FISH FINDER: A High-Throughput Tool for Analyzing FISH Images

Sereyvathana Ty[1], James Shirley[2], Xiuwen Liu[1] and David M. Gilbert[2]

[1]Department of Computer Science, Florida State University, Tallahassee, FL 32306

[2]Department of Biology, Florida State University, Tallahassee, FL 32306

ABSTRACT

Motivation: Fluorescence *in situ* hybridization (FISH) is used to study the organization and the positioning of specific base pair sequences of DNA within the nucleus of a cell. Many FISH image analysis software tools have been developed capable of nuclei extraction from the image background as well as fluorescent signal identification. However, previously developed tools use a threshold-based method for extracting nuclei from the image background; therefore, they are highly inflexible, and thus manual extraction of the nuclear boundary is often required. In addition, manual extraction is a subjective task due to the investigator's cognitive bias. As a result, we propose a graphical software tool called FISH FINDER to automatically analyze FISH images. Furthermore, FISH FINDER is based on a statistical approach called Compound Bayesian Classifier, which enables investigators to analyze FISH images efficiently and objectively; additionally, FISH FINDER is capable of both 2D and 3D FISH image analysis.

Availability: FISH FINDER is a standalone MATLAB application and platform independent software. The program is freely available to download from http://www.cavis.fsu.edu/?page=7

Contact: ty@cs.fsu.edu, jshirley@bio.fsu.edu

1 INTRODUCTION

Fluorescence *in situ* Hybridization (FISH) is a technique used to visualize the location of a specific DNA base pair sequence within a nucleus. FISH incorporates fluorescently labeled probes that bind only to the segment of the genome with which they have a high degree of sequence similarity. Thus, FISH provides a way to visually locate a gene within the nucleus using fluorescent microscopy. The manual process of data acquisition is time consuming and subjective due to inconsistency of the investigator's decision. Therefore, to achieve high throughput and objectivity, the method of data acquisition should be standardized as well as automated so less manual work is required from the investigator and data is objectively acquired.

The simplest automated approach for extracting a cell from the background is achieved by setting a threshold value above the background level of fluorescent light intensity (Shopov *et al.*, 2000). All pixels with light intensity values higher than the threshold are considered as part of the valid nucleus, while pixels with light intensity lower than the threshold are excluded as background noise. Another similar approach is to set a fixed signal-to-noise ratio, which incorporates a fixed threshold in order to isolate and extract the valid nuclear boundary (Pernthaler *et al.*, 2003), (Heintzmann *et al.*, 2004). However, fluorescent light intensity values vary greatly between experiments because fluorescent and cytological debris are usually present in FISH experiments. Due to non-uniform intensities found within the background of FISH images, the threshold-based segmentation methods appear to have intrinsic problems. Thus, both of the common threshold-based methods used by existing software programs require manual modification from one sample to the next within the same experiment (Iannuccelli *et al.*, 2010). Moreover, other advanced segmentation methods such as dynamic programming (DP) or pattern recognition require user interaction and/or training data for producing optimum output (Gudla *et al.*, 2008), (McCullough *et al.*, 2008), (Nandy *et al.*, 2009). Therefore, it is essential for us to turn our attention to a more flexible and adaptive method that is independent of the threshold-based notion and requires minimal user interaction.

In this paper, we have developed a highly automated graphical user interface software tool called FISH FINDER, which is specifically designed to analyze FISH images effectively and objectively. The main contributions of FISH FINDER are 1) It segments cells automatically using a statistical approach called Compound Bayesian Classifier. Using this method, we are able to produce highly satisfactory results of segmentation (Yokochi *et al.*, 2009), which remains the critical issue of automatically analyzing FISH images. 2) FISH FINDER automatically identifies FISH signals and can compute the shortest distance from fluorescent signals to nuclear landmarks such as the nuclear boundary using the Euclidean metric, or compute the shortest distance between alternately colored signals for analysis of DNA compaction; FISH FINDER also allows for selection of up to ten different fluorescent signal channels to be processed. 3) Processed images can be opened for editing or re-editing (i.e. add or remove FISH signals, remove unwanted cells, and add or adjust cell boundary.) 4) Finally, it allows investigators to save the results in comma-separated values (CVS) file format that can be imported into other computational software such as R or Microsoft Excel.

2 PROGRAM FEATURES

2.1 Input

FISH FINDER uses LOCI Bio-Format toolkit for reading in different types of input files such as Tagged Image File Format (.tiff) image stack, or DeltaVision (.dv) images stack. Furthermore, FISH FINDER allows investigators to import a list of data folders containing multiple FISH images for processing. Investigators must specify the number of fluorescent channels to be analyzed before starting the image analysis process. However, the investigator does not need to specify or set any predefined values for nuclei segmentation and FISH signal detection. FISH FINDER automatically determines optimal parameters for segmenting foreground from background.

2.2 Segmentation via Classification

The main advantage of FISH FINDER is our method of nuclei segmentation. Given a FISH image, the program classifies pixels into either of two classes, the background or the nucleus. We de-

termine each class based on the pixel value distribution of the image. The condition of the determined classification is obtained by using the general statistical approach called Bayesian Decision Theory. This method requires us to estimate each class conditional probability distribution, which is achieved by computing the histogram of the FISH image; thus, the segmentation problem becomes a classification one.

For reliability, instead of classification based on individual pixel value, we use a Compound Bayesian Classifier to enhance the classification efficiency (Duda *et al.*, 2000). Afterward, we iteratively re-estimate the class conditional probability distribution based on the current classification to improve the boundary accuracy.

2.3   Fluorescent Probe Signal Extraction

A fluorescent probe signal can be determined objectively by analysis of pixel intensity values within an extracted nucleus. First, we select the most in focused image using the DAPI staining channel from the image stack; this is usually the image with the highest overall intensity. Then, we compute the average light intensity of the fluorescent signal channel, which then serves as the threshold for selecting fluorescent probe signals. Any notable bright spot identified with higher intensity than the threshold will serve as candidates for FISH signals. Then we reduce the number of candidate signals by analyzing the size of the signal as well as comparing the intensities of candidate signals.

Therefore, given the location of a FISH signal, we determine the distance between the signal and the nearest point on the determined nuclear boundary using Euclidean distance measurement. FISH FINDER then computes the radial distribution ratio, which is the peripheral distance of the FISH signal divided by the average radius from the determined nuclear boundary. The average radius of a segmented nucleus is determined by finding the radius of a circle with the same area as the segmented nucleus. Furthermore, the program can also compute the shortest distance between alternately colored FISH signals for analysis of DNA compaction.

2.4   Result Verification

After processing image datasets, investigators can immediately choose to open FISH FINDER's editing screen, which allows investigators to import folders of processed data. The editing screen enhances the investigator's ability to edit the image data by allowing them to add or remove nuclei, to adjust a nuclear boundary, and to add or remove FISH signals. Finally, investigators can then use FISH FINDER to export the results to comma-separated values files, which can be imported into many computer software programs for further analysis.

3   USER INTERFACE

FISH FINDER's graphical interface was created using the Math Works MATLAB programming environment. FISH FINDER was designed for investigators who have little or no experience in using MATLAB. In addition, FISH FINDER assumes that users have no background knowledge in computer programming languages. Furthermore, the functionality of FISH FINDER's graphical interface is divided into two main phases. The first phase is the automatic processing mode, and requires the investigators to specify input data for processing. Once files have been specified, the program produces the necessary outputs or results for the second phase,

which allows the investigators to verify, modify, and export results. However, because MATLAB is a commercial produce, a licensed copy of MathWorks MATLAB is needed in order to use FISH FINDER
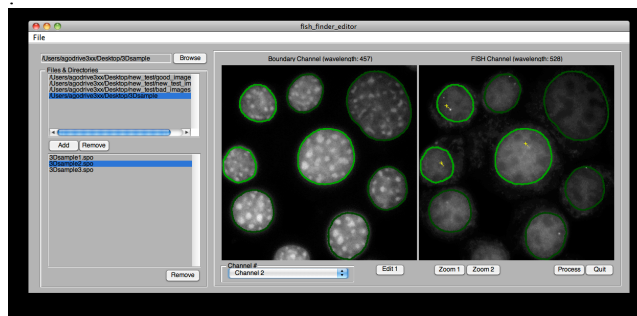.

Fig. 1 FISH FINDER Editor Window showing result from the segmentation process

4   CONCLUSION

FISH FINDER is an important analysis tool capable of automatically extracting nuclei boundaries via Compound Bayesian Classifier, localizing FISH signals, enhancing investigators' ability to verify data, and saving data in the comma-separated values format for future reference. Consequently, FISH FINDER minimizes errors due to manual FISH analysis, by reducing the cognitive bias of an investigators' judgmental process. More importantly, FISH FINDER is a highly efficient user-friendly software tool which allows investigators to obtain satisfactory and objective results.

REFERENCES

Duda, R. *et al.* (2000). *Pattern Classification.* Wiley-Interscience Publication.
Gudla, P. R. *et al.* (2008). A High-Throughput System for Segmenting Nuclei Using Multiscale Techniques. *Cytometry Part A* , 73A: 451-466.
Heintzmann, R. *et al.* (2004). Double-pass fourier transform imaging spectroscopy. *Opt. Express* , *12* (5), 753-763.
Iannuccelli, E. *et al.* (2010). NEMO: a tool for analyzing gene and chromosome territory distributions from 3D-FISH experiements. *Bioinformatics* , *26* (5), 696-697.
McCullough, P. D. *et al.* (2008). Segmentation of Whole Cells and Cell Nuclei From 3-D Optical Microscope Images Using Dynamic Programming. *IEEE Transactions on Medical Imaging* , *27* (5), 723-734.
Nandy, K. *et al.* (2009). Automatic Nuclei Segmentation And Spatial FISH analysis For Cancer Detection. (pp. 6718-6721). Engineering in Medicine and Biology Society.
Pernthaler, J. *et al.* (2003). Automated enumeration of groups of marine picoplankton after fluorescence in situ hybridization. *Applied and Enviromental Microbiology* , *69* (5), 2631-2637.
Shopov, A. *et al.* (2000). Improvements in image analysis and fluorescence imcroscopy to discriminate and enumerate bacteria and viruses in aquatic sample. *Aquatic Microbial Ecology* , *22* (2), 103-110.
Yokochi, T. *et al.* (2009). G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proceedings of the National Academy of Sciences* , 19363-19368.