PARALLEL IMAGE ANALYSIS
OF MORPHOLOGICAL YEAST CELLS*

Laurent Manyri [†]
LAAS-CNRS, 7 Avenue du Colonel Roche, 31077 Toulouse, FRANCE

Andrei Doncescu
LAAS-CNRS, 7 Avenue du Colonel Roche, 31077 Toulouse, FRANCE

Laurence T. Yang
Department of Computer Science St. Francis Xavier University,
Antigonish, NS, B2G 2W5, Canada

Jacky Desachy
GRIMAAG University of French West Indies, 97157 Pointe-q-Pitre, FRANCE

Abstract   Fermentation is a critical process for the production simpler substances from or-
ganic molecules or could be used to obtain the ethanol by the anaerobic breaking
down of sugar. In this work we present different image analysis steps in order
to characterize the cell morphology during the biotechnology process. The cell
morphology is an important element of the stress which could disturb the pro-
duction of the biomass or a metabolite. For this purpose we develop a Java
software dedicated to an automatic analysis. The software allows us to have in-
formation about the growth cells, morphometric analysis (volume/surface) and
morphology (budding cells). We want to have the information in biological real
time to be able to modify the control parameters, therefore the image is cut into
small slices which are analyzed separately by a parallel algorithm.

Keywords: Image Processing, JavaMPI, cells, snakes

1.   Introduction

Recently, the american researchers showed the human yeast carry out synthetically the otherwise different tasks of producing the human biosynthesis protein. The recent results allow us that a 'humanised yeast' could simplify drug manufacture by the introduction of human gene inside yeast chromosome. Since yeast grow faster and need less tending than mammalian cells, could be a possible solution to produce proteins cheaper and easier. One of the most famous example is the production of the human insulin for diabetics, which does not need added sugars, is brewed from yeast in this way. Moreover, it is difficult to produce these proteins to a commercial scale due to the incapacity of the biologist to keep the cell on a specific pathway therefore the necessity to develop control adaptive tools. Today, the pace of progress in fermentation is fast and furious, particularly since the advent of genetic engineering and the recent advances in computer sciences and process control. The high cost associated with many fermentation processes makes optimization of bioreactor performance trough command control very desirable. Clearly, control of fermentation is recognized as a vital component in the operation and successful production of many industries.

The last results indicated that the improvability and the control of the bioethanol production by alcoholic fermentation needs information about cell morphology.

The bioethanol is obtained by a chemical reaction :

$$C_6H_{12}O_6 \longrightarrow 2CH_3CH_2OH + 2CO_2$$
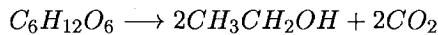
The operator of this reaction is a yeast, *Saccharomyces Cerevisiae* . The speed of this reaction and the production of ethanol depend of the physiological state of the cells (quality) and may be studied by morphometry (number of cells, surface and volume) and by morphology (classes). The yeasts develop by budding. Four classes of cells appear in the process : the single alive, the single dead, the budding cell (a mother and a child) and the aggregate of cells. This analysis is non based on a study of viability but we are interested in sizing and classifying the cells.

For this characterization, we develop different tools for microscopy image analysis in order to create an optical sensor which can be

- *off line in the case of separated treatment*

- *in situ for direct information on the process*

The aim of this study is to try to give the evolution of a cell during a process. Different kinetic models describe the steps of a alcoholic fermentation; this work have to approximate the steps of :

- adaptation : there's a change of the size of the cells

- exponential growth where the number of cells increase : increasing of the number of budding cells

- limit growth

- decline of the cells

2.   Parallel Image Processing

In this paper we describe a new method of edge extraction of the yeast cells by active contours. To differentiate the budding cells of the single cells we have used the curvature analysis. The last algorithm used allowing to approximate each cell (budding or not) by an ellipse using a method based on least square method. All these steps are englobed in an unique algorithm which is applied in parallel for each cell or conglomerate of cells.

Image acquisition

One of the difficulty of the Cells Image Analysis is the modification of the size by take out of a sample. Two methods has been developed to solve this problem:

  1 to examine under the microscope the cells

  2 in-situ microscope

In this paper we present the first one. Therefore the growth of the cells during the fermentation is analysed by a microscope coupled to a camera which snaps several images. The microscope used is Olympus and the camera is a Nikon. A specific method of optical microscopy based on 'black background' allows to have only the contour of the cells. This method has been selected to reduce the influence of the background into the algorithm. The images obtained contain only the cell's edges with certain fuzziness.

Segmentation for parallel treatments

One of the control parameter for the 'health cells' is the modification of the ellipsoidal form which seems to be a stress indicator. If the computing time of this task is important the cell's culture risks to stop the proposed evolution. Due to the image size and time computation, the image is divided in small area : each area corresponds to one cell. This decomposition allows to use different system of processing. The final scheme chosen is a parallel implementation based on master/slave scheme. The master will segment the image and send a small image containing one cell to a slave. The computation of the slave is to
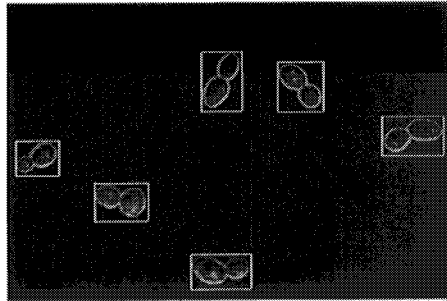
Figure 1.   Separation of the cells

approximate the cell by a model and the master will recover the correspondent data for each cell in order to produce the stats. A simple threshold and separates the cells as it's visible on Figure 1.

JAVA MPI Interface

The treatment is divided by mean of the segmentation that identify region of interest (ROI). For these regions some subroutines have to estimate the contour of the cells by ellipses defined in section IV. As these subroutines used JAVA, we would like to use an approach to parallel programming in Java, defined as the most common object-oriented language.

For parallel processing, mpiJava (Carpenter, 1997) is an object-oriented Java interface to the standard Message Passing Interface (MPI, 1995); this interface was developed as part of the HPJava project. Indeed we need a good message passing API and Java presents various package for communication used in this interface that's why it's a well-designed, functional and efficient Java interface to MPI (Kanevsky, 1998).

Our parallel implementation is based on a master/slave scheme guided by mpi-Java.

Image segmentation by Snakes

One of the difficulty of the segmentation of this type of images is the diffraction on the cell membrane producing a fuzzy edge and of course the low level of luminance. In order to have a better approximation of the contour, we have ameliorated the convergence of the parametric snake, based on active contours proposed by Kass *et al.* (Kass, 1988). The snakes are used in computer vision and image analysis to detect and locate objects, and to describe their shape (Cohen, 1991).

A snake is a curved $x(s) = [x(s), y(s)]$, $s \in [0, 1]$, that moves through the

spatial domain of an image to minimize the functional

$$E = \int_0^1 (\alpha \|x'(s)\|^2 + \beta \|x''(s)\|^2) + E_{ext}(x(s)) ds \qquad (1)$$

- α and β are parameters that control the snake's tension and rigidity

- $x'(s)$ is the first derivative of $x(s)$

- $x''(s)$ is the second derivative of $x(s)$

- E_{ext} is the external energy function derived from the image, it rule is to move the snake toward the important values of the edges; for example, the gray value of the image, the negative first derivative may be used.

In our implementation, the external force is the gradient vector flow (Xu, 1997) (Xu, 1998). This field is computed as a spatial diffusion of the gradient of an edge map derived from the image. It's powerful than traditional snakes because they cannot move toward objects that are too far away and that snakes cannot move into boundary concavities or indentations.

How the cells are separated, each slave have to force the snake to lock on one contour. Different external forces have been used and to conclude, the gradient vector flow is the appropriate external force because of his fast convergence to concavities in the case of budding cell. This is the main condition to make the difference between the single cell and the budding cell.

At the end on the process a contour corresponds to a cell that is described in the map by (x_n, y_n) with $n \in [0, N]$.

3. Analysis

If the extraction of the cell has been done we need to find a robust algorithm allowing to measure the size of the single cell or the morphometric parameters of the budding cell. This involves that we are able to discriminate the mother cell and the daughter cell. The originality of this work is based on the curvature of the contour.

A single cell has a regular curvature; it should be approximate by a circle. In the other hand, a bud has an invariant curvature. For the curvature, we compute the radius as follow

$$R = \frac{\sqrt{d_{x_n}^2 + d_{y_n}^2}^{-\frac{3}{2}}}{d_{x_n} d_{y_n}^2 - d_{y_n} d_{x_n}^2}.$$

But the change of the curvature may be appreciated by the sign of

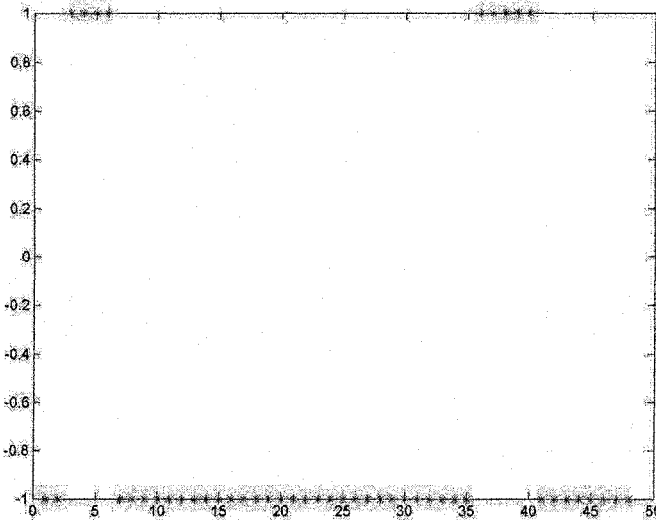$$d_{x_n} d_{y_n}^2 - d_{y_n} d_{x_n}^2 \quad \text{with}$$

Figure 2.   Curvature radius computation

- d_{x_n} and d_{y_n} the first derivatives of (x, y)

- $d_{x_n}^2$ and $d_{y_n}^2$ the second derivatives of (x, y)

To compute the first and the second derivative, the choice of the step allows or not to find the change of curvature. More it's small more there will be changes while more it's large less a change in the curvature can be detected. In this case of images, a step of 10 has been chosen.

We can evaluate the results of this computation on the Figure 2.

The X-axis correspond to the pixels and the Y-axis is the value (-1 or 1) of the radius. For a simple cell, there is no change of curvature but for this contour two peaks appear. When the curvature value is equal to -1 it's correspond to a cell (mother or child). Indeed the pixels corresponding to the first at the second peak is a cell; moreover the pixels between the second and the first peak belong to the second cell.

4.   Cell Modeling

The biological knowledge allows us to modeling the cell shape by an ellipse. After the analysis of the contours there is no information a priori to distinguish a single cell of a budding one. This comparison is possible using the size of these cells; that's why this contours are approximated by models. A cell is considering an ellipse defined by five parameters : a center (x_0, y_0), a minor

axis a, a major axis b and an orientation θ. It can be described by several equations :

- explicit $\frac{((x-x_0)+E(y-y_0))^2}{a^2(1+E^2)} + \frac{((y-y_0)+E(x-x_0))^2}{b^2(1+E^2)} - 1 = 0$

- implicit $F(a, x) = ax^2 + bxy + cy^2 + dx + ey + f = ax$ with $a = (a, b, c, d, e, f)$ and $x = (x^2, xy, y^2, x, y, 1)^T$

- parametric $< \overline{x} - \overline{v} > E < \overline{x} - \overline{v} >^T = r^2$

Different global and local methods have been evaluated, a global estimation of their efficiency is presented in Table .

Method	Time (sec)	Precision
Hough Transform (Hough)	360	Lower
Fuzzy c-shells (Dave, 1992)	2	Lower
Differential Evolutionary (Storn, 1995)	188	Fine
Stochastic gradient	0.01	Lower
Direct Least Square (Fit, 1999)	0.01	Very fine

Table 1. Precision and time computation for estimation

This table presents five methods that have been experimented in order to approximate a simple contour and the more efficient is Direct Least Square; it's described in the next section. The precision of one method is difficult to be appreciated; it represents the best value of the minimum of the sum of squared algebraic distance. Only the Differential Evolutionary Algorithm, a class of stochastic search and optimization methods including Genetic Algorithm, and the method Direct Least Square are efficient in time for this problem.

Fitting Ellipse

This is an efficient method for fitting ellipses from scattered data. Other algorithms either fitted general conics or were computationally expensive. By minimizing the sum of squared algebraic distance

$$D_A(a) = \sum_{i=1}^{N} F(x_i)^2$$

of the curve F to the points using the implicit equation of the conic :

$$F(a, x) = ax^2 + bxy + cy^2 + dx + ey + f = ax$$

To avoid the solution $a = 0$ some constraints are applied to the parameters :

- linear : $c.a = 1$

- quadratic $a^T C a = 1$ where C is a 6x6 constraint matrix.

As a quadratic constraint is applied, the minimization can be solved by : $D^T Da = \lambda$ where $D = [x_1...x_N]^T$ and λ is the Lagrange multiplier. Moreover, as the conic is an ellipse, we impose $b^2 - 4ac > 0$.
Then $a^T Ca = 1$ as

$$a^T \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} a = 1$$

Then the minimization is about $E = \|Da\|^2$ with the constraint $a^T Ca = 1$. Introducing Lagrange multiplier the system becomes

$$2D^T Da - 2\lambda Ca = 0 \qquad (2)$$
$$a^T Ca = 1 \qquad (3)$$

Then

$$Sa = \lambda Ca \qquad (4)$$
$$a^T Ca = 1 \qquad (5)$$

where $S = D^T D$. The system is solved by considering the generalized eigenvectors of (4).
If (λ_i, u_i) solves (4) then $(\lambda_i, \mu u_i)$ for any μ from (5), we can find μ_i as $\mu_i^2 u_i^T Cu_i = 1$ giving

$$\mu_i = \sqrt{\frac{1}{u_i^T Cu_i}} = \sqrt{\frac{1}{u_i^T Su_i}}$$

5.   Results and discussion

The image's segmentation, the analyze of the contours and the approximation of elliptic model allow to count the number of singles and budding cells. Note that after the analyze of contours, the cell may correspond to a budding cell i.e. a mother and a child or it may correspond to two singles. Indeed, by modeling and having the parameters, the two major axises are compared. Consider $a_{max} > a_{min}$ the two axes then :

- if $a_{max}/a_{min} > 1/3$ there are two cells

- if $a_{max}/a_{min} \leq 1/3$ it's a budding cell i.e. a mother and a child

The cells (single and budding cells) have been counted in the Figure 3. As describe in the first section, the traditional schemes show an exponential growth
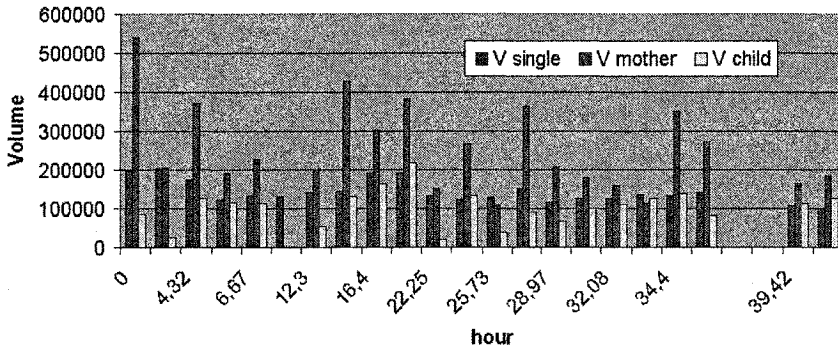
Figure 3. Percentage of single and budding cells

: an increasing of the number of buds. These methods have been implemented on a Local Area Network using MPI, interfaced by JavaMPI. Our strategy uses laboratory clones (Pentium III 450 Mhz, 128 Mo).

Slaves	1 image	1500 images
1	40sec	60000sec
2	21sec	31500sec
3	16sec	24000sec

Table 2. Time computation for slaves

This table presents time and estimated time execution for different slaves number. We can estimate the time execution for N images because it's a linear function. For off line treatments, the time is not important but in the case of on line estimation, we'll need more powerful clusters. Indeed if we implement these methods on a Pentium M (1.3GHz, 256Mo) the execution time is 10 sec for an image; it will be implemented in future works. Our goal for the fermentation process for a real time sensor is to characterize the cells physiological states by treating 100 images in 3 minutes.

If we consider p the number of processors or clones, T_s the sequential time execution, $T(p)$ the time execution for p processors, $A(p)$ the speed up or the acceleration and the efficiency $E(p)$ we have :

$$A(p) = \frac{T_s}{T(p)} \text{ and } E(p) = \frac{A(p)}{p}$$

This measures, applied to this implementation are presented in the following table

Slaves	Speed up	Efficiency
2	1.9	0.95
3	2.5	0.83

Table 3. Efficiency of parallelism

The efficiency has to be near to 1 and the speed up has to be near to the number of slaves. In the case of 2 slaves we have good results, i.e 1.9 for the speed up and 0.95 for the efficiency. There is no efficiency in increasing the number of processors but the time execution between 2 and 3 processors is more important. We have to make a trade off between the desired time execution and the efficiency of the implementation.

6.   Conclusion

The databases concerning the yeast increase each year with new results about the discovered of new proteins and new behavior due some expressed genes. The mathematical modeling of the eukaryote cells therefore the yeast is always an open question which excites the scientific community. One of the parameters which could be used by the biologist to understand the yeast evolution is the morphometry of it. These parameters could be introduced in a structural model allowing to analyze the cell cycle.

We present in this paper a robust algorithm for morphological yeast analysis allowing to have the information about the budding and stress modification in biological real time. This is due to our parallel software which has been based on a master/slave scheme guided by JavaMPI. In future works

7.   Acknowledgments

References

R. N. Dave and K. Bhaswan, *"Adaptive fuzzy c-shells clustering and detection of ellipses"*, IEEE Transactions on Neural Networks 3, pp. 643-662, September 1992.

Andrew W. Fitzgibbon, Maurizio Pilu, and Robert B. Fisher, *"Direct least-squares fitting of ellipses"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5), 476–480, May 1999

P. V. C. Hough, *"Method and means for recognizing complex patterns."*, U. S. Patent 3,069,654, 1962.

Michael Kass, Andrew Witkin, and Demetri Terzopoulos, *"Snakes: Active contour models"*, International Journal of Computer Vision, pages 321-331, 1988

Storn R. and Price K., *"Differential Evolution - A simple and efficient adaptative scheme for global optimization over continuous spaces"*, Technical Report TR-95-012,ICSI 1995.

C. Xu and J.L. Prince, *"Gradient Vector Flow: A New External Force for Snakes"*, Proc. IEEE Conf. on Computer Vision Pattern Recognition (CVPR), Los Alamitos: Comp. Soc. Press, pp. 66-71, June 1997.

C. Xu and J. L. Pince, *"Generalized gradient vector flow external forces for active contours"*, Signal Processing , 71 (1998) 131Ŭ139.

L. D. Cohen, *" On active Contour Models and Balloons"*,CVGIP: Image Understanding, 53(2):211-21, March 1991

Bryan Carpenter, Yuh-Jye Chang, Geoffrey Fox, and Xiaoming Li, *" Java as a language for scientific parallel programming"'*, In 10th International Workshop on Languages and Compilers for Parallel Computing, volume 1366 of Lecture Notes in Computer Science, pages 340-354, 1997.

Arkady Kanevsky, Anthony Skjellum, Anna Rounbehler, *" MPI/RT - An Emerging Standard for High-Performance Real-Time Systems"*, HICSS (3) 1998: 157-166

Message Passing Interface Forum, , *" MPI: A Message-Passing Interface Standard"*, University of Tenessee, Knoxville, TN, June 1995. http://www.mcs.anl.gov/mpi.