Journal of Physics:
Conference Series

PAPER • **OPEN ACCESS**

Visualization and Segmentation of Lung Tissue in DICOM-format CT Images

To cite this article: Jennifer Ortega 2024 *J. Phys.: Conf. Ser.* **2796** 012012

Visualization and Segmentation of Lung Tissue in DICOM-format CT Images

Jennifer Ortega

Physics Department, Escuela Politécnica Nacional, Quito, Ecuador

E-mail: `jennifer.ortega@epn.edu.ec`

Abstract. In this project, two tools were developed as a preliminary step towards the creation of future Computer-Aided Diagnosis systems for the automatic detection of lung diseases. The first methodology aimed to extract information from DICOM-format files using Python and the Jupyter Notebook programming environment. The implemented process enables the generation of images in various anatomical planes, using the pixel matrix in Hounsfield units. Additionally, a custom tool was developed for anonymizing and visualizing file information. The techniques employed are independent of the grayscale image generated and are not reliant on the type of window used. The second tool introduces a lung segmentation technique for chest CT images. The segmentation uses the pixel matrix in Hounsfield units to initially identify lung tissue based on its linear X-ray attenuation. Subsequently, the Watershed algorithm based on markers is employed as the primary foundation of the technique. Various morphological operations and transformations on axial slices are applied, resulting in successful lung segmentation.

1. Introduction

In contemporary medicine, the use of technologies based on computer image analysis for diagnostic tasks is indispensable, facilitating the visualization and assessment of various anomalies in patients [1]. Among these technologies, chest CT scans play a pivotal role in diagnosing lung diseases, such as lung cancer, which according to "World Cancer Research Fund International", is the most common cancer (excluding skin cancer) and represents 12% of all cancers. For this reason, in recent years a large number of CAD (Computer Aided Diagnosis) systems have been developed in order to achieve rapid detection through the use of chest CT images, since this helps to reduce the death rate [2].

Image segmentation is an important step prior to the development of most CAD systems for the identification of lung diseases. Depending on the type of segmentation used, the resulting outcomes could be affected.[13]. In a CT image of the chest, the lungs are surrounded by other parts of the human body, which makes segmentation a challenging task. For this project, images in DICOM format were used, since, in addition to storing medical images, it stores patient details and image acquisition parameters, making it possible to produce slices in different anatomical planes, as well as to generate 3D structures, by means of image processing techniques [4].

2. Methodology

2.1. Datasets

The initial database consists of 35 examinations corresponding to different patients in different areas of the body, in the CT modality. These images were acquired using iQ-VIEW / PRO

Journal of Physics: Conference Series **2796** (2024) 012012 doi:10.1088/1742-6596/2796/1/012012

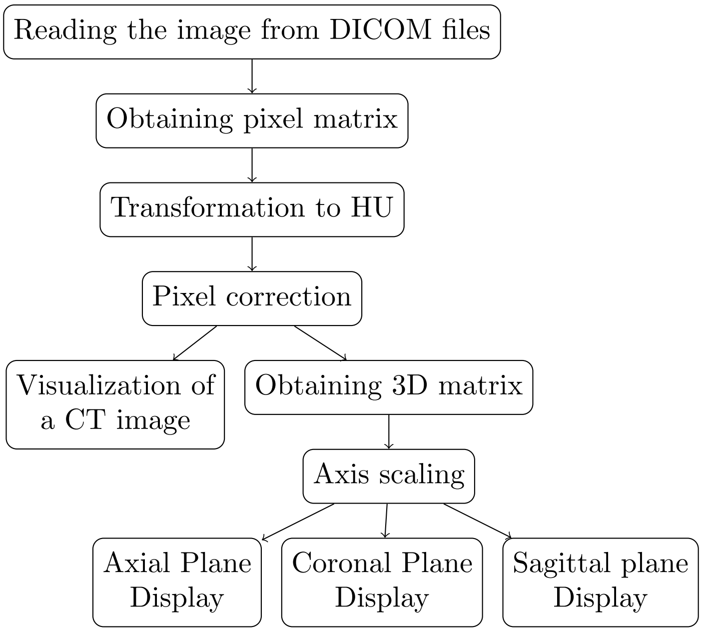


Figure 1: Workflow for the visualization of a CT image and examination in different anatomical planes.

software version 3.0.0. [5]. The database used for visualization consisted of 35 CT scans from different patients, while 17 scans were used for segmentation.

The following are the main features of the data system.

- The number of images per patient is variable, with a mean of 97, a minimum of 55 and a maximum of 771.
- The size is 512 x 512 pixels for each image in all cases.
- The scan resolution dimension (Slice Thickness) is $5mm$ for 34 patients and $2.5mm$ for one patient.
- The type of study (StudyDescription) is Therapy Planning for 34 patients and Body CT for one patient.
- The patient position corresponding to HFS (Head First-Supine), head first and with the face and torso upward. [6], in all cases.
- The distance between pixels, also known as PixelSpacing for all studies is $0.976562mm$ for the x and y axis.

2.2. Image Processing
Image processing consists of three main phases: data reading, processing (in this case low level) and finally visualization. For this project two types of output images are considered. The first one is of a single CT image and the second one is based on a complete examination, where in turn, the representation of the slices in different anatomical planes will be obtained. The workflow for the visualization, in both cases, is as shown in Figure 1.

For the generation of the images in different anatomical planes, different attributes of the DICOM images were used, such as: Pixel Spacing that specifies the physical distance between the centers of each two-dimensional pixel of the axial CT image, the Image Position Patient element to find the distance between each slice, PixelData that contains a matrix of N rows

Journal of Physics: Conference Series **2796** (2024) 012012 doi:10.1088/1742-6596/2796/1/012012

by M columns, which represents a body slice of each two-dimensional pixel of the axial CT image, the Image Position Patient element to find the distance between each slice, PixelData containing a matrix of N rows by M columns, representing an axial slice of the body, among others. In addition, the change of the pixel matrix to Hounsfield Units was performed using the transformation given by equation $m \times SV + b$. The SV parameter is the matrix with the pixel values given by the PixelData element and the b and m constants obtained from the DICOM attributes. [7].

2.3. Lung segmentation on CT images

The segmentation process is an important step prior to the development of CAD (Computer Aided Diagnosis) systems, and is used to locate objects in an image or to find edges. Segmentation is based on the characteristics of the pixels in the image and delimits an area of interest, in terms of voxels or pixels. [8, 9, 10, 11, 12].

The process applied for segmentation is an automatic method that mainly uses the Watershed transformation based on markers and different morphological operations implemented in the high-level language Python. [13]. First, the selection of examinations corresponding to the Protocol Name THORAX RADIOTHERAPY and ABDOMEN AND PELVIS, which include the areas where the lungs are located, was performed according to the database. Then, we proceeded to read the DICOM files, using the designed code, which allows us to extract the pixel matrix, and convert each pixel to Hounsfield units, in order to achieve the filtering of the pulmonary region, we used about $-400HU$ to $-600HU$, to determine in first instance the region where the lungs are located. [14]. Subsequently, we conducted low-level processing, which involved generating the gradient image and removing artifacts such as the trachea and stretcher. This entire process relies on morphological operations previously implemented in Python. Following this, markers are generated, and the Watershed algorithm is applied to refine segmentation using various transformations and morphological operations. The workflow is illustrated in Figure 2.

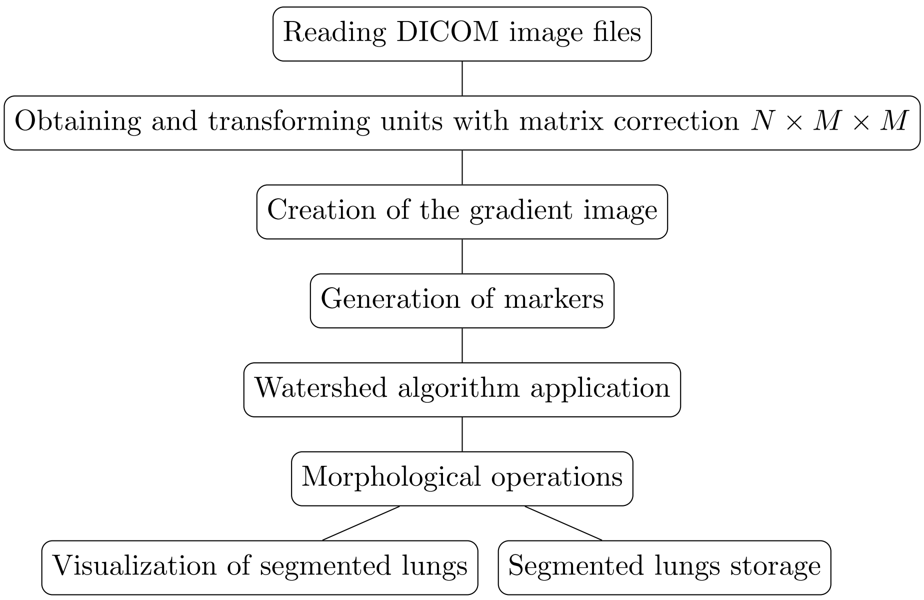
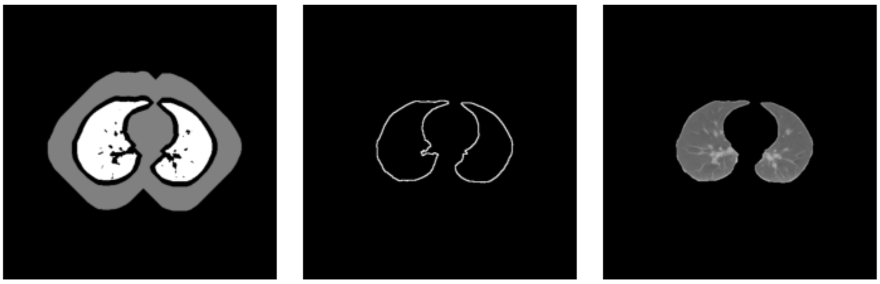


Figure 2: Workflow for lung tissue segmentation in chest CT images.

The lung segmentation process involves filling and smoothing the edges of the segmented image. This step is crucial because the application of the Watershed algorithm might exclude certain dense structures that are part of the lungs but fall outside the typical values of linear attenuation for lung tissues. This scenario often occurs with pulmonary nodules located at the lung edges. Finally, the segmented lungs are visualized, as demonstrated in Figure 3.



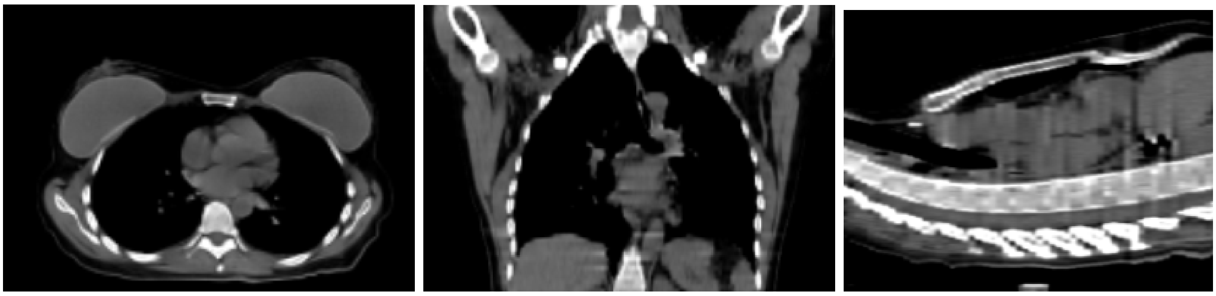
(a) Markers (b) Gradient image (c) Segmentation

Figure 3: Application of Watershed algorithm for segmentation in an axial slice. (a) Image with the labeled markers for the application of the Watershed algorithm. (b) Morphological gradient image, applied to the Watershed segmentation result. (c) Segmented lungs.

3. Results and Discussion

3.1. Visualization of a CT image

The designed code allows the user to enter the percentage of images to be obtained as output. The visualization uses the Matplotlib library to obtain the CT image, with an automatic scaling of Hounsfield units in grayscale, the results of a sample are presented in the Figure 4, the algorithm was applied for all 35 tests with no errors in the total number of tests.



(a) Axial Plane (b) Coronal Plane (c) Sagittal Plane

Figure 4: Visualization of a chest CT examination with 63 axial slices of a 48-year-old female patient, using the full range of Hounsfield units and automatically adapting them to the grayscale. (a) Visualization of a chest CT image with axial slice, using the window of $[-1000, 1000]HU$. (b) Coronal plane with window between $-100HU$ to $300HU$. (c) Sagittal plane with window between $-100HU$ to $300HU$.

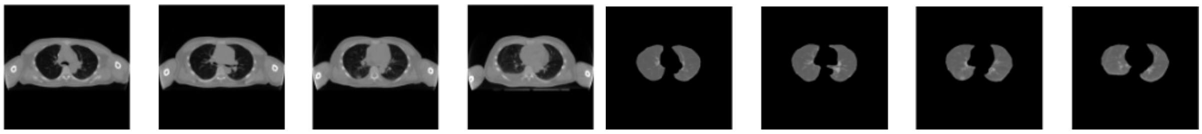
3.2. Lung Segmentation Results

Seventeen patients were studied in this investigation, with an average of 53 slices per exam, constituting a total of 907 chest CT images in all. Of the 907 segmented slices, 249 images of lungs with contour problems and imperfect identification of lung tissue were recorded, representing 26.68% of the total tests performed, as shown in Table 1.

The final data set testing, consistently yielding favorable segmentation results in the majority

Journal of Physics: Conference Series **2796** (2024) 012012 doi:10.1088/1742-6596/2796/1/012012

of cases. In all tests conducted, the algorithm accurately distinguishes stretcher areas from lung tissue, avoiding any confusion between the two see as Figure shows 5. In addition, the identification of trachea and lungs is performed in a better way, unlike other codes that use the Watershed algorithm as a basis for segmentation. [15].



(a) Axial slices (b) Segmentation

Figure 5: Application of Watershed algorithm for axial slice segmentation to a sample of images from a chest radiotherapy CT scan. (a) Visualization of axial slices using the window with values between $-1000HU$ to $1000HU$. (b) Segmented lungs.

Table 1: Lung segmentation results in different examinations. N.° Exam (Exam Identifier), N. ° Cuts (Number of cuts per exam), N.° Segmented cuts (with errors) (Number of cuts with errors in the segmentation of lungs).

Exam Number	Number of Cuts	**N**° Segmented cuts (with errors)	N. ° Exam	**N**° Cuts	**N**° Segmented cuts (with errors)
1	25	7	10	50	10
2	29	5	11	60	15
3	29	7	12	45	6
4	44	8	13	60	9
5	50	12	14	40	7
6	9	2	15	61	18
7	240	82	16	38	6
8	47	28	17	30	12
9	50	8	-	-	-

Based on the tests performed, the following comments can be extracted:

- The position of the patient must be HFS (Head First-Supine), otherwise the algorithm will not work accurately in the first slices (slices where the trachea appears), because the identification of the trachea depends on the location of the trachea in the image.

3.3. Conclusions
The DICOM format provides a large amount of data from the computed tomography process, which helps to support the information given only by the image. In addition, it allows the development of slices in different anatomical planes, as well as obtaining images under different windows and gray levels, through manipulation, obtaining the pixel matrix (matrix with the attenuation coefficients of each voxel) and other attributes, which make possible a good interpretation of the patient's anatomical structures.

Lung segmentation in axial slice images of chest CT was developed using a technique that involves the following concepts: Hounsfield units, Watershed's marker-based algorithm, morphological operations on images. In addition, the algorithm generates good results when

the patient is in the HFS (Head First-Supine) position, which is the most common position for CT examinations. The tests performed yielded visually acceptable results for 73.3% of the total tests performed.
This project served as the culmination of my Bachelor's degree in Physics at Escuela Politécnica Nacional. For detailed information, you can refer to my GitHub repository: `https://github.com/jennifer95/visualizacion-segmentacion-pulmonar-DICOM-python`.

References

[1] S. A. Kane y B. A. Gelman, "Introduction to Physics in Modern Medicine," CRC Press, Taylor and Francis Group, 3rd edition, 2020, pp. 189-262, Chapter 5.
[2] World Cancer Research Fund International, *Worldwide Cancer Data*, WCRF International, `http://www.wcrf.org/dietandcancer/worldwide-cancer-data/`, [Online; accessed 1-July-2021], May 2021.
[3] R. Shojaii, J. Alirezaie, P. Babyn, *Automatic Lung segmentation in CT images using watershed transform*, IEEE International Conference on Image Processing 2005, DOI: 10.1109/icip.2005.1530294, 2005.
[4] NEMA, *Does DICOM work with other standards-development organizations?*, Virginia, USA, `http://dicom.nema.org/dicom/geninfo/Brochure.pdf`, [Online; accessed 15-April-2021].
[5] IMAGE Information Systems Ltd. — IMAGE Information Systems Europe GmbH, *iQ-LITE USER MANUAL*, London, UK, ver. 3.0.0, 2016.
[6] Innolitics, LLC, *DICOM Standard Browser*, *Patient Position Attribute – DICOM Standard Browser*, `https://dicom.innolitics.com/ciods/raw-data/general-series/00185100`, [Online; accessed 20-July-2021].
[7] NEMA, *DICOM PS3.3 2018d - Information Object Definitions*, `http://dicom.nema.org/medical/Dicom/2018d/output/chtml/part03/sect_C.11.html`, [Online; accessed 10-May-2021].
[8] Candace Makeda Moore, *Segmentation: Radiology Reference Article*, *Radiopaedia Blog RSS*, `https://radiopaedia.org/articles/segmentation`, [Online; accessed 20-May-2021].
[9] Yuheng Song, Hao Yan, *Image Segmentation Techniques Overview*, *2017 Asia Modelling Symposium (AMS)*, DOI: 10.1109/ams.2017.24, 2017.
[10] Gagandeep Kaur, Jyotirmoy Chhaterji, *A Survey on Medical Image Segmentation*, *International Journal of Science and Research*, vol. 6, pages 1305-1311, 2017.
[11] Sujata Saini, Komal Arora, *A study analysis on the different image segmentation techniques*, *International Journal of Information & Computation Technology*, vol. 4, no. 14, pages. 1445-1452, 2014, International Research Publications House.
[12] S. Umaa Mageswari, M. Sridevi, C. Mala, *An Experimental Study and Analysis of Different Image Segmentation Techniques*, *Procedia Engineering*, vol. 64, pages. 36-45, 2013, ISSN: 1877-7058, DOI: 10.1016/j.proeng.2013.09.074, `https://www.sciencedirect.com/science/article/pii/S1877705813015889`.
[13] R. Shojaii, J. Alirezaie, P. Babyn, *Automatic Lung segmentation in CT images using watershed transform*, *IEEE International Conference on Image Processing 2005*, DOI: 10.1109/icip.2005.1530294, 2005.
[14] Juan Carlos Ramirez Giraldo, Carolina Arboleda Clavijo, Cynthia H. McCollough, *TOMOGRAFÍA COMPUTARIZADA POR RAYOS X: FUNDAMENTOS Y ACTUALIDAD*, *Revista Ingeniería Biomédica*, vol. 2, págs. 54-66, ISSN: 1909-9762, December 2008, `http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1909-97622008000200008&nrm=iso`.
[15] Ankasor, *Improved Lung Segmentation using Watershed*, *Kaggle*, January 2017, `https://www.kaggle.com/ankasor/improved-lung-segmentation-using-watershed`, [Online; accessed 01-Jul-2021].