

Digital processing of in situ hybridization images: Identification and spatial allocation of specific labels

Mariela A. Gonzalez †‡, Melina Rapacioli †, Virginia L. Ballarin ‡, Sara Fiszer de Plazas # y Vladimir Flores †,#

† Interdisciplinary Group in Theoretical Biology, Department of Biostructural Sciences, Favaloro University.

Institute of Cell Biology and Neurosciences "Prof. E. De Robertis", U.B.A.

‡ Signal Processing Lab., Department of Electronics, Universidad Nacional de Mar del Plata / CONICET

ABSTRACT

In situ hybridization (ISH) method allows to reveal specific genes expression, identify specific cell types and detect areas or tissues, displaying differential gene expression. This work describes a standardized procedure of digital image processing that allows detailed analyses of ISH preparations. We have developed a software that allows through a graphical interface (a) to reliably identify and quantify ISH labels, (b) to locate each label within the image reference system (c) to assemble the total series of images obtained from a complete histological sections of a biological structure, and d) to locate all the labels within a unique reference system that corresponds to the complete biological structure.

As each ISH label is positioned within a spatial coordinates system coinciding with an intrinsic biological reference axis, this software can be used to analyze the spatial pattern of distribution of specific genes expression during the embryonic development. The software allows the construction of numerical space series that can be used to analyze the variability and the dynamics of genes expression as a function of space and/or time during the embryonic development.

Keywords: Segmentation, Digital image processing, In situ hybridization, Developmental Biology.

1. INTRODUCTION

The development of pluricellular organisms implies an increase in complexity, from the initial stage, the ovum, to the final stage of terminal differentiation. Although this increase in complexity is a continuous process, it can be considered as an ordered sequence of stages; corresponding a typical set of structural and functional changes to each specific stage. Developmental changes are the net result of several developmental cell behaviors (DCBs) operating simultaneously and interactively. DCBs are temporally and spatially

organized. Cell determination, cell differentiation and patterning are central aspects of the Developmental Biology [1]. The former refers to the processes by which different cell types are generated; the latter refers to processes by which the different cell types are temporally and spatially organized originating the different tissues and organs of complex organisms.

The cell diversity is explained by successive processes of differential or selective gene expression. Although the whole population of cells of a pluricellular organism possesses the same basic genetic information, different cell types express different patterns or combinations of cell-type specific genes. In situ hybridization (ISH) method allows detection of genes expression and qualitative and quantitative descriptions[2].

Most of the commercial image processing softwares does not precisely determine some basic characteristics and/or details of the images obtained from an ISH preparation. This has led to the development of several types of software.

Performance analysis of the software designed by several authors indicate that, apart from its obvious utility, they also display some disadvantages. Some of them deserve consideration: (a) they do not integrate the information of several images; (b) they do not capture the spatial information; (c) they exhibit poor ability to discriminate overlapped objects [3]-[8]; (d) they usually require excessive time of image pre-processing [3]; (e) they usually applied filters and threshold that are very sensitive to variations in the images properties thus decreasing the results reproducibility [3]-[8].

The present paper aims at designing and implementing registration and segmentation algorithms in order to improve image analysis of ISH preparations and eliminate the former mentioned disadvantages.

2. MATERIALS, METHODS AND RESULTS

Algorithms were implemented in C++ [9].

The first step was devoted to get an automatic identification of specific labels taking as reference the diagnostic made by a biologist experienced in ISH method.

The second step was directed to obtain an automatic counting of specific labels followed by their positioning within the image reference system.

The third step was dedicated to obtain a unique image by assembling successive serial images obtained from a complete biological structure along a defined spatial axis. Simultaneously, the specific labels location of all images were re-assigned into a unique global reference system that corresponds to the complete biological specimen. This procedure allows to analyze the complete spatial pattern of labels taking as reference a spatial axis that coincides with a defined intrinsic reference system of the biological specimen. Finally, spatial coordinates datum of each specific label within global reference system were stored.

A. Specific label identification

The biological material under analysis correspond to histological sections of the developing chick embryo optic tectum (an associative area of the midbrain of the central nervous system) processed according to standardized procedures of ISH methods. A detailed description of the ISH method used in this work is found in reference [2]. ISH preparations analyzed in this work reveal the expression of GABA_A receptor subunits mRNA. This receptor complex is expressed by local circuit neurons in the layer “i” of the developing optic tectum where GABA functions as the main inhibitory neurotransmitter [1][2]. Images were captured with a color digital scanner (Axiocam HR color) from a light microscopy (Axioplan 2 imaging) at a final magnification of 5000 X. Then, images were converted to a gray scale by averaging their color components.

Figure 1 shows a typical image of developing neurons processed by ISH method. Round to ovoid-shaped neurons can be observed. They are encircled by clear zones that correspond to the extracellular matrix. The cytoplasm and the nuclei exhibit a wide range of gray levels with irregularly shaped dark zones which will be called “gray-dark bodies” (g-d bodies). The specific labels appear as well-defined,

approximately round in shape and intensely dark granules.

Figure 2 shows the profile of gray levels corresponding to a row of a negative image. These profiles typically exhibit three kind of oscillations of the gray level: (a) low frequency waves with an amplitude of approximately 60-80 levels that correspond to adjacent neurons separated by intercellular space; (b) high frequency variations with an amplitude of approximately 20-30 levels that correspond to the g-d bodies and (c) peaks of very high frequency with an amplitude of more than 100 levels of gray values. These peaks correspond to the specific labels. Small oscillations with amplitude below 10 levels, appear as a noise along the entire record.

First, a threshold which varies as a function of the gray level of the image is calculated. The value of this variable threshold (VT) is calculated for each pixel of the negative images. This VT is composed of two terms: (a) a variable statistical threshold (VST) calculated as the mean value of the gray levels of defined area around each pixel and (b) a constant term “C” empirically determined by the operator.

The value of VT for each pixel corresponding to the row “r” and column “c” [VT(r,c)] is calculated as

$$VT(r,c) = VST(r,c) + C \quad \text{where}$$

$$VST(r,c) = \left[\frac{I}{(2 \times j + 1) \times (2 \times i + 1)} \sum_j^j \sum_i^i \text{image}(r+i, c+j) \right]$$

where “i” and “j” defined the area around each pixel and “image(r+i, c+j)” is the gray value of the pixel positioned at row “r+i” and column “c+j” of the negative image.

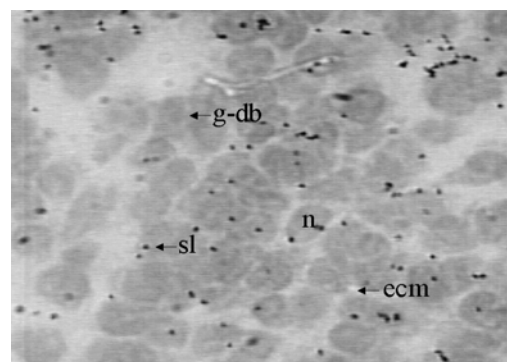


Fig. 1. Light micrograph of an ISH preparation obtained from a developing optic tectum. n: neuron; ecm: extracellular matrix; g-db: gray-dark body; sl: specific ISH label. 1000X

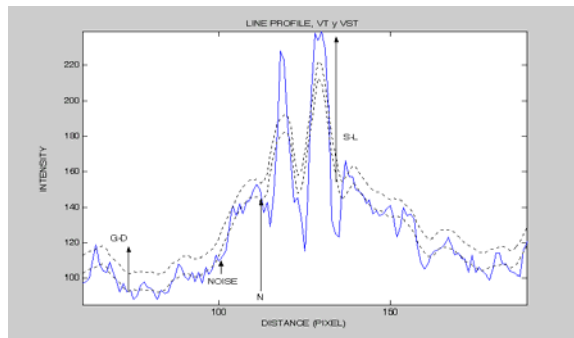


Fig. 2. Upper Dotted Line: VT, Lower Dotted Line: VST, Continuous Line: Single Line Image Gray Level

As the mean value of a set of pixels with significantly different gray values is minor than the maximal value and major than the minimal one, VST results in a smoothing of gray levels profile. An appropriate selection of the number of pixels to be averaged, which depends on the size of the area around each pixel, allows a reliable identification of specific labels which corresponds to sharp peaks. Figure 2 also shows VST and VT profiles. It can be seen that labels, g-d bodies and noise gray values are greater than VST values. However, addition of an appropriate “C” value allows to obtain a VT value that eliminates noise and also smaller g-d bodies variations. Thus labels gray values significantly surpass the VT values, in comparison with irrelevant components. Moreover due to the statistical nature of VT, this is independent of global characteristic variations such as brightness, contrast and resolution.

The second step identifies the specific labels. It can be noted (Fig. 2) that the area of the plot defined between the record of gray values of the negative image and the VT profile is significantly larger for the specific labels than for the g-d bodies and others irrelevant components. These areas are calculated along the each row by adding the gray values of successive pixels (conforming (a) sets of immediately adjacent pixels with (b) gray values higher than their corresponding VT value). A frequency histogram of these areas shows a bimodal spectrum. Thus it is possible to set a threshold, AT, to distinguish specific labels using the areas previously calculated. Whenever the calculated area is higher than the AT, the set of pixels corresponds to a specific label. This procedure allows to obtain a binary image. The segmented image preserves only specific labels.

B. Specific label allocation within the image reference system

In this step, a file containing the coordinates of each specific label in segmented image reference system is obtained. The position of each label is determined by calculating the mean value of the coordinates of pixels that define its contour. The contour of each label was obtained by implementing a standard algorithm of contour tracing named “Moore” [10]. The algorithm outcome is a file containing the coordinates of the pixels that outline each label and the values of their perimeters. Partially overlapped labels, which are rarely found in ISH images, cannot be discriminated by the segmentation algorithm. In these cases an index named as Roundness Factor (RF) is used to approximate the number of labels which are overlapped. The RF depends on the object shape: values near to 1 correspond to circular objects and values near to 0 correspond to very anisometric objects [10][11]. An FR value can be established as threshold to discriminate between partially overlapped labels (with a low FR) and non overlapped ones (with a FR near to one). Afterwards, the number of labels occupying each area with FR value below the threshold is estimated. The value of each area resulting from label superposition is divided by the value of a reference area that is slightly smaller than the mean area of the non-overlapped labels.

Then the coordinates of the labels that compounds each group of overlapped labels are set equidistant using the number of labels determined in the prior procedure.

All the coordinates are store in a file. From this file we are able to obtain a sintetic image which contains standarts labels correctly separated and positioned.

C. Assembly of serial images obtained from a complete histological preparation

A procedure was designed in order to re-locate the labels of several serial images, corresponding to a complete histological preparation, within a unique reference system. The reference system of the first image sets the coordinates origin for the complete series of images. The labels of the first one remain within their own reference system, i.e. their coordinates remain unchanged. The labels of the following images are submitted to the reference of the first one, i.e. their coordinates change according to their position respect

to the reference system of the first image. This procedure is performed step-by-step. Every new image is assemble to the preceding one by determining the position of its coordinates origin respect to the coordinates origin of the preceding one. This procedure allows the integration of the series of images into a single image of the complete histological preparation. It also integrates the labels by means of coordinates transformation submitting them to the biological specimen reference system. Afterwards, the coordinates of the specific labels within the unique reference system are stored as a single text file and these data are used to create a new image. This image displays the spatial pattern of distribution of the whole label population of the biological specimen. The result of this procedure is illustrated in figure 3. It can be noted that an area of border overlapping of adjacent images is needed in order to perform an appropriate assemble of images. This allows to apply the cross-correlation between the overlapped areas. An optimal automatic assembly was obtained by cross-correlating the borders of the negatives of the original non-processed images. We also delete repeated labels during this process.

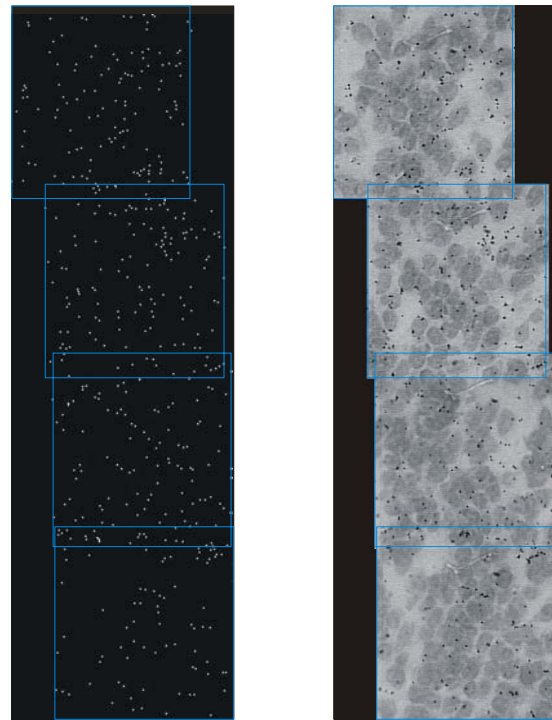


Fig 3. Righth Image: Overlapped ISH Images by the expert. Lef. Image: Overlapped ISH Images by the software developed

D. Fidelity

The segmentation algorithm performance was evaluated by determining the percentages of misleading identification. False positives (F+) and false negatives (F-), as a function of the AT and of the “C” values, are calculated. F+ are defined as labels preserved in the segmented image and rejected by an expert. F- are identified by an expert but are not detected by the algorithm. Images with a size of 256 x 256 pixels and a 72 dpi resolution were used to this purpose.

Figure 4 shows that the number of F+ significantly decreases when the AT increases up to 28 pixels. At that value the number of both F+ and F- are negligible. It can be seen in figure 5 that the number of F+ significantly decrease while the number of F- remains almost unchanged (within a range of 1% +/- 0.5% of the total number of label) as a function of the increments in the “C” value. In this case also exists a range of “C” values with a negligible percentage of F+ and F-. Due to these behaviors, the AT and the “C” values can be increased obtaining a significant decrease in the percentage of F+ without increasing the percentage of F-. Once the appropriate values of AT and “C” are selected they are used for processing the complete series of images since they results in a similar

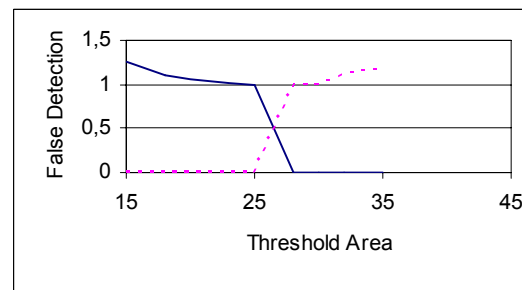


Fig 4. Percentage of F+(continuos line) and F- (dotted line) as a function of the AT values. There is range of AT values with almost null percentage of errors.

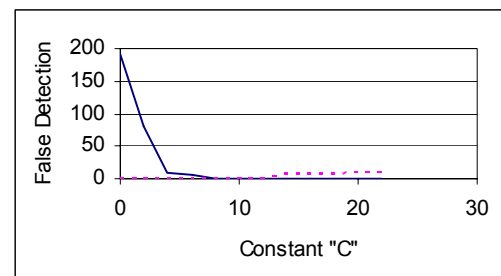


Fig. 5. Percentage of F+ (continuos line) and F- (dotted line) as a function of the constant “C” values. There is range of “C” values with almost null percentage of errors.

satisfactory performance for a particular ISH preparation. The application of these algorithms to images obtained with different resolutions, brightness and contrasts gives segmentation results with a percentage of error fewer than 1%.

3. CONCLUSIONS

The genesis of complexity that takes place during the embryonic development requires the simultaneous operation of several DCBs. The DCBs are spatially and temporally organized. Any attempt to understand and formalize the developmental effects of a particular DCB requires a comprehensive description of its dynamics of operation in terms of scaling coefficient. These coefficients can be obtained by analyzing the variability of a biologically significant parameter as function of the time and/or the space. This paper introduces a computational procedure that can be useful to record the variability or differential expression of specific genes as a function of the space. These kinds of records are suitable for the above mentioned purpose.

A limiting step to apply this kind of quantitative analyses to images obtained from developing organisms resides on: (a) the poor precision of the conventional (commercial) segmentation algorithms, (b) the difficulty to apply them automatically and (c) the large number of images that must be analyzed to record a significant amount of data. This explains the scarcity of mathematical analyses and the abundance of qualitative descriptions in the field of the developmental biology.

This paper presents a software with high precision, automaticity and versatility to identify specific labels of gene expression. Once the parameters of the algorithm are appropriately selected by the expert this software allows: (a) identification of more than 99% of the labels identified by an expert; (b) allocation of each label within the reference system of the image and (c) assembly of serial images obtained from a biological specimen and the re-assigning of defined coordinates to each specific label within the global reference system of the complete histological section. It can be noted that if the histological section coincides with an intrinsic reference system of the biological specimen, this procedure allows to obtain a replica of the spatial pattern of cells that at a defined stage are expressing a specific gene. In other words, the software preserves the spatial information.

The automaticity of the algorithms allows the processing of a great number of images in a short lapse of time and the high performance is due to the strategy to differentiate amongst objects displaying

different ranges of gray levels. In fact, the selection of appropriate values for the parameters VST, C and AT successively eliminates the gray levels variations due to the intercellular space, neurons, g-d bodies and noise and allows the identification and preservation of the specific labels.

With regards to the versatility or adaptability, it can be mentioned that: (a) the expert has the possibility to change the VST value for the detection of other important objects with different sizes and/or different spatial resolution; in fact, the VST is independent from the image resolution. The user can modify the area around each pixel depending on the number of pixels occupying by the specific object; (b) since the VST varies with the brightness and contrast of the image this parameter eliminates the difficulties resulting from abrupt changes in those image characteristics; (c) as a consequence, no image improvement is required avoiding excessive times of image pre-processing and decreasing the need for excessive memory; (d) AT can be changed in order to reveal or to preserve labels out of the microscope focus; (e) the application of RF allows the detection and separation of overlapped labels.

This software can also be used to determine the number and position of each cell and this allows to obtain relevant biological indexes: number of cells, percentage of labeled cells, intensity of expression (mean number of labels / cell) etc.. Although this software was implemented for the detection of ISH preparation it can be used advantageously for the detection, location and determination of the spatial pattern of specific labels obtained by several other methods such as histochemistry, in situ immunolabeling, in situ zimography and others.

Acknowledgements. This paper was supported by grants from CONICET (Argentina).

4. REFERENCES

- [1] S. Gilbert, *Developmental Biology*, Sinauer, 1997.
- [2] D. J. Rodriguez Gil, M. Vacotto, M. Rapacioli, G. Scicolone, V. Flores and S. Fiszer de Plazas. "Development and Localisation of GABAA Receptor $\alpha 1$, $\alpha 2$, $\beta 2$ and $\gamma 2$ Subunit mRNA in the Chick Optic Tectum". *Journal of Neuroscience Research*, Vol. 81, 2005, pp. 469-480
- [3] W. H. Schuette, S. Chen, S. J. Occhipinti, H. S. Mujagic, S. E. Shackney, "Automated radioautographic grain counting", *Cell Tissue Kinet*, Vol. 3, 1983, pp. 221-7.
- [4] M. Masseroli, A. Bollea, C. Bendotti, G. Forloni, "In situ hybridization histochemistry quantification: automatic count on single cell in digital image", *J Neurosci Methods*, Vol. 2, 1993, pp. 93-103.

- [5] R. Mize, C. Thouron., L. Lucas, R. Harlan, "Semiautomatic image analysis for grain counting in situ hybridization experiments", *Neuroimage*, Vol. 3, 1994, pp. 163-172.
- [6] W. Stolz, K. Scharffetter, W. Abmayr, W. Koditz, T. Krieg, "An automatic analysis method for in situ hybridization using high-resolution image analysis", *Arch Dermatol Res.*, Vol. 5, 1989, pp. 336-341.
- [7] A. Ambesi-Impiombato, G. D'Urso, G. Muscettola, A. de Bartolomeis, "Method for quantitative in situ hybridization histochemistry and image analysis applied for Homer1a gene expression in rat brain", *Brain Res Brain Res Protoc.*, Vol. 3, , 2003, pp. 189-196.
- [8] W. Zhao, J. Truettner, R. Schmidt-Kastner, L. Belayev, M. D Ginsberg, "Quantitation of multiple gene expression by in situ hybridization autoradiography: accurate normalization using Bayes classifier.", *J Neurosci Methods*, Vol. 1, 1999, pp. 63-70.
- [9] H.M. Deitel, P.J. Deitel, *Como programar en C++*, Prentice hall, 1994.
- [10] C. Russ, *The Image Processing Handbook*, CRC Press, 1992.
- [11] R. González, R. Woods, *Tratamiento Digital de imágenes*, Addison Wesley, 1996.