

A New Automatic System of Cell Colony Counting

U. Bottigli, M. Carpinelli, P.L. Fiori, B. Golosio, A. Marras, G. L. Masala, and P. Oliva

Abstract—The counting process of cell colonies is always a long and laborious process that is dependent on the judgment and ability of the operator. The judgment of the operator in counting can vary in relation to fatigue. Moreover, since this activity is time consuming it can limit the usable number of dishes for each experiment. For these purposes, it is necessary that an automatic system of cell colony counting is used. This article introduces a new automatic system of counting based on the elaboration of the digital images of cellular colonies grown on petri dishes. This system is mainly based on the algorithms of region-growing for the recognition of the regions of interest (ROI) in the image and a Sanger neural net for the characterization of such regions. The better final classification is supplied from a Feed-Forward Neural Net (FF-NN) and confronted with the K-Nearest Neighbour (K-NN) and a Linear Discriminative Function (LDF). The preliminary results are shown.

Keywords—Automatic cell counting, neural network, region growing, Sanger net.

I. INTRODUCTION

THIS work presents software developed for the project CRORAD (CRYopreserved cells and IONizing RADiation). The purpose of this project is to estimate the increase of the damage which occurs to cells when they are cryogenically preserved and exposed to ionized radiations, as opposed to the case of non-cryogenically preserved ones.

Therefore, it is necessary to have a reliable instrument counting the cellular colonies grown on petri dishes that provides the least possible error. Moreover, this error must be known to the researchers. The manual count of colonies is a long, laborious process that is dependent on the ability of the operator [1-2]. The number of colonies to count for a single slab can amount to the hundreds with the necessity to count

This work was supported by the Italian National Institute of Nuclear Physics (INFN) with the project CRORAD.

G.L. Masala is with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari and Sezione INFN di Cagliari, Italy, Via Vienna 2, Sassari, 07100, Italy (corresponding author; phone:+39079229486; fax: +39079229482; e-mail: giovanni.masala@ca.infn.it).

U. Bottigli, was with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari and Sezione INFN di Cagliari. He is now with the Dipartimento di Fisica dell'Università di Siena and Sezione INFN di Cagliari, Italy (e-mail: bottigli@uniss.it).

B. Golosio, A. Marras, P. Oliva are with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari and Sezione INFN di Cagliari, Italy (e-mail: golosio@uniss.it; marras_anton@tiscali.it; oliva@uniss.it).

P.L. Fiori is with Dipartimento di Scienze Biomediche dell'Università di Sassari and Sezione INFN di Cagliari, Italy (e-mail: fioripl@uniss.it).

M. Carpinelli is with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari and Sezione INFN di Pisa, Italy (e-mail: massimo.carpinelli@df.unipi.it).

hundreds of plates. In these situations, the operator has a difficult task to perform counts which are then susceptible to miscalculation and erroneous identifications of the objects under study, due to the mental and visual distress caused by the routine of the process itself. Counting ability can vary due to fatigue and stress. This activity is time consuming and can limit the usable number of dishes for each experiment. For this purpose, automatic software can perform an objective analysis of whose criteria of colony classification are non-variable. If the software makes an error, it is systematic for all the dishes and is identified in the moment of realization. The counting is completed quickly providing the possibility to produce greater numbers of dishes for experiments and supplying better validation of data statistics. Software that perform automatic colony counting already exist [1-4] but the error of the system is often unknown or the system is inadaptable to the characteristics of the CRORAD project.

In this work, an automatic system of cell counting based on the analysis of digital images was created. The images of the dishes are in color produced through a linear scanner with 8 bit and a resolution of 300 dpi.

In this article the main steps for the extraction algorithms of the regions of interest (ROI) and their characterization are described. The results obtained through a classifier based on neural nets in respect to k-nearest neighbours and linear discriminative function is shown.

II. MATERIALS AND METHODS

A. Cell Cultures

V79 cells, a cell line derived from a Chinese hamster lung fibroblasts, were maintained in continuous cultures in Dulbecco' modified MEM (DMEM), supplemented with 10% fetal calf serum, 103 UI/ml penicillin, 10 mg/ml streptomycin, and 200mM Glutamin, and grown in 75-cm² sterile plastic flasks (T75) in a humid atmosphere containing 5% CO₂.

Cells were grown until semi-confluence (70-80% area) and then trypsinized and centrifuged ten minutes at 4°C, 1000 rpm in a phosphate buffer (PBS). Cells were then re-suspended in complete DMEM and a total of 100 cells were incubated in DMEM in 6 cm sterile petri dishes. Cell cultures were then incubated 7 days at 37 °C, then gently washed in PBS, and finally fixed and stained using a solution of gentian violet. Plates were gently washed in water. Colonies were observed using optical microscopes.

B. Image Analysis

After acquisition by the scanner, the image was split in three channels (each one of 8 bit) of the main colors that form it [5-11] according to the outline RGB. All the successive analysis of the image is made on the blue channel. In fact, coloring used for the cells renders them blue-violet, as in Fig. 1-A. The coloration of the colony is more intense towards the center part of the same one and diminishes as it proceeds toward the margins.

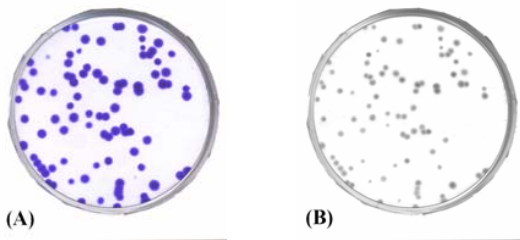


Fig. 1 Images of the dishes (A) Image of three channels (B) Only the blue channel

The separation of the blue channel degrades the perimeter part of the colony, which begins the process of cluster

separation that is unavoidably present. This also involves the elimination of a great part of the noise present in the image. Therefore, after this phase the external edge of the slab remains with many colonies already separated. Finally, the images are converted into a scale of greys.

The complete procedure is illustrated in the following block-diagram (Fig. 2).

III. ROI HUNTING

The search of the regions of interest happens through the application of an algorithm of region growing [5-7] with optimized thresholds, as shown in Fig. 3. The main steps of the procedure are:

- Start by choosing an arbitrary *seed pixel* in the image (using an automatic raster scan inside the border of dish) and compare it with neighboring by a threshold
- Region is *grown* from the seed pixel by adding in neighboring pixels that are up the optimized threshold, increasing the size of the region.
- When the growth of one region stops we simply choose another seed pixel which does not yet belong to any region and start again.
- This whole process is continued until all pixels belong to some region.

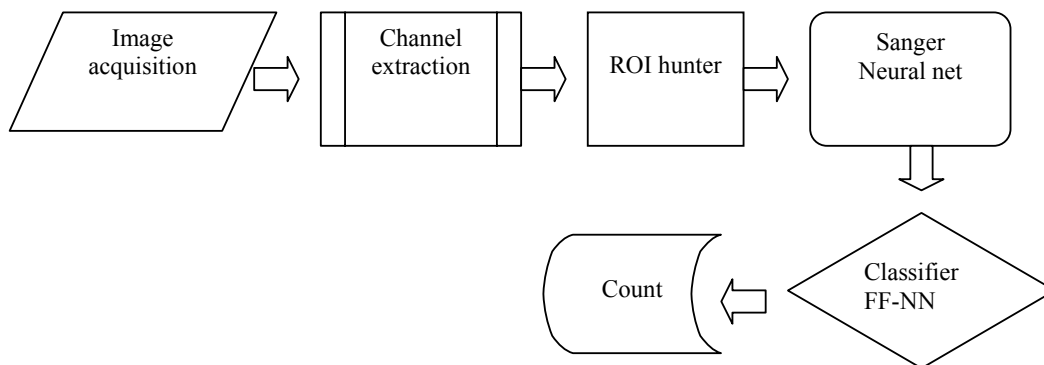


Fig. 2 Block diagram. The image is acquired from the scanner. The blue channel of the image is separated and passed to the ROI hunter that extracts the possible candidates. The candidates are analyzed through a Sanger neural net that extracts the eight principal components. With such features the candidates are recognized from the classifier before the counting

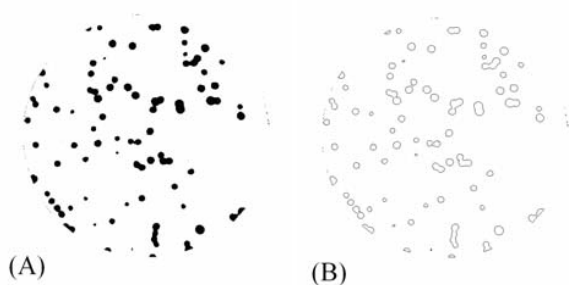


Fig. 3 Image after the region growing (A) All the candidates (B) Their perimeters

To this point of the procedure many, but not all, the clusters of colonies have been resolved in the single forming colonies. In this operation loss of colonies is not had. Every object that satisfies the candidate criteria has been stored in an image of 70X70 pixels, which will constitute the regions of interest (ROI dataset) database.

IV. DATASET

The dataset of the regions of interest is constituted from representative images of ROI extracted from 20 petri dishes as described in table I. In particular, it is shown the composition of the dataset: ROI that do not have to be identified as colonies (class 0); single colonies (class 1); and the cases in which more cellular colonies are overlapped: double colonies (class 2), triple (class 3), quadruple (class 4). At the moment there are no cases of clusters more numerous than four colonies.

TABLE I
DATASET COMPOSITION

Number of total samples		Training set	Test set
		1132	1187
False colonies	(class 0)	240	245
Single colonies	(class 1)	801	838
Double colonies	(class 2)	75	82
Triple colonies	(class 3)	12	18
Quadruple colonies	(class 4)	4	4

The various classes are illustrated that range from one until to a maximum of four overlapping colonies; moreover, one class of objects with characteristics similar to the colonies is present but are not considered itself as such.

V. FEATURES EXTRACTION AND CLASSIFICATION

Every example of the dataset is supplied as input to an auto-encoder (a Sanger's neural network) [12-14,17], that having an input 70X70 pixels correspondent to the images it supplies in output the 8 principal components of the image. So an important features reduction is made. For every ROI, also the area of the region and the perimeter (Fig. 3b) are extracted.

In order to decide if a cluster is constituted from a single colony or more colonies, the recognition problem is transformed in a problem to 5 classes as evidenced in Fig. 4.

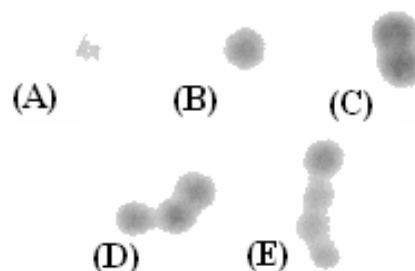


Fig. 4 The five classes. (A): False colony. (B): Single colony. (C): Cluster of two colonies. (D): Cluster of three colonies. (E): Cluster of four colonies

A validation system [12,14] as a k-folder cross-validation ($k=5$), 5x2, supplies the training and test sets.

On the basis of these sets and the described classes the classifiers [5,12-14]: K-Nearest Neighbours (K-NN) with $K=15$, a Feed-Forward Neural Network (FF-NN) with 10 input, 5 hidden, 5 output and a Linear Discriminative Function (LDF) are tested respectively.

VI. ANALYSIS

The preliminary results as a mean of the validation system 5x2 are shown in table II. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity [15-16].

In Table III, the confusion matrix for the best classifier based on neural nets is shown; the rows represent the proportion % of the corresponding row that is classified by a classifier in the category of the column.

TABLE II
MEAN ACCURACY OF THE CLASSIFIERS

Classifiers	Accuracy	Errors
K-NN	93	-2 +1
FF-NN	94	-2 +1
LDF	93	-2 +1

Mean accuracy of the Feed Forward Neural Networks, the K-Nearest Neighbours, and the Linear Discriminative Function. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity

TABLE III
CONFUSION MATRIX FOR THE FF-NN

	0	1	2	3	4	OK	BAD	TOT	ACC
0	93.9	6.1	0	0	0	232	15	247	0.94
1	0.8	98.5	0.7	0	0	829	13	842	0.98
2	0	38.3	51.9	9.9	0	42	39	81	0.52
3	0	0	53.9	46.2	0	6	7	13	0.46
4	0	0	25	50	25	1	3	4	0.25
						1110	77	1187	0.94
								Accuracy	0.94

The rows represent the proportion % of the corresponding row that is classified by a classifier in the category of the column.

In general, it is possible to note that even as the classifiers are varied the mean accuracy is always around to 93%. The protocol 5x2 used together with confidence limits of 95% supplies a good estimate of the performance of the system on the considered samples.

However, it is necessary to notice that the medium accuracy is low because of the minimal performances on the multiple colonies, as shown in Table III. In fact, in the dataset there are few cases of this type in order to train the classifiers effectively; this represents a development point that could be improved through more statistics.

As for other automatic systems [2] the curve of regression between machine count and human is reported in Fig. 5. As noted, there is a strong linear association between the two parameters.

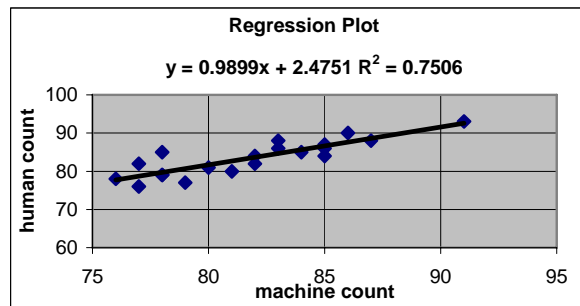


Fig. 5 Diagram of the linear regression between the automatic count and that manual counting based on FF-NN; the result is relative to counts of 20 petri dishes

VII. CONCLUSION

In conclusion, this article introduced an automatic system of cell counting based mainly on algorithms of image analysis and neural nets. The results show a good accuracy limited mainly from statistics of more complex cases. The results shown in each case are comparable with other systems [2][18]. Therefore, this is an optimal point of departure for future developments.

ACKNOWLEDGMENT

Dr. Roberto Cherubini, Dr. Silvia Gerardi, and Dr. Chiara Zilio, from the National Laboratories of INFN in Legnaro (IT), for their collaboration in setting up the growth procedures for the cell cultures.

REFERENCES

- [1] N. Blackburn, A. Hagström, J. Wikner, R. Cuadros-Hansson, P. K. Bjørnsen, "Rapid Determination of Bacterial Abundance, Biovolume, Morphology, and Growth by Neural Network-Based Image Analysis", *Appl Environ Microbiol.* 1998 September; 64(9): 3246–3255.
- [2] V. Piuri, F. Scotti, "Morphological Classification of Blood Leucocytes by Microscope Image", *CIMSA 2004 – IEEE International Conference on Computational Intelligence for Measurement Systems and Applications* Boston, MD, USA, 14-16 July 2004.
- [3] ImageJ, <http://rsb.info.nih.gov/ij/>
- [4] MACE colony counting software, <http://www.colonycount.com/index.html>
- [5] S. Serpico, G. Vernazza: "Teoria e tecniche del riconoscimento", ed. CUSL "il gabbiano", 1997.
- [6] W. Pratt, "Digital image processing", Wiley & Sons, 1978.
- [7] D.H. Ballard, C.M. Brown, "Computer vision", Prentice Hall, 1982.
- [8] Ian T. Young, Jan J. Gerbrands, Lucas J. van Vliet, "Fundamentals of Image Processing", ISBN 90-75691-01-7, NUGI 841 Printed in The Netherlands at the Delft University of Technology.
- [9] Gonzales R. C.; Wintz P., "Digital image processing", Addison Wesley Ed., 1987.
- [10] E. R. Dougherty, J. Astola, "An introduction to Nonlinear Image Processing" vol. TT16 SPIE PRESS.
- [11] E. R. Dougherty, "An Introduction to Morphological Image Processing" vol. TT9 SPIE PRESS.
- [12] O. Duda, P. E. Hart, D. G. Stark, "Pattern Classification", second edition, A Wiley-Interscience Publication John Wiley & Sons, 2001.
- [13] S. Haykin "Neural Networks – A comprehensive foundation", second edition, 1999, Prentice Hall.
- [14] M. Buscema, "Reti neurali artificiali e sistemi sociali complessi", 1409.1, FrancoAngeli, 1999.

- [15] Newcombe, Robert G. "Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods," *Statistics in Medicine*, 17, 857-872 (1998).
- [16] Wilson, E. B. "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, 22, 209-212 (1927).
- [17] M.A. Kramer, "Nonlinear principal component analysis using auto associative neural networks", *AIChE Journal* February 1991, Vol. 37, No. 2, 233-243.
- [18] Automated counting of mammalian cell colonies by means of a flat bed scanner and image processing. *Cytometry A*. 2004 Aug;60(2):182-8.