Image analysis platform for yeast biologists

Tleis, M.; Anemaet, I.G.; Heusden, G.P.H. van; Verbeek, F.J.

Image Analysis Platform for Yeast Biologists

Mohamed Tleis*, Ginny Anemaet†, Paul van Heusden†, Fons Verbeek*
*Section Imaging and BioInformatics, LIACS, Leiden University, Leiden, The Netherlands.
Email: mtlais@liacs.nl, fverbeek@liacs.nl
†Institute of Biology, Leiden University, Leiden, The Netherlands.

***Abstract*—The yeast *Saccharomyces cerevisiae* is used to understand gene expression and genetic networks. To study the levels and localization of individual proteins, tagging with fluorescent proteins like GFP is widely used. An important issue is quantification of the fluorescent signal in microscopic images. In addition to microscopy, Image Analysis has developed as an established complement allowing routine quantification of microscopic observations. Gene expression is visualized by microscopy through specific fluorescence, hence image analysis was probed to accomplish further progress. This research promotes a yeast analysis platform that assists yeast molecular biologists in analysing their experiments. The major components of the platform are options to segment yeast images, measure chosen features and generating statistical reports of the measurements. This system is currently used to quantify microscopic images. More options to segment cells at different stages of the cell cycle and choosing an automatic segmentation method based on a ground-truth image, are marked as our future work.**

I. Introduction

In molecular genetics the yeast *Saccharomyces cerevisiae* is used to understand gene expression and genetic networks. The yeast is then studied over a range of conditions. An important issue is a quantification of the yeast cell in microscopic images.

Yeast cells expressing fluorescently tagged proteins can be viewed by (confocal laser) fluorescence microscopy, for example by a Zeiss LSM5 Exciter confocal microscope [1]. Images are acquired as two or three channel images. One or two channels are of the reporter construct (GFP, YFP, CFP. Excitation at 488nm for GFP or 514nm for YFP, and emission at 500-550 for GFP or 530-600nm for YFP) [2]. The last channel being a bright-field image depicting the structure of the cells.

In addition to microscopy, image analysis has developed as an established complement, allowing routine quantification of microscopic observations [3]. Currently flow cytometry is one of the methods used to measure levels of fluorescent proteins; this can, however, not provide the quantification that image analysis can. Hence the use of image analysis was probed to accomplish further progress. This should be part of a comprehensive Image Analysis Platform. Such platform must have the following components: (1) Segmentation, (2) Measurement (3) Data Visualization and (4) GUI (graphical user interface).

Image analysis has been successfully applied in a wide variety of fields ranging from astronomical observations to cells analysis. To name few other fields: nuclear medicine, medical diagnostics, industry, lithography, microscopy, lasers, biological imaging, remote sensing, law enforcement, radar images, geological exploration [4], sports management [5], chemical Imaging [6], etc...

Crucial to image analysis is image segmentation which is defined as subdividing an image into its constituent regions or objects [4]. The result of image segmentation are objects representing connected regions of similar intensity. From these regions, measurements can be conducted. Some practical applications of image segmentation are: image processing, computer vision, face recognition, medical imaging, digital libraries, image and video retrieval, and cell image analysis to measure gene expression [7], [8].

From the segmentation results, binary masks and contours of cells can be extracted to perform measurement on the cells for their features. Measurement is defined as the assignment of numbers to objects or events [9] and it is a cornerstone of most natural sciences.

After performing measurement, statistics and data visualization are necessary to make the measurement meaningful for humans. Statistics refers to quantities concluded from a set of data [10] while data visualization aims to communicate information clearly and effectively through graphical means [11].

In this paper we describe a system that offers segmentation, measurement and data visualization tools to molecular yeast biologists (cf. Fig. 1). The paper is structured as follows. First, it discusses some of the segmentation methods and algorithms used, then the features measured, and the visualization report of the data. In the results section, a sample experiment is performed to show the benefits of this system.

II. Segmentation

For a different merits of images, certain set of segmentation methods and parameters can give different results. Hence, this system provides an option to choose between a number of different segmentation methods that have proven to give good results for the type of images under study.

One segmentation method uses sigma filter [12], [13] to smooth and blur images and is specially useful with fluorescence images (cf. Fig. 2). Despeckle tool [14] was also used to remove impulsive (salt-and-pepper) noise [4] (cf. Fig. 3).

Another method uses sobel filter [4], [14] to detect edges, and specially useful on bright-field images. Next, a recently developed method using Hough transform and minimal path algorithm is also available. The idea of the new method is to use Hough Transform to locate the circles that form part of the cell contours in a gradient skelentonized image, create a gray scale parametric plane of the pixels surrounding the cells, and apply minimal path algorithm to find the path (full contour) of every cell (cf. reference [15] for more details on this method). The extracted contours are then used to obtain a binary mask of every cell to measure it in the overlaid fluorescence channel.

978-1-4799-0251-4/13/$31.00 ©2013 IEEE

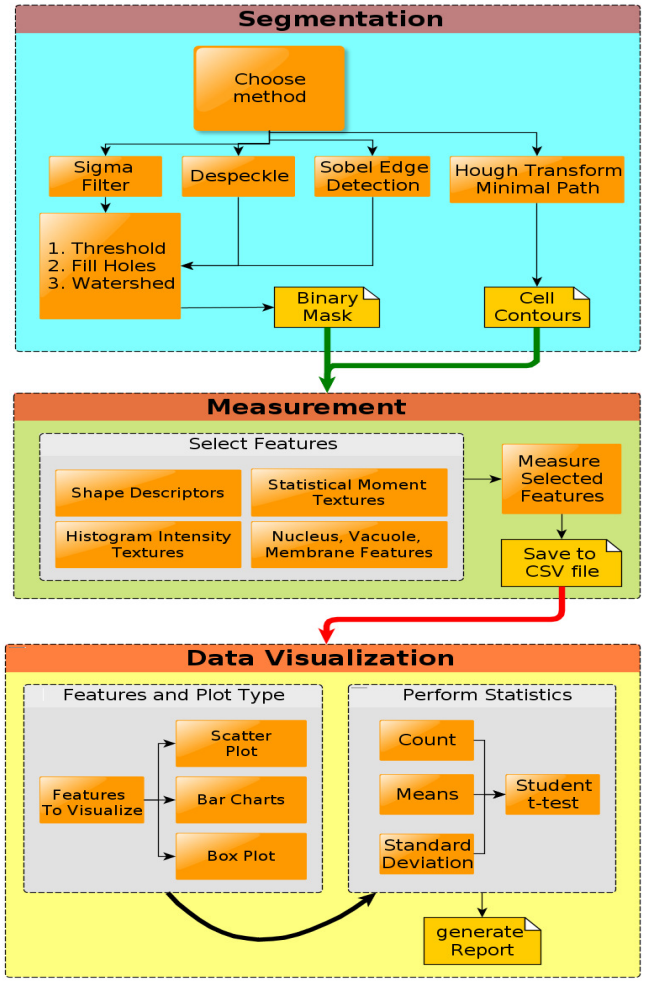
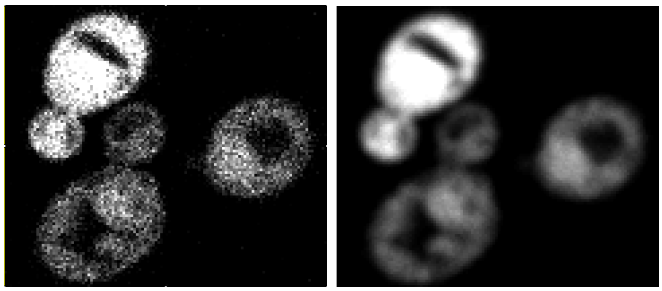


Fig. 1. Workflow of the Image Analysis Platform



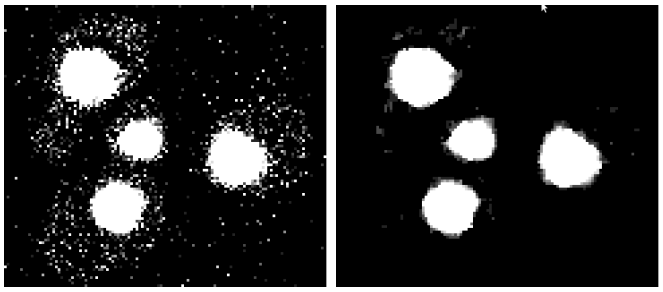
(a) Fluorescence Image (b) Applying Sigma Filter

Fig. 2. Sigma Filter

In region growing segmentation methods that results in a binary mask of images, manual or automatic threshold such as triangle auto-threshold [16], [14] were used (cf. Fig. 4(a)). For separation of clumped cells, Watershed algorithm [4], [14] is used (cf. Fig. 4(b)).

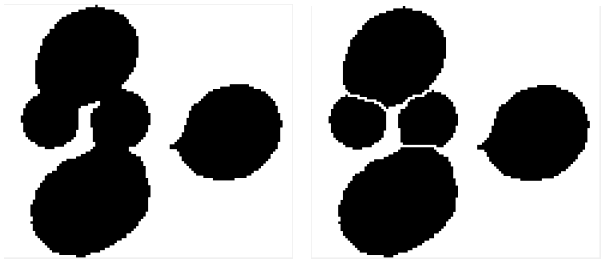
III. Selected Measurement

This system provides an option to choose from a list of features to be measured; moreover, it provides an option to exclude outliers based on the circularity and size of the measured cells.



(a) impulsive noise (b) Applying Despeckle

Fig. 3. Despeckle



(a) Applying Triangle Auto-threshold (b) Applying Watershed Algorithm

Fig. 4. Thresholding Image and Separating Cells

The features that can be measured are: area, intensity, intensity standard deviation, perimeter, density, estimation of vacuole size, textures based on statistical moments (variance, skewness), and textures based on the intensity histogram (smoothness, uniformity and entropy) [4].

Moreover, for those images that contain a DAPI stain channel in the nucleus, the nucleus is measured for its intensity, area, and other features.

The surface area of a cell (object) is simply the number of pixels occupied by that cell in the image, and for SI unit, it is multiplied by the area of one pixel in μ-meters (width * height of pixel).

The intensity of a cell is the average gray-level value of the pixels occupied by this cell, which ranges between 0 and 255 for the gray images used [14]. In addition, standard deviation of the intensity is calculated.

The perimeter representation used was that of Vossepoel and Smeulders [17]. This representation is based on chain-codes:

$$L_{vs} = (0.980).N_e + (1.406).N_o - (0.091).N_c \qquad (1)$$

where N_e, N_o and N_c represents the number of even codes, odd codes and corners in the chaincode of cell boundaries [4].

The density of a particle object is its area multiplied by the average mean of gray-level values:

$$d = A * \mu \qquad (2)$$

where d represents density, A represents area and μ represents mean of intensity of a measured cell.

Estimation of vacuole size can be computed if necessary, for proteins expressed by genes known to have expression only in the cytoplasm and nucleus of yeast cells without its vacuoles. This estimation is computed in two ways, either by manually thresholding each of the region of interests ROIs representing each cell with a predefined threshold value, or by using a vacuole filter algorithm that looks for dark spots in the ROI.

The second statistical moment of intensity values represents its variance:

$$\mu_2(z) = \sum_{i=0}^{L-1} (z_i - m)^2 . P(z_i) \tag{3}$$

where z_i is the intensity value of the histogram at location i, m the mean intensity value, and L the total number of intensity levels (histogram range).

The variance is a measure of intensity contrast that were also used to establish the descriptor of relative smoothness:

$$R(z) = 1 - \frac{1}{1 + \sigma^2(z)} \tag{4}$$

The third statistical moment represents skewness of the intensity histogram:

$$\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 . P(z_i) \tag{5}$$

Two more features based on histograms of the cell ROI (region of interest in an image) were calculated. The first is the uniformity, which has a maximum value for a cell image in which all intensity levels are equal:

$$U(z) = \sum_{i=0}^{L-1} P^2(z_i) \tag{6}$$

The second feature being the Entropy which is a measure of variability. The Entropy is zero for constant images:

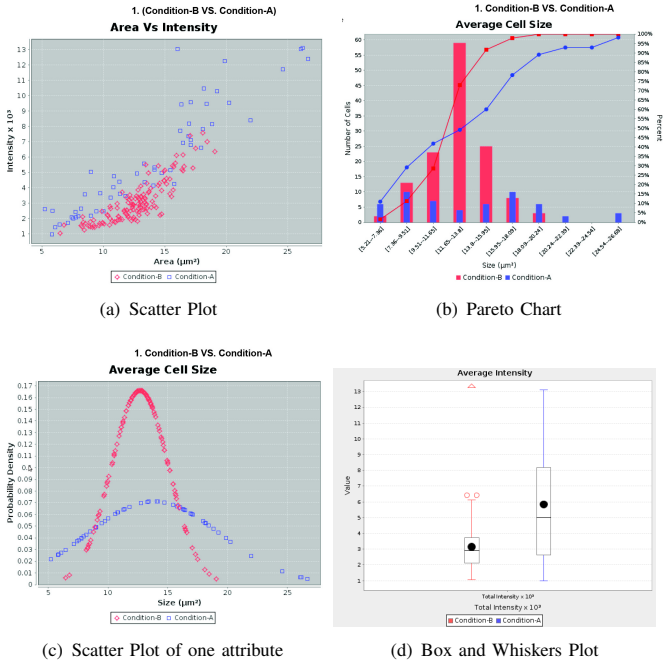
$$e(z) = -\sum_{i=0}^{L-1} P(z_i) . log_2 P(z_i) \tag{7}$$

All these measurements of features and simple feature textures are saved automatically into a CSV (comma-separated values) file that is used at the following step to generate a report or inspected manually.

IV. Statistics and Data Visualization

After measurement, usually molecular biologists would like to compare two set of cells to study any significant changes in cell size or fluorescence intensity in cells having a certain gene mutation or cultured in a different medium (ex. 2M NaCl, 50mM KCl, etc..). The changes of cells in different conditions is compared usually to wild-type cells.

To assist in achieving this goal of comparison, our system generates automatically a report in pdf format holding basic statistics about the experiment and graph charts to visualize the results.



(a) Scatter Plot (b) Pareto Chart

(c) Scatter Plot of one attribute (d) Box and Whiskers Plot

Fig. 5. Graph Charts to visualize measurement

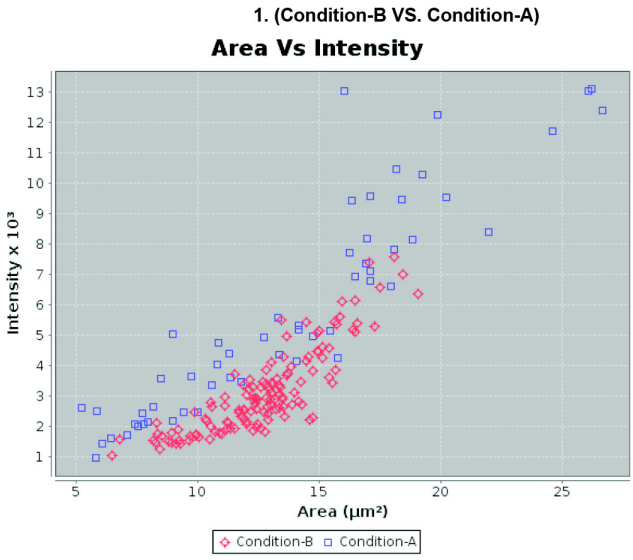
The statistical information created into the report includes counts of the cells belonging to different cultures. The mean value of their surface area and intensity of the fluorescence signal along with their standard deviations. The unpaired student t-test is performed to report the t-value and p-value to assist in recognizing the significant of the difference between two different cell cultures.

The most meaningful graph chart for the biologists are those scatter plots that visualize the different values of cell surface area on one axis against fluorescence intensity on the other with a different label for each cell culture (cf. Fig. 5(a)). Other visualization options available are charts visualizing one attribute of the cells in one chart, as a Pareto chart (cf. Fig. 5(b)), a scatter plot visualizing the Gaussian density distribution of the data for the two cell cultures (cf. Fig. 5(c)), or a box and whisker plot to facilitate the analysis of the range of data (cf. Fig. 5(d)).

V. Results

To check results of our system, a small experiment was performed to compare two cell cultures (labelled as condition-A and condition-B). 18 images were acquired in this experiment, and 188 cells were detected. Then the cells were measured and all the measurement are output into a CSV file, which is analysed automatically to generate a report holding a graph chart and statistics(cf. Fig. 6).

Having a look at the chart, and the statistics (specially p-values of the cell surface area and intensity). It is clearly visible that cells belonging to condition-B are smaller and has less GFP fluorescence than the cells in condition-A. The results of the flow cytometry test (cf. Fig. 7) also agrees with these conclusions.



(a) Scatter Plot

Un-paired Student's t-test:
Area t-Value : -1.46
Area P-Value (Assuming null hypothesis): 0.15
Intensity t-Value : -5.54
Intensity P-Value (Assuming null hypothesis): 0

Condition-B : 133 Cells.
Area Mean : 12.67
Area Stdev : 2.4
Intensity Mean : 3.14 x 10^3
Intensity Stdev : 1.39 x 10^3

Condition-A : 55 Cells.
Area Mean : 13.82
Area Stdev : 5.61
Intensity Mean : 5.84 x 10^3
Intensity Stdev : 3.5 x 10^3

(b) Statistics

Fig. 6. Report of Experiment

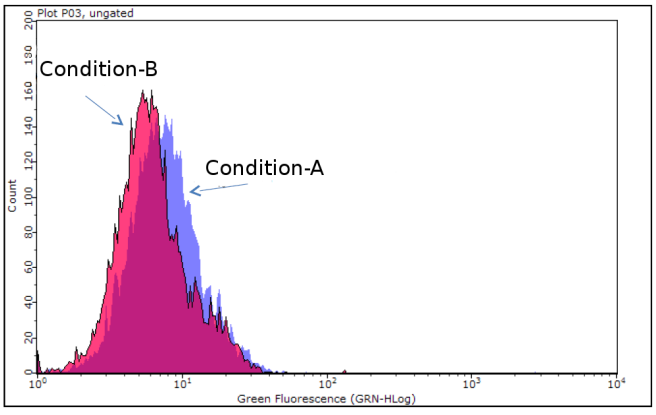


Fig. 7. Results of the Flow Cytometry test

VI. Conclusion

In this research, a new system to assist yeast biologists analysing yeast cell images, was proposed. The results show the benefits of using this system to quantify the microscopic images. Hence, this newly proposed system is used by our collaborators.

As future work, we are going to proceed in developing more algorithms specific in detecting and distinguishing between different cells in various stages of the cell cycle, and create a system that automatically chooses the appropriate segmentation method based on a ground truth image or some manual information. Moreover, we are going to work on more complex feature texture measurements and shape descriptors, in addition to automatic classification of data.

Acknowledgment

We would like to acknowledge that this research is partially supported by the Erasmus Mundus, JOSYLEEN program.

References

[1] Z. Inc., "Zeiss 5 exciter laser scanning microscope manual," 2012.

[2] J. Zahradka, G. P. H. van Heusden, and H. Sychrova, "Yeast 14-3-3 proteins participate in the regulation of cell cation homeostatis via interaction with nha1 alkali-metal-cation/proton antiporter," *Biochimica et Biophysica Acta*, vol. 1820, pp. 849–858, 2012.

[3] C. R. Thomas and G. C. Paul, "Applications of image analysis in cell technology," *Current Opinion in Biotechnology*, vol. 7, no. 1, pp. 35–45, February 1996.

[4] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed. Pearson Education, 2008.

[5] A. Ferrand and M. Pages, "Image management in sport organisations: the creation of value," *European Journal of Marketing*, vol. 33,3/4, pp. 387–402, 1999.

[6] M. Schaeberle, C. Karakatsanis, C. Lau, and P. Treado, "Raman chemical imaging: Noninvasive visualization of polymer blend architecture," *Analytical chemistry*, vol. 67(23), pp. 4316–4321, December 1995.

[7] M. Tleis and F. J. Verbeek, "Yeast cell features extraction plugin," *In Proceedings, ImageJ Conference 2012*, 2012.

[8] P. van der Putten, L. Bertens, J. Liu, F. Hagen, T. Boekhout, and F. J. Verbeek, "Classification of yeast cells from image features to evaluate pathogen conditions," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 6506, Jan. 2007.

[9] J. Pedhazur and L. Schmelkin, *Measurement, Design, and Analysis: An integrated approach*, 1st ed. Lawrence Erlbaum Associates, 1991.

[10] "Merriam-webster online dictionary." [Online]. Available: www.merriam-webster.com

[11] V. Friedman, "Data visualization and infographics," *Graphics, Monday Inspiration*, January 2008.

[12] J. Lee, "Digital image smoothing and the sigma filter," *CVGIP(24)*, vol. No. 2, pp. 255–269, 1983.

[13] M. Schmid and T. Collins, "Sigma filter plus," 2007, http://rsbweb.nih.gov/ij/plugins/sigma-filter.html.

[14] W. Rasband, "Imagej," 2012, u. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/ 1997-2012.

[15] M. Tleis, P. van Heusden, and F. Verbeek, "Yeast cell contour detection using hough transform and minimal path algorithm," submitted.

[16] G. Zack, W. E. Rogers, and S. A. Latt, "Automatic measurement of sister chromatid exchange frequency," *Journal of Histochemistry and Cytochemistry*, vol. 25 (7), pp. 741–753, 1977.

[17] A. M. Vossepoel and A. W. M. Smeulders, "Vector code probability and metrication error in the representation of straight lines of finite length," *Computer Graphics and Image Processing*, vol. 20, pp. 347–364, 1982.