

Segmentação Automática de Candidatos a Nódulos Pulmonares em Imagens de Tomografia Computadorizada

Maria J. C. de Moura¹, Alcilene D. de Sousa¹, Ivo A. de Oliveira¹, Laércio N. Mesquita¹, Patrícia M. L. de L. Drumond¹

¹Campus Senador Helvídio Nunes de Barros - Universidade Federal do Piauí (UFPI)
Picos – PI – Brasil

{maria.jaquelane, ivoalves.analista, laercio.hag}@gmail.com,
alcilene@ufpi.edu.br, patymedy@hotmail.com

Abstract. *This paper presents an algorithm for automatic segmentation of pulmonary nodules candidates in chest computed tomography images. The methodology includes acquisition images, noise elimination, segmentation of pulmonary parenchyma and segmentation pulmonary nodules candidates. The use of the filter wiener and the application of ideal threshold ensures to the algorithm a significant improvement in results, allowing to detect a greater number of nodules on the images. The tests were conducted using a set of images of the base LIDC-IDRI, containing 708 nodules. The test results showed that the algorithm located 93.08% of the nodules considered.*

Resumo. *Este trabalho apresenta um algoritmo para segmentação automática de candidatos a nódulos pulmonares em imagens de Tomografia Computadorizada do tórax. A metodologia empregada inclui aquisição das imagens, eliminação de ruídos, segmentação do parênquima pulmonar e segmentação dos candidatos a nódulos pulmonares. O uso do filtro wiener e a aplicação do limiar ideal garante ao algoritmo uma melhora significativa nos resultados, permitindo detectar um maior número de nódulos nas imagens. Os testes foram realizados utilizando um conjunto de imagens da base LIDC-IDRI, contendo 708 nódulos. Os resultados do teste mostraram que o algoritmo localizou 93,08% dos nódulos considerados.*

1.Introdução

O câncer de pulmão é o mais comum de todos os tumores malignos e uma das principais causas de morte, apresentando aumento de 2% por ano na sua incidência mundial. Na maioria dos casos (cerca de 90%), o câncer de pulmão está associado ao consumo de tabaco. A sobrevida média cumulativa total em cinco anos varia entre 13 e 21% em países desenvolvidos e entre 7 e 10% nos países em desenvolvimento [INCA 2015]. Dessa forma, o câncer de pulmão se destaca por apresentar uma elevada taxa de incidência e de mortalidade.

A Tomografia Computadorizada (TC) é a modalidade de imagem utilizada para diagnosticar e caracterizar nódulos pulmonares. Comparada à convencional radiografia do tórax, a imagem de TC apresenta alta resolução e permite detectar nódulos pequenos e/ou com baixo contraste [Lee *et al.* 2012]. Devido a boa qualidade, a TC do tórax facilita a detecção e o acompanhamento da evolução dos nódulos pulmonares.

O diagnóstico precoce do câncer de pulmão através de TC em grupos de risco, como fumantes e ex-fumantes com níveis elevados de tabaco, pode afetar os resultados,

contribuindo para aumentar as chances de sobrevivência dos pacientes [Santos *et al.* 2014]. No entanto, é demorado para os radiologistas identificarem um grande número de lesões anormais a partir das imagens de TC. Portanto, o problema de reconhecer lesões automaticamente em imagens de TC do pulmão para auxiliar os radiologistas no diagnóstico de doenças pulmonares tem recebido bastante atenção nos últimos anos [Ma *et al.* 2015].

Várias pesquisas surgiram com o objetivo de desenvolver sistemas para a Detecção Assistida por Computador (*Computer-Aided Detection* - CAD) e para o Diagnóstico Assistido por Computador (*Computer-Aided Diagnosis* - CADx) do câncer de pulmão. Esses sistemas ajudam os especialistas a lidar com um grande volume de informações dos pacientes e fornecer um diagnóstico eficiente.

A detecção de nódulos pulmonares depende da segmentação, que é responsável por fornecer as regiões da imagem que serão analisadas e classificadas como nódulos. Uma boa técnica de segmentação deve ser capaz de incluir o maior número possível de nódulos presentes nas imagens e ao mesmo tempo excluir regiões desnecessárias. Este trabalho apresenta um algoritmo para segmentação automática de candidatos a nódulos pulmonares em imagens de TC do tórax. A metodologia empregada inclui aquisição das imagens, eliminação de ruídos, segmentação do parênquima pulmonar e segmentação dos candidatos a nódulos pulmonares.

O artigo está organizado da seguinte forma: na Seção 2, são apresentados os principais trabalhos relacionados; na Seção 3 são descritos os métodos empregados; na Seção 4, são analisados os resultados obtidos; e a Seção 5 apresenta as conclusões e trabalhos futuros.

2.Trabalhos Relacionados

A literatura possui diversos trabalhos voltados para a análise de imagens médicas com o objetivo de detectar doenças pulmonares. Um algoritmo de detecção automática de nódulos pulmonares solitários é proposto por Shao *et al.* (2012), contendo as seguintes etapas: pré-processamento, segmentação do parênquima pulmonar, localização dos candidatos a nódulos, extração das características e classificação por meio de SVM (*Support Vector Machine*). Esse sistema obteve uma sensibilidade de 89,5%, precisão de 90,4% e especificidade de 90,5%.

Netto *et al.* (2012), apresenta uma metodologia para segmentação automática de nódulos pulmonares, que consiste em extrair o parênquima pulmonar, aplicar o algoritmo de aprendizagem competitiva *Growing Neural Gas* (GNG) para obter as estruturas internas ao parênquima pulmonar (nódulos, vasos sanguíneos, brônquios, etc.), separar as estruturas semelhantes a nódulos pulmonares de outras estruturas, tais como vasos e brônquios, e por fim, classificar as estruturas utilizando SVM. Os resultados experimentais conferiram a essa metodologia sensibilidade de 86%, especificidade de 91%, precisão de 91% e taxa de falso positivo por exame de 0,138.

Em Jacobs *et al.* (2013) é apresentado um sistema de detecção automática para nódulos pulmonares semi-sólidos. As etapas que compõe esse sistema são a segmentação dos candidatos a nódulos, o cálculo de características dos candidatos (como intensidade, textura, características de forma e contexto) e o uso de classificadores para classificar os candidatos em nódulos e não-nódulos. Os resultados mostraram uma sensibilidade de 80% e uma média de 1 falso positivo por exame.

Em seu trabalho, Carvalho Filho (2013), descreve uma metodologia para a detecção automática de nódulos pulmonares, utilizando técnicas de segmentação, extração de características e classificação. A segmentação dos candidatos a nódulos é feita utilizando o algoritmo *Quality Threshold Clustering* (QT) e a classificação é realizada com base na extração de características, como forma e textura e o uso de SVM. Esse algoritmo alcançou sensibilidade de 85,91%, especificidade de 97,70%, precisão de 97,55% e uma taxa de falso positivo de 1,82 por exame.

Zhou *et al.* (2014) propõe um método automático para incluir os nódulos justapleurais e vasos presentes nas bordas do pulmão, que muitas vezes são perdidos durante a segmentação dos nódulos pulmonares. O algoritmo possui os seguintes passos: pré-processamento das imagens, extração do tórax, identificação do pulmão e correção do contorno do pulmão. Os resultados mostraram que esse algoritmo é capaz de incluir todos os nódulos justapleurais e obter uma taxa de sobreposição de volume médio de $95,81 \pm 0,89\%$ e uma média de distância de borda média de $0,63 \pm 0,09$ mm comparados com os resultados de segmentação manuais.

Ramalho *et al.* (2014) relata um sistema de detecção de doenças pulmonares utilizando extração de características e aprendizagem de máquina extrema. Neste trabalho, foi empregado o método *Adaptive Crisp Active Contour Models* (ACACM) para segmentar as estruturas internas do pulmão (nódulos, vasos e etc.) e os descritores SIM (*Spatial Interdependence Matrix*), GLCM (*Gray Level Cooccurrence Matrix*) e VIF (*Visual Information Fidelity*) para extrair as características das imagens segmentadas de TC, sendo que o SIM obteve melhor desempenho. A aprendizagem de máquina extrema é utilizada para realizar a classificação em três classes: fibrose pulmonar, doença pulmonar obstrutiva crônica ou pulmão saudável. Os resultados experimentais alcançaram uma precisão de 96%.

Messay *et al.* (2015) apresenta um método de segmentação de nódulos pulmonares em exames de TC do tórax, composto de três novos algoritmos de segmentação: um sistema totalmente automático (*Fully-automated*), um sistema semiautomático (*Semi-automated*) e um sistema híbrido. A aplicação sequencial dos três algoritmos e o uso de uma rede neural de regressão para calcular uma série de características de cada candidato a nódulo pulmonar, permite melhorar os resultados da segmentação. Além disso, a comparação com outros dez métodos de segmentação existentes, mostrou que este trabalho apresenta um melhor desempenho pela sobreposição média de 69,23%.

Vivanti *et al.* (2015) desenvolveu um algoritmo de segmentação automática de tumores pulmonares em exames de TC, obtidas durante o acompanhamento de pacientes. O processo de segmentação consiste em registrar a imagem TC de linha de base e a delimitação dos tumores com as imagens de acompanhamento, realizar a segmentação inicial dos tumores nas imagens de acompanhamento, detectar e corrigir os vazamentos da segmentação dos tumores de acompanhamento, e regularizar o limite do tumor. Este algoritmo apresenta resultados significativos, como indicado pelo erro médio de sobreposição DICE de 14,5% e fornece medições relevantes para os tumores pulmonares.

Em todos os métodos descritos nesta seção, é observado que a segmentação aparece como uma etapa fundamental para a detecção de doenças pulmonares, influenciando o resultado das etapas seguintes. Um novo método de segmentação utilizando técnicas mais simples é descrito na próxima seção.

3. Metodo Proposto

Esta seção descreve o método proposto para segmentação automática de candidatos a nódulos pulmonares em exames de TC. A método compreende quatro estágios, a saber: a aquisição das imagens, eliminação de ruídos, segmentação do parênquima pulmonar e segmentação dos candidatos a nódulos. A base de imagens utilizada é a LIDC-IDRI [Armato III I *et al.* 2011]. A eliminação de ruídos tem como objetivo melhorar a qualidade das imagens. A segmentação ocorre em dois momentos, primeiro é segmentado o parênquima pulmonar e depois é segmentado os candidatos a nódulos. A Figura 1 ilustra esses quatro estágios.

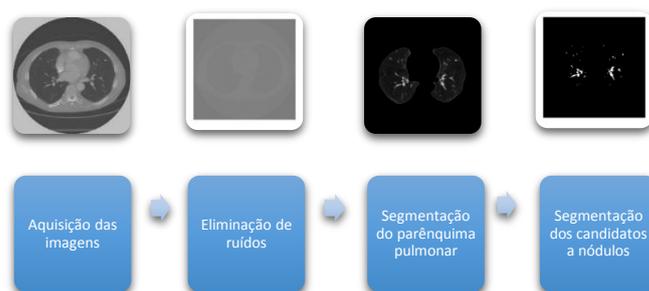


Figura 1. Estágios da metodologia para a segmentação de candidatos a nódulos pulmonares.

3.1. Aquisição das Imagens

No desenvolvimento deste trabalho foi utilizado a coleção de imagens *Lung Image Database Consortium - LIDC-IDRI* [Armato III I *et al.* 2011], por ser uma base que preserva a identidade dos pacientes e apresenta a avaliação dos exames realizada por quatro especialistas. A base é um recurso internacional acessível via *web* para o desenvolvimento, treinamento e avaliação do diagnóstico de sistemas para a detecção e diagnóstico de câncer de pulmão.

A base LIDC-IDRI é disponibilizada pelo *National Cancer Institute of EUA* (NCI), sendo resultado da associação entre o consórcio de instituições conhecido como *Lung Image Database Consortium* (LIDC) e o *Image Database Resource Initiative* (IDRI). A base contém 1012 exames com uma quantidade variável de fatias por exame e um arquivo individual para cada exame contendo as marcações feitas pelos especialistas. O arquivo especifica as coordenadas dos nódulos, as fatias que o compõe e as características dos nódulos. As imagens estão no formato DICOM (*Digital Imaging Communications in Medicine*), possuem 16 bits por voxel (representação tridimensional de um pixel) e dimensões de 512 X 512. O desafio da detecção se deve ao fato das imagens serem obtidas por tomógrafos distintos.

3.2. Eliminação de Ruídos

Após a aquisição das imagens, o próximo passo é realizar um pré-processamento para eliminar os ruídos que podem existir nessas imagens, o que facilita a detecção dos candidatos a nódulos. Alguns filtros foram testados como os filtros da mediana e gaussiano, porém o filtro *wiener* apresentou um melhor resultado.

O filtro *wiener* é um filtro de suavização que realiza a filtragem com base no cálculo da média e da variância da vizinhança de cada pixel. A Figura 2 mostra o resultado da aplicação do filtro *wiener* na imagem original.

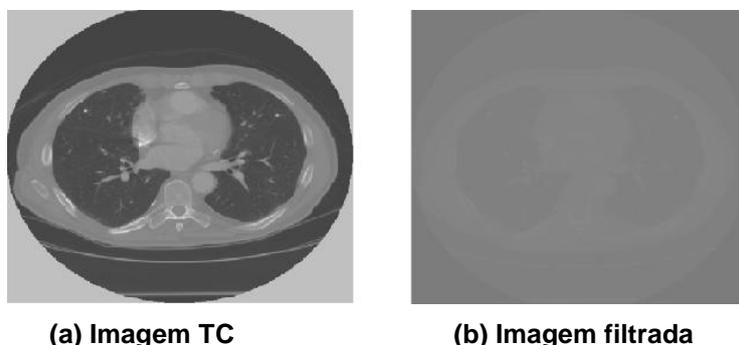


Figura 2. Resultado da aplicação do filtro *wiener*.

3.3.Segmentação do Parênquima Pulmonar

A próxima etapa é a segmentação do parênquima pulmonar. Nessa fase, é necessário limiarizar a imagem filtrada gerando uma imagem binária como ilustrado na Figura 3. O processo de limiarização consiste em separar as regiões de uma imagem utilizando um determinado limiar. No trabalho de Shao *et al.* (2012), o limiar é calculado a partir de um método iterativo, enquanto que neste trabalho é empregado um limiar ideal, cujo valor é 0,51 para imagens com níveis mais elevados de intensidade e 0,01 para imagens com níveis de intensidade mais baixos. Vários valores de limiar foram testados, porém obtemos os melhores resultados com esses valores, garantindo a segmentação completa do parênquima pulmonar.

Inicialmente a imagem filtrada é binarizada utilizando o limiar 0,51. Depois é calculado o maior valor de intensidade de uma coluna central da imagem. Se esse valor for igual a 0, significa que a imagem não foi limiarizada corretamente. Então, a imagem filtrada é binarizada novamente utilizando o limiar 0,01. Assim, o limiar é definido de acordo com a imagem considerada. A partir da imagem binária é obtido uma máscara como ilustrado na Figura 4 (a). A máscara é gerada da seguinte forma: segmentando a caixa torácica, segmentando a região exterior a caixa torácica e aplicando a operação *not* nesta imagem. Cada um desses passos gera uma nova imagem que é a entrada do passo seguinte. A operação *not* em uma imagem binária gera uma nova imagem atribuindo aos pixels cujo valor é igual a 1 o valor 0 e aos pixels cujo valor é igual 0 o valor 1.

A Figura 4 (b) mostra a segmentação realizada com a máscara, eliminando a região exterior a caixa torácica. Essa imagem é então transformada em uma imagem binária e depois é invertida por meio da operação *not* gerando uma nova imagem. Em seguida é aplicado a operação *and* com essa imagem e a máscara, resultando em uma imagem segmentada do parênquima pulmonar. A operação *and* é aplicada a duas imagens binárias para gerar uma nova imagem contendo regiões comuns a essas duas imagens. Uma operação de fechamento utilizando um elemento estruturante circular com raio de dez unidades é feita para reconstruir o parênquima pulmonar como mostrado na Figura 4 (c). O parênquima pulmonar é então obtido a partir desta imagem, como mostrado na Figura 4 (d).

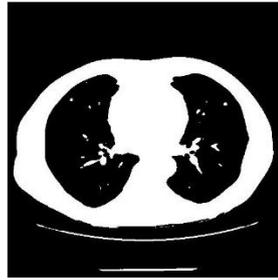


Figura 3. Imagem binária.

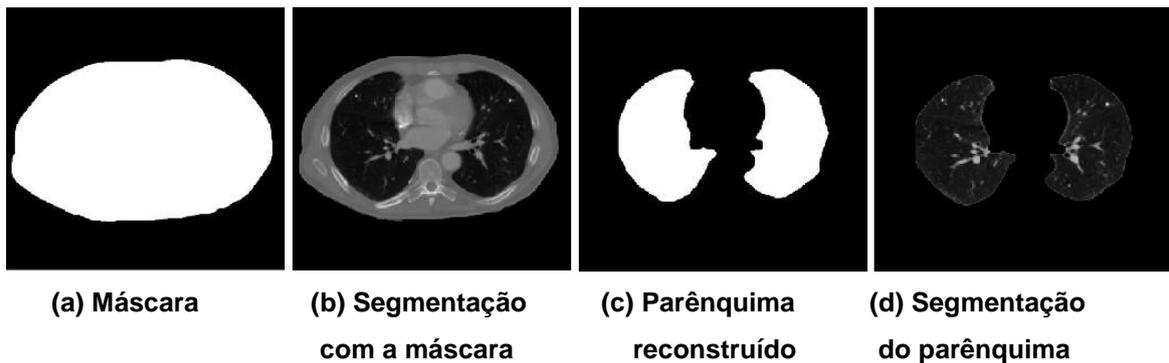


Figura 4. Resultado da segmentação do parênquima pulmonar.

3.4.Segmentação dos Candidatos a Nódulos Pulmonares

A etapa final deste trabalho é a segmentação dos candidatos a nódulos pulmonares, que consiste em segmentar as estruturas interiores ao parênquima pulmonar (nódulos, traqueia, vasos sanguíneos e etc.).

Neste trabalho, a segmentação dos candidatos a nódulos é feita de forma simples, convertendo a imagem resultante da seção 3.3 em uma imagem binária, utilizando um limiar menor do que o valor definido na etapa de segmentação do parênquima pulmonar. A partir de testes realizados foi observado que diminuir esse limiar permite detectar um maior número de nódulos, que seriam perdidos por apresentar um nível de intensidade semelhante ao do pulmão, ou estar localizado muito próximo a borda. Assim, o limiar 0,51 foi atualizado para 0,504 e o limiar 0,01 foi atualizado para 0,005. A imagem binária resultante da aplicação desses limiares apresenta as bordas do pulmão. Por isso, o passo seguinte consiste em eliminar essas bordas. A eliminação da borda é feita segmentando a borda do parênquima e subtraindo da imagem segmentada a imagem que contém a borda do pulmão, resultando em uma imagem com os candidatos a nódulos pulmonares como mostra a Figura 5 (b).

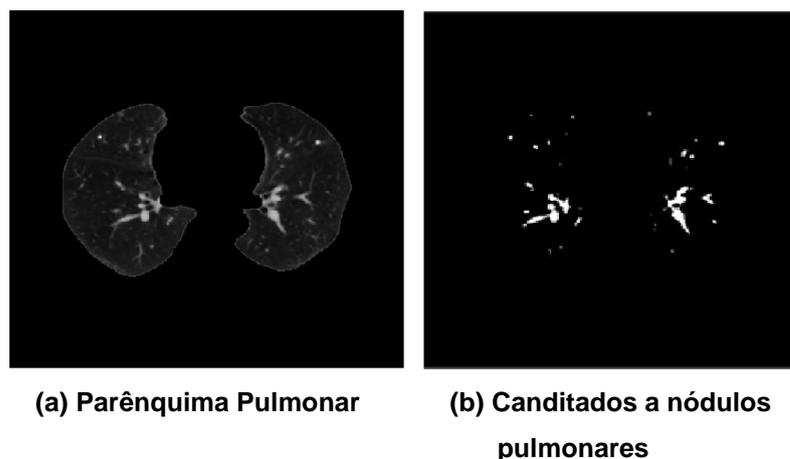


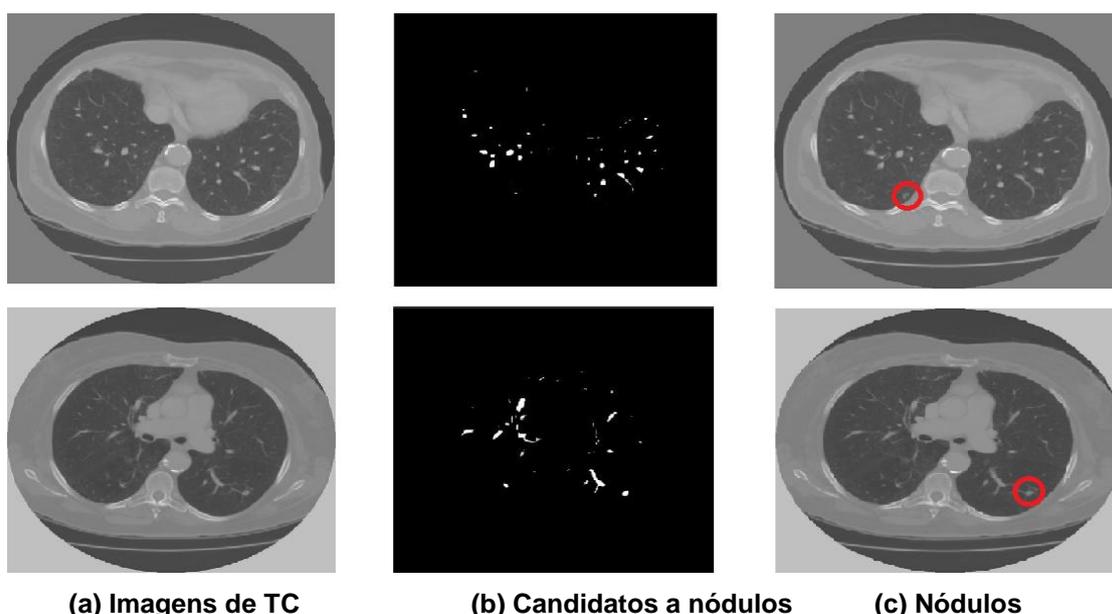
Figura 5. Resultado da segmentação dos candidatos a nódulos pulmonares.

4.Resultados Experimentais e Discussão

Para demonstrar a eficácia do método de segmentação de candidatos a nódulos pulmonares proposto foram realizados testes utilizando 18 exames da base LIDC-IDRI, contendo 708 nódulos. A base LIDC-IDRI contém informações das coordenadas do contorno dos nódulos e as fatias que compõem esses nódulos. Conhecendo a localização exata do nódulo em determinada fatia, o teste consiste em aplicar o método de segmentação nesta imagem e verificar se dentre as estruturas segmentadas alguma corresponde ao nódulo considerado.

Os resultados dos testes mostraram que o algoritmo localizou 659 nódulos e perdeu 49, ou seja, o algoritmo localizou 93,08% dos nódulos especificados. O objetivo dos testes é comprovar que o método é capaz de segmentar os nódulos presentes nas imagens. Uma parte dos nódulos foi perdida durante a segmentação por apresentar um nível de intensidade muito próximo ao do parênquima pulmonar ou por estar ligado a borda pulmonar.

A dificuldade de validar o método proposto se deve ao fato de não haver na literatura os resultados referentes a fase de segmentação que possam ser comparados aos resultados obtidos nesta metodologia. Os trabalhos relacionados não apresentam os resultados isolados das etapas de segmentação, extração de características e classificação, mas os resultados finais da detecção dos nódulos pulmonares. A Figura 6 mostra os resultados do algoritmo de segmentação dos candidatos a nódulos pulmonares presentes nas imagens de TC do tórax.



(a) Imagens de TC

(b) Candidatos a nódulos

(c) Nódulos

Figura 6. Resultado do algoritmo de segmentação de candidatos a nódulos pulmonares.

A Figura 6 (a) mostra duas imagens de TC do tórax, a Figura 6 (b) contém as regiões que correspondem aos candidatos a nódulos e a Figura 6 (c) mostra as mesmas imagens da Figura 6 (a) com a indicação do nódulos. É possível observar que os nódulos indicados estão presentes nas imagens segmentadas.

5. Conclusões e Trabalhos Futuros

Este trabalho apresenta uma metodologia para a segmentação automática de candidatos a nódulos pulmonares em Tomografias Computadorizadas do tórax. Essa metodologia compreende a aquisição das imagens, a eliminação de ruídos, segmentação do parênquima pulmonar e a segmentação dos candidatos a nódulos pulmonares.

O algoritmo tem como entrada uma imagem de TC e a partir dela gera uma imagem contendo as regiões segmentadas que correspondem aos candidatos a nódulos. O conjunto de teste é composto de 708 nódulos, dos quais o algoritmo conseguiu localizar 93,08% dos nódulos considerados. O uso do filtro *wiener* e a aplicação do limiar ideal garante ao algoritmo uma melhora significativa nos resultados, permitindo detectar um maior número de nódulos nas imagens. Com o filtro foi possível suavizar a imagem e o limiar ideal permitiu identificar os candidatos. Em alguns casos, não foi possível localizar os nódulos devido ao nível de intensidade ser semelhante ao do parênquima pulmonar ou por estar localizado muito próximo a borda. Os resultados obtidos demonstram a eficiência do método proposto.

A vantagem desta metodologia sobre outros métodos de segmentação de candidatos a nódulos, é a facilidade de implementação, devido a menor complexidade e que resulta em uma execução mais rápida. Em outros trabalhos por exemplo, são utilizadas técnicas como crescimento de região, uso de histogramas para cálculo de um limiar ideal, algoritmos de agrupamento, redes neurais, entre outros. Além de ser simples, o método proposto fornece bons resultados de segmentação, comprovado pela porcentagem dos nódulos segmentados nos testes.

Para trabalhos futuros pretende-se aperfeiçoar o método de segmentação para recuperar os nódulos que foram perdidos, testar mais imagens da base LIDC-IDRI e de outras bases de imagens, dar continuidade a construção de um sistema CAD, desenvolvendo um método que permita extrair as características das estruturas segmentadas com o algoritmo proposto neste trabalho, e classificar essas estruturas como nódulos e não-nódulos.

6.Referências

- Armato III, S. G.; et al. (2010) “Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study”, *Medical Image Analysis*, p. 707-722.
- Carvalho Filho, A. O. de. (2013) “Detecção automática de nódulos pulmonares solitários usando quality threshold clustering e mvr”, *Dissertação de Mestrado na área de Ciência da Computação, (Programa de Pós-Graduação em Engenharia de Eletricidade), Universidade Federal do Maranhão, São Luís.*
- Instituto Nacional do Câncer – INCA (2015), *Mistério da Saúde, Câncer de pulmão*, <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pulmao/definicao>, abril.
- Jacobs, C.; Rikxoort, E.M.V.; Twellmann, T.; Scholten, E. Th.; Jong, P. A.; De. Kuhnigk, J. M.; Oudkerk, M.; Koning, H. J.; De Prokop, M.; Prokop, C. S.; Ginneken, B. V.(2013) “Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images”, *Medical Image Analysis*, p. 374-384.
- Lee, S. L. A.; Kousani, A. Z.; Hu, E. J. (2012) “Automated detection of lung nodules in computed tomography images: review”, *Mach.Vis.Appl.*23 (1), p. 151–163.
- Ma. L.; Liu, X.; Song, L.; Zhou, C.; Zhao, X., Zhao, Y. (2015) “A new classifier fusion method based on historical and on-line classification reliability for recognizing common CT imaging signs of lung diseases”, *Computerized Medical Imaging and Graphics* 40, p. 39-48.
- Messay, T.; Hardie, R. C.; Tuinstra, T. R. (2015) “Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the Lung Image Database Consortium and Image Database Resource Initiative dataset”, *Medical Image Analysis*, p.48-62.
- Netto, S. M. B; Silva, A. C; Nunes, R. A; Gattass, M. (2012) “Automatic segmentation of lung nodules with growing neural gas and support vector machine”, *Computers in Biology and Medicine* 42, p. 1110–1121.
- Ramalho, G. L. B.; Filho, P. P. R.; Medeiros, F. N. S.; Cortez, P. C. (2014) “Lung disease detection using feature extraction and extreme learning machine”, *Revista Brasileira de Engenharia Biomédica*, p. 207-214.
- Santos, A. M.; Filho, A. O. C.; Silva, A. C.; Paiva, A. C.; Nunes, R. A.; Gattass, M. (2014) “Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM”, *Engineering Applications of Artificial Intelligence* 36, p. 27-39.

- Shao, H.; Cao, L.; Liu, Y. (2012) "A detection approach for solitary pulmonary nodules based on CT images", 2nd International Conference on Computer Science and Network Technology (ICCSNT).
- Shen, S.; Bui, A. A. T.; Cong, J.; Hsu, W. (2015) "An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy", *Computers in Biology Medicine* 57, p. 139-149.
- Vivanti, R.; Joskowicz, L.; Karaaslan, O. A.; Sosna, J. (2015) "Automatic lung tumor segmentation with leaks removal in follow-up CT studies", *Int J CARS*.
- Zhou, S.; Cheng, Y.; Tamura, S. (2014) "Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonar vessels on chest CT images", *Biomedical Signal Processing and Control*, p. 62-70.