

PAPER • OPEN ACCESS

Pulmonary adenocarcinoma characterization using computed tomography images

To cite this article: Y Huérfano *et al* 2019 *J. Phys.: Conf. Ser.* **1408** 012004

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Pulmonary adenocarcinoma characterization using computed tomography images

Y Huérfano¹, M Vera², M I Vera¹, O Valbuena², E Gelvez-Almeida², and J Salazar-Torres²

¹ Grupo de Investigación en Procesamiento Computacional de Datos, Universidad de Los Andes, San Cristóbal, Venezuela

² Facultad de Ciencias Básicas y Biomédicas, Universidad Simón Bolívar, San José de Cúcuta, Colombia

E-mail: yoleidyfismat@gmail.com

Abstract. Lung cancer is one of the pathologies that sensitively affects the health of human beings. Particularly, the pathology called pulmonary adenocarcinoma represents 25% of all lung cancers. In this research, we propose a semiautomatic technique for the characterization of a tumor (adenocarcinoma type), present in a three-dimensional pulmonary computed tomography dataset. Following the basic scheme of digital image processing, first, a bank of smoothing filters and edge detectors is applied allowing the adequate preprocessing over the dataset images. Then, clustering methods are used for obtaining the tumor morphology. The relative percentage error and the accuracy rate were the metrics considered to determine the performance of the proposed technique. The values obtained from the metrics used reflect an excellent correlation between the morphology of the tumor, generated manually by a pneumologist and the values obtained by the proposed technique. In the clinical and surgical contexts, the characterization of the detected lung tumor is made in terms of volume occupied by the tumor and it allows the monitoring of this disease as well as the activation of the respective protocols for its approach.

1. Introduction

The lungs are very spongy and conical organs, they are responsible for breathing and they are located inside the rib cage. Its mission is to capture oxygen from the outside to incorporate it into the blood and eliminate carbon dioxide, a waste product of the body's cells. They are separated from each other by the mediastinum, an area that contains the heart with the large blood vessels, the trachea, the esophagus and the lymph nodes [1].

The world health organization (WHO) has ranked lung diseases among the leading causes of death in the world. Within lung diseases, lung cancer is present. Cancer is an uncontrolled growth and spread of abnormal cells in the body that invade and damage tissues and organs. When some agents called carcinogens (such as, for example, tobacco) act on the body, causing damage to the genes of the healthy cell. As a result, it can modify the mechanism of cell growth and function, causing cancerous cells.

Lung cancer is classified into two main categories: non-small cell lung cancer, which is more frequent and is subdivided into three types: squamous cell carcinoma, adenocarcinoma and large cell carcinomas; and small cell lung cancer (also called small cell cancer) whose cells look like oat grains when viewed under a microscope, grow rapidly and similarly spread to other organs [2,3].



Lung cancer, specifically adenocarcinoma, begins at the periphery of the lungs and under the lining of the bronchi, but most commonly occurs in the peripheral lung parenchyma. It is the most common form of lung cancer in women and in people who have never smoked [2].

Smoking is the major risk factor for the development of lung cancer [2]. Smokers are between 15 and 25 times more likely than non-smokers to suffer from lung cancer. The risk is related to the number of cigarettes, the duration of the habit, the age of onset and the amount of nicotine in cigarettes. On the other hand, occupational exposure to tar, soot, arsenic, chromium and nickel and, above all, to asbestos, increases the risk of cancer and lung cancer.

The imaging tests that are currently used in the diagnosis of lung cancer are simple chest radiography, computed tomography (CT), magnetic resonance and the hybrid approach PET-CT (Positron emission tomography). They are the imaging methods of choice currently used in the diagnosis, staging, prognosis and evaluation of the therapeutic results in lung cancer. This is due to the great capacity of spatial resolution and contrast that these methods possess (in addition to the possibility of obtaining multiplanar three-dimensional images). In the present investigation, CT modality will be used [4].

On the other hand, there are a large number of works related to the segmentation of lung cancer, which are presented at next. In this sense, Charbonnier, *et al.* [5], propose an automatic technique, based on computational voxel classification, to detect adenocarcinoma nodules in computed three-dimensional (3D) tomography images. They report an average Dice coefficient of 0.8123. Likewise, Kubota, *et al.* [6], report a computational method that uses morphological and diffusion operators for the segmentation of pulmonary nodules. They report an average Dice coefficient of 0.7100. On the other hand, Alilou, *et al.* [7], presented an integrated scheme of segmentation and classification based on the way to distinguish adenocarcinomas from granulomas in pulmonary CT. They applied an active contour gradient vector flow model of (ACGVFM) for the extraction of nodule boundaries. The coefficient of Dice between the nodules segmented automatically by ACGVFM and the manual delineations by the expert radiologists was 0.8400. Finally, Wang, *et al.* [8], propose a comprehensive analysis of the pathological images of lung cancer to discover the tumor shape and the characteristics of the limits that predict the survival outcome. They made use of a deep convolutional neuronal network, obtaining a great correspondence.

In this paper, the main contributions are: a) propose a robust automatic segmentation algorithm to segment the tumor called adenocarcinoma, this algorithm contains the stages of preprocessing, segmentation and postprocessing. b) Consider the relative percentage error (RpE) and the Dice coefficient (Dc) for performing a comparative study between manual and automatic segmentations.

2. Materials and methods

2.1. Dataset description

The dataset (DS) used was acquired through the CT modality. In addition, manual segmentation of the tumor (reference segmentation) generated by a pneumologist is available. This reference segmentation will be used to evaluate the performance of the proposed technique.

2.2. Computational proposed technique

Figure 1 shows a schematic diagram that synthesizes the computational algorithms that make up the computational technique for segmenting adenocarcinoma.

From the algorithms that appear in Figure 1, a brief description will be made since these algorithms was described, in detail, in the references [9-11].

2.2.1. Gauss filter. The Gauss filter, also known as the Gaussian filter, simulates a multivariable Gaussian distribution, which can be expressed by means of a Gaussian mask or arbitrary size kernel. The maximum value appears in the central voxel and decreases towards the extremes in function of the standard deviation parameter σ [12]. The result will be a set of values between 0 and 1. If you want to

filter a three-dimensional image, the scalars that make up the aforementioned kernel can be obtained using Equation (1).

$$G(i, j, k) = \frac{1}{(\sqrt{2\pi}\sigma)^3} e^{-\left(\frac{i^2+j^2+k^2}{2\sigma^2}\right)}, \quad (1)$$

being: σ the standard deviation (in each spatial dimension) and i, j, k the gaussian size kernel.

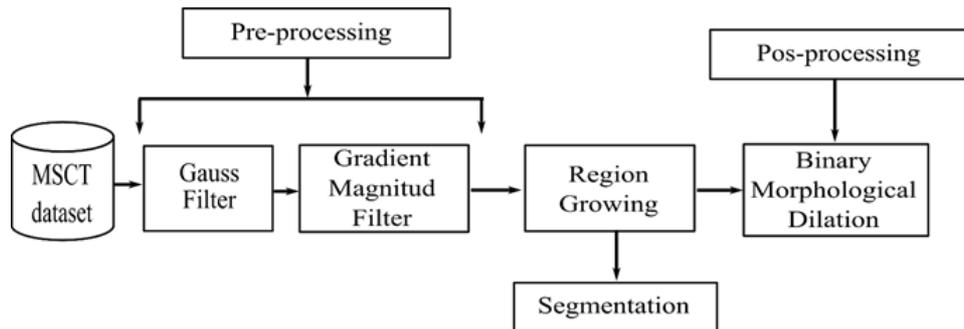


Figure 1. Block dia gram of the proposed technique.

2.2.2. Gradient magnitude filter. In the digital image processing, the gradient magnitude filter is used to identify the objects contours (edge detection) and the homogeneous regions separation. Edge detection is the significant discontinuities identification in the gray level or color of the image [11]. This technique is based on the gradient of an image $[f(x, y, z)]$, using the first directional partial derivatives, given by Equation (2).

$$\left[\left(\frac{\partial f(x, y, z)}{\partial x} \right), \left(\frac{\partial f(x, y, z)}{\partial y} \right), \left(\frac{\partial f(x, y, z)}{\partial z} \right) \right] \quad (2)$$

The classical 3D mathematical model, to obtain a gradient magnitude image $[|\nabla f(x, y, z)|]$, is presented by Equation (3).

$$|\nabla f(x, y, z)| = \sqrt{\left(\frac{\partial f(x, y, z)}{\partial x} \right)^2 + \left(\frac{\partial f(x, y, z)}{\partial y} \right)^2 + \left(\frac{\partial f(x, y, z)}{\partial z} \right)^2} \quad (3)$$

2.2.3. Region growing technique. This technique allows a search of pixels or voxels that belong to the objects that compose an image according to a predefined criterion. The image elements neighbors that have similar features (similar gray levels, for example) are considered for forming the segmented information. In the 3D context, the method starts with a voxel, which is automatically selected or provided by the user and then examines the neighboring voxels to decide if they have similar features. If so, the neighbor voxel that meets such similarity criterion is grouped together with the previous ones to form a region [13].

The region growing technique requires a seed voxel that can be selected manually or automatically, to extract all the voxels connected to the seed [13]. Additionally, this method has the following parameters: initial neighborhood size (r) and the parameter (m) 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.

2.2.4. Binary morphological dilation. The purpose of the binary morphological dilation is to compensate the images edges transformations introduced by the Gaussian filter. The binary dilation (\oplus), considering

an image (A) and a structuring element (B), over a x image element, is defined by the mathematical model given by Equation (4) [14]:

$$A \oplus B = \{x | (B)_x \cap A \neq \emptyset\} \quad (4)$$

On the other hand, to make a comparative study between manual and automatic segmentations, the relative percentage error (RpE) will be used in such a way that its performance can be established when the tumor volume is obtained automatically (A_v). During the comparison, the volume obtained by the manual method, applied by a pneumologist, is taken as the reference volume (R_v). The RpE is calculated using the mathematical model given by Equation (5).

$$\text{RpE} = \frac{100(R_v - A_v)}{R_v} \quad (5)$$

Additionally, the dice coefficient (Dc) is a metric, used in this paper, to compare the segmentations of the 3D image, obtained by different methodologies. In the medical context, it is generally considered that the Dc establishes how similar, spatially, are manual segmentation and automatic segmentation [15].

3. Results

3.1. Qualitative results

Figure 2 shows a two-dimensional (2D) view of both the original tumor and the processed versions after applying the proposed technique to the dataset considered.

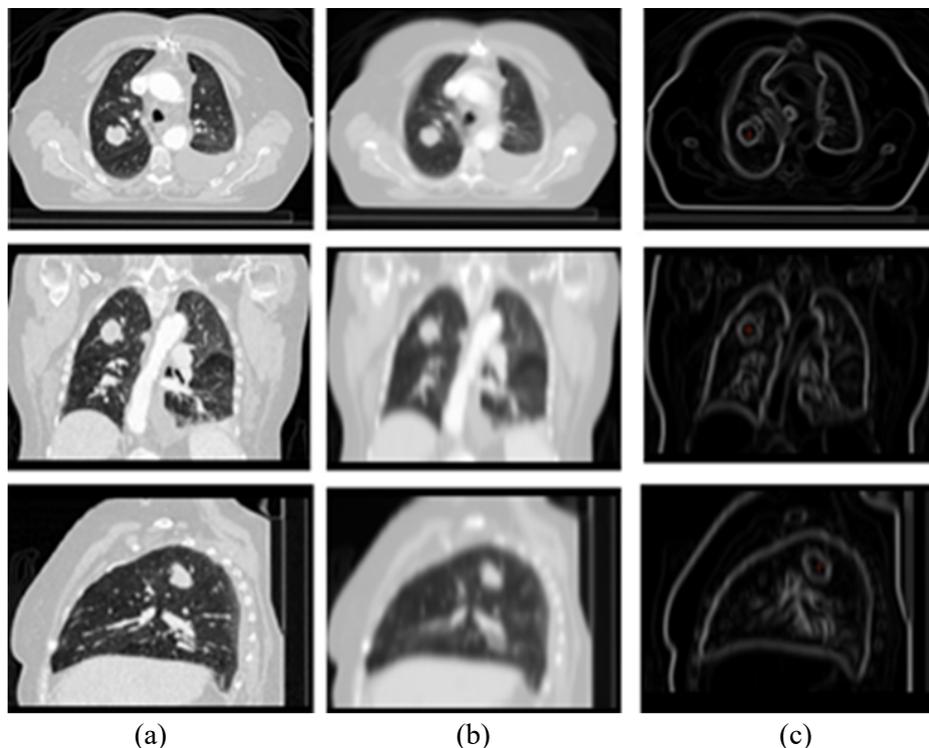


Figure 2. Results obtained from the preprocessing stage: axial view (top), coronal view (middle) and sagittal view (bottom). (a) Original. (b) Gauss. (c) Gradient magnitude.

On the other hand, Figure 3 illustrates the adenocarcinoma tumor segmentation; whereas, in Figure 4 the post-processing stage is shown.

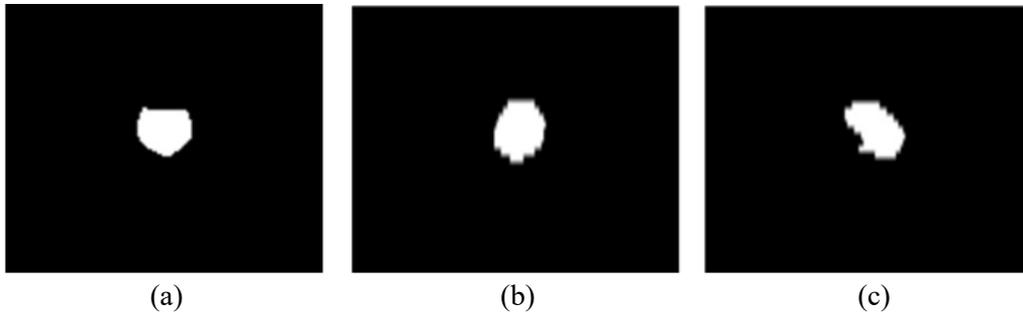


Figure 3. Segmentation stage bidimensional results. (a) Axial view. (b) Coronal view. (c) Sagittal view.

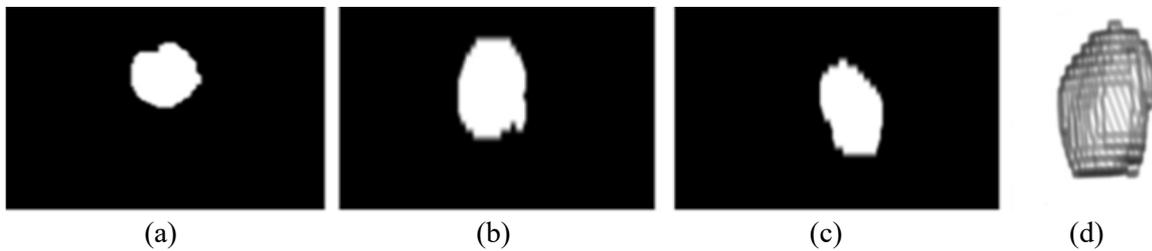


Figure 4. Postprocessing stage results. (a) Axial view. (b) Coronal view. (c) Sagittal view. (d) 3D view.

3.2. Quantitative results

For a given filter, the tuning process stops when the optimum value for its parameters is obtained, that is, when the values for which the best segmentation, considering the RpE, is generated. The volume occupied by the segmented adenocarcinoma was 17.08 cm^3 ; while the volume reported by the clinical expert, obtained considering the manual segmentation, was 16.66 cm^3 .

The RpE, considering these volumes, was 2.7% and the maximum dice coefficient generated was 0.8734, and allowed to establish the optimal parameters, of the computational algorithms that make up the proposed technique, which are presented in Table 1.

Table 1. Optimal parameters for the proposed technique.

Parameters for Gauss filter		Region growing technique		Dilation filter
Kernel size	σ	r	m	Kernel size
3x3x3	2.0	3.0	4.5	5x5x5

4. Conclusions

A semi-automatic technique available for detecting lung adenocarcinoma, in a precise and efficient manner, has been presented.

The three-dimensional representation of this type of lung cancer is useful for the detection and monitoring of lung 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 lungs. Likewise, support the planning of therapies and associated surgical processes, in general, with pulmonary pathologies.

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] Guyton J 2006 *Textbook of medical physiology* (USA: Elsevier Saunders)
- [2] Alberg A, Samet J 2003 Epidemiology of lung cancer *Chest* **123(1)** 21s
- [3] Ait B, El Hassani A, Majda A 2018 Lung ct image segmentation using deep neural networks *Procedia Computer Science* **127** 109
- [4] Mingjie X, Shouliang Q, Yong Y, Yueyang T, Lisheng X, Yudong Y, Wei Q 2019 Segmentation of lung parenchyma in ct images using cnn trained with the clustering algorithm generated dataset *Biomed Eng Online* **18(2)** 1
- [5] Charbonnier J, Chung K, Scholten E, Van Rikxoort E, Jacobs C, Sverzellati N, Silva M, Pastorino U, Van Ginneken B, Ciompi F 2018 Automatic segmentation of the solid core and enclosed vessels in subsolid pulmonary nodules *Sci Rep.* **8** 646
- [6] Kubota T, Jerebko A, Dewan M, Salganicoff M, Krishnan A 2011 Density and attachment agnostic ct pulmonary nodule segmentation with competition-diffusion and new morphological operators *Multi modality state-of-the-art medical image segmentation and registration methodologies* ed A. El-Baz (Boston: Springer)
- [7] Alilou, M, Beig N, Orooji M, Rajiah P, Velcheti V, Rakshit S, Reddy N, Yang M, Jacono F, Gilkeson R, Linden P, Madabhushi A 2017 An integrated segmentation and shape-based classification scheme for distinguishing adenocarcinomas from granulomas on lung ct *Med Phys.* **44(7)** 3556
- [8] Wang S, Chen A, Yang L, Cai L, Xie Y, Fujimoto J, Gazdar A, Xiao G 2018 Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome *Scientific Reports* **8(1)** 10393
- [9] Yang B, Xiang D, Yu F, Chen X 2018 Lung tumor segmentation based on the multi-scale template matching and region growing *Proc. SPIE, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging* **10578** 105782Q
- [10] Huérfano Y, Vera M, Mar A, Bravo A 2019 Integrating a gradient-based difference operator with machine learning techniques in right heart segmentation. *J. Phys. Conf. Ser.* **1160** 012003
- [11] Pratt W 2007 *Digital image processing* (Unite State of America: John Wiley & Sons Inc)
- [12] Meijering H 2000 *Image enhancement in digital x ray angiography* doctoral dissertation (Netherlands: Utrecht University)
- [13] González R, Woods R 2001 *Digital image processing* (New Jersey: Prentice Hall)
- [14] Saénz F, Vera M, Huérfano Y, Molina V, Martínez L, Vera MI, Salazar W, Gelvez E, Salazar J, Valbuena O, Robles H, Bautista M, Arango J 2018 Brain hematoma computational segmentation. *J. Phys. Conf. Ser.* **1126** 012071
- [15] Dice L 1945 Measures of the amount of ecologic associationn between species *Ecology* **26(3)** 29