion**

Clients invoke IS by constructing the special URL containing parameters which identify the operands (i.e. images), the set of operations and their parameters. It is possible to perform the following operations on a single image: 1) scaling, the scaling coefficients are specified by a client; 2) cut of a rectangular area, the coordinates of the desired area are to be specified by a client; 3) filtering of fluorescence intensity, i.e. extraction of areas, where the expression level of at least one (or each) gene exceeds the predefined threshold or lies in the predefined interval, the threshold or interval of intensities is specified; 4) contrast enhancement; 5) background removal; 6) etc.

There are the following operations performed on a set of images: masking of one image by another, combination of up to three gray-scaled images into the color one, generation of an absolute value of difference between two images, registration of several images. It is possible to combine several operations in a single request.

The operation set described above allows to solve the following tasks: 1) superposition of a nuclear mask, i.e. extractions of segments (nuclei), in which a given gene is expressed; 2) generation of an image, which displays the expression patterns of all genes scanned in an embryo, each gene represented with its own color; 3) evaluation of the difference between two images; 4) estimation of the quality of quantitative data by examination of nuclear masks; 5) the accuracy of registration can be evaluated by comparing gene expression data before and after registration; 6) the level of background signal can be estimated from the image obtained as a result of subtraction of the background free image from the source one; 7) the variability of the expression of a given gene and changes of this variability in time or space can be estimated by subtraction of two images of the same age or by combining of up to three images.

There are several possible directions of further development of the system: increase of a number of basic and subject domain specific operations on images; the possibility to embed image processing operations into a SQL query; expansion of the system to a distributed multiagent system for on-line processing and analysis of images.

**Acknowledgements**

This work is supported by NIH grant R01 RR07801 and GAP award RBO-1286. We thank John Reinitz and David Kosman for key contributions to this work.

**References**


