

Mass detection in lung CT images by using Graph classification

Atiyeh Hashemi, Software Engineering Department, Islamic Azad University, Malayer,
atiyehhashemi20044@gmail.com

.Abdol Hamid Pilevar, Department of Computer Engineering, Bu-Ali Sina University, Hamedan ,
pilevar@basu.ac.ir

Abstract- Lung cancer is the most common fatal malignancy in both men and women. Early detection and treatment of lung cancer can drastically improve the survival rate of patients. Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. In this area, Computerized tomography (CT) is considered as the best imaging modality for identification of anatomy's defect. There are a lot of medical image processing software tools for research and diagnostic aims, including data gathering in image and algorithms for analysis of the new images and eventually for testing the systems result. To achieve this and to diagnose the area of the cancer mass, two methods are presented in this paper, and the area of abnormality space is calculated. One of the techniques for diagnosing unusual mass is Graph classification method. In this paper, component of images are classified and displayed as a set of points and edges with graph method. In the other method, changes in colors in the surrounding area are examined. In the end, the results of the two techniques are compared and the accuracy of the algorithm is tested. Experiments show that the proposed method has improved significantly in mass detection and the sensitivity of the proposed system is up to 90%, which shows that the proposed method can help the radiologists to increase their diagnostic confidence.

Keywords: Medical Image Processing, Lung cancer, Graph classification, Calculating the mass area.

1. Introduction

1.1. Lung Cancer

Lung cancer is considered to be as the mass epidemic and the main cause of cancer death worldwide. It is difficult to detect in its early stages because symptoms appear only in advanced stages. Medical statistics shows that the number of the cases breast, prostate and colon cancer is totally less than those of lung cancer[1]. More than one third of the British population will be diagnosed with the cancer in their lifetime and one fourth of the cases will die from the disease [2]. Approximately 60% of patients are men and 33% of all death-leading cancers in men are due to lung abnormality. However, awareness of the common symptoms of this disease makes the treatment possible in the primary phases. As a

type of disease in which the malignant tissue grows in one or both lungs, lung cancer kills more than 80 % of patients within 5 years of the diagnosis. Many of them notice their disease, when it's too late and the surgery is not simply possible. In USA, lung cancer has been the main cause of deaths from cancers with 14% of cases in 2011. In the recent three decades, we have failed to increase much the live of the patients suffering from the cancer. While, in the late 1970s, around 37% of patients had the chance of survived for one year or longer after the diagnosis, today 42% of them have the chance6-3][. Today several techniques are used to diagnose the disease, such as Computerized Tomography (CT), Chest Radiograph (x-ray), Magnetic Resonance Imaging (MRI scan). Nevertheless, these techniques can detect the disease only in its advanced stages, which leads to the death of a number of the patients. Hence, we need the technique to diagnose the disease in its early stages. This will certainly enhance the speed and the quality of the diagnosis. Different software's for representing and processing have been designed, such as 3D-Doctor, workstation, analyze, vitrea2-fusion7D, 3dviewnix, 3D slicer and Julius. Some of these software tools as image viewers support some visual formats like Dicom and analyze. Some of the other tools provide features such as processing and visualization for instant registration, noise suppression, analyzing images for diagnosis purposes, applying conventional segmentation methods and representing 2D and 3D data using regular visualization methods. One of the most important feature of medical images is its quick and easy use of data7][.

This article presents the image processing technique leading to an insight to lung tissues and detection of abnormal parts. The most challenging chore of this technique is the auto-detection of abnormal areas which can be a sign of the lung cancer. The development of "Computer-Assisted Detection" (CAD) techniques has an important role in the diagnosis of the cancer. In this field, Computerized Tomography (CT) is considered as the best imaging modality.

Computerized Tomography (CT) images are used to detect the anatomy defects of human body especially head, legs and etc [8] . This type of imaging takes advantages of a particular X-ray attenuation and provides a variety of images of the parts inside the body, which can be viewed on the screen monitor or printed as photos. By using a special device for the accurate interpretation of CT images of the body, radiologists can diagnose disorders such as cancer, heart disease, infections, muscular and skeletal problems more easily.

To achieve this, our method presented in this article comprises two steps. In the first step, Graph Classification is used and images are represented as a set of points and edges between them. In the second step, image processing techniques are used. This method has been applied on a variety of images in the MRI center in Hamedan, some of which are displayed in Figure 1. These images are taken from HRCT (high-resolution computerized tomography) scanning, made by Siemens.

The organization of the article will be as following. In Section 2.1, Graph classification technique is presented. In the next section, Image processing technique is introduced. Section 2.3 deals with the analysis phase, the two steps are compared and the accuracy of the method is determined. Finally, in Section 3, the conclusion and future works are given.

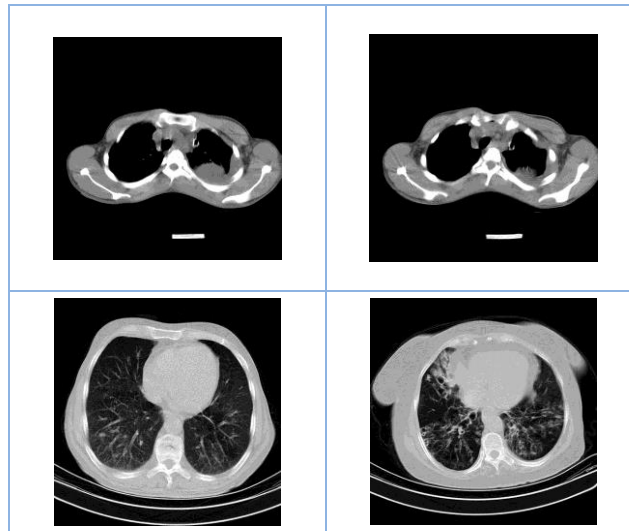


Figure 1. Examples of MRI image and the mass in the lung

1.2. Previews works

In a recent study, three tracking algorithms are given to find the best path between the nodes of the graph and a comparison is made with others. The study concludes that the method based on the Minimum Cost Path (MCP) method is the best and creates the most optimal route [3]. Despite the great need for such a technique to evaluate large 3D datasets, unfortunately little study has been done in this area. Another study has been carried out on the segmentation of the wire frame of the stent in 2D slices [9]. In another study the stent's frame is found by means of a statistical model. The researcher report that the threshold at 2000 Hounsfield units is appropriate to obtain a graph consisting of points related to the stent. Unfortunately, due to the difference in the applied doses or the stent frame materials, our data had a lot of noise. In this regard, further studies have been carried out to segment blood vessels in 3D, which have similarities with the wire of the frame of the stent. These investigations may be important to our study (see Lesage [5-7]). The two-step approach is a method [13-10] which segments the vessel by a vessel measure [14] and then performs the centerline extraction. The methods which fit a number of ellipsoids or spheres to the vessel, and the methods which segment the contour in slices perpendicular to the vessel centerline suppose a solid vessel with a diameter of several vertices. Due to the small diameter of the stent's wire (1-3 vertices) and its sharp corners, these methods are also inappropriate. Region growing methods have problems with leaks and gaps and need a second stage to find the geometry from the segmented vertices. Another study focuses on two types of stent grafts commonly used in our institution to treat patients with AAA, made of nitinol and stainless steel, respectively. This study uses endovascular aortic replacement (EVAR) and stent grafts [7]. In another study a complete system for computer-aided detection of GGNs is presented consisting of initial segmentation steps, candidate detection, feature extraction and a two-stage classification process. A rich set of intensity, shape and context features is constructed to describe the appearance of GGN candidates and apply a two-stage classification approach which makes use of a linear discriminant classifier and a Gentle Boost classifier to classify candidate regions efficiently [15]. In another article, a computer-aided detection (CAD) system has been developed for the identification of small pulmonary nodules in low-dose and thin-slice CT scans. The FROC curve obtained on this data set exhibits high sensitivity to lung nodules (80-85% range) at an acceptable level of false positive findings as for each patient [16]. This study analyzes four geo-statistical functions-semivariogram, semivariogram, covariogram, and correlogram. The aim was to characterize lung nodules as malignant or benign in computerized tomography images. Stepwise discriminant analysis was applied to see which measures could distinguish the benign and malignant nodules. ROC curve was presented for evaluating the results with an accuracy between good and excellent [17]. In this study a computer assisted automatic diagnostic system was used for lung cancer. The system consists of two steps: analytic and diagnostic. In the analysis step, it extracts the lung and the

blood vessel regions by the fuzzy clustering algorithms and analyzes the features of these. In the diagnosis step, the system defines diagnostic rules, using the extracted features which support the determination of the nodule locations[18].

2. Method

Development of a computational decision aid for a new medical imaging modality is a long and complicated process typically. It consists of collecting data in the form of images and annotations, development of image processing and pattern recognition algorithms for the analysis of the new images and finally testing of the resulting system[19]. In this article two methods have been used, graph classification and image processing, and the results are compared and the accuracy of algorithm is calculated. In the following, different stages of each method will be presented in details. The diagram of the proposed method is presented in Figure2.

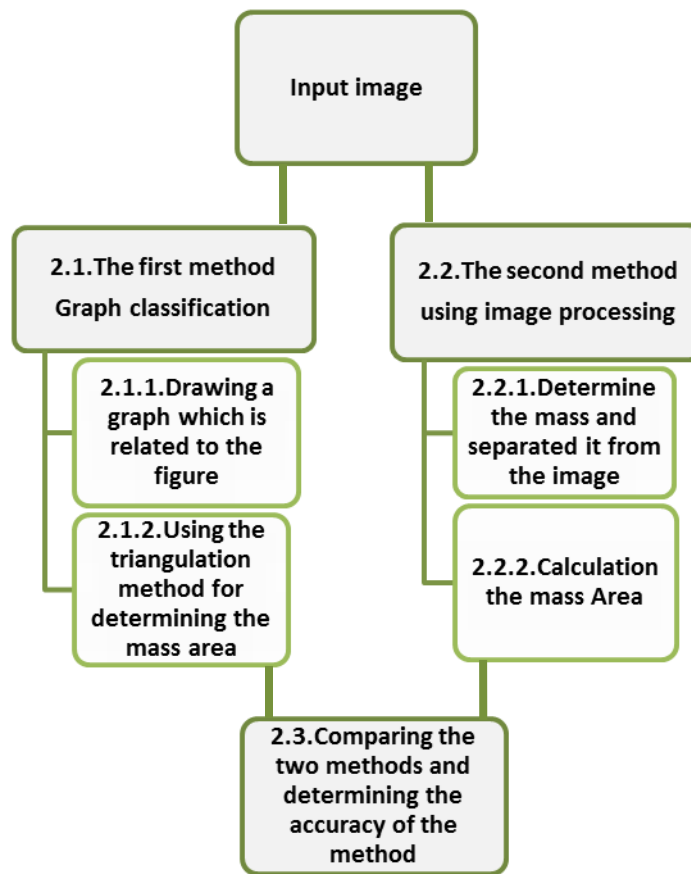


Figure2. The diagram of the proposed method

2.1. The first method, Graph classification

2.1.1. Drawing a graph related to the figure

To be able to gather information in the form of a graph exactly, we propose an automatic method to segment grafts from CT images, which consists of two stages (Figure3): firstly, the detection of seed points by using the points which are

likely to be on the edges, and secondly, finding the connections between them to produce a graph by using a modified version of the MCP (minimum cost path) method. It is worth noting that each image is divided into two sections (right and left lungs), as shown in Figure 4. All of these stages are done on both sections of the image and the sum of both results will be the final result.

At the end of this step, we will have a set of seed points $P = \{p_1, p_2, \dots, p_N\}$ (with $p_i = (x, y)$, $i \in [0 \dots N - 1]$, where N is the number of nodes), and the costs $C(x) \Rightarrow R$ between each two points calculated by using Algorithm1.

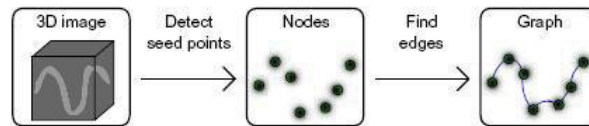


Figure 3. Processing steps required for the extraction of a geometric model from CT data

a. Detecting the points

The key points, which are on the edges of the figures, are marked on both figures (the normal lung and the abnormal one). At first, all the boundary points are identified and then the points in key parts are extracted. This stage will be performed non-automatically by an expert, which can preferably be performed automatically by using some methods like artificial neural networks or genetic algorithm.

b. Finding the edges between the nodes

The identified nodes from the previous step are connected by using different algorithms based on MCP (minimum cost path) in this step. A form of Dijkstra's algorithm that finds the shortest path in a graph is used in this article. The algorithm used is presented below (Algorithm 1). Finally the points are considered as the vertices of the graph and the distance between any two points is taken as the weight of its edges. The final result after applying the algorithm is shown in Figure 4.

Algorithm 1. The proposed version of the Minimum Cost Path method.

```
01 function MCP (C):
04 insert the points in list L
05 while not is empty (L):
06 take voxel m from L
07 Compute its distance to remaining points
08 Connect it to the nearest point
09 return 0
```

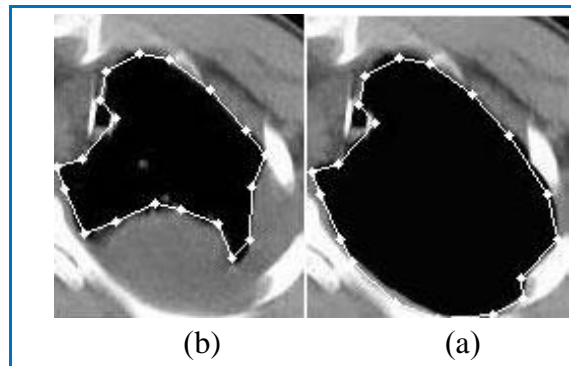


Figure 4. The final result after applying the algorithm (a) in a normal lung and (b) in an abnormal lung.

2.1.2. Using the triangulation method for determining the mass area

In this step the resulting graph, in the previous step, is processed and consists of two stages. In the first stage the mass boundary is determined by a dotted line and in the second, the area of the determined part is calculated. These two stages are discussed in the following.

a. Determining the mass boundary in the figure

This stage, too, can be divided into two steps. The center of the image is determined first and its distance to the vertices of the graph is, then, measured. The distances are, next, placed in an array, respectively, so that, the highest point is considered as the starting point and its value is placed in the first array (moving clockwise). The distance to the nodes is determined in Figure 5. To explain further, the distances are displayed with white lines. At the end of this step, we will have two arrays $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$ (with $a_i, b_i = d \Rightarrow R$), where elements are the distance between the graph vertices to the center of both images (the normal and the abnormal lung).

The two arrays created in the first step are compared in the second step. If the difference of both peer to peer elements is greater than the threshold value, there is a sign of an abnormal mass in the lung and these points are marked by a dotted line. The threshold value is obtained through trial and error method. Finally the threshold value is considered equal to 5 in our experiments. The location of the abnormality is determined as Figure 6.

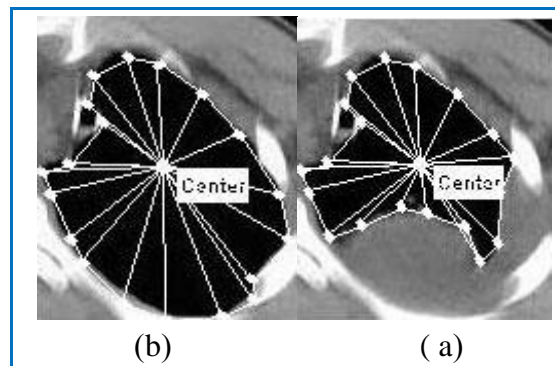


Figure 5. Showing the distance between the vertices of the graph to the center (a) in a normal lung (b) in an abnormal lung

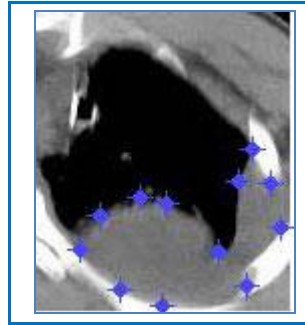


Figure 6. Determining the mass boundary by a dotted line

b. Calculation of the mass area

The triangulation method is used in this stage to calculate the mass area. Each shape is divided into a number of triangles, as shown in Figure 5. Then, we can calculate the area of the triangles and the total area will be the sum of the areas, so that the area of a triangle is equal to $\{\sqrt{s(s-a)(s-b)(s-c)}\}$, where a,b and c are the lengths of the sides of the triangle and $\{s=(a+b+c)/(2)\}$. The area is calculated for both normal and abnormal shapes and the difference between the two areas is the area of the mass.

2.2. The second method, using image processing

In previous steps, the graph classification method was used and the area of the mass was calculated. We go through all those steps by using the conventional method in image processing to calculate the mass area and determine the accuracy of the calculated area in the first method. To achieve this, the results of both methods are compared and the accuracy rate is measured. The proposed method includes two stages for this step. The boundary of the masses is determined first, so that the mass is separated from the other parts of the image completely and the area of the separated part is calculated in the second stage.

2.2.1. Determining the mass and separating it from the image

Color difference is used for determining the mass and separating it from the image, so that the changes in colors in the surrounding area are examined. Then, the color difference between two images (the normal and the abnormal shapes) can determine the mass. As it can be seen in Figure 7, the dotted area in Figure 6 is separated from the other parts completely.

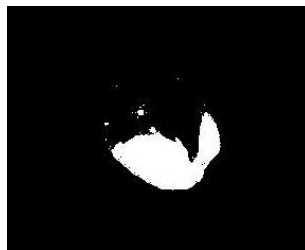


Figure 7. Determining the mass by color difference

2.2.2. Calculation of the mass Area

Considering that the central colors are adjusted, all the values of array elements in Figure 7 are black and white pixels and the gray nodes are eliminated. So, the mass area is obtained by counting the number of white pixels.

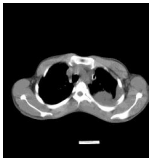

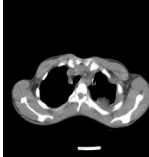
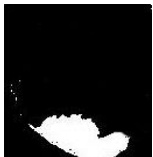

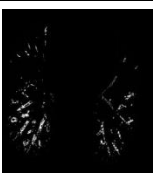
2.3. Comparing the two methods and determining the accuracy of the method


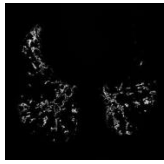
In this section, we will present the results obtained through our proposed method and its accuracy will be measured. For this purpose, the calculated areas of two methods are compared. The results of the images shown in figure 1 are presented in Table 1. The following formula is used to calculate the accuracy.

$$accuracy = (the\ area\ obtained\ in\ the\ first\ method) / (the\ area\ obtained\ in\ the\ second\ method)$$

The accuracy of the proposed method was seen to be 90%. This difference is due to the few numbers of the points which are considered as the vertices of the graph in the first method. By increasing the number of points and reducing the distances between them, the calculated result will be more accurate. The algorithm works for noises and small variations, does not require much memory according to existing standards, and is fast enough to be used in a clinical setting. Of course, as it is clear in Table 1, the larger the size of the mass, the more the accuracy. The ROC plots are also shown in Figure 8 and, as the figure illustrates, the graph of the first method (triangulation method) is more uniform. This indicates that the first method is more efficient than the second method and shows significant improvement.

Table 1. The result of the proposed method

Segmented image	Masse size	Masse gray level	Decision	Error (%)
		1699.3	15193	11
		1413.5	15062	9
		78.8800	986	8

		214.8979	3031	7
---	---	----------	------	---

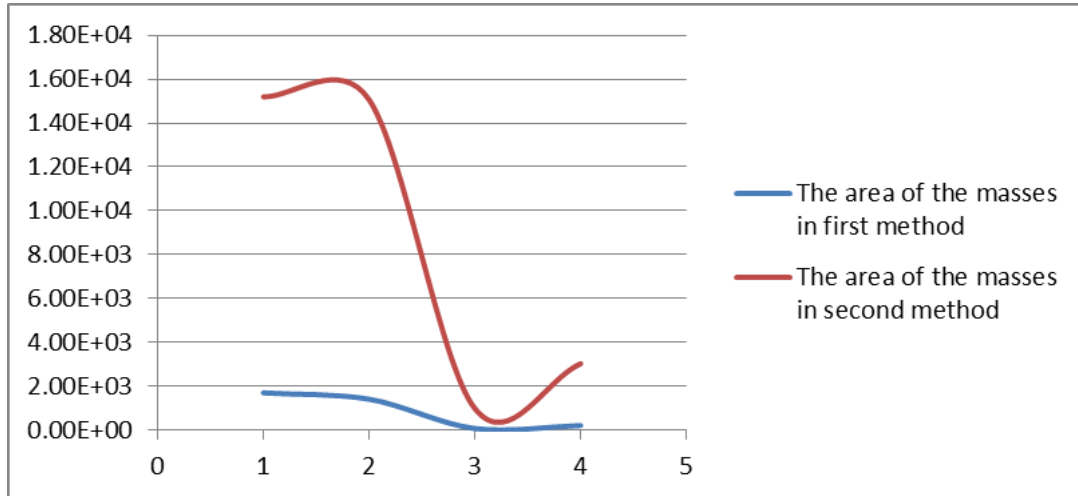


Figure 8. ROC plots for the proposed method

3. Conclusion

Considering the necessity of early and in-time detection of lung cancers, we have presented a new method, based on CT image processing that consists of three steps. In the first step, stent grafts were segmented from CT data, so that, the location of the mass will be determined and its area will be calculated by using the graph classification method. This step is performed non-automatically by an expert. Nevertheless, through some methods like neural network or genetic algorithm, it can be performed automatically in future. In the second step, the location of the mass will be detected by using the image processing techniques. Then, the mass is separated from the other parts of the image completely and the area of the separated part will be calculated. In the final step, the result of two previous steps is compared and the accuracy of the method is obtained.

Our proposed method is highly effective in decreasing human errors in the diagnosis of the masses in images. Additionally, the algorithm does not require much memory and is fast enough to be used in a clinical setting. Some CT images, received from an MRI center, were reviewed and analyzed by the proposed method. The results were compared with the diagnosis of the expert radiologists. The speed and the accuracy of the results revealed a significant improvement. Experiments show that the sensitivity of the proposed software system is more than 90%. In future studies, we will try to decrease the human intervention and to increase the accuracy. Accordingly, by using automatic detection of the points in the graph and designing better grafts, the errors will decrease. By better understanding of the reasons for graph failures, we can also design a better graph in future.

References

- [1] G. DeNunzio, A. Massafra, R. Cataldo, I. DeMitre, M. Peccarisi, M. E. Fantacci, *et al.*, "Approaches to juxta-pleural nodule detection in CT images within the MAGIC-5 Collaboration," *Nuclear Instruments and Methods in Physics Research*, pp. 5103-5105, 2011.
- [2] X.-Y. Wang and J. M. Garibaldi, "Simulated Annealing Fuzzy Clustering in Cancer Diagnosis," *Automated Scheduling, Optimisation and Planning (ASAP) Research Group, Department of Computer Science & Information Technology, The University of Nottingham, Jubilee Campus, Wollaton Road, United Kingdom*, vol. 29, pp. 61-70, 2005.
- [3] M. F and F. CHG, "Molecular Biology of Cancer. Springer-Verlag New York, Incorporated," 1997.
- [4] K. KW and V. B, "Genetic Basis of Human Cancer. McGraw-Hill," 2002.
- [5] L. MS and S. GV, "Genetics of Cancer: Genes Associated With Cancer Invasion, Metastasis and Cell Proliferation. Academic Press , London.," 1997.
- [6] K. R and T. M, "Molecular Biology in Cancer Medicine. Pearson education, Harlow," 1999.
- [7] S. E. Mahmoudi, A. Akhondi-Asl, R. Rahmani, S. Faghih-Roohi, V. Taimouri, A. Sabouri, *et al.*, "Web-based interactive 2D/3D medical image processing and visualization software," *computer methods and programs in biomedicine*, vol. 98, pp. 172-182, 2010.
- [8] M. Safdari, A. Karimian, and M. Yazdchi, "Metal Artifact Reduction in CT images of Head by Image Processing Techniques," *physics Procedia*, vol. 22, pp. 209-211, 2011.
- [9] A. Klein, W. K. Renema, L. J. Oostveen, L. J. S. Kool, and C. H. Slump, "A segmentation method for stentgrafts in the abdominal aorta from ECG-gated CTA data," *Proceedings of SPIE*, p. 69160R, 2008.
- [10] J. N. Kaftan, H. Tek, T. Aach, J. P. W. Pluim, and B. M. Dawant, "A two-stage approach for fully automatic segmentation of venous vascular structures in liver CT images," *Proceedings of SPIE*, p. 725911, 2009.
- [11] M. Freiman, L. Joskowicz, J. Sosna, M. I. Miga, and K. H. Wong, "A variational method for vessels segmentation: algorithm and application to liver vessels visualization," *Proceedings of SPIE*, p. 72610H, 2009.
- [12] A. Manduca, J. G. Fletcher, R. J. Wentz, R. C. Shields, T. J. Vrtiska, H. Siddiki, *et al.*, "Reproducibility of aortic pulsatility measurements from ECG-gated abdominal CTA in patients with abdominal aortic aneurysms," *Proceedings of SPIE*, p. 72620L, 2009.
- [13] S. Worz, W. J. Godinez, K. Rohr, J. P. W. Pluim, and B. M. Dawant, "Segmentation of 3D tubular structures based on 3D intensity models and particle filter tracking," *Proceedings of SPIE*, p. 72591P, 2009.
- [14] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," *Lect. Notes Comput. Sci*, vol. 1496, pp. 130-137, 1998.
- [15] C. Jacobs, C. I. Sánchez, S. C. Saur, T. Twellmann, P. A. de Jong, and B. van Ginneken, "Computer-aided detection of ground glass nodules in thoracic CT images using shape, intensity and context features.," *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 14, pp. 207-214, 2011.
- [16] A. Retico, P. Delogu, M. E. Fantacci, I. Gori, and A. Preite Martinez, "Lung nodule detection in low-dose and thin-slice computed tomography," *Computers in Biology and Medicine* vol. 38, pp. 525-534, 2008.
- [17] Aristofanes, c.silva, p. cezar, p.carvalho, and m. gattass, "Analysis of spatial variability using geostatistical functions for diagnosis of lung nodule in computerized tomography images," *Springer-Verlag London Limited 2004-short paper*, pp. 227-234, 2004.
- [18] K. Kanazawa, Y. Kawata, H. S. N. Niki, H. Ohmatsu, R. Kakinuma, M. Kaneko, *et al.*, "Computer-aided diagnostic system for pulmonary nodules using helical CT images," *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, vol. 1496, pp. 449-456, 1998.
- [19] M. A. Mazurowski, J. Y. Lo, B. P. Harrawood, and G. D. Tourassi, "Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis," *Journal of Biomedical Informatics*, vol. 44, pp. 815-823, 2011.