# Mammographic Images Segmentation using Superpixel

Glenda Botelho Computer Science Department Tocantins Federal University 109-Norte, Av. NS-15, ALCNO 14, Palmas, Tocantins, Brazil Alexandre Tadeu Rossini Computer Science Department Tocantins Federal University 109-Norte, Av. NS-15, ALCNO 14, Palmas, Tocantins, Brazil Ary Henrique M. Oliveira Computer Science Department Tocantins Federal University 109-Norte, Av. NS-15, ALCNO 14, Palmas, Tocantins, Brazil

### ABSTRACT

The images segmentation is a very important step for the area of image analysis and has the objective of separating regions of an image according to the objects represented in it. However, this is a challenging step in image processing area, since most traditional segmentation techniques have a high computational cost, which difficult their application in high resolution images. Based in this context, this paper proposes the use of the superpixels extraction technique, known as Speeded-up Turbo Pixels, to segment high resolution mammographic images with satisfactory processing time.

#### **General Terms**

Image Analysis, Superpixel, Health Informatics, Medical Image Processing.

#### **Keywords**

Image Segmentation, Superpixel, Mammographic Image, Medical Image Processing, Heath Informatics.

# 1. INTRODUCTION

The image analysis process consists of extracting characteristics or information from an image by automatic or semi-automatic methods [1]. In general, the first step in the image analysis corresponds to the segmentation process, which consists of dividing the image into regions or objects. Due to the enormous diversity of applications, there are different methods of digital images segmentation that are usually classified in [2, 3]: thresholding, edge-based segmentation and region-based segmentation. Some of the main segmentation methods have a high computational cost and, consequently, are not suitable for application in high resolution images. However, there are techniques that perform image preprocessing through pixel grouping. It is known that the grouping of data is highly related to the images segmentation, since the common objective is to assign a label to each object (or pixel, in the case of images), such that similar objects have the same label [4]. Therefore, it is possible to consider grouping methods as a way of identifying objects in images [5].

The possibility of easy storage and acquisition of images in the digital format due to the presence of low cost systems favored the evolution of several activities within the image processing area. The researches for more accurate and lower computational methods have become increasingly interesting and useful for different situations involving images. Segmentation techniques are currently used to contribute to the resolution of different problems that require methods capable of simplifying the visual information for analysis and interpretation. In medicine, for example, image analysis is strongly present in the medical diagnostic phase, helping in the early pathologies detection and obtaining more accurate diagnoses. With the improvement of acquisition techniques of medical image, such as magnetic resonance, the capacity and fidelity of the diagnostic by image were amplified. A medical image database can benefit from processing techniques to classify images and enable pattern recognition. The images segmentation is an efficient area used in this application type, because it allows isolating regions in the image that have characteristics in common, helping to classify them according to the structures that compose them. For example, superpixels extraction techniques, based in k-means clustering algorithm, can be used to segment appropriately medical images with low computational cost [6, 7, 8].

The superpixels extraction technique for the image segmentation of digital mammography allows delimiting areas that may contain possible nodules, contributing to the diagnosis, even the disease being in its initial state. According to National Cancer Institute (http://www.inca.gov.br), breast cancer is the most common cancer type among women in the world and in Brazil, after non-melanoma skin, accounting for about 28% of new cases each year. According to the World Health Organization (WHO), in the 1960s and 1970s there was a significant increase of 10 times in age-adjusted incidence rates in population-based cancer registries from several continents [9].

To aid in the early detection of breast cancer, this paper presents the application of the Speeded-up Turbo Pixel (STP) algorithm [6], based on superpixels extraction, for the segmentation of mammography images, making possible the nodules detection. In this paper, the images are divided into groups of pixels with low computational cost, considering the application in high resolution images. In addition, the definition of adequate values for the parameters used in the STP technique was performed, which allowing the detection of nodules in digital mammography images, contributing to a precise and rapid diagnosis.

# 2. MATERIALS AND METHOD

Speeded-up Turbo Pixels (STP) algorithm is technique for superpixels generation based on K-means grouping algorithm [6]. This algorithm generates uniforms superpixels with low computational cost. Initially, the image is divided in rectangular regions according to desired pixel number. Thus, each superpixel has the rectangular form with equidistant centroids. This is considered the initial step for the superpixels extraction. Then, the pixels at the segment edges are attributed to new segments, by minimization of the following cost function.

$$C_{x,y}(i) = \lambda_1 * |I(x,y) - I_i| + \lambda_2 * |(x - C_x^i)^2 + (y - C_y^i)^2|$$

In the equation,  $I_i$  indicates the mean intensity of the segment *i*, x and y are coordinates of the pixels tested between different segments and  $C_x^i$  and  $C_y^i$  are centroids do segment *i*.

Values of  $\lambda_1$  and  $\lambda_2$  correspond respectively to coefficient of similarity and convexity constraint. In the equation, the first term ensures intensity similarity between the pixels that will be incorporated, while the second term ensures that the superpixels will have convex shapes, preventing distant pixels being embedded. The superpixels convexity can be changed by modifying the parameter value  $\lambda_2$ , generating more convex superpixels, that fit well to the objects edges of the images.

This technique provides a fast algorithm, given that a pixels limited number is tested between neighboring segments (consider the 4-connected neighborhood). As soon as all edges are tested, the mean intensities and centroids are updated considering the new pixels grouped or removed. The iterations number of this algorithm is limited and defined according to the application. In this work, this algorithm was used as a final segmentation technique in mammography images, but it can also be used as a pre-segmentation technique.

Figure 1 presents the pixels update process in each iteration of the STP algorithm. The algorithm receives a high-resolution image (superior than 1000x1000) as input (Figure 1(a)). Then, a regular grid is generated by marking the pixels group in the image according to the desired pixels number (Figure 1(b)). After this process, the pixels at the segment edges are assigned to new segments by minimizing the cost function proposed by the STP algorithm. Finally, after the execution of each iteration, the superpixels will converge even better to fit the objects edges of the image (Figure 1(c)).

#### 3. RESULTS

Experiments were performed with the objective of evaluating the most important aspect of the segmentation approach proposed in this work, that is, the commitment to obtain a satisfactory segmentation of high resolution mammographic images in a computationally efficient way, to allow the detection of possible nodules and to collaborate with the medical diagnosis. The following subsections present the definition of the parameter values used in the STP algorithm and the experiments performed on the mammographic images

#### 3.1 Parameter Values

During the experiments to segment the mammographic images, it was observed that the results quality depended exclusively of the images quality and the values passed as parameters. The experiments demonstrated that the determinant parameters are the values of  $\lambda_2$  and the initial size of the superpixels. So, in the experiments, the values passed as parameters were changed so that different results were obtained. Two experiments sets were performed with different parameters, as presented in the following subsection.

#### **3.2 Experiments**

In the first experiments set, a high-resolution mammographic image was used, as shown in Figure 2(a). This image presents a nodule, has a size equal to  $1496 \times 2003$  and was segmented with the parameter values shown in Table 1.

For the expected result to present high precision in the segmentation, it is necessary the correct convergence of the superpixels with the edges of the interest objects of the image, that is, the nodule. As shown in Table 1, only the value of the superpixels initial size parameter was changed, maintaining the values of the other static parameters. The result in Figure 2(b) shows that the superpixels converged well with the edges of the interest object (nodule) and maintained uniform relatively sizes. In addition, a small value for the superpixels

initial size (10x10) generated satisfactory result because the superpixels converged well and the processing time was low (25 seconds).



(a) Original image



(b) Regular grid definition



(c) Segmented image

Fig 1: Superpixels generation process using Speed-Up Turbo Pixels algorithm

International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 11, August 2018

experiments set.							
Figure	Size	$\lambda_{I}$	$\lambda_2$	Iteration	Processing time (sec.)		
2(b)	10	0.1	0.001	20	25		
2(c)	20	0.1	0.001	20	20		
2(d)	50	0.1	0.001	20	17		
2(e)	70	0.1	0.001	20	17		

Table 1. Parameters used in the first segmentation

Figure 2(c) shows superpixels of size 20x20. In this figure a segmentation improvement is observed, as well as the reduction of the processing time (20 seconds). In Figure 2(d), the superpixels size is 50x50 and continues to fit well to the nodule edges and the processing time has also decreased (17 seconds). Finally, superpixel size of 70x70 was analyzed and the results improved even more, with convex superpixels delimiting perfectly to the edge around the nodule, with low computational cost.



(d) Superpixel size 50x50

(e) Superpixel size 70x70

#### Fig 2: Application of the STP algorithm in mammographic image with nodule, varying the superpixels size.

In this second experiment was used a mammographic image with size equal to 1024x1024 and the same parameters values of the previous experiment, modifying only the superpixels initial size. In this experiment, the image does not have a nodule, however this image type is tested to show that the superpixels maintain a uniform relatively size, without significant discrepancies in the superpixels format, that could indicate a lesion. Table 2 presents the parameters values used in this experiment and Figure 3 shows the results.

Table 2. Parameters used in the second segmentation experiments set

experiments set.							
Figure	Size	$\lambda_l$	$\lambda_2$	Iteration	Processing time (sec.)		
3(b)	10	0.1	0.001	20	11		
3(c)	50	0.1	0.001	20	6		
3(d)	70	0.1	0.001	20	4		

International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 11, August 2018



(a) Original image

(b) Superpixel size 10x10

(c) Superpixel size 20x20

(d) Superpixel size 50x50

#### Fig 3: Application of the STP algorithm in mammographic image without lesion, varying the superpixels size.

A third experiments set was performed with the parameter values presented in Table 3. Figures 4 and 5 present the segmentation of mammographic images with lesions and Figure 6 shows the segmentation in a mammographic image that does not have any lesion. In this experiments can be observed that superpixels adjusted well the edge of the interest object (lesions) in the mammographic images. In mammographic images without nodules, the superpixels remained convex, maintaining a relatively regular and uniform size.

 Table 3. Parameters used in the third segmentation experiments set.

Figure	Size	$\lambda_{I}$	$\lambda_2$	Iteration	Processing time (sec.)
4	70	0.1	0.001	20	12
5	20	0.5	0.02	20	6
6	20	0.5	0.06	20	6



(a) Original image

(b) Super pixel 70x70

Fig 4: Segmentation results in a mammographic image with lesions.



(a) Original image



(b) Superpixel 20x20

Fig 5: Segmentation results in a mammographic image with lesions.



(a) Original image



(b) Superpixel 20x20

Fig. 6: Segmentation results in a mammography image without lesions.

#### 4. CONCLUSIONS

The results demonstrated that the Speeded-up Turbo Pixels technique performs the segmentation of high resolution mammographic images with low processing time and satisfactory results, because it allows the detection of possible nodules, that is, the adequate identification of the lesions edges. In addition, results showed that the parameters that most influence in the segmentation are the  $\lambda_2$  and the superpixels initial size. For example, for mammographic images, as soon as the superpixels size increases, the

computational cost decreases, but maintain the segmentation quality, that is, the superpixels continue to adjust the objects edges properly.

As future work, the results obtained from the segmentation based on superpixels can be used together with classification techniques of the Artificial Intelligence area to automatically differentiate mammographic images with and without lesions, aiding in the early diagnosis of breast cancer.

#### 5. ACKNOWLEDGMENTS

Our thanks to the Federal University of Tocantins for the financing of the execution of this work and related researches.

#### 6. REFERENCES

- [1] Gonzalez, R. C. and Woods, R. E. 2002. Digital image processing. 2<sup>nd</sup> Edition: Prentice Hall.
- [2] Fu, K. S. and Mui, J. K. 1981. A survey on image segmentation. Pattern recognition, 13(1), pp. 3-16.
- [3] Gonzalez, R. C. and Woods, R. E. 2010. Digital image processing. Pearson Prentice Hall.
- [4] Theodoridis, S. 2003. Konstantinos koutroumbas pattern recognition. Elsevier.
- [5] Cuadros, O., Botelho, G., Rodrigues, F. and Neto, J. B. Segmentation of large images with complex networks. IEEE Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference; 2012: 24-31.
- [6] Cigla, C. and Alatan, A. A. Eficient graph-based image segmentation via speeded-up turbo pixels. In Image Processing (ICIP), 2010 17th IEEE International Conference; 2010: 3013-16.
- [7] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and S. Susstrunk. SLIC superpixels compared to state-of-the-Art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):2274–2282, 2012.
- [8] Linares, O. C., Botelho, G. M., Rodrigues, F. and Batista Neto, J. Segmentation of large images based on superpixels and community detection in graphs. IET Image Processing, v. 11, p. 1219-1228, 2017.
- [9] World Health Organization (WHO) [homepage on the Internet]. World health statistics annual: 1997-1999.
   WHO; 2001. Available from: http://www.who.int/en/