

# Improved Automatic ES Cells Counting Method in Fluorescence Microscopy Images

Geisa M. Faustino, Marcelo Gattass and Carlos J. P. de Lucena  
Department of Informatics  
Pontifícia Universidade Católica do Rio de Janeiro  
Rio de Janeiro – RJ, Brazil  
gfaustino@inf.puc-rio.br, mgattass@tecgraf.puc-rio.br,  
lucena@inf.puc-rio.br

Priscila B. Campos and Stevens K. Rehen  
Institute of Biomedical Sciences  
Universidade Federal do Rio de Janeiro, UFRJ  
Rio de Janeiro – RJ, Brazil  
brittopris@anato.ufrj.br, srehen@anato.ufrj.br

**Abstract**—In this paper we present an improved method for counting embryonic stem cells (ES cells) in fluorescence microscopy images. By applying a graph-clustering algorithm, we increased the precision of the original technique [1], making it suitable for images of the migration of embryoid body cells. The improved method was extensively tested on more than a hundred images and the results demonstrated its superiority.

**Keywords**-*automatic cell counting, fluorescence microscopy images, graph-based image representation, graph mining, graph clustering.*

## I. INTRODUCTION

Stem cell research has been hailed for the potential to revolutionize the future of medicine with the ability to regenerate damaged and diseased organs. Embryonic Stem cells (ES cells) can generate all three germ layers and also specialized themselves in adult cells of the human body. This characteristic of pluripotency and the real possibility of directing their differentiation *in vitro* to specific cell types make them a promising alternative for cell-based treatments of various diseases.

Cell counting plays an important role in the statistical analyzes and enables specialists to understand and validate experiments. Using different cell markers they are able to determine, by manual counting, the total number of cells, how many specialized themselves in specific mature cells and how many cells have died. However, manual counting is tedious, time-consuming and yields subjective results. Thus, the automation of the cell counting process is a crucial step for streamlining the evaluation of culturing methods and consequently accelerates the stem cell therapy process.

With the discovery of the stem cell potential, many researchers have been dealing with the segmentation, tracking and quantification of these kinds of cells. See [1] and [3] for a review. However, these works deal with only one type of stem cell in their images. The first work to deal with several types of pluripotent stem cells in the same microscopy image was presented in [1]. Although it works well, its performance is significantly reduced in images where there is more than one brighter point per cell.

In this paper, we present an improved method for counting ES cells in fluorescence microscopy images. We have extended and improved the method presented

in [1], which was designed to handle images of stem cell sections from embryoid bodies (EB cryosections images). As in that previous work, our method uses the luminance information to generate a graph-based image representation. The cell pattern is defined as a subgraph and then a graph mining process is used to detect the cells. The improvement is achieved by applying a graph-clustering algorithm subsequent to the mining process. It increases the precision for images of EB cryosections and makes it suitable for images of the migration of embryoid body cells. Both, the original method [1] and its improved version here proposed were extensively tested on a database of 141 images and specialists validated the results. We improved the precision by 2.4% for images of EB cryosections and increased, from 67.33% to 89.74%, for the images of embryoid body cell migration with a zoom of 40 $\times$ .

The remainder of the paper is organized as follows: Section 2 details the image features and the acquisition process; Section 3 provides an overview of the original method; Section 4 describes the improvement method; Section 5 presents the experimental results, and; Section 6 presents our conclusion and some future works.

## II. IMAGE CHARACTERIZATION

The embryoid body (EB) images used in this work were collected in the Institute of Biomedical Sciences at UFRJ/Brazil. We deal with embryoid bodies obtained from ES cells cultured *in vitro*. ES cells are induced to form floating spheres structures called embryoid bodies. EBs are cultured for eight days under a neural differentiation process. In this procedure, the EBs are stimulated to differentiate into a neural phenotype by incubation during the last four days, with retinoic acid at final concentration of 2 $\mu$ M. In order to obtain EB cryosections images, a fraction of these EBs were fixed and processed for cryoprotection. Sections of 10 $\mu$ m are obtained on cryostat and nuclei were counterstained with DAPI (4'-6'-Diamidino-2-phenylindole) for five minutes. Images of embryoid body cells migration (EB cells migration) are obtained from the other EB fraction, which is plated onto laminin/fibronectin coated glass slides for four days to promote cell migration. Cells were fixed with 4% *paraformaldehyde* and stained as before. The acquisition system consists of a Nikon Eclipse TE300 inverted epi-fluorescence microscope, a MagnaFire Digital CCD camera and the Image Pro

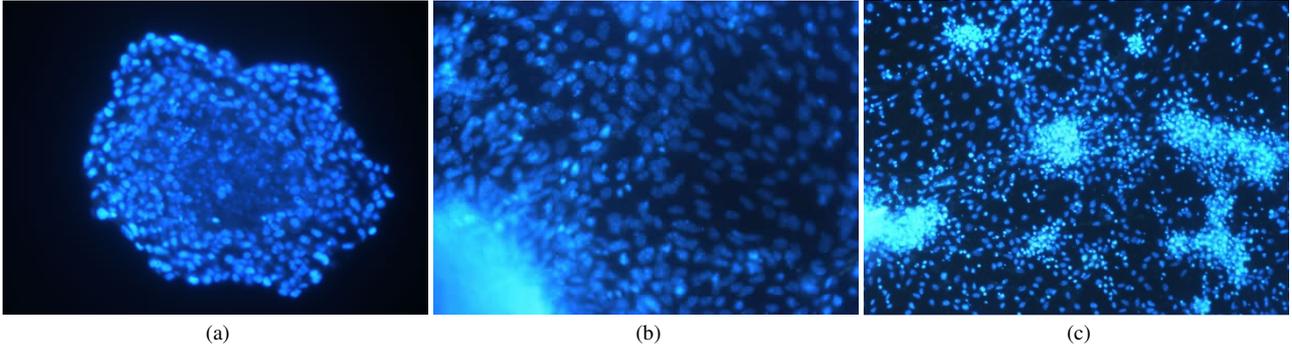


Figure 1: Captured stem cell images: a) EB cryosections; b) EB cell migration with a zoom of 40 $\times$  and; c) EB cell migration with a zoom of 20 $\times$ .

Express software. The acquisition procedure is semi-automatic and the specialist controls some parameters, such as magnification, time exposure and focus. The resolution of captured images is fixed in 1392 $\times$ 1040 pixels. A 40 $\times$  zoom was used for images of EB cryosections. The images of EB cell migration were captured using a 40 $\times$  and 20 $\times$  zoom. All images were stored using Tagged Image File (.tif) format with lossless LZW compression. Fig. 1 shows examples of captured images.

### III. OVERVIEW OF THE ORIGINAL METHOD

In this section we briefly review the main steps of the original method. For more details, please see [1].

A Gaussian filter is applied to reduce noise and then the cells are separated from the background by using a simple threshold. Next, the horizontal axis of the image histogram is partitioned into intervals of fixed size ( $\epsilon$ ). The foreground image is then divided into 8-connected regions of pixels that belong to the same partition in the histogram. The regions are then labeled with natural consecutive numbers so that higher luminance has the smaller numbers. The result of this step is a matrix  $M$ , with the same dimensions of the input image, where each entry  $M(i, j)$  contains the component label value to which the pixel  $p(i, j)$  belongs.

A region adjacency graph  $G = (V, E)$  is constructed based on  $M$ . Each vertex  $v_i \in V$  represents a region and its index corresponds to its label. The edges in  $E$  connect a pair of 4 $\times$ 4 adjacency regions. Each ES cell is represented, in the  $G$ , by a *simple path*  $S$ , whose vertices are in ascending order. Thus, a graph mining process is applied to detect the cells.

Although the method proposed in [1] works well, it is prone to errors when too many cells contain more than one brighter point, as shown in Fig. 2. The model of the original technique assumes that cell images have only one brighter point and that the luminance decreases monotonically as it goes from this point to the cell boundary. However, due to phenomena such as the DNA condensation and picnotic nuclei [2] the assumption is not always true. Thus, the method proposed in [1] does not work properly for images of EB cell migration where those phenomena are more recurrent.

In order to improve the original method and solve this shortcoming, after the graph mining process we apply a hierarchical graph-clustering algorithm as

described in next section. More details about definitions and methods for graph clustering can be found in [4].

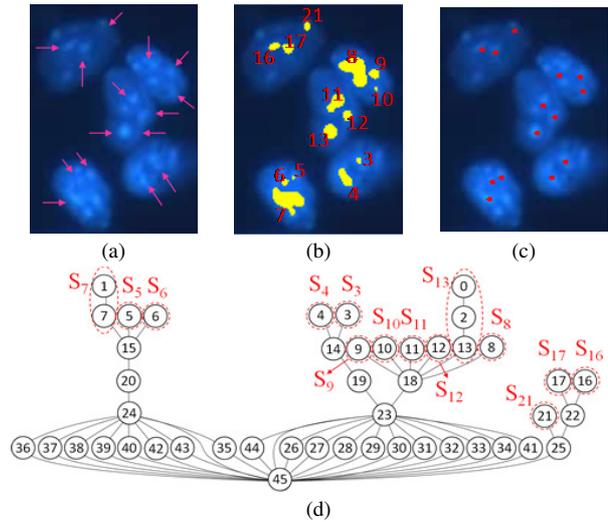


Figure 2: Original method result: a) input image with visible DNA condensation phenomenon (the pink arrows point out the brighter points); b) and c) result of the mining process (the yellow regions represent the simple paths with the correspondent biggest vertex index in red; the red dots represent the cells detected) and; d) graph-based image representation with the simple paths ( $S_i$ ), detected over the mining, marked by red dashed lines. Note that, although the mining process detects the simple paths correctly, the method failed. For example, the simple paths  $S_7$ ,  $S_5$  and  $S_6$  should represent only one cell. As a result, it detected 14 cells instead of 5.

### IV. IMPROVED METHOD

To decrease the number of false positives, i.e. the elements that are wrongly classified as a cell, we merge the simple paths that represent the same cell in clusters (supersets). We use the Euclidian distance as the metric and the fact that, if a set of simple paths represents the same cell, they should have a common neighbor, as shown in Fig. 2d (simple paths  $S_5$ ,  $S_6$ , and  $S_7$  for example). To facilitate the description of the algorithm consider the follows definitions:

- the **distance** between two simple paths  $S_i$  and  $S_j$ , is given by  $d_E(c_i, c_j)$  where  $c_i$  and  $c_j$  are the center point of the bounding box of the regions which have the biggest label in  $S_i$  and  $S_j$  respectively;
- the **neighborhood** of a simple  $S_i$  is defined as the neighbor vertices of the vertices of  $S_i$  that are not in  $S_i$ ;

- two simple paths  $S_i$  and  $S_j$  are **neighbors** if there is at least one edge connecting a vertex of  $S_i$  to a vertex of  $S_j$ ;
- a **superset** is a set of vertices made up of simple paths and additional vertices;

The basis of the clustering algorithm is that two simple paths  $S_i$  and  $S_j$  represent the same cell if  $d_E(c_i, c_j) < \lambda$ , where  $\lambda$  is one-half of the average cell's diameter. To implement this, consider the following algorithm:

Given the graph  $G$  and the list  $L = \{S_1, S_2, \dots, S_n\}$  of simple paths detected through the mining process, in order to group the simple paths which represent the same cell in clusters do the following:

1. Group the simple paths of  $L$  that have a common neighbor into sets of vertices.
2. For each set, merge those paths that represent the same cell ( $d_E < \lambda$ ) into supersets. If all simple paths, of a given set, are merged and the common vertex does not represent the background, add this vertex to the superset. Simple paths that are not merged in this step become a superset.
3. Evaluate the neighborhood of each superset and add the vertices that: a) do not belong to the other superset; b) are not neighbors of other supersets; and c) do not represent the background image.
4. Merge the supersets that represent the same cell (neighbor supersets or the ones which have a common neighbor, and  $d_E < \lambda$ ).
5. Repeat steps 3 and 4 until no supersets can be merged and no vertex can be added to some superset.

The result is a collection of supersets (clusters) in the graph, each one representing a cell. Fig. 3 presents the result of the clustering algorithm for the graph and image showed in Fig. 2.

## I. EXPERIMENTS AND RESULTS

In this section, we assess the proposed method and analyze the experimental results. Both, the original counting method [1] and its improved version here proposed were tested. Although, the cell counting is a well known problem, we are not aware of any other work that handles with several types of pluripotent stem cell in the same microscopy image. For that reason, we not provide comparative studies related to others work.

A database with 141 images, such as those described in Section II, was constructed and divided into four sets: set 1, with 69 images of EB cryosections with an acceptable level of noise; set 2, formed by 23 images of EB cryosections with strong presence of noise; set 3, with 32 images of EB cell migration with a zoom of 40x, and; set 4, composed of 17 images of EB cell migration with a zoom of 20x. The sets 1 and 2 are the same used in [1]. The values for the input parameters were obtained through experimental tests. The value of the Gaussian radius  $\sigma$  was 2 for image sets 1, 3 and 4,

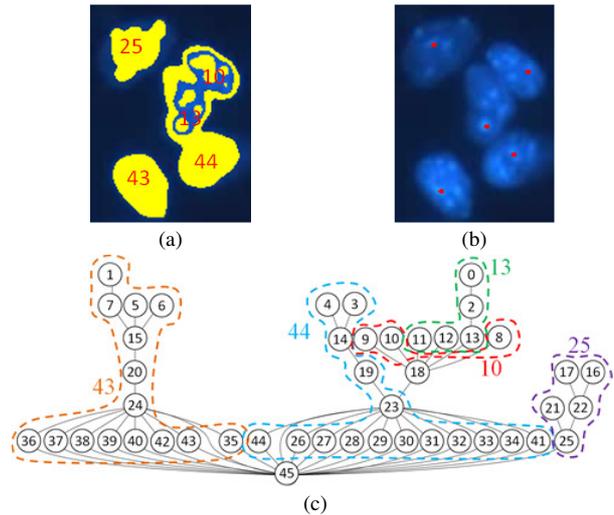


Figure 3: Improved method result: a) and b) result from the clustering algorithm on the input image (the yellow regions represent the supersets with correspondent biggest vertex index in red; the red dots represent the cells detected); and c) the clustering algorithm outcome on the graph. Clusters (supersets) are marked by dashed colored lines and the colored numbers represent the biggest index of each cluster.

and 3 for image set 2. The threshold was the same used in [1] for all groups. The histogram was partitioned into intervals of size  $\epsilon = 16$  for sets 3 and 4, and  $\epsilon = 8$  for image sets 1 and 2. The estimated value for the cell diameter was  $\lambda = 12$  for set 1 and 2. For sets 3 and 4, we use  $\lambda = 30$  and  $\lambda = 15$ , respectively.

In order to compare the quality of both methods, we used the measures of precision ( $P$ ), recall ( $R$ ) and F-measure ( $F$ ) defined as follows:  $P = tp/(tp+fp)$ ,  $R = tp/(tp+fn)$  and  $F = (2*P*R)/(P+R)$ , where  $tp$  (true positives) represents the number of items correctly labeled as cells,  $fp$  (false positives) represents the items incorrectly classified as cells and  $fn$  (false negatives) represents items that were not classified as cells but should have been. The values of  $fp$  and  $fn$  were obtained by specialists from the Institute of Biomedical Sciences at UFRJ/Brazil. They evaluated each image counted using both methods by pointing out cells that were not counted (false negatives) and artifacts that were incorrectly classified as cells (false positives). The measures were calculated for each image and then averaged over all images. The results of the improved method are presented in Table 1 and Fig. 4 compares both methods.

TABLE 1: EXPERIMENTAL RESULTS OF THE IMPROVED METHOD. ALL NUMBERS ARE AVERAGE VALUES OVER ALL IMAGES.

	Precision (%)	Recall (%)	F-measure (%)
Image set 1	96.29	93.13	94.60
Image set 2	96.96	91.48	94.08
Image set 1 and 2	96.47	92.68	94.46
Image set 3	89.74	83.70	86.09
Image set 4	97.07	82.74	89.24

The original technique [1] was designed for images of EB cryosections (image sets 1 and 2) and most counting errors were due to the existence of more than one brighter point per cell. By applying the graph-

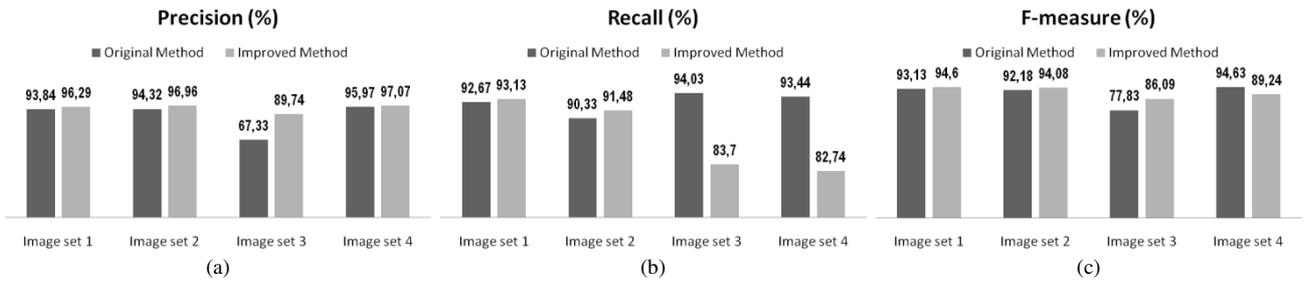


Figure 4: Comparison of original method [1] and its improved version here proposed: a) Precision, b) Recall and c) F-measure.

clustering algorithm described in the previous section, we improved the precision, recall and F-measure by 2.49%, 0.64% and 1.58% respectively for these images.

As for images of the migration of EB cells, we improved the precision by 22.41% for images with a zoom of 40x (image set 3). However, some cells that were very close (a distance less than  $\lambda$ ) are detected as only one, increasing the number of false negatives. Nevertheless, the F-measure increased from 77.83% to 86.09% for those images. When it comes to images of EB cell migration with a zoom of 20x (image set 4), we see that the improved method does not work properly. Due to the zoom factor, the cells present in these images are very small and as a consequence the brighter points are not visible. Thus, the cells appear with just one bright point as supposed by the original technique. As a consequence, when the clustering process is applied, it groups some simple paths that represent different cells. Fig. 5 shows examples of three images, one for each image type, processed by both methods.

II. CONCLUSION AND FUTURE WORKS

In this work a refined method for counting ES cells in fluorescence microscopy images was described. It uses the same principle of cell detection proposed in [1] and incorporates a graph-clustering algorithm to improve the

efficiency of the resulting system. Both methods were extensively tested in a database of 141 images and specialists validated the results. We improved the precision by 2.4% for images of EB cryosections and increased it from 67.33% to 89.74% for images of EB cell migration with a zoom of 40x. Nevertheless, the original technique works better than the one presented previously in this article for images of EB cell migration with a zoom of 20x, since they have the cell pattern proposed in [1].

Future works involves extending the proposed methodology to deal with images of ES cells under Mouse Embryonic Fibroblast (MEF). Although the current method is able to detect the cells, it is not able to distinguish between ES cells and MEF.

REFERENCES

- [1] G. Faustino, M. Gattass, S. Rehen, C. J. P. de Lucena, Automatic embryonic stem cells detection and counting method in fluorescence microscopy images, in: ISBI, 2009, pp.799-802.
- [2] M. Sell, Stem Cells Handbook, Human Press, 2003.
- [3] Z. Theodosiou, I. Kasampalidis, G. Livanos, M. Zervakis, I. Pitas and K. Lyroudia, Automated analysis of FISH and immunohistochemistry images: a review, Cytometry Part A 71A, 2007, pp. 439-450.
- [4] S. E. Schaeffer, Graph Clustering, Computer Science Review, 1(1): pp. 27-64, 2007.

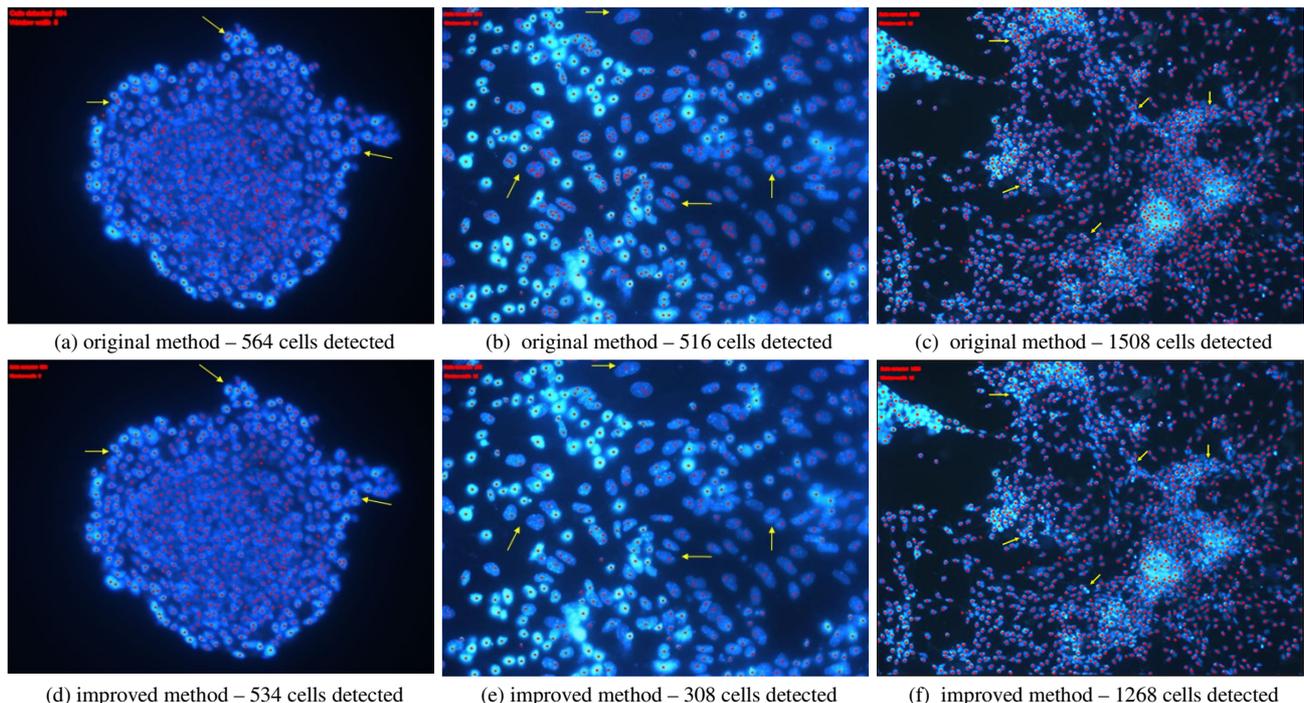


Figure 5: Results of original and improved method: the red dots represent the cells that were detected and the yellow arrows point out some image regions where the graph-clustering algorithm led different cell detection (counting).