Band Detection in DGGE Images Using Connectionist Models and Evolutionary Computation

Carola Figueroa[1], **Cristopher Alvear**[2], **Angélica Caro**[1]

[1]Department Computer Science and Information Technology
Universidad del Bío-Bío – Chile

[2]School of Computer and Information Engineering
Universidad del Bío-Bío – Chile

{cfigueroa,mcaro}@ubiobio.cl

Abstract. *Current research shows the development of a first approximation algorithm of Denaturing Gradient Gel Electrophoresis (DGGE) band image identification. The proposed algorithm considered the use of artificial neural networks based on deformable connectionist models subjected to an evolutionary algorithm.*

Resumo. *A pesquisa atual mostra o desenvolvimento de um primeiro algoritmo de aproximação da identificação da imagem da banda da eletroforese de gel de gradiente desnaturante (DGGE). O algoritmo proposto considerou o uso de redes neurais artificiais baseadas em modelos de conexão deformáveis submetidos a um algoritmo evolutivo.*

1. Introduction

Digital images are two-dimensional representations corresponding to a set of pixels arranged in a matrix for digital processing. If the representation is in a grayscale, each pixel is represented by a numerical value from 0 to 255, which illustrates the different shades of gray ranging from black (0) to white (255). On the other hand, segmentation [Mao and Jain 1992, Rathod et al. 2010] is an operation that divides an image into parts or regions with some particular characteristic. Segmentation algorithms are based on two fundamental properties that depend on changes in intensity. The first property is related to the similarity between pixels whereas the second focuses on the continuity related to gray levels. The latter approach involves a family of edge detection algorithms that consider strong intensity that varies between neighboring pixels. The image type used in this paper was a DGGE (Denaturing Gradient Gel Electrophoresis) image; this technique is based on DNA and generates a genetic profile which identifies the members of a microbial community [Muyzer and Smalla 1998, Cebron et al. 2004]. The DNA sample was deposited on a gel to which an electrical current was applied; DNA molecules were denatured, so that their strands ran on the gel to form an image with lanes (depending on DNA molecular weight) and bands (depending on the number of DNA bases). Once the DGGE image was obtained in digital format, the analysis or processing stage began; in the case of some Biotechnology Centers, this task was carried out with the Quantity-One software tool. However, automatic detection of these bands by these applications occurs inefficiently and with a significant percentage error due to imprecise parameters for identifying bands; detection is therefore carried visually and manually in most cases [Cebron et al. 2004, Prabhakar and Bishop 2014]. Providing an efficient solution to this problem requires automatic segmentation of the bands present in the DGGE images. A new segmentation methodology was therefore used, which has been initially tested as efficient in different images and based on connectionist models (artificial neural networks, ANNs) for modeling the required image segmentation [Sierra et al. 2013, Novo et al. 2012]. This methodology implies that ANNs are applied on each image pixel and receive the local ANN information in thispixel to determine its segmentation. The same ANN is applied on all the image pixels, thus acting as a segmentation operator such as Sobel or Canny Matlab. The iteration along the image and in different temporal iterations provide an emerging image segmentation, thus taking advantage of the generalization property of the ANNs which can correctly segment entries not considered during their training. The ANNs are automatically obtained and trained through evolutionary computation methods. They especially focus on a current method with tested effectiveness and robustness as differential evolution [Graves and Pedrycz 2009, Salman et al. 2007, Omran et al. 2009, Noktehdan et al. 2010]. The main objective of this article was to show if it was possible to obtain automatic DGGE image processing by connectionist models and evolutionary computation. Since this type of processing is necessary for image analysis in many biotechnology centers for wastewater and drinking water treatment, microbially induced corrosion analysis, or species biodiversity measurement in a given environment [Muyzer and Smalla 1998, Kim et al. 2008, He et al. 2010, Durand et al. 2013] Therefore, the results obtained in this study provide new image segmentation techniques and adapts them to the application of DGGE imaging. Section 2 describes the problem in more detail by highlighting the usefulness of this type of image and the positive impact it could have in finding solutions in the field of biology. Section 3 is related to the processing and treatment of DGGE images with some information about connectionist models and evolutionary algorithms in the image segmentation process. Section 4 describes the proposed solution while Section 5 gives details about how the proposed solution was developed, describing the methodology implemented and the algorithms and heuristics used. Section 6 describes the experiments performed on DGGE images and artificial images to test and illustrate the efficacy of the solution. Finally, Section 7 presents conclusions and future work.

2. Description of the Problem

The development of techniques related to the study of DNA provides comprehensive and rapid knowledge about biodiversity, evolution, and animal and plant genetics. The high sensitivity of these techniques along with the discovery of regions with a high degree of variability have led to significant advances in studies of population genetics, biogeography, and polymorphism in human populations; applications have been developed in forensic medicine, paternity determination, and the diagnosis of hereditary diseases [Durand et al. 2013]. The main advance is that these techniques are able to directly access an individual?s genotype, thus preventing phenotypic expression and environmental influence [Muyzer and Smalla 1998, Durand et al. 2013]. The basic principle of molecular marker acquisition and detection by gel electrophoresis involves the electrochemical separation of molecules. As indicated in the previous section, the DGGE imaging technique is based on the differential migration of molecules with different size loads when

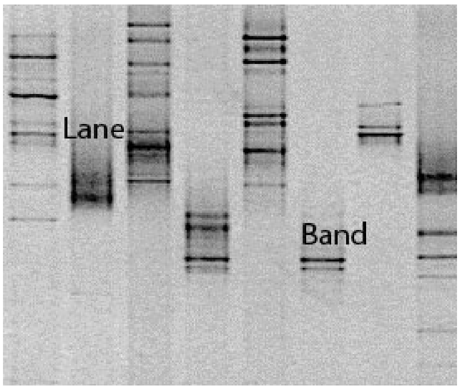

Figura 1. DGGE image [Nicolaisen and Ramsing 2002]

they are subjected to an electric field (electrophoresis). An example of an electrophoresis image is shown in Figure 1, The vertical rails are called lanes (each lane is a sample). The material in the lanes are called bands; they are the positions at which the molecules stop. The analysis of the electrophoresis results is performed by comparing the paths between the bands. However, this comparison is usually a complex and tedious process due to the subjectivity of human visual perception. Therefore, two people visualizing the same material can reach different conclusions [Muyzer and Smalla 1998].

Researchers from several biotechnology centers in Chile and the world are currently visually analyzing electrophoresis images and building a similarity matrix. This matrix is used as input for statistical software such as GENES [Cruz 2013], STATISTICA [Statsoft 2014], or Quantity-One [Quantity-One 2016] that perform cluster analysis and data dispersion. Digital electrophoresis image analysis is emerging as an important application for reducing human error and improving data evaluation speed [Ye et al. 1999]. The automatic band pattern analysis of a lane could evaluate many parameters that are generally ignored by human analysts [Machado et al. 1997]. Given that electrophoresis image analysis is a manual, subjective, and tedious task for researchers, the automated analysis of such images is interesting because it would reduce human errors, the interpretation process would be faster, and it could even be used to study large databases of samples. Therefore, a tool that is able to perform an accurate and efficient automated analysis would be of great value in the field of general biology and genetics because it could reduce research time and the costs involved.

3. Related Research

Digital image processing methods are applied in two main areas. The first is improving visual information for human interpretation, and the second is processing image data for perception through automatic machines [Benoit et al. 2014]. The field of image processing has steadily grown since the 60s. The same techniques are used to solve a variety of problems where visual information needs to be improved for analysis and human interpretation. The following are some of the areas that support image processing: space research, medicine, geography, archeology, physics, astronomy, biology, as well as supporting the law or defense. Some related research studies about DGGE image processing are described below.

In [Figueroa et al. 2009] defined a technique that starts by extracting lanes and

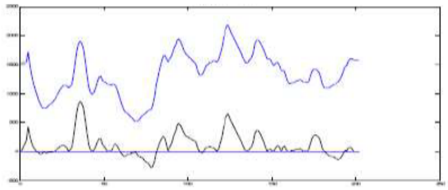

Figura 2. Normalized histogram. External Source [Figueroa et al. 2009]

then obtains a histogram sum of pixels, which is subjected to the use of morphological operators and smoothing. After obtaining the smooth curve of the histogram?s sum of pixels normalization occurs, which functions through partial derivatives on minimum points of each peak present in the histogram. This leads them to the same base or zero level as shown in Figure 2, where the upper curve corresponds to smoothing whereas the lower curve corresponds to normalization. This process improves band detection and hence noise elimination.

The above mentioned process leads to the coloring of the position of each histogram peak in Figure 2; this results in the bands that are labeled in Figure 3 because it is assumed that each band corresponds to a peak.

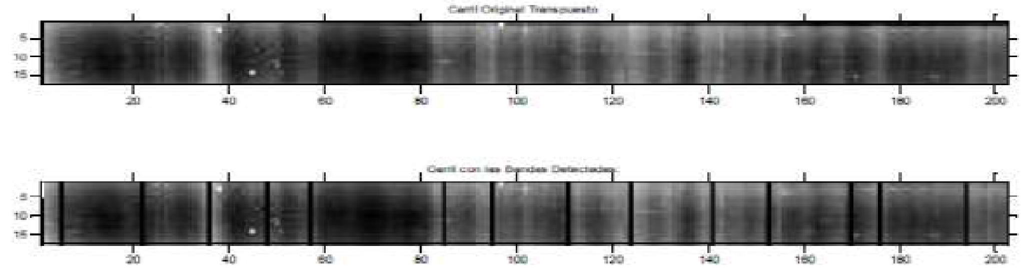

Figura 3. Identified and highlighted bands. External Source [Figueroa et al. 2009]

This technique made it possible to perform automatic band detection with 0.20 accuracy error [Figueroa et al. 2009]. However, it was not possible to obtain correct band segmentation for further quantification because when it was normalized and the noise was eliminated, useful information for DNA experts was lost, and it was almost impossible to pinpoint when a band ended and started.

In this second approach [Figueroa 2012], work was related to edge detection algorithms based on ants and proposed by nombrar autor [Dorigo and Di Caro 1999]. In these algorithms, virtual ants simulate the stigmergy process when they deposit pheromones that provide information to the others so as to collectively find short paths (optimized) to food sources. It was possible to adapt this process to optimize band detection to trace the contour of each ant; this required implementing two algorithms, that is, the Elitist Ant System and Ant System [Pinninghoff et al. 2012]. The best parameters to find the best possible results were established in these studies, which provided support for the subsequent band quantification stage. Some results obtained after applying these algorithms are reflected in Figure 4.

Certain bands in Figure 4 were difficult to detect by virtual ants and hence the resulting segmentation was not successful. Alternative aspects of this work are described

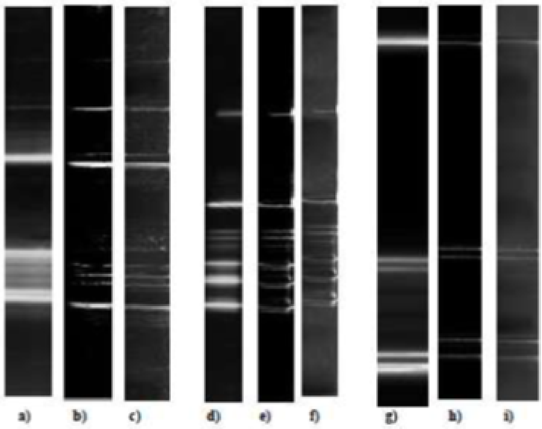

Figura 4. Comparison of the values obtained from the Ant System (b, e, h) and Elitist Ant System algorithms. External Source [Figueroa 2012]

in other studies [Figueroa 2012, Pinninghoff et al. 2012]. Some research studies can be highlighted that are related to DGGE image treatment; one study describes some methods to improve the quality of DGGE image presentation with respect to band appearance using pure genetic algorithms and genetic algorithms with tabu search [Pinninghoff et al. 2014]. Other studies only correct how the lanes and bands appear as described by nombrar autor [Gárate et al. 2011]. However, these authors only used RAPD images to evaluate how the technique performs and combined the genetic algorithm technique with tabu search and thresholding. On the other hand, the image segmentation is on of the main areas in automatic image processing. This process is responsible for subdividing an image into its constituent parts or objects; for example, the separation of the objects from the background. Thus, image segmentation is usually the first step in performing the analysis [Ademek 2006]. Automatic image segmentation is usually one of the most difficult tasks in image processing [González and Gasull Llampallas 2008] and [Zeng 1999]. This step determines the success or failure of the analysis; when it is effectively carried out, success is almost guaranteed. To perform robust automatic segmentation and given image brightness variations, many algorithms are based on selecting cutoff values in terms of grayscales in the image, which is usually done by analyzing the histogram [Otsu 1979]. Approaches for segmenting monochromatic images are generally based on two basic properties of grayscale values: discontinuity and similarity. In the first category, methods are aimed at dividing the image based on abrupt changes in grayscales, offering image lines and edges. In the second category, methods try to group the image pixels that have similar values for a specific set of characteristics. Likewise, nombrar los autorespropose a new technique that involves neural networks and a deformable model (Topological Active Nets, TAN) [Sierra et al. 2013, Tsumiyama and Yamamoto 1989], an extension of the classic snake model [Kass et al. 1988]. The TAN model integrates characteristics of the two described segmentation techniques by providing information about the image contours (by external mesh or networknodes) and internal characteristics of the image (by internal nodes, see Figure 5). The deformation over time of the TAN nodes is defined by an artificial neural network (ANN), which learns to move each node of the segmentation model by using local image information in the position of each node. The ANN is applied to each of the nodes in different time intervals up to the final segmentation as shown in Figure

5. Therefore, this type of model has a dynamic behavior that allows precise adjustments on local topological changes to find all the objects of interest in the image. The deformation of the model or mesh is controlled by energy functions such that the energy mesh (TAN) has a minimum value when the model correctly segments the objects of the scene. Thus, the segmentation process is converted into a minimization task and ANN receives the energy changes as information; these occur when each node of the mesh is moved (which in turn is local image information in the node position). This proposal was tested in different artificial and real images, which showed the abilities and advantages of this type of methodology. An example is illustrated in Figure 5 where the TAN was initially established within the limits of the image and all nodes were moved until the correct segmentation was achieved [Sierra et al. 2013, Novo et al. 2011].

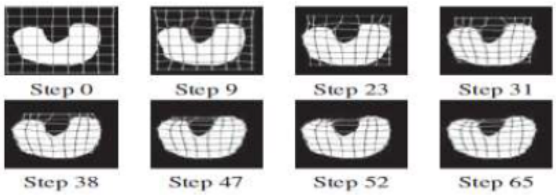

Figura 5. Emerging segmentation provided by TAN (Topological Active Net). External Source [Novo et al. 2011]

Therefore, the main objective of this article was to obtain automatic processing of these images, which was necessary for their subsequent analysis by connectionist models and evolutionary computation; accuracy was thus increased in band detection.

4. Proposed Solution

As previously defined, the purpose of this article was to obtain automatic processing in the detection and identification of DGGE image bands. An ANN was used to define an emerging segmentation over time, deformable models, and evolutionary computation. By adapting the methodology to DGGE images and our main objective, the ANN input was the pixel neighborhood around the pixel to which the ANN was applied while the output indicated a grayscale (trying to indicate which pixels had gradient transitions). The same ANN was applied to all pixels in various iterations, so that an emerging segmentation was once more obtained by defining the band contours. The ANN defining the segmentation was obtained automatically by an evolutionary method. We proposed differential evolution [Novo et al. 2011, Storn and Price 1995] because it is a method with proven robustness in optimizing problems that are encoded in a genotype of real numbers (as in the present study where the genotype of the individuals of the genetic population defined numerical connection weights between ANN nodes) and with an automatic balance between exploration and exploitation in the search. Therefore, the ANN behaved as a segmentation operator, but unlike a classic predetermined segmentation operator (Sobel, Prewitt, or Kirsch) the ANN was iteratively applied in time and also automatically optimized for segmenting images. In other words, the search for an appropriate segmentation operator was automated. In addition, an ANN was obtained in an evolutionary way with an image or set of images (training set); their generalization could be tested with images (DGGE or another type) not considered during training/development (validation set) using the classic validation methodology of connectionist models.

5. Implementation

Specific characteristics of DGGE images convert the band detection process into a difficult task. Due to the low gradient scale range, a solution based on a deformable mesh with a unidimensional vertical movement that can return band recognition through an ANN behavior was proposed. The proposed algorithm is based on image processing [Sierra et al. 2013]. The algorithm proposed an image segmentation based on TAN formed by multiple nodes distributed into a bidimensional space (image) and which were adjusted to fit the shape of the object being identified after the algorithmic evolution. The following briefly describes some concepts related to the solution.

i) Differential evolution: classic evolutionary algorithm that consists in subjecting a population of algorithmic structures to evolution or improvement given a crossing percentage , mutation, and fitness parameter selection.
ii) Lifetime: Iterations where each structure improves before generational change.
iii) Fitness: In an evolutionary algorithm, this is a value defining an absolute score by arbitrary parameters of an individual or algorithmic structure that is evolving in order to compare it with other individuals.
iv) ANN: Artificial neural networks. This consists of neurons or algorithmic structures working together to achieve a common aim. For the algorithm in the present study, the ANN corresponds to the deformable mesh.
v) Node: A structure forming the ANN.
vi) Deformable Mesh: An ANN subjected to evolution and selection.
vii) Capture: This was understood as the action of a node or node set established in a band contour.
viii) Pixel value difference: In a node, this is the absolute value indicating the gradient of the current pixel vs. the future pixel.
ix) Resistance: In a node, this defines the lower limit of the range of differences;this range allows moving toward the node only within a limited grayscale. When resistance is lower, the node captures mostly smooth contours whereas the node captures abrupt contours for higher resistance. Its value is an integer from 0 to 255 (pixel grayscale value range).
x) Vertical width: In a mesh, this is the length from the row with the highest node to the row with the lowest node.
xi) Dispersion: This is the maximum length of the vertical width.

5.1. Methodology

The development of the algorithm was focused on the grayscale DGGE images and only the already segmented lanes were processed in order to simplify testing.The first stage began by developing image visualization architecture, drawing structures and representing results (solutions). First, the image matrix and visualization through the screen was implemented. The drawing functions over the image, coordinates (pixels), and complete lines and meshes were then developed. In this way, the unit tests were easily performed. The second stage consisted of implementing the logical structure of the DGGE image processing algorithm, which first addressed node implementation followed by the mesh and the evolutionary algorithm (differential evolution). Regarding the node structure, it was based on a two-dimensional coordinate because it operates over a bidimensional image; a direction variable was also incorporated and measured in degrees. It was therefore possible to test node behavior. Node movement was implemented over the image by considering that it could not pass the margins or be located on its other nodes. Four states were also defined for the node; these are described in detail in the following sections. The mesh was the second structure that was created. This structure has a header referring to neighboring nodes from the same mesh. Every node belonging to the mesh has the same resistance and direction. The fitness function was defined according to the node states and a function to calculate its maximum vertical width (dispersion) in order to report its range to the other meshes. Finally, the differential evolution algorithmic structure was implemented. Its development was not overly difficult but some particular characteristics for the specific problems were added, which are specified in the following section.

5.1.1. Heuristics

The heuristics used in implementing the solution belong to the present research study. They were useful to narrow down the size of the problem, especially reducing the differential evolution algorithm randomness in mesh insertion and generation.

i) Select the range image differences: For mesh insertion into the image, the vertical minimum and maximum between pixels were calculated. In this way, node resistance was calculated as a value within the described range.
ii) Ignore occupied mesh positions: In mesh insertion, the positions occupied by the vertical width of active meshes in the image were ignored.
iii) Eliminate scattered nodes: When the nodes are fixed and highly scattered from the rest of the nodes in the mesh, they are eliminated.
iv) Insert empty positions by normal distribution: Mesh insertion into the empty image positions was performed by a normal distribution.

5.1.2. Mesh Generation

Mesh generation produced by the above heuristics is shown in Figure 6 .

Each mesh consisted of a set of nodes that was determined by a horizontal density parameter, a decimal value within the 0 to 1) range. This corresponded to the percentage of the horizontal area of the mesh to be occupied by the nodes, and which allowed calculating their number given the image?s horizontal length. In order to place the nodes within the horizontal range, its number was calculated and then equidistantly inserted from left to right.

Each netmesh at the time of insertion was randomly assigned a direction, allowing to it to move either 90ï¿$\frac{1}{2}$ (up) or 270ï¿$\frac{1}{2}$ (down). This was the direction of all its nodes. In addition, each mesh was assigned a resistance that depended on the heuristic range image differences, and which was assigned to all the nodes.

5.1.3. Adaptation of Differential Evolution to the Problem

Given that each mesh (solution) is positioned in a distinct space within the image, thus capturing a possible band, crossing characteristics did not provide adequate heuristics. It

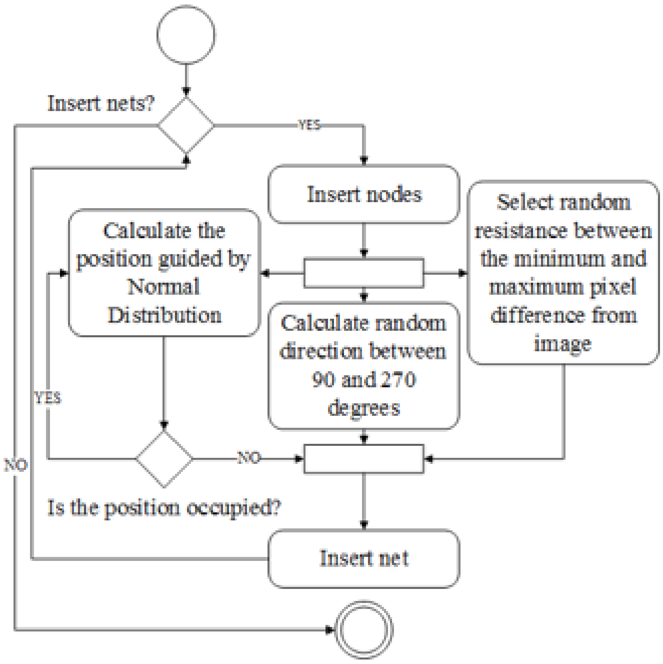

Figura 6. Shows the mesh generation process previously described

was therefore necessary to implement the evolutionary algorithm with a subtle difference in order to process and insert the new mesh generations. The crossing, which combined the characteristics of the best individuals, was rejected to generate new individuals in the next generation. This characteristic was replaced by a lifetime in which the mesh attempted to capture a band contour and improve fitness that is consistent with the adaptation concept.

5.2. Algorithm Structure and Functioning

The previously described proposal tried to use a connectionist model subjected to evolution by adapting the differential evolution algorithm to achieve band recognition with an automated sensitivity. This was achieved by adapting and selecting solutions generated by the algorithm. When the main structure of the study was introduced, it behaved like a deformable mesh with unidimensional movement (vertical). It consisted of many guide nodes, that is, a set of coordinates that share information from their environment to achieve vertical advanced behavior and capture the band contour. The main structures of the algorithm are described below.

5.2.1. Node

The node represents the basic mesh unit.It simulates an intelligent coordinate (on a given pixel), which communicates its state to other mesh nodes , and moves vertically depending on its resistance (see Figure 7).

The node states are:

- FREE (state 0): the node is free and it can continue to move through the image.

- READY (state 1): the node has captured a favorable contour according to its resistance.
- BLOCKED (state -1): the node cannot move because another node is blocking its advance. This state is not permanent and can return to the FREE state to the extent that the node that is blocking it advances.
- WASTE (state -2): is rubbish. The node is permanently blocked by another fixed node, (because the other node can be in the READY state), or has reached the limits of the image without capturing a contour

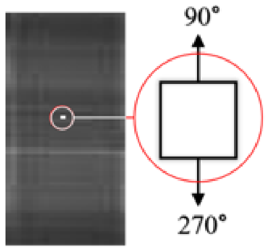

Figura 7. Node in a DGGE laned

The behavior of each node is illustrated in Figure 10.

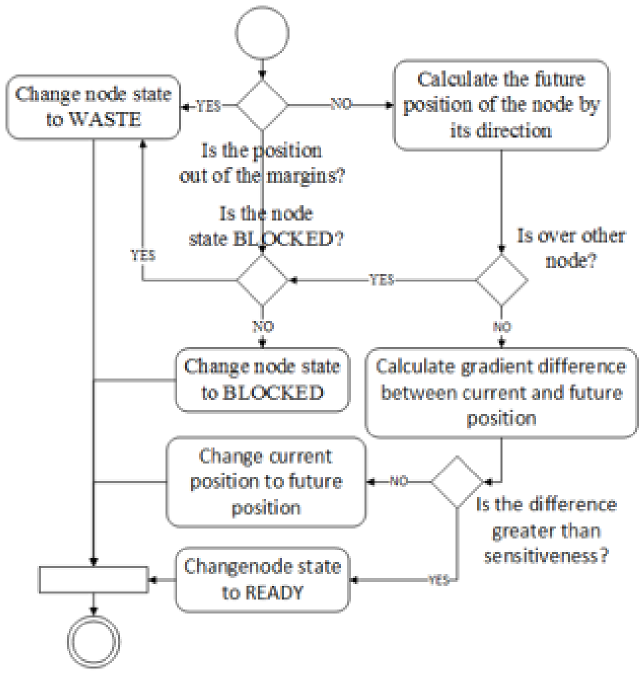

Figura 8. Node behavior diagram

Algorithmic implementation details are also established in the following pseudo-code:

5.2.2. Deformable Mesh

The structure was an ANN proposal based on a connectionist and deformable model, a mesh of nodes, which advanced performing vertical scans (90ï¿$\frac{1}{2}$ and 270ï¿$\frac{1}{2}$) on a

Algorithm 1 Node Behavior

1. Node defines four states:
```
FREE = 0;
READY =1;
BLOQUED =-1;
WASTE = -2;
```
2. Calculate grayscale value of future (next adjacent) vertical coordinate, depending of the node direction.
3. Verify if the value calculated is valid and isn't other net position
```
IF future position is invalid or is in other next positin
       IF node state is BLOCKED
             set node state to WASTE$
        ELSE
              set node state BLOCKED
        ENDIFELSE
ENDIF
```
4. Verify if the future position is an edge
```
IF difference between current pixel and future pixel
is greater than resistance
   set node state to READY
   set the future position to the current position
ENDIF
```

DGGE image lane. The mesh was a set of nodes arranged horizontally at a uniform distance. This structure handled information about its vertical width (vertical dispersion of nodes); although mesh and node insertion occurred on a perfect horizontal line (slope 0), nodes were dispersed when this behavior was executed as a consequence of the capture or blocking of some nodes before their peer nodes. This structure managed this information so the meshes were not interposed with each other, and a band contour was not captured again. The mesh had the same resistance for all nodesand a maximum dispersion measured in pixels that indicated maximum mesh width.

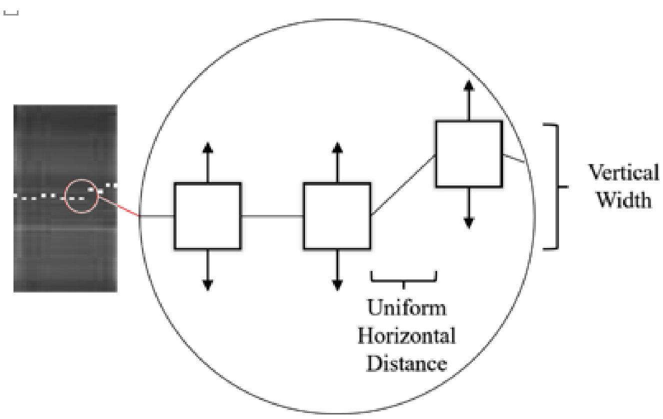

Figura 9. Deformable mesh

The mesh constantly verified if any of its nodes had exceeded maximum dispersion; if this was the case, it was removed.

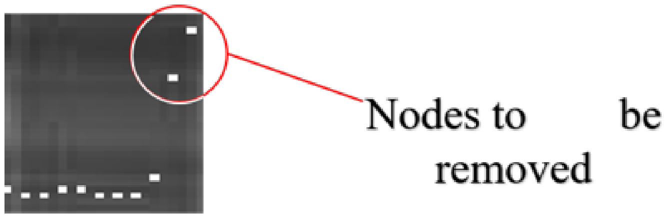

Figura 10. Node dispersion

The proposed solution defined that the mesh algorithm (ANN) should be subjected to evolution. It was therefore necessary to implement the fitness calculation within the structure, considering the key aspects of optimality that establish measurement parameters between the different meshes in an image. The main characteristics of the fitness calculation are defined below.

i) Dispersion score (vertical width): node dispersion is critical to recognize good capture because the band contour slope is very close to 0 in most cases; therefore, adequate dispersion should not exceed two pixels.

ii) Score by node state: the mesh is by definition a set of nodes; therefore, their state is the main measure of fitness. Each of the node state values are added to the fitness value. (FREE: 0, READY: 1, BLOCKED: -1, WASTE: -2).

iii) Empty mesh: extreme case. If the mesh does not contain any nodes, fitness becomes extremely negative and it is removed.

Therefore, the fitness score determines in the evolutionary algorithm which meshes survive and move on to the next generation.

The following pseudocode details the fitness calculation.

Algorithm 2 Fitness calculation

1. Verify if array of nets hasn't nodes
```
IF array of nets hasn't nodes
   return -255
ENDIF
```
2. Calculate score (fitness) of the nets
   i) Initialize fitness value in 0 and do:
```
FOREACH node in the net do
  add the state value of the node to the fitness value
  {FREE =0; READY = 1; BLOCKED = -1; WASTE=-2}
ENDFOREACH
```
   ii) Calculate the thickness score
```
IF  thickness is equal or lower than dispersion
    thickness score will dispersion allowed-thickness
ELSE
    thickness*-2
ENDIFELSE
```
   iii) Add the thickness score to the fitnes value

5.2.3. Differential Evolution

General steps and decisions of the algorithm are summarized in Figure 11

Initial heuristics and evolution through lifetime were considered essential to the implemented evolutionary algorithm, The insertion of each mesh generation was performed by heuristics that calculated the range of differences that allowed defining the resistance of each node and band capture in the image. Heuristics vertically scanned the image and calculated the maximum and minimum values of the differences between the current and future pixel. Replacement by crossing for lifetime allowed meshes to adapt and improve and then undergo selection without the need to mix their characteristics throughout generations. The flowchart (Figure 12) shows the steps of mesh selection after each generation.

The algorithm works with the following parameters:

i) Population density (vertical) (0-1): percentage of image vertical filling with meshes. For example, if the value is 1, the image is fully covered by meshes in all vertical positions; however, there is not a single mesh on the image if the value is 0.
ii) Node density (horizontal) (0-1): the same as in (i) but applied to the horizontal filling of nodes in the mesh.
iii) Mutation (0-1): probability of individuals with random resistance characteristics (outside the range calculated by heuristics) in each generation.
iv) Selection (0-1): percentage of the population of the current generation selected for the next generation.
v) Sensitivity (0-1): percentage of resistance range calculated by heuristics formesh insertion. When sensitivity is higher, capture occurs in more demarcated contours.

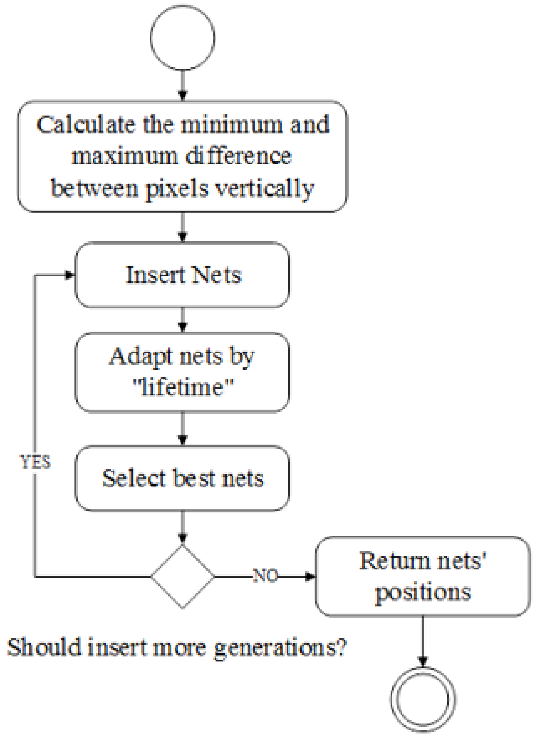

Figura 11. General diagram of differential evolution

vi) Dispersion (0-255): maximum value (in pixels) of node dispersion. The following pseudocode provides the details of the implementation of the evolutionary algorithm.

6. Experiments and Analysis of Results

To carry out the experiments, lanes segmented from DGGE images were used. This section includes the results of tests performed on a DGGE image with different preprocessing levels when applying the algorithm; results exhibited meshes with better fitness and capture. The tests were performed with a lane from an already segmented image to observe only the band detection behavior. All tests were run with a total of 999 generations. The image (see Figure 13) size was 22 pixels (width) by 221 pixels (height).

6.1. Lane without Preprocessing

The algorithm was executed with the following parameters and values:

- Population density (vertical): 0.1
- Node density (horizontal): 0.3
- Mutation: 0.1
- Selection: 0.4
- Sensitivity: 0.4

Figure 14 shows raw processing of a DGGE lane. Meshes captured enough band contours without error, but because the gradient in many band contours was very subtle, the meshes could not capture this contour. Therefore, the meshes failed in this case. Sensitivy was increased to 0.8; results are shown in Figure 15.

Increasing sensitivity to 0.8 produced greedy contour detection. Although it captured all the band contours, it also erred with noise gradients outside the bands.

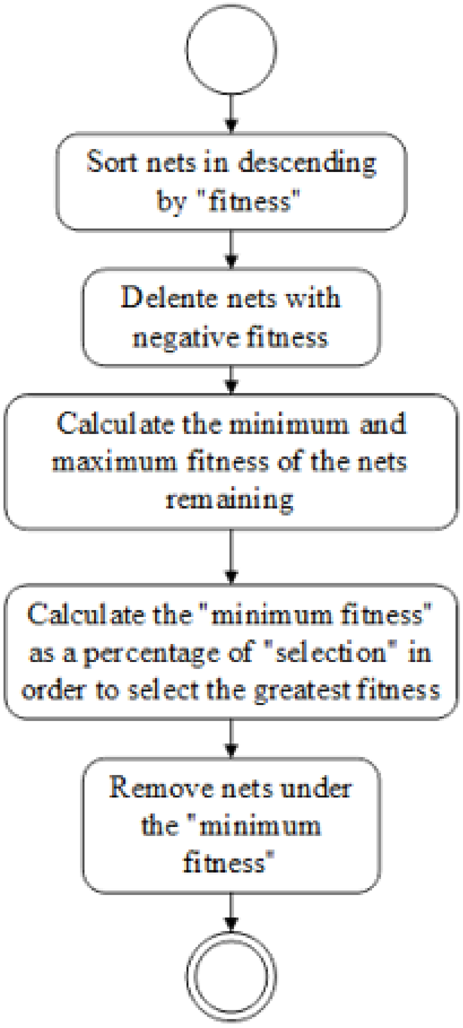

Figura 12. Net selection diagram

Figura 13. DGGE image lane used in the tests

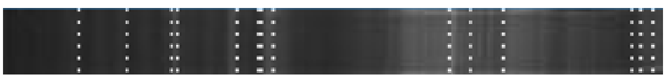

Figura 14. Lane without preprocessing after band detection with 0.6 sensitivity

Figura 15. Lane without preprocessing after band detection with 0.8 sensitivity

6.2. Laplacian Filter Lane

The Laplacian filter was applied as image preprocessing to see what would happen if the algorithm was executed after this. Results are shown below.

Algorithm 3 Differential evolution applied to DGGE

1. Calculate score (fitness) of the nets
```
popdensity: Percentage of vertical (0,1)
ndensity: Percentage of horizontal (0,1)
mutation : [0,1]
nets: Array of nets empty
```
2. Calculate the minimum an the maximum pixel vertical value difference of the image
3. Read hoy many generations should be generated, and to this:
```
FOREACH generation do
  Calculate net specific features.
  Calculate lifetime.
  Generate the new nets of the current generation.
  Move the nodes by direction.
  Fit the population
ENDFOREACH
```

The values for each of the algorithm parameters were as follows:

- Population density (vertical): 0.1
- Node density (horizontal): 0.4
- Mutation: 0.1
- Selection: 0.4
- Sens1itivity: 0.4

Figure 16 shows the results of processing after applying the Laplacian filter to a DGGE lane.

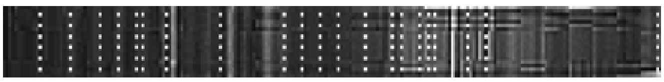

Figura 16. Lane with Laplacian filter after band detection

The filter caused the meshes to capture incorrect differences because the filter acted as deep contouring without eliminating noise. This tends to distort algorithm behavior.

6.3. Lane with Truncated Color Range

As a final preprocessing test on the DGGE lane in Fig. 13, attempt was made to truncate the color range to define and focus the processing of the algorithm and eliminate noise from the DGGE image. The information (grayscale) for each pixel in an image ranges from 0 to 255 bytes. Preprocessing tries to reduce this range to emphasize the DGGE image bands. The parameters and their values used for all executions are as follows:

- Population density (vertical): 0.1
- Node density (horizontal): 0.3

- Mutation: 0.1
- Selection: 0.4
- Sensitivity: 0.6

6.3.1. Truncated Range to 128 bytes

Figura 17. Lane truncated to 128 bytes after band detection

The results in Figure 17 are almost exactly the same as those in the lane with no processing. Therefore, truncating the color range in half is not very useful for this problem where images have low resolution and are very noisy.

6.3.2. Truncated Range to 64 bytes

Figura 18. Lane truncated to 64 bytes after band detection

The results observed in Figure 18 wereare better for detecting contours with noise and with a very smooth gradient.

6.3.3. Truncated Range to 32 bytes

Figura 19. Lane truncated to 32 bytes after band detection

Finally, the results graphed in Figure 19 after implementing a lane with the truncated color range to 32 bytes is a very good result when detecting the band contours previously detected in other executions. However, it also detects contours with very smooth gradients, which had been omitted in previous executions with other types of DGGE image preprocessing.

6.3.4. Artificial Image: Truncated Range to 32 bytes

As a final experiment, an artificial image was created to corroborate the effectiveness of the algorithm (see Figure 20).

The parameters and their values used for all executions are as follows:

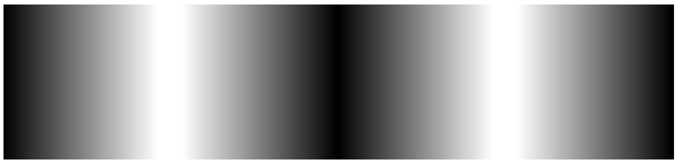

Figura 20. Artificial image truncated to 32 bytes

- Population density (vertical): 0.6
- Node density (horizontal): 0.4
- Mutation: 0.1
- Selection: 0.2
- Sensitivity: 0.1

The result of this experiment is displayed in Figure 21

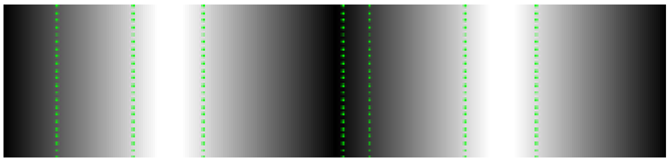

Figura 21. Artificial image truncated to 32 bytes after contour detection

7. Conclusions and Future Work

The execution of the algorithm in the image with no processing provides unfavorable results when there are high noise levels or very diffuse band gradients. Mesh behavior at the insertion level is performed by normal distribution, which makes the algorithm faster. However, it avoids important characteristics of the problem, which could serve as heuristics and generate mesh insertion in accurate positions by taking into account information related to different gradients. By applying preprocessing filters to the image, band contour recognition markedly improved, as well as mesh behavior on noisy areas and diffuse gradients. In particular, truncating the color range allowed lowering noise levels and obtaining good results. The algorithm in the early phase of the study shows appropriate performance with the help of preprocessing. In the near future, we intend to implement new heuristics related to mesh insertion and fitness providing greater intelligence to every node and increased communication with its environment (analysis of pixels in a radius). Current behavior is blind and only the mesh provides the flow of information between nodes that enables band contour recognition. There is much to do in the field of automation because it is a difficult and often unsystematic task to define appropriate parameters in evolutionary algorithm . Therefore, implementing previous analysis algorithms to adjust these parameters would greatly facilitate their usability and effectiveness.

8. Acknowledgments

This research is part of DUB 150919 3 / I project promoted by the General Directorate for Research, Development and Innovation of the Universidad del Bï¿½o Bï¿½o

Referências

Ademek, T. (2006). *Using contour Information and Segmentation for Object registration , Modeling and Retrieval*. PhD thesis, Dublin City Universtiy.

Benoit, L., Rousseau, D., Belin, É., Demilly, D., and Chapeau-Blondeau, F. (2014). Simulation of image acquisition in machine vision dedicated to seedling elongation to validate image processing root segmentation algorithms. *Computers and Electronics in Agriculture*, 104:84–92.

Cebron, A., Coci, M., Garnier, J., and Laanbroek, H. J. (2004). Denaturing gradient gel electrophoretic analysis of ammonia-oxidizing bacterial community structure in the lower seine river: Impact of paris wastewater effluents. *Applied and Environmental Microbiology*, 70(11):6726–6737.

Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum. Agronomy*, 35(3):271–276.

Dorigo, M. and Di Caro, G. (1999). The Ant Colony Optimization Meta-Heuristic. *New Ideas in Optimization*, 2:11–32.

Durand, N., El Sheikha, A. F., Suarez-Quiros, M.-L., Oscar, G.-R., Nganou, N. D., Fontana-Tachon, A., and Montet, D. (2013). Application of PCR-DGGE to the study of dynamics and biodiversity of yeasts and potentially OTA producing fungi during coffee processing. *Food Control*, 34(2):466–471.

Figueroa, C. (2012). *Identificación y Cuantificación de Bandas en Imágenes DGGE con Algoritmos Basados en Hormigas Identification and Quantification of bands in DGGE Images with algorithms based on Ants.* PhD thesis, Universidad de Concepción.

Figueroa, C., Gutiérrez, T., and Urrutia, H. (2009). *Indentificación y Cuantificación de Bandas en Imágenes DGGE*. PhD thesis, Universidad del Bío Bío.

Gárate, O. C., Pinninghoff, J. M. A., and Contreras, A. R. (2011). Band correction in random amplified polymorphism dna images using hybrid genetic algorithms with multilevel thresholding. In *Proceedings of the 4th International Conference on Interplay Between Natural and Artificial Computation: New Challenges on Bioinspired Applications - Volume Part II*, IWINAC'11, pages 426–435, Berlin, Heidelberg. Springer-Verlag.

González, H. and Gasull Llampallas, A. (2008). *Herramientas para la indexación de vídeo: extracción de imágenes relevantes y análisis de imágenes de agencia*. PhD thesis, Universitat Politècnica de Catalunya.

Graves, D. and Pedrycz, W. (2009). Multivariate Segmentation of Time Series with Differential Evolution. *IFSA/EUSFLAT Conf.*, pages 1108–1113.

He, S., Chen, J., Li, J., Xu, Y., and Yang, H. (2010). Dynamics of nitrobacteria community in biological contact oxidation process by DGGE and FISH. *2010 4th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2010*, (2007):0–3.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4, pp 321-331).

Kim, B., Lee, S.-E., Song, M.-Y., Choi, J.-H., Ahn, S.-M., Lee, K.-S., Cho, E., Chon, T.-S., and Koh, S.-C. (2008). Implementation of artificial neural networks (anns) to analysis of inter-taxa communities of benthic microorganisms and macroinvertebrates in a polluted stream. *The Science of the total environment*, 390(1):262â274.

Machado, A., Campos, M., and Siqueira, A. M. (1997). An Iterative Algorithm for Segmenting Lanes in Gel Electrophoresis Images. *Computer Graphics and Image Processing, 1997. Proceedings., X Brazilian Symposium on*, pages 140 – 146.

Mao, J. and Jain, A. K. (1992). Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recogn.*, 25(2):173–188.

Muyzer, G. and Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (dgge) and temperature gradient gel electrophoresis (tgge) in microbial ecology. *Antonie van Leeuwenhoek*, 73(1):127–141.

Nicolaisen, M. H. and Ramsing, N. B. (2002). Denaturing gradient gel electrophoresis (dgge) approaches to study the diversity of ammonia-oxidizing bacteria. *Journal of Microbiological Methods*, 50(2):189 – 203.

Noktehdan, A., Karimi, B., and Husseinzadeh Kashan, A. (2010). A differential evolution algorithm for the manufacturing cell formation problem using group based operators. *Expert Syst. Appl.*, 37(7):4822–4829.

Novo, J., Penedo, M. G., and Santos, J. (2011). Optimization of Topological Active Nets with Differential Evolution. *Advances in Computational Intelligence*, 669:pp 282–290.

Novo, J., Santos, J., and Penedo, M. G. (2012). Topological active models optimization with differential evolution. *Expert Syst. Appl.*, 39(15):12165–12176.

Omran, M. G. H., Engelbrecht, A. P., and Salman, A. (2009). Bare bones differential evolution. *European Journal of Operational Research*, 196(1):128–139.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

Pinninghoff, M. A., Figueroa, C., and Urrutia, H. (2012). Using Ant Colony Systems for Edge Detection in DGGE Images. In Herons, B., editor, *Scientific Computing, Communicability and Cultural Heritage: Future Trends in Software and Interactive Design*, pages 127–133, Venecia, Italia.

Pinninghoff, M. A., Valenzuela, M., Contreras, R., and Mora, M. (2014). Automatic Lane Correction in DGGE Images by Using Hybrid Genetic Algorithms. In *Hybrid Artificial Intelligence Systems. 9th International Conference*, pages 221–232. Springer International Publishing, Salamanca, Spain.

Prabhakar, A. and Bishop, A. H. (2014). Comparative studies to assess bacterial communities on the clover phylloplane using mlst, dgge and t-rflp. *World Journal of Microbiology and Biotechnology*, 30(1):153–161.

Quantity-One (2016). Quantity-One.

Rathod, V. R., Alaknanda, and Anand, R. S. (2010). Comparison of different segmentation techniques in detection of flaws in weldments. In *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, volume 2, pages 673–677.

Salman, A., Engelbrecht, A. P., and Omran, M. G. (2007). Empirical analysis of self-adaptive differential evolution. *European Journal of operational research*, 183(2):785–804.

Sierra, C. V., Novo, J., Santos, J., and Penedo, M. G. (2013). Emergent segmentation of topological active nets by means of evolutionary obtained artificial neural networks. In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,*, pages 44–50.

Statsoft (2014). Statsoft.

Storn, R. and Price, K. (1995). Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces. pages 1–12.

Tsumiyama, K. and Yamamoto, K. (1989). Active net: Active net model for region extraction. *IPSJ SIG notes*, pages 89(96):1–8.

Ye, X., Suen, C. Y., Cheriet, M., Wang, E., Gm, S., West, D. M. B., Hg, Q., Supérieure, É. D. T., Université, D., Road, C. S.-c., and Ht, Q. (1999). A Recent Development in Image Analysis of Electrophoresis Gels. pages 19–21.

Zeng, Y. (1999). Perceptual Segmentation Algorithm and Its Application To Image Coding. *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, 2:820 – 824.