A computational methodology for the staging of lung tumors considering geometric descriptors

Una metodología para la estadificación de tumores pulmonares considerandos descriptores geométricos

Miguel Vera, MSc, PhD[1], https://orcid.org/0000-0001-8573-0974, Yoleidy Huérfano, MSc[2], https://orcid.org/0000-0003-0415-6654, Antonio Bravo, MSc, PhD[1], https://orcid.org/0000-0001-8572-5868
[1]Facultad de Ciencias Básicas y Biomédicas, Universidad Simón Bolívar, Cúcuta, 540004, Colombia.
*Corresponding e-mail: m.avera@unisimonbolivar.edu.co
[2]Grupo de Investigación en Procesamiento Computacional de Datos, Universidad de Los Andes, San Cristóbal 5001, Venezuela.

Abstract

Lung diseases diagnosis, specifically the presence of lung tumors, is usually performed with the support of radiological techniques. Computed tomography is the most widely used imaging technique to confirm the presence of this disease. When several researchers require identifying the morphology of these tumors, they deal problems related to the poor delimitation of the borders associated with the anatomical structures that compound the lung, Poisson noise, the streak artifact and the non-homogeneity of gray levels that define each object in the chest images. In this paper, a methodology has been presented to identify in which stage (staging) the mentioned tumors are. For this, first, anisotropic diffusion filter and magnitude of the gradient filter are used in order to address the aforementioned problems. Second, a smart operator and the level set algorithm are used to segment lung tumors. Finally, considering these segmentations, a set of geometric descriptors is obtained, and it allows staging of such tumors to be precisely established, generating results that are in high correspondence with the reference data, linked to the analyzed tagged images.

Keywords: Computerized tomography, lung tumors, segmentation, geometric descriptors.

Resumen

El diagnóstico de enfermedades del pulmón, específicamente la presencia de tumores pulmonares, suele efectuarse con el apoyo de técnicas radiológicas. La tomografía computarizada es la técnica imagenológica más utilizada para confirmar la presencia de esta enfermedad. Cuando diversos investigadores desean generar la morfología de esos tumores se enfrentan a problemas relativos a la pobre delimitación de las fronteras asociadas con las estructuras anatómicas que conforman el pulmón, el ruido poisoniano, el artefacto tipo escalera y la no-homogeneidad de los niveles de gris que definen cada objeto en las imágenes de tórax. En el presente artículo, se presenta una metodología para identificar en cual estadio (estadificación) se encuentran los mencionados tumores. Para ello, en primer lugar, los filtros de difusión anisotrópica con curvatura y magnitud del gradiente son utilizados a fin de abordar los mencionados problemas. En segundo lugar, un operador inteligente y el algoritmo denominado conjuntos de nivel son utilizados para segmentar los tumores pulmonares. Finalmente, considerando estas segmentaciones se obtiene un conjunto de descriptores geométricos que permite establecer con precisión la estadificación de tales tumores, generando resultados que están en alta correspondencia con los datos de referencia, vinculados con las imágenes etiquetadas analizadas.

Palabras clave: Tomografía computarizada, tumores pulmonares, segmentación, descriptores geométricos.

www.revhipertension.com

Revista Latinoamericana de Hipertensión. Vol. 15 - N° 3, 2020

T he diagnosis of lung diseases, specifically the presence of lung tumors, is usually carried out with the support of imaging techniques, among which the computed tomography currently used in the diagnosis, staging, prognosis and evaluation of therapeutic results in lung cancer stands out. This is due to the high spatial and / or contrast resolution typical of this technique[1].

When various researchers want to generate the morphology of these tumors, they face problems such as: poor delimitation of the borders associated with the anatomical lung structures, poison noise is characterized by being dependent on the intensity of each voxel in the image, as well as the voxel of higher intensity are at risk of being more affected by noise[2], ladder-type artifact, the possible causes of this artifact are due to inadequate collimation management and errors in reconstruction[3] and non-homogeneity of levels of gray that define each object in the chest images[4].

Lung cancer represents one of the leading causes of death in the world. 85% of cases of this type of cancer correspond to non-small cells and 15% to small cells. Hoyos et al.[5], presents a synthesis of the classification of the most frequent epithelial lung tumors[5].

In this classification, 6 major categories of pathologies are presented. However, in general, all of them can be classified as lung tumors. Additionally, the seventh edition of the system based on the descriptors Tumor-Nodules-Metastasis (TNM system), for staging lung cancer was established on a retrospective analysis of a large database (81,495 patients) collected by the International Association for the Study of Lung Cancer (IASLC) and analyzed by cancer research and biostatistics (CRAB), a non-profit organization which goal is to prevent and cure cancer[6]. In the TNM system, the descriptor or stage T refers to the characteristics of the primary tumor, N to regional lymph node involvement and M to intra and extra-thoracic metastases.

This retrospective analysis had its limitations, so the IASLC proposed in 1999 a new international prospective database for the study of the staging and prognostic factors of lung cancer that ended in 2010 and which has served to prepare the eighth edition of the TNM classification of lung cancer. In this new database from 16 countries 94,708 patients with lung cancer were initially registered and, after the exclusions made by the CRAB, only 77,156 patients were finally studied, of whom 70,967 had non-small cell lung cancer (NSCLC) and 6,189 had small cell lung cancer (CPCP)[7].

In this article, a methodology to identify the stage of these tumors is presented. In this sense, first, anisotropic diffusion filter, based on curvature, and magnitude of the gradient filter were used in order to address the aforementioned problems. Secondly, the Hough transform and the level sets algorithm are used to segment lung tumors and, from them, geometric descriptors are obtained to establish the tumors staging.

Data Source

Table 1 shows the characteristics of the multislice computed tomography (MSCT) databases (DB), considered in this work, which are composed of chest images containing squamous cell lung tumors.

Additionally, we have the manual segmentations or ground truth, carried out by a clinical expert, which will serve as a reference to develop the tuning process of the parameters that operationally control all the techniques involved in the proposed strategy.

Table 1. Numerical characteristics of the computed tomography images.

	DB dimensions	Voxel spatial size (mm³)	Patient Age (years)
DB1	512x512x129	0.977x0.977x3	81
DB2	512x512x104	0.977x0.977x3	78
DB3	512x512x136	0.977x0.977x3	87

Computational strategy proposed

Using figure 1, a schematic diagram is presented. In this figure, it is synthesized the computational algorithms that make up the proposed methodology for staging lung tumors.

Figure 1. Block diagram of the computational strategy proposed.
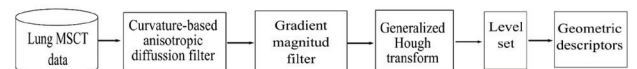
Curvature-based anisotropic diffusion filter.

This type of filter is iterative, it has a solid mathematical foundation and was developed by Álvarez et al[8]. In general, the curvature anisotropic diffusion filter incorporates, in a hybrid way, a parabolic nonlinear differential equation, level sets and elements of the diffusion models of Witkin[9], Koenderink[10] and Perona-Malik[11].

Furthermore, anisotropic diffusion is characterized by performing effectively against the noise problem that exhibits medical images and, additionally, this filter has 3 parameters that control its operation. They are number of iterations (iter), time base (t) and diffusion threshold (k)[12].

Gradient Magnitude Filter

The gradient magnitude filter generates a smoothed version of the anisotropic diffusion image based on the calculation of the magnitude of the three-dimensional gradient of such image, using a given mathematical model based on finite differences[13,14].

Variational Algorithm
Algorithm Initialization

The generalized Hough transform (GHT) is an algorithm for detecting the centroid of objects, with an arbitrary shape, in an image considering the training and detection stages. In training, a prototype is extracted from a previously selected image.

Figure 2 it is used to present an example of a prototype in which certain geometric elements have been incorporated, which are identified below: a) (xc, yc): centroid of the prototype. b) (x, y): arbitrarily chosen contour point. c) r: radial distance between the centroid and the arbitrary point. d) α: angle that r forms with the horizontal line (line drawn in green). e) θ: angle formed by the horizontal with a line tangent to the arbitrary point (line drawn in blue). θ represents the direction of the gradient[4].

Figure 2. Representation of a prototype used to describe GHT.
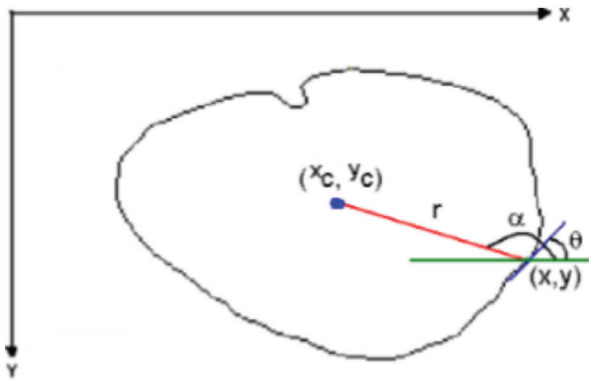
On the other hand, in the detection stage, the parameters of Table-R are considered to automatically identify the centroid of the structure to be detected, in images other than the one used during training.

In this article, GHT is used to obtain the coordinates of the seed voxel required by the level sets (LS) to initiate tumor segmentation.

Level sets technique (LS)

LS are iterative methods that allow controlling the evolution of snake-like active contour models for the segmentation of objects present in an image, for example, the tumors present in the DB described[15].

The aforementioned evolution is controlled by a differential equation that depends on 2 parameters: curvature (α) and smoothness (β)[4]. From the seed voxel, an iso-sphere evolves iteratively considering α and β parameters until the iterations of the level set, reaches a value equal to 512, that is, the size of the horizontal image dimension (see table 1).

Additionally, in this work, the geometric descriptors of the segmented tumors coincide with their: perimeter (Pe), area (Ar) and diameter (Di). They allow precisely establishing the stage in which such tumors are classified, according to the criteria established[16], which are presented using table 2.

Table 2. Ranges to determine the stage in which a lung tumor is classified, considering the descriptors Di, Pe and Ar.

|	Stage I	Stage II	Stage III	Stage IV
Diameter (cm)	5 to 7	11 to 21	21 to 42	21 to 42
Perimeter (px)	18 to 25	25 to 38	38 to 80	80 to 177
Area (px)	0 to 57	57 to 100	100 to 352	1446

Dice Coefficient

The Dice coefficient is a metric to evaluate the correspondence between the results of two procedures. In this paper, the Ds let us compare the manual and automatic segmentation of lung tumors. The expected values for the Dice score (Ds) are real numbers between 0 and 1. The main rule using Ds is: the performance of a particular procedure is better if its score is closer to 1[17].
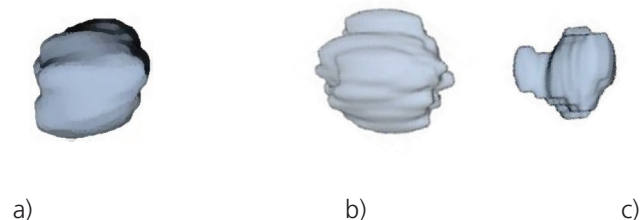
Qualitative results

Figure 3 presents axial views showing the effect of applying both the filters and the GHT and LS to the original images that make up the all DB described.

a)   b)   c)   d)
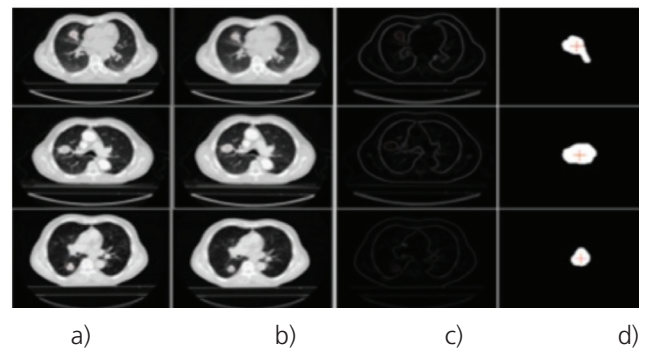
Figure 3. Axial view of tumors in the image: a) Original. b) Anisotropic diffusion. c) Magnitude of the gradient. d) Segmented with LS.

In this figure, we can appreciate a good edges definition which was an important step for faciliting the tumor segmentation process. In addition, figure 4 shows an excellent 3D representation about the 3 segmentated lung tumors.

a)   b)   c)

Figure 4. Three dimensional images about lung tumors in: a) DB1. b) DB2. c) DB3.

246

www.revhipertension.com

Revista Latinoamericana de Hipertensión. Vol. 15 - N° 3, 2020

Quantitative results

It is important to note that the parameters that control the performance of the entire strategy were adjusted by means of cross validation[18], considering the comparison between ground truth and automatic segmentations which was assessed using the Dice coefficient (Ds) as a metric. In this sense, the optimal values reported in this section correspond to the maximum value obtained for the Ds, which was 0.9081.

Table 3 shows the results related to the tuning process about the parameters that, in this research, control the performance of the techniques that make up the methodology proposed.

Table 3. Optimal parameters derived from the intonation process of the proposed methodology.

Technique	Anisotropic diffussion			Level set	
	iter	t	k	α	β
Parameter	3	0.0625	10	0.15	0.1

Table 4 shows the values of the geometric descriptors and the stage established for segmented tumors.

Table 4. Staging of segmented tumors.

	Pe (px)	Ar(px)	Di (cm)	Stage
DB1	27.68	55.12	10.52	II
DB2	54.90	123.86	14.97	III
DB3	24.07	44.41	6.73	I

According to the values obtained for the descriptors, showed in the previous table, the patients present tumors that go from the first to the third stage. This implies the adoption of different therapeutic, clinical or surgical protocols when planning, establishing or deciding on the most appropriate procedure to improve the health of patients suffering from the consequences of having these tumors lodged in their lungs.

Conclusions

An automatic methodology for the segmentation and staging of lung tumors has been presented. In it, the filter bank was used to minimize the problems associated with MSCT images.

Using the automatically generated segmentations, certain geometric descriptors were calculated, accurately and efficiently, which are useful to establish the stage in which each of the segmented tumors is.

The information derived from the aforementioned descriptors is of vital importance to define the clinical or surgical behavior to be followed, regarding the approach to detected tumors.

References

1. Anita C, Sonit SS. Lung cancer detection on CT images using image processing. IEEE Comput Society, 4, 142-46, 2012.

2. Zanella R, Boccacci P, Zanni L, Bertero M. Inverse Problems in Computational Methods for Inverse Problems in Imaging, 25(4) 1, 2009.

3. Barrett J, Keat N Artifacts in CT: Recognition and avoidance1 Radiographics 24(6) 1679, 2004.

4. Vera M 2014 Segmentación de estructuras cardiacas en imágenes de tomografía computarizada multi–corte. Tesis doctoral. Universidad de Los Andes-Mérida-Venezuela.

5. Hoyos N, Montoro F, García J, Morales B, Pavón M Cáncer de pulmón: ¿qué hay de nuevo? Rev Patol Respir. 20(2) 47, 2017.

6. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer. (2) 706, 2007.

7. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE The IASLC Lung Cancer Staging Project: Proposals for revisión of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol (11) 39, 2016.

8. Álvarez, L., Guichard, F., Lions, P.L., Morel, J.M. (1993). Axioms and Fundamental Equations of Image Processing. Arch. Ration. Mech. Anal. 123, 199-257.

9. Witkin, A. (1983). Scale-space filtering. Int'l Joint Conf. Artificial Intelligence, 1019- 1021.

10. Koenderink J. The structure of images. Biological Cybernetics, 50, 363-370, 1984.

11. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intell. 12, 629-639, 1990.

12. Vera M, Gonzalez E, Húerfano Y, Gelvez E, Valbuena O. New anisotropic diffusion operator in images filtering. Journal of Physics: Conf. Series 1448-012019 doi:10.1088/1742-6596/1448/1/012019, 2020.

13. Pratt W 2007 Digital image processing (New York: John Wiley & Sons Inc).

14. Sapiro G. Geometric Partial Differential Equations and Image Analysis. Cambridge, UK,: Cambridge University Press, 2001.

15. G. Dharanibai y J. P. Raina, Gaussian mixture model based level set technique for automated segmentation of cardiac MR images. International Journal of Environmental Science and Technology, vol. 3, no. 4, pp. 2970–2976, 2011.

16. Lavanya M, Muthu P. Lung lesion detection in ct scan images using the fuzzy local information cluster means (FLICM) automatic segmentation algorithm and back propagation network classification. Asian Pacific Journal of Cancer Prevention 18 3395, 2017.

17. Dice, L. Measures of the amount of ecologic association between species. Ecology. 26: 297-302, 1945.

18. Zhu H, Rohwer R. No free lunch for cross-validation. Neural Computation, 8(7):1421-1426, 1996.

247