Large cells cancer volumetry in chest computed tomography pulmonary images

Large cells cancer volumetry in chest computed tomography pulmonary images

Y Huérfano[1], M Vera[2], E Gelvez-Almeida[2], M I Vera[1], O Valbuena[2], and J Salazar-Torres[2]
[1] Facultad de Ciencias, Universidad de Los Andes, San Cristobal, Venezuela
[2] Facultad de Ciencias Básicas y Biomédicas, Universidad Simón Bolívar, San José de Cúcuta, Colombia

E-mail: m.avera@unisimonbolivar.edu.co

Abstract. Lung cancer is the leading oncological cause of death in the world. As for carcinomas, they represent between 90% and 95% of lung cancers; among them, non-small cell lung cancer is the most common type and the large cell carcinoma, the pathology on which this research focuses, is usually detected with the computed tomography images of the thorax. These images have three big problems: noise, artifacts and low contrast. The volume of the large cell carcinoma is obtained from the segmentations of the cancerous tumor generated, in a semi-automatic way, by a computational strategy based on a combination of algorithms that, in order to address the aforementioned problems, considers median and gradient magnitude filters and an unsupervised grouping technique for generating the large cell carcinoma morphology. The results of high correlation between the semi-automatic segmentations and the manual ones, drawn up by a pulmonologist, allow us to infer the excellent performance of the proposed technique. This technique can be useful in the detection and monitoring of large cell carcinoma and if it is considering this kind of computational strategy, medical specialists can establish the clinic or surgical actions oriented to address this pulmonary pathology.

1. Introduction

In the human beings, the mammary glands are two anatomical structures localized on anterior and superior thorax's face [1]. The benign pathologies in this kind of glands deepen of the patient symptoms such as tumor, secretion, pain or shape mamma alterations [2]. On the other hand, the breast cancer is the accelerated and uncontrolled growing of cells in the glandular epithelium, which link it with the normal lobules and shot lobules [2].

In multi-layer computed tomography (MSCT) images, the presence of noise, artifacts and low contrast is a problem. Additionally, the three-dimensional (3D) segmentation of the lungs is a problem of great interest because: (i) Lung cancer is the leading oncological cause of death in the world [1] and (ii) the symptomatology that it produces appears late, once its existence is suspected, it is obligatory to request the computed tomography of the thorax as an initial imaging study [2].

However, MSCT images have very challenging imperfections such as Poisson noise, low contrast and artefacts that affect the images quality. The stair-step artifact is the most common artifact presents in tomography images. It is produced by an overlap of the images during the reconstruction process [3]. When the stair-step artifact is present in the MSCT images it can be visualized because it produces abrupt alterations in the continuity of the contours that delimit one or more objects present in such

images [3]. This type of artifact dramatically deteriorates the appearance of such objects and can affect the dimensions (diameter, surface or volume) especially of small structures present in the images acquired under this modality [4]. For example, Figure 1 shows a typical artifact in a chest MSCT image.

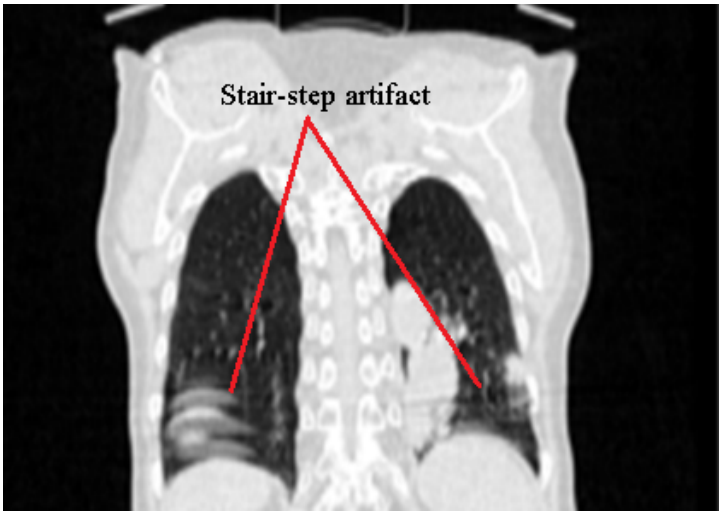


Figure 1. Stair-step artifact presents in MSCT images.

On the other hand, several works related to the segmentation of lung tumor, which are presented at next. In this sense, Kubota, *et. al.* report a computational method that uses morphological and diffusion operators for the segmentation of pulmonary nodules. They report an average Dice score (Ds) of 0.71 [5].

Yang, *et. al.* proposes lung tumor segmentation based on the coincidence of templates and growth of regions with three-dimensional positron emission tomography images. The proposed method obtained a Ds of 0.90 [6].

Ait, *et. al.* proposes a segmentation of pulmonary in computed tomography images using the convolutional neuronal network architecture (U-net). The experimental results show a precise segmentation with a Ds of 0.95 [7].

In this paper, the main purpose is to generate a semi-automatic technique (SAT) for the segmentation of the large cell lung carcinoma (LCLC) in computed tomography images. This technique is based on the application of a filter bank (median and gradient magnitude filters) and region growing technique in order to generate the 3D LCLC morphology. The filter bank is necessary for address the problems of the aforementioned imperfections of MSCT images. This technique can be useful in the detection and monitoring LCLC with the objective of establishes the clinic or surgical actions oriented to address this type of pathology.

2. Materials and methods

2.1. Dataset

A three-dimensional MSCT database was used, which has a voxel number of 512 x 512 x 95. Manual segmentation of the LCLC by a pulmonologist is also available.

2.2. Computational strategy proposed

In Figure 2, a schematic diagram is presented. It synthesizes the computational algorithms that make up the semi-automatic technique for the segmentation of the LCLC. In this figure, the lung MSCT dataset matches with the dataset described in the 2.1 section of this paper. The other elements shown in Figure 1 will be explained in the sections 2.2.1 and 2.2.2.

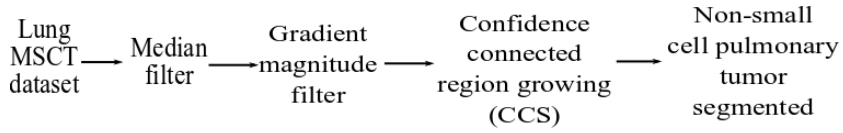


Figure 2. Block diagram of the proposed strategy called SAT.

2.2.1. Pre-processing. The main steps of this stage are: Median filter: A median image (M_I) is obtained processing each original image (O_I) with a median filter [8]. The role of this filter is to address the noise present in the images. The tuning parameter of this filter is the size of the neighborhood.

Gradient magnitude filter (GMF): In this work, an approach based on finite differences was used for GMF computational implementation [9]. This filter generates a smoothed version, called GM_I, calculating the three-dimensional gradient magnitude of M_I, using the mathematical model given by Equation (1).

$$\mathrm{GM_I} = \left(\left(\frac{\partial \mathrm{M_I}}{\partial \mathrm{i}}\right)^2 + \left(\frac{\partial \mathrm{M_I}}{\partial \mathrm{j}}\right)^2 + \left(\frac{\partial \mathrm{M_I}}{\partial \mathrm{k}}\right)^2\right)^{1/2}, \tag{1}$$

being: i, j, k the spatial directions in which the gradient is calculated and $\left(\frac{\partial \mathrm{M_I}}{\partial \mathrm{i}}, \frac{\partial \mathrm{M_I}}{\partial \mathrm{j}}, \frac{\partial \mathrm{M_I}}{\partial \mathrm{k}}\right)$ the partial derivatives of M_I. It is important to note that the computational cost (time of calculations) of the continuous model for GMF, given by the Equation (1), is very expensive. By this reason an approach based on central finite differences is used, in this paper, for modelling computationally the GMF [10].

2.2.2. Segmentation. In this stage the region growing (RG) technique is applied. The RG partitions an image (f) into regions (Ri) whose voxels are connected according to certain predefined criteria based on connectivity and similarity of the image. The most popular predefined criterion is given by Equation (2) [11]. The RG needs a seed voxel into an initial neighborhood (Iv).

$$\left|\mathrm{F}(i, \mathrm{j}, \mathrm{k}) - \mu_{\mathrm{R_i}}\right| < \mathrm{km}, \tag{2}$$

being: f(i, j, k) the gray levels of Iv, $\mu_{\mathrm{R_i}}$ the average gray levels of a Iv of arbitrary shape and size, m an arbitrary scalar and $\sigma_{\mathrm{R_i}}$ the standard deviation of an arbitrary neighborhood of Iv. The Equation (2) represents the most popular predefined criterion due its simplicity and effectiveness.

The RG tuning parameters are the initial neighborhood size i and k parameter that controls the amplitude of the range of intensities considered to accept or reject a voxel in a region. Such parameters must undergo a tuning process.

During the tuning process, the LCLC segmented is compared with the manual segmentation traced by a pulmonologist. The Ds is used in order to estimate the difference or matching between these structures [12].

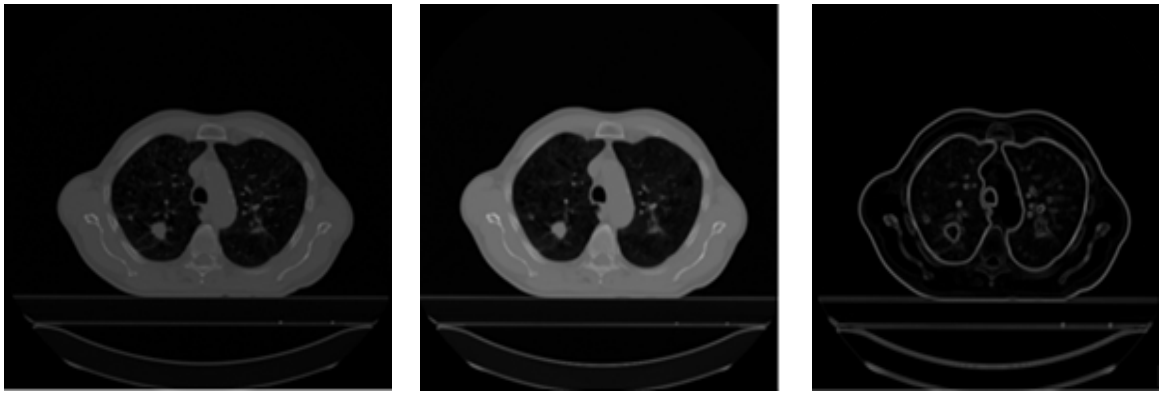
3. Results

First, we present qualitative results using original and preprocessed images in two-dimensional (2D) and 3D views. Then the quantitative results are shown considering mainly the Dice score and optimal parameters obtained by the algorithms present in the proposed SAT.

3.1. Qualitative results

Figure 3 shows a two-dimensional (2D) view of both the original image and the processed images after applying the proposed technique to the dataset considered. In Figure 3, we can observe the effects generated by filter bank application. A qualitative analysis revels that the considered filters generated

images of high quality reducing the noise and producing a high contrast between the anatomical structures. In this figure, we can observe the excellent performance of gradient magnitude filter for generating the contours of aforementioned structures. Particularly, the median filter acts to produce an image with a higher contrast compared to the original image (see Figure 3(a) and Figure 3(b)); while, in the Figure 3(c), we can see the image borders clearly defined as a consequence of having applied the gradient magnitude filter.



(a) (b) (c)

Figure 3. Preprocessing stage results; (a) original image, (b) median image, (c) gradient magnitude image.

On the other hand, Figure 4 illustrates the results of the segmentation process developed using region growing technique. In this figure, we can observe an excellent representation of segmented LCLC morphology, in both 2D and 3D views.

(a) (b)

Figure 4. LCLC segmented: (a) 2D view, (b) 3D view.

3.2. Quantitative results

When the tuning process was performed the mammary tumor was characterized considering its volume. So, the volume occupied by the automatically segmented LCLC was 1.06 cm^3; while the volume, reported by the clinical expert, (obtained considering the manual segmentation) was 1.10 cm^3.

The percentage relative error, considering these volumes, was 3.63% and the maximum Dice score generated was 0.89. These results allowed establishing the optimal parameters of the computational algorithms that make up the proposed technique. They were: (i) Median filter: the size of the kernel was corresponding with (3x3x3). (ii) Region growing: r = 1 and k = 2.5.

Additionally, the Table 1 shows comparative information about the Dice score obtained both in this paper and in others researches, reported in the specialized literature. Here, it is necessary pointed that the Ds is a metric with values between zero and 1. This metric is better when its value is closest to 1 [12]. In a medical image segmentation context, this means that the manual segmentation and the

automatic one matching when the Ds is 1 and they no matching at all when the Ds is zero. In this sense, normally, values of Ds over 0.75 are accepted, in the medical routine.

Table 1. Comparison using Dice score values for LCLC segmentation**.**

Authors	Technique	Dice score
Kubota, *et. al.* [5]	Morphological and diffusion operators	0.71
Yang, *et. al.* [6]	Region growing	0.90
Ait, *et. al.* [7]	Deep learning (U-net)	0.95
Huérfano, *et. al.* (Our technique)	SAT	0.89

The information analysis presented in Table 1, let us to affirm that SAT is outperform by the results reported by Yang, *et. al.* [6] and Ait, *et. al.* [7]; whereas the computational technique proposes in this paper generated a higher value for the Ds than that reported by Kubota, *et. al.* [5].

It is important to notice that the technique mentioned in [7] is based on smart operators that belong to the deep learning context. According with the literature, these techniques exhibit an excessive computational time during the training stage; while our technique is very efficient and let us obtain results with a similar precision that the deep learning techniques when it is intend performing the LCLC segmentation.

Additionally, we can observe that the result reported in [6] is comparable with our result considering the Dice score. However, the Ds value reported by us outperforms the Kubota, *et. al.* results.

4. Conclusions

A semi-automatic technique, available for detecting a LCLC in a precise and efficient manner, has been presented. This technique was based in the application of median and gradient magnitude filters in order to address the noise problem and the region growing technique for generating the 3D LCLC morphology, presents in MSCT images.

The three-dimensional representation of LCLC is useful for the detection and monitoring of pulmonary diseases; as well as for the planning of medical treatments and clinical-surgical procedures linked to this pathology.

On the other hand, it is expected that the segmentation generated by the proposed method can be useful to promote, deepen and potentiate the study of the real anatomy of the structures linked to the LCLC.

In the immediate future it is planned to validate the proposed technique with a significant number of databases in order to estimate the robustness of the aforementioned technique.

References

[1] Latarjet M and Ruiz A 2004 *Anatomía humana* (Barcelona: Médica Panamericana)
[2] Webb W and Higgins C 2005 *Thoracic imaging: pulmonary and cardiovascular radiology* (Philadelphia: Lippincott Williams and Wilkins)
[3] Barrett J and Keat N 2004 Artifacts in CT: Recognition and avoidance *Radiographics* **24(6)** 1679
[4] Wang G and Vannier M 1994 Stair–step artifacts in three-dimensional helical ct: An experimental study *Radiology* **191** 79
[5] Kubota T, Jerebko A, Dewan M, Salganicoff M and Krishnan A 2011 Density and attachment diagnostic CT pulmonary nodule segmentation with competition-diffusion and new morphological operators *Multi Modality State of the Art Medical Image Segmentation and Registration Methodologies* ed El-Baz A, Acharya U R, Mirmehdi M and Suri J (Boston: Springer) Chapter 6 p 143
[6] Yang B, Xiang D, Yu F and Chen X 2018 Lung tumor segmentation based on the multi-scale template matching and region growing *Society of Photo-Optical Instrumentation Engineers (SPIE) Medical Imaging* (Houston) (Houston: SPIE) p. **10578**
[7] Ait B, El Hassani A and Majda A 2018 Lung ct image segmentation using deep neural networks *Procedia Computer Science* **127** p 109
[8] Petrou M y Bosdogianni P 2003 *Image processing the fundamentals* (New York: John Wiley & Sons Inc)

[9] Pratt W 2007 *Digital image processing* (New York: John Wiley & Sons Inc)
[10] Burden R and Faires D 2010 *Numerical analysis* (Mexico: Cengage Learning)
[11] Huérfano Y, Vera M, Mar A and Bravo A 2019 Integrating a gradient–based difference operator with machine learning techniques in right heart segmentation *Journal of Physics: Conference Series* **1160** 012003
[12] Dice L 1945 Measures of the amount of ecologic association between species *Ecology* **26(3)** 29