IL NUOVO CIMENTO Vol. 30 C, N. 6 Novembre-Dicembre 2007
DOI 10.1393/ncc/i2007-10273-3

Automatic cell colony counting by region-growing approach

G. L. Masala([1])([2])(*), U. Bottigli([3])([4]), A. Brunetti([1])([2]), M. Carpinelli([1])([2]), N. Diaz([5])([2]), P. L. Fiori([5])([2]), B. Golosio([1])([2]), P. Oliva([1])([2]) and G. Stegel([1])([2])

([1]) *Struttura Dipartimentale di Matematica e Fisica, Università di Sassari - Sassari, Italy*
([2]) *INFN, Sezione di Cagliari - Cagliari, Italy*
([3]) *Dipartimento di Fisica, Università di Siena - Siena, Italy*
([4]) *INFN, Sezione di Pisa - Pisa, Italy*
([5]) *Dipartimento di Scienze Biomediche, Università di Sassari - Sassari, Italy*

(ricevuto il 29 Ottobre 2007; revisionato il 22 Aprile 2008; approvato il 29 Aprile 2008; pubblicato online l'1 Luglio 2008)

Summary. — This paper 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 region-growing algorithms for the recognition of the Regions Of Interest (ROI) in the image and Sanger's neural network for the characterization of such regions. Moreover a recognition of the most important filters is made in alternative respect to region-growing approach. The new Graphics Users Interface is introduced. The better final classification is supplied from a Feed-Forward Neural Net (FF-NN) and compared with the K-Nearest Neighbour (K-NN). The results on large dataset of ROIs are shown.

PACS 87.57.U- – Nuclear medicine imaging.
PACS 87.57.N- – Image analysis.

1. – Introduction

This work presents software developed for the project CRIORAD (CRyopreserved cells and IOnizing RADiation) funded by the Italian National Institute of Nuclear Physics (INFN).

The purpose of this project is to estimate the increase of the damage which occurs in 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 be counted

(*) E-mail: giovanni.masala@ca.infn.it

for a single slab can amount to the hundreds with the necessity to count 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 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 performing automatic colony counting already exists [1-4] but the error of the system is often unknown or the system is inadaptable to the characteristics of the CRIORAD project.

In this work, an automatic system of cell colony counting based on the analysis of digital images was created. The main steps for the extraction algorithms of the Regions Of Interest (ROI) and their characterization are described. We make a short comparison with respect to other important filters in the literature [5-8] for such recognition problems as the watershed filters or methods based on Laplace filter. The results obtained through a classifier based on neural nets with respect to K-nearest-neighbours function on large dataset are shown.

2. – Materials and methods

2˙1. *Cell cultures*. – V79 cells, a cell line derived from Chinese hamster lung fibroblasts, were maintained in continuous cultures in Dulbecco's modified MEM (DMEM), supplemented with 10% fetal calf serum, 103 UI/ml penicillin, 10 mg/ml streptomicin, and 200 mM Glutamin, and grown in 75 cm^2 sterile plastic flasks (T75) in a humid atmosphere containing 5% CO_2.

Subconfluent cells were tripsinized, and centrifuged (1000 r.p.m. for 10 minutes); pellet was resuspended in complete medium. Cells concentration was evaluated by using a Burker hemocytometer, and cells were resuspended in DMEM (low-glucose) supplemented with 10% heat-inactivated foetal bovine serum, at final concentration 10^3 cells/ml. Aliquots of 0.1 ml were finally resuspended in 5 ml of complete medium, cultured in 6 cm diameter Petri dishes (100 cells per plate) and incubated 7 days in a humidified incubator at 37 °C in atmosphere of 5% CO_2 in air. After incubation, liquid medium was discharged and adherent cellular colonies gently washed twice in PBS. Cells were finally ethanol-fixed and stained with a crystal violet solution for 5 minutes. The plates were washed in H_2O. Finally 150 dishes are manually counted and used to develop the cell colony counting tool.

2˙2. *Image analysis*. – The images of the dishes are produced in color through a linear scanner with 8 bit and a resolution of 300 dpi. After acquisition, the image was split in three channels (each one of 8 bit) the main colors that form it [9-15] according to the RGB outline. All the further analysis of the image is made on the blue channel. In fact, coloring used for the cells renders them blue-violet, as in fig. 1A. The coloration of the colony is more intense towards the center and diminishes as it proceeds toward the margins.

The separation of the blue channel degrades the perimeter part of each colony, giving rise to the process of cluster separation that is unavoidably present. This also involves

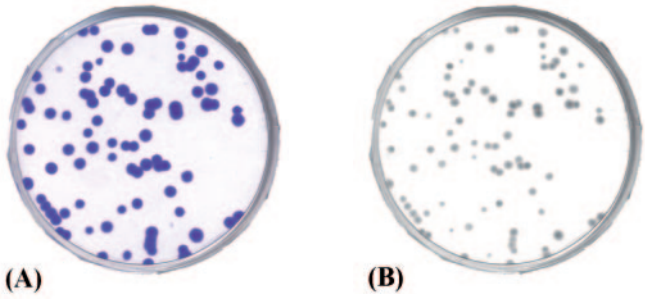

Fig. 1. – (Colour on-line) Images of the dishes. (A) Color image. (B) Only the blue channel.

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 grays.

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

3. – ROI hunting

The search of the regions of interest takes place through the application of an algorithm of region growing [9-11] 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 above optimized treshold, 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.

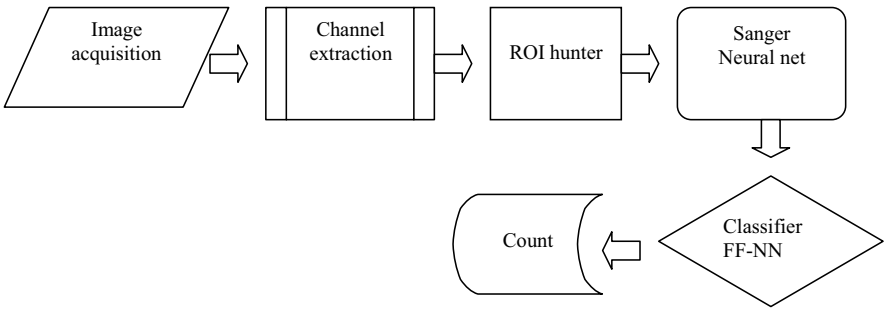

Fig. 2. – Block diagram of the proposed method. The image is acquired from the scanner. The blue channel of the image is separated and passed to the ROI hunter that extracts the possible candidates. The candidates are analyzed through a Sanger neural net that extracts the eight principal components.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

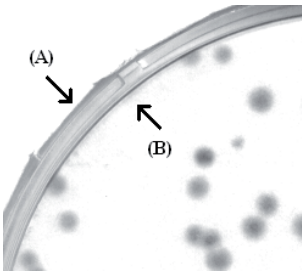

Fig. 3. – (Colour on-line) The two perimeters of the edge of the slab: (A) External perimeter. (B) Inner perimeter.

– This whole process is continued until all pixels belong to some region.

It is necessary to characterize the zone in the image where the colonies (inside the slab) are present and to eliminate the plastic edge of the Petri dish.

The edge does not appear in the image as a uniform object because of its transparency. For this reason it is not a single region that one can fill up and eliminate with the technique of the region growing.

An inner perimeter and another perimeter on the exterior of the edge can be characterized as shown in fig. 3.

Operating with the region growing, with a seed pertaining to the inner area of the dish, a region is obtained whose perimeter corresponds to the inner perimeter of the edge. But using only inner perimeter (fig. 4A), we lose the colonies on the edge of the dish. We solve such problem through the convex polygon constituted from points that define the perimeter with the Graham algorithm [10]. Eliminating from the original image all the objects that are external to the convex polygon we can eliminate the edge of the Petri dish (fig. 4B). So the search for the regions of interest is performed through the application of an algorithm of region growing.

To this point of the procedure many, but not all, clusters of colonies have been resolved in the single forming colonies. In this operation there was no loss of colonies. Every object that satisfies the candidate criteria has been stored in an image of 70×70 pixels, which will constitute the regions of interest (ROI dataset) database.

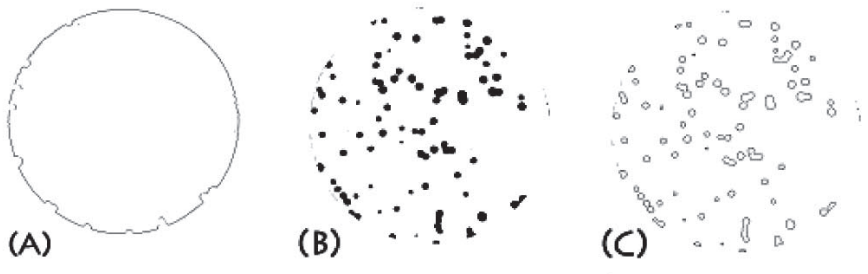

Fig. 4. – Inner perimeter of the edge (A). Image after the region growing: (B) All the candidates. (C) Their perimeters.

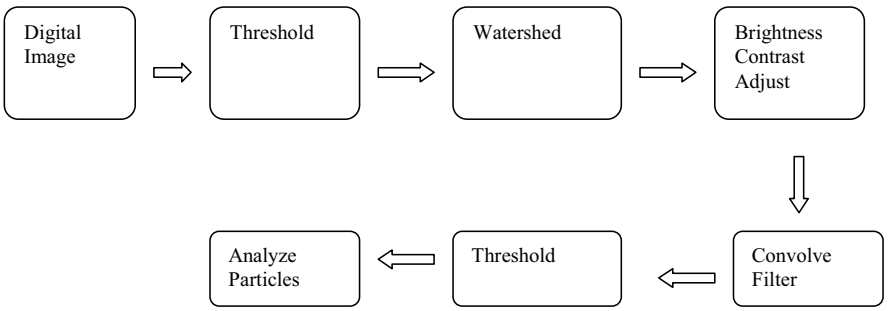

Fig. 5. – Block diagram of the watershed-convolve filter. The image is acquired from the scanner. The blue channel of the image is separated and passed on the watershed filter. After optimizing the contrast value, the image is passed to a convolve filter. After a threshold, the Analyze Particles routine of ImageJ detect and count the ROI.

4. – Comparing other filters

In this part we make a comparative analysis of our own approach with respect to different filters described in the literature [6,7]. In particular we uses a subset of 20 digital images of Petri dishes and apply watershed-median, watershed-convolve, a Laplace-based filter (Soft imaging System) (figs. 5-7).

The effects of the methods described now on the original blue channel image are shown in fig. 8. We can see how the Laplace filter lost quality and several details of the dishes. Furthermore it is possible to see that the best quality is obtained with region-growing methods or watershed-convolve filter but with differences between the approaches. In particular in some region (signed with red circle) of the image passed to watershed-convolve filter we lost some colonies.

To make a correct comparison we apply the different approaches without a classification system and counting all the colonies found with Analyze Particles routine of Image-J [3]. Without a classifier Analyze Particles detect all objects in the image. The total number of objects detected is shown and the difference with respect to human count is calculated.

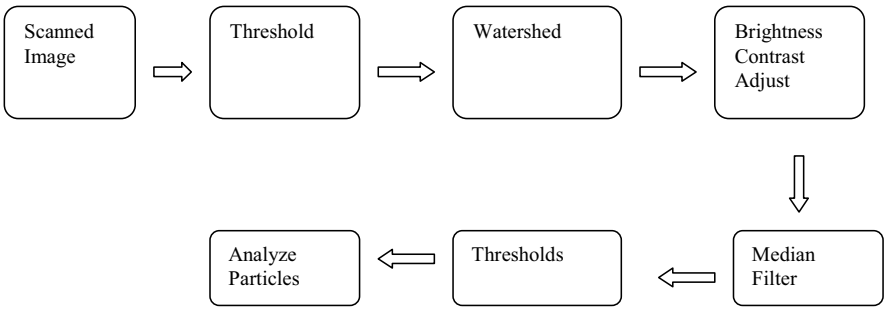

Fig. 6. – Block diagram of the watershed-median filter. The image is acquired from the scanner. The blue channel of the image is separated and passed to the watershed filter. Optimizing the contrast value, the image is passed to a median filter. After a threshold, the Analyze Particles routine of ImageJ detect and count the ROI.

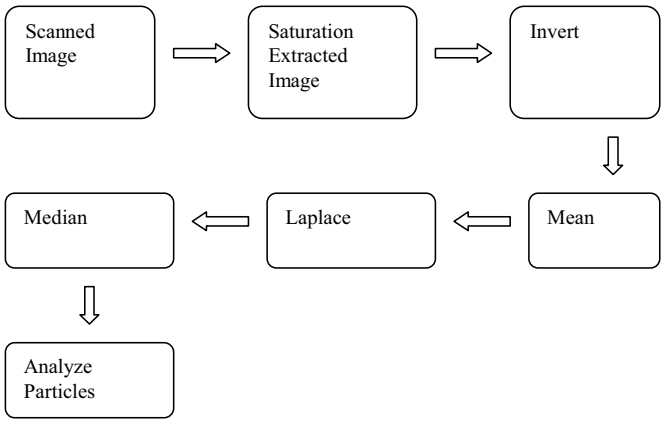

Fig. 7. – Block diagram of the Laplace filter. The image is acquired from the scanner. After some operation of saturation, inversion and mean, a Laplace filter followed by median filter. The Analyze Particles routine of ImageJ detects and counts the ROI.

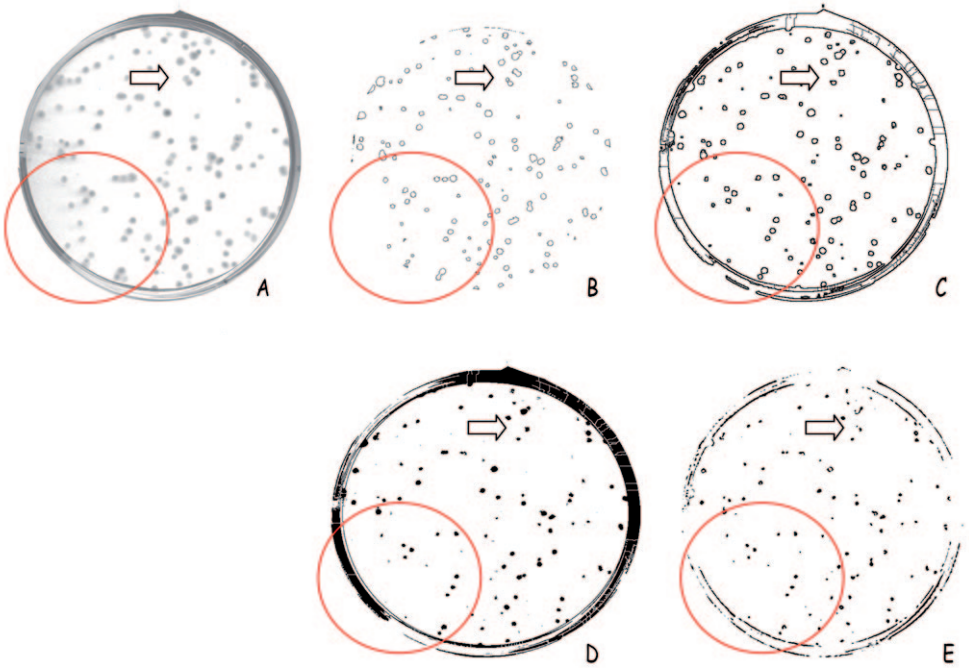

Fig. 8. – Application of the various approaches. In order, starting from the left, we have: A) the original image (blue channel), B) region-growing, C) watershed-convolve, D) watershed-median and E) Laplace. The circles indicate a region where watershed and Laplace lost some colonies while the pointer indicates different effects of filter on the same group of colonies.

TABLE I. – *Comparing different approaches.*

	Region-Growing	Laplace	Watershed-Median	Watershed-Convolve	Human count
TOT	2488	1861	2673	2182	2324
DIFFERENCE	164	−463	349	−142	
DIFFERENCE %	8.2	−23.15	17.45	−7.1	

In the table we find on the columns the different approaches and the correct results (human count) while in the rows we found the total number of colonies counted by Analyze Particles of ImageJ and the difference and relative differences with respect to the correct result.

Looking at table I we can highlight that the results obtained by Laplace and watershed-median are poor. The best results are comparable with region-growing and watershed-convolve because are near to the human correct count of colony. In each case we prefer region-growing because we do not loss colony taken some false positives. We resolve the number of false positives subsequently with the trainable classifier.

In the following graphs (fig. 9) the linear regressions of the methods are shown.

It is possible to see from the graphs of fig. 9 that region-growing, using also in this case Analyze Particles of ImageJ, has the best regression linear curves (best value of R^2).

In the next section, using region-growing, we show the result with trainable classifiers.

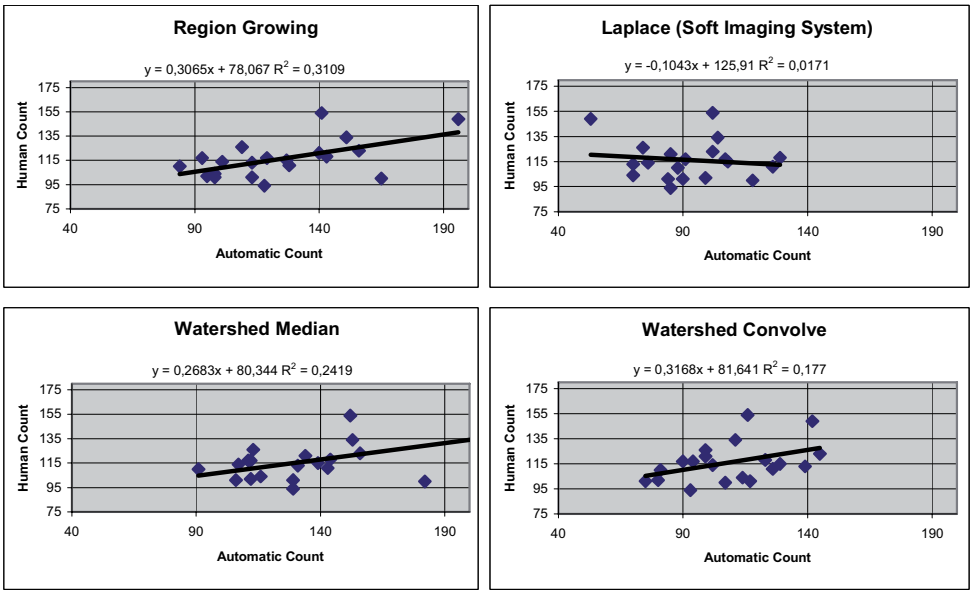

Fig. 9. – Regression curves using Analyze Particles of ImageJ with respect to human count for various methods.

TABLE II. – *Dataset composition.*

Number of total samples		Training set 8175	Test set 8221
False colonies	(class 0)	1208	1206
1 colonies	(class 1)	6230	6300
2 colonies	(class 2)	590	582
3 colonies	(class 3)	112	95
4 colonies	(class 4)	21	23
5 colonies	(class 5)	7	6
6 colonies	(class 6)	4	4
7 colonies	(class 7)	3	4

Various classes are illustrated ranging from one to a maximum of seven overlapping colonies; moreover, there is a class of objects that share some characteristics of colonies but are not regarded as such.

5. – Dataset

The dataset of the regions of interest is constituted of representative images of ROI extracted from 150 petri dishes as described in table II. In particular, we show the composition of the dataset: ROI that must not to be identified as colonies (class 0); single colonies (class 1); and the cases in which more cellular colonies are overlapped: 2 colonies (class 2), 3 colonies (class 3), up to 7 colonies (class 7). There are no cases of clusters containing more than seven colonies.

6. – Features extraction and classification

Every example of the dataset is supplied as input to an auto-encoder (a Sanger neural network) [16-19], that having an input 70×70 pixels corresponding to the images supplies in output the eight principal components of the image. So an important features reduction is made.

In order to decide whether a cluster is constituted of a single colony or several colonies, the recognition problem turned into a 8 classes problem as shown in fig. 10.

In the following four graphs (figs. 11) Sanger's components are shown. It is possible to see how important some components are, in order to discriminate a class, with respect to the others. For every ROI, the area of the region and the perimeter (fig. 3B) are also extracted and added to the features vector. So we have a 8 component vector of features for each record of dataset.

Fig. 10. – The eight classes problem. Starting from left it is shown: false colony, single colony and cluster of two colonies, three colonies, four colonies, five colonies, six colonies, seven colonies.

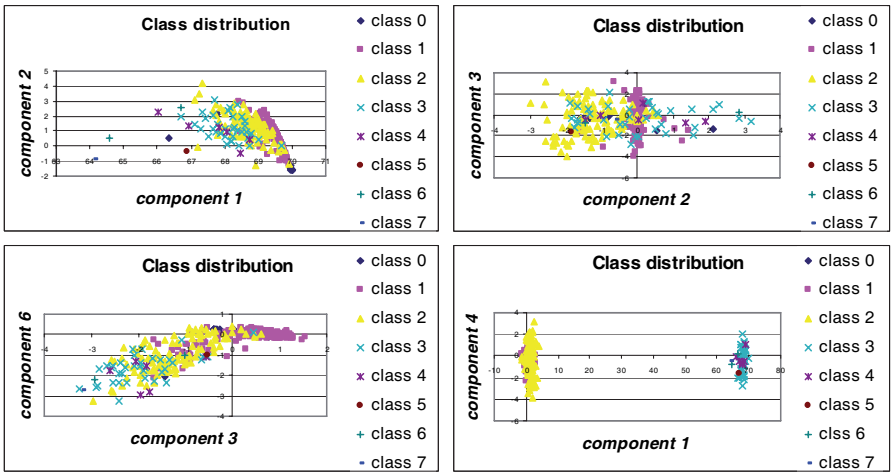

Fig. 11. – (Colour on-line) Class distribution of Sanger's component. Some components are shown and it is possible to see how some components well discriminate the classes.

A validation system [16,18] as a k-folder cross-validation ($k = 5$), 5×2, supplies the training and testing sets. We used for such purpose a two-dimensional Self-Organizing-Map (SOM) [17] with 10×10 nodes, which maps the input data in 10×10 subsets, each subset including ROIs close to each other in the feature space. At each iteration, the negative cases for the training and for the validation set are extracted by randomly sampling a percentage of each SOM subset, yielding 8175 cases for the training set and a similar number (8221) for the testing set. So very representative random subsets are extracted for the procedure of validation.

On the basis of these sets and the described classes, the classifiers [9, 16-18]: K-Nearest Neighbours (K-NN) with $K = 15$, a Feed-Forward Neural Network (FF-NN) with 10 input, 5 hidden, 5 output are respectively tested.

7. – Analysis

The results as a mean of the validation system 5×2 are shown in table III. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity [20,21].

TABLE III. – *Mean accuracy of the classifiers.*

Classifiers	Accuracy (%)	Errors (%)
K-NN	91	−1.7 +1.4
FF-NN	92	−1.7 +1.4

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

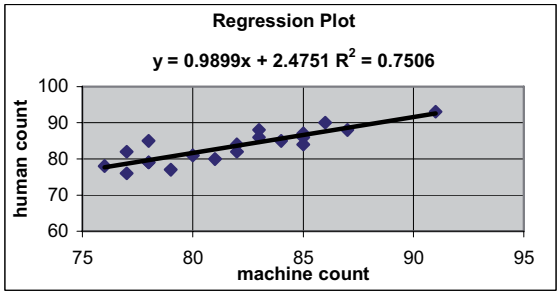

Fig. 12. – Diagram of the linear regression between the automatic count and manual counting based on FF-NN; the result is relative to counts of 20 petri dishes.

In general, we note that even as the classifiers are varied the mean accuracy is always around 92%. The protocol 5×2 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 average accuracy is low because of the minimal performances on the multiple colonies, as shown in table II. In fact, in the dataset there are few cases of this type in order to train the classifiers effectively; this is a point which could be improved through more statistics.

In any case this accuracy does not condition the performances of the method. In fact most colonies are of type 0, 1, 2 and some errors are compensated. Therefore for 150 slabs considered the error on counts turns out to be of 1.5%.

As for other automatic systems [2] the curve of regression between machine count and human is reported in fig. 12. As noted, there is a strong linear association between the two parameters.

Comparing it with fig. 9, we can see the improvement of the curve after the classifier as shown in fig. 12.

The system is implemented with a graphical users interface as shown in fig. 13.

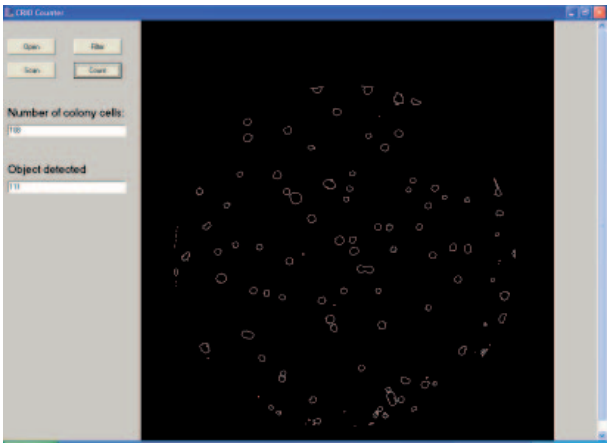

Fig. 13. – (Colour on-line) A graphical users interface of the program. The output of the filter and final count is shown. It is possible to read the number colony detected with (number of colony cells) or without classifier (object detected).

8. – Conclusion

In conclusion, this paper introduced an automatic system of cell colony counting based mainly on algorithms of image analysis and neural nets. The results show a good accuracy limited mainly by statistics of more complex cases. The results shown in each case are comparable with other systems and with respect to important filters, as watershed or Laplace [2, 5-8].

* * *

We thank Dr. R. Cherubini, Dr. S. Gerardi, and Dr. C. Zilio, from the National Laboratories of INFN in Legnaro (IT), for their collaboration in setting up the growth procedures for the cell cultures. This work was supported by the Italian National Institute of Nuclear Physics (INFN) with the project CRIORAD. A part of the results reported in this paper have been produced by the Cybersar project managed from the Cosmolab Consortium. A project partially funded from the Ministero dell'Università e della Ricerca (MIUR) by the Piano Operativo Nazionale "Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione" (PON 2000-2006). More information is available on web page: `http://www.cybersar.com`.

REFERENCES

[1] Blackburn N., Hagström A., Wikner J., Cuadros-Hansson R. and Bjørnsen P. K., *Appl. Environ. Microbiol.*, **64** (1998) 3246.
[2] Piuri V. and Scotti F., *Morphological Classification of Blood Leucocytes by Microscope Image*, in *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] Dahle J., Kakar M., Steen H. and Kaalhus, *Cytometry A.*, **60** (2004) 182.
[6] Rolland J. P., Bon P. and Thomas D., *J. Exp. Ther. Oncol.*, **5** (2005) 9.
[7] Suris Di Jasjit S., Setarehdan S. Kamaledin and Singh Sameer, *Advanced Algorithmic Approaches to Medical Image Segmentation: State-of-the Art* (Springer) 2002, pp. 153-154, ISBN 1-85233389-8.
[8] Chen Q., Yang X. and Petriu E., *Watershed Segmentation for Binary Images with Different Distances Transforms*, in *HAVE'2004 - IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications, Ottawa, Ontario, 2004*, pp. 111-116.
[9] Serpico S. and Vernazza G., *Teoria e tecniche del riconoscimento* (CUSL "il gabbiano") 1997.
[10] Pratt W., *Digital Image Processing* (Wiley & Sons) 1978.
[11] Ballard D. H. and Brown C. M., *Computer Vision* (Prentice Hall) 1982.
[12] Young Ian T., Gerbrands Jan J. and van Vliet Lucas J., *Fundamentals of Image Processing*, ISBN 90-75691-01-7, NUGI 841, printed in The Netherlands at the Delft University of Technology.
[13] Gonzales R. C. and Wintz P., *Digital Image Processing* (Addison Wesley) 1987.
[14] Dougherty E. R. and Astola J., *An Introduction to Nonlinear Image Processing*, Vol. **TT16** (SPIE press) 1992.
[15] Dougherty E. R., *An Introduction to Morphological Image Processing*, Vol. **TT9** (SPIE press) 1992.
[16] Duda O., Hart P. E. and Stark D. G., *Pattern Classification*, second edition (Wiley-Interscience Publication John Wiley & Sons) 2001.

[17] Haykin S., *Neural Networks – A Comprehensive Foundation*, second edition (Prentice Hall) 1999.
[18] Buscema M., *Reti neurali artificiali e sistemi sociali complessi*, Vol. 1409.1 (FrancoAngeli) 1999.
[19] Kramer M. A., *AIChE J.*, **37** (1991) 233.
[20] Newcombe and Robert G., *Stat. Med.*, **17** (1998) 857.
[21] Wilson E. B., *J. Am. Stat. Assoc.*, **22** (1927) 209.