Identification of Cell Viability by an In-situ Dark-field Microscopic Sensor

Ning Wei[1]*, Axel Saalbach[2], Thorsten Twellmann[2], Tim W. Nattkemper[2], Erwin Flashel[1]

1 Fermentation Group, Technical Faculty, Bielefeld University, Bielefeld, 33615, Germany

2 Neuroinformatics Group, Technical Faculty, Bielefeld University, Bielefeld, 33615, Germany

* Corresponding author. E-mail: nwe@fermtech.techfak.uni-bielefeld.de

Abstract

Cell viability is a key factor in biotechnological industries and researches in which cells of all kinds of organisms are dealing with. As final products are strongly influenced by the cell performance, monitoring the cell viability during the fermentations is a necessity. A real time in-situ microscopic sensor using dark-field technology is developed for this purpose. Cell images taken at different time points are processed for automatically identification of living and dead cells. This procedure is based on supervised machine learning technique. Support vector machine is applied and satisfying results show promising applications in fermentation industries.

Keywords: Viability, Time-lapse image, Dark field, In-situ, Machine learning, Support vector machine

1. Introduction

The characterization and monitoring of vital status of biological cells is a very essential part in the process of biological research and production, as it has a direct effect on the quality of the final results or products. In this sense, many methods have been developed for rapidness, accuracy and reliability in identifying the vital status of biological cells. Of these different approaches, two major categories can be made, one being off-line assays and the other being in-situ (in-line) inspection. The conception of off-line assays is collecting a fraction of biomass out of the fermenter, so that all kinds of possible physical, chemical or biological tests can be undertaken easily upon the sample, which is considered to be in the similar situation as the biomass in the fermenter and to have the similar parameters. Despite of the advantages of simplicity and diversity, off-line assays are still considered to give way to in-situ (in-line) methods in some specific situations like batch production in biological industries, because the latter are believed to be more easily conducted to automatic controlling and management, and better for feed-back and real time analysis.

In order to meet the industrial demand, an in-situ microscopic sensor for determination of biomass parameters have been devised in the Technical Chemistry Institute of Hannover University[1,2], from which some important parameters of the biological cells can be determined, like cell size, cell density, and so on. However, due to the adoption of bright-field illumination, the pictures of the cells are of quite low contrast. What can be clearly seen in these pictures are rather the cell contours than the details of the inner structure. Because of this disadvantage, no information about the cell vital status, namely viability (whether the cells are dead or living) and vitality (how vigorous the cells are) can be further given.

In order to overcome this shortcoming, the manner of illumination has been suggested by our research group be changed into the dark-field type, which has been early proved to have higher contrast and suitable for observation of the inner structures of a cell. According to a former research done in our research group, significant intracellular movement can only be seen in living cells rather than in dead cells, which provides a persuading evidence for assertion of cell viability and vitality. Schematic set-up of our dark-filed microscopy device is shown in Figure 1.

This device comprises three major modules. One is an optical system comprises an light source unit which provides with dark-field illumination, and an imaging unit for taking pictures of the cells; the second is a mechanical system, which is designed to hold all the optical components and to drive the movable parts; the third, an electronic system in charge of supplying power to the sensor, and providing interface of communication with a personal computer.
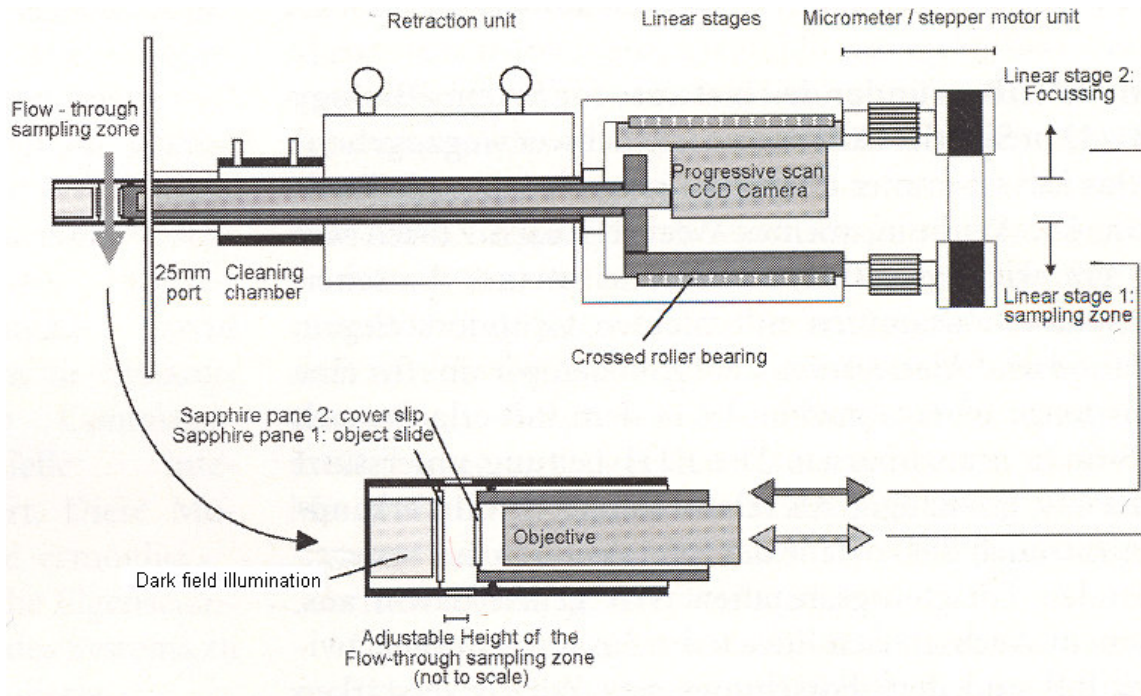
Figure 1. Set-up of the microscopic sensor developed for determination biomass in fermenters, from Technical Chemistry Institute, Hannover University.

2. Methodology

In the meantime of hardware development, we have also carried out the software development. Both work was started almost at the same time; therefore the dark-field images of the cells required for software development were not taken with any in-situ sensor, but under normal laboratory microscope. As mentioned before, according to a former research done in our research group, significant intracellular movement can only be seen in living cells rather than in dead cells. On the basis of this research, we have taken series of pictures of yeast cells with laboratory microscope under dark-field condition like this: the suspension of yeast cells are picked up with a pipette and put onto the slide, and then a cover glass is placed upon the drop of sample to form a thin layer of liquid with yeast cells. What followed is sealing the cover glass at the rim with the oil used for oil immersion observation, so that the thin layer of liquid with yeast cells would not get dry quickly. In order to get high quality images of the cells, we have to wait for the cells to "calm down" in the liquid. We took not only one picture of the cells, but also a series of pictures with an equal time interval, for example, 5 seconds; and a series may contain up to 12 pictures, in which the intracellular details of the cells are recorded at different time point. In these pictures, we can see there exists two different patterns of intracellular movement: for some cells, they appear almost the same in all these 12 pictures without noticeable changes of the intracellular pattern, which are believed to be dead cells; for the other cells, significant movement of intracellular substance can be clearly observed, and this is expected to be an indication of the cells being still alive.

In order to realize automatic classification of living and dead cells in the pictures, we have decided to make use of a pattern classification algorithm based on machine learning and neural networks as suggested by Nattkemper et al. [3,4]. For the purpose of training the neural networks, some supervising data set must be given in advance to teach the networks where are living cells and where are dead cells. So we have labeled one of the twelve images with a red dot representing living cells and a green dot representing dead cells. Our determination of living or dead is made through browsing the intracellular movement in the time-lapse image sequences in accordance with the criterion described before.

We also need to give the information of what the living cells or dead cells look like. So we have carried out a PCA (principle component analysis) operation upon the a data volume containing the score of each pixel of all the 12 images, and extracted the principle component (PC) that is for describing the score changes of each pixel during the specific time span, because the scores changes of the pixels are strong evidence of intracellular movement. In this case we have chosen the second PC to represent the different intracellular pattern in a time span for classification of living and dead cells.

After that we can extract patches around the center of those labeled cells to form the training data set of the neural networks. For example, if we

choose patch size of 31pixels x 31pixels around the cell center, we get patches as shown in Figure 2. The three columns of images at the left side are patches extracted from the post-PCA image (the 2nd PC) as obtained in last step; while at the right side, four columns contain those extracted from dead cell images. It is clear that due to two different patterns of intracellular movement, the living cells can be separated from the dead cells without much effort by eyes.
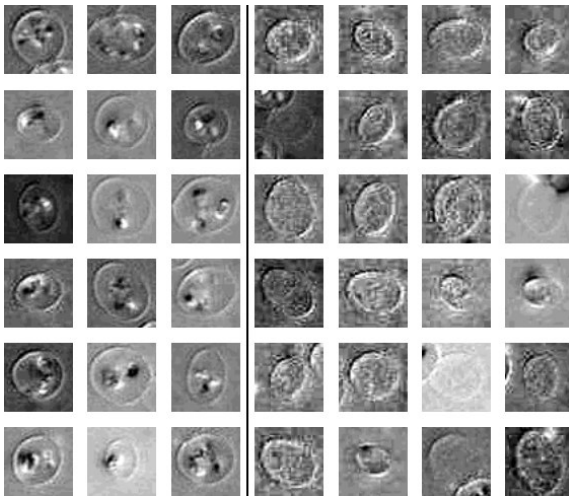

Figure 2. Patches with size of 31pixels x 31 pixels extracted from the 2nd PC of living cells and dead cells. The three columns at the left side contain the patches of living cell images, and the four columns at the right side contain those from dead cell images.

Another work worthy of doing is, try to reduce the dimensionality of the training dataset before we start to train the neural networks. That is to say, it is better to extract further "features" out of the patches shown in Figure 2 to represent the essential pattern of intracellular movements. This is a kind of reduction of dimensionality normally done in the pattern recognition field as the pretreatment of the data sets. For a patch with size of 31 x 31, the dimensionality is 961, namely, we use a 961-dimension vector to represent a cell. That is too much for data processing with the neural networks, especially when the size of the whole training data set is very large. So it would be of great benefit to make use PCA again to reduce the dimensionality, thus reducing the complexity of the computation.

Application of PCA to those patches shown in Figure 2 yields a distribution of the first ten principle components (PC's) as shown in Figure 3. According to the $1/e^2$ criterion, the 6th PC, whose variance is just greater than $1/e^2$ of the 1st PC, might be adopted as the cut-off, namely, those PC's whose variances are less than that of 6th PC should be considered as noise and removed from the data set. That is to say, we can use only 6 PC's to represent the original data set, thus reducing the

dimensionality from 961 to 6, and the computational complexity might be significantly reduced, perhaps with several orders of magnitude. That is very important for high-throughput or real-time analysis demanded by the biological industry.
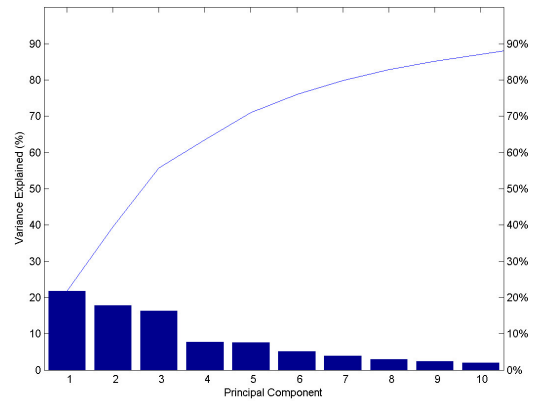

Figure 3. Dimensionality reduction for the patches shown in Figure 2 by applying PCA. The top ten principle components have been shown in descending order of the variance value.

What follow are classification trials on the dataset we have so far, fully or reduced dimensional, with neural networks based classifiers. The datasets we have used for training and testing are denoted as shown in Table 1.

Table 1. Denotation of dataset that has been used.

patches of 31x31		feature patches of 6 x 6 via PCA		feature patches of 3 x 3 via PCA	
R1	0˚	R1'	0˚	R1"	0˚
R2	90˚ rotation of R1	R2'	90˚ rotation of R1'	R2"	90˚ rotation of R1"
R3	180˚ rotation of R1	R3'	180˚ rotation of R1'	R3"	180˚ rotation of R1"
R4	270˚ rotation of R1	R4'	270˚ rotation of R1'	R4"	270˚ rotation of R1"

3. Results and Conclusions

Support Vector Machine (SVM) classifier is applied to eight trials with different datasets and parameters. Two typical kernels are used for mapping: linear kernel and Gaussian kernel. The results are given in Table 2. The major indication of success of identification is the testing error rate TE. We can see that Gaussian kernel always gives low error rates that is enough for identification of living and dead cells; while the lineal kernel is applicable only in some limited cases.

Table 2. Summary of the results of the trials.

	1	2	3	4	5	6	7	8
Training Dataset	$R_1 \cup R_2 \cup R_3$	R_1	$R_1 \cup R_2 \cup R_3$	R_1	R_1'	R_1'	R_1''	R_1''
Testing Dataset	R_4	R_4	R_4	R_4	R_4'	R_4'	R_4''	R_4''
Linear/ Gaussian Kernel	L	L	G	G	L	G	L	G
TE_{min}(%)	19.048	23.81	4.762	2.381	33.33	4.762	11.905	9.524
σ_{TEmin}	-	-	6.96	41.0	-	0.62	-	1.0
$SV\#_{TEmin}$	83 (126)*	35 (42)	127 (126)	34 (42)	23 (42)	43 (42)	-	-

TE: Testing error rate
σ: Shape parameter of Gaussian kernel
SV#: number of support vectors
*: The figure in the brackets is the total number of vectors

From Table 2 we can also conclude that:
1.) Gaussian kernel based SVM outpaces the Linear SVM in all cases.
2.) For Linear SVM, the bigger the training datasets, the better the results; on the other hand, dimensionality reduction should be done beforehand in favor of improving the accuracy.
3.) Different training datasets yield different Gaussian-shape parameter of σ. The smaller the training datasets, the larger the σ; the less the dimensionality, the smaller the σ.
4.) Increasing the size of the training datasets and reducing the dimensionality seem not to play an important role in Gaussian kernel SVM.

4. Future Works

Up to now, the identification trials have been done only with original images being training datasets and their rotational transformation being testing datasets. Next, a more rigorous cross-testing trial for small-scale datasets will be undertaken to prove the accuracy of the software as mentioned. Furthermore, automatic segmentation and labelling of the same micrographs on basis of the neural networks obtained by supervised training will be done to give another evidence of the feasibility of the software. After that, the framework of this software will be applied to the image sequences obtained from in-situ experiments, and more statistical approaches should be carried out.

Acknowledgements

References

[1] Bittner C., Wehnert G. and Scheper T. (1998) In Situ Microscopy for On-Line Determination of Biomass. *Biotechnology and Bioengineering*, **60**(1): 24-35.
[2] Schilling J. (2004) *Bioworld*, **1**: 8-9.
[3] Nattkemper T.W., Ritter H. and Schubert W. (2001) A Neural Classificator Enabling High-Throughput Topological Analysis of Lymphocytes in Tissue Sections. *IEEE Trans. on Inf. Techn. in Biomed.*, **5**(2): 138-149.
[4] Nattkemper T.W., Twellmann T., Schubert W. et al. (2003) Human vs. Machine: Evaluation of Fluorescence Micrographs. *Computers in Biology and Medicine*, **33**(1): 31-43.