

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

PEDRO HENRIQUE BANDEIRA DINIZ
ANÁLISE TEMPORAL DE NÓDULOS E MASSAS PULMONARES
UTILIZANDO ÍNDICES DE SIMILARIDADE

São Luís
2014

PEDRO HENRIQUE BANDEIRA DINIZ

**ANÁLISE TEMPORAL DE NÓDULOS E MASSAS PULMONARES
UTILIZANDO ÍNDICES DE SIMILARIDADE**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Eletricidade na área de concentração Ciência da Computação.

Orientador: Prof. Dr. Aristófanés
Corrêa Silva

São Luís
2014

Diniz, Pedro Henrique Bandeira

Análise temporal de nódulos e massas pulmonares utilizando índices de similaridade / Pedro Henrique Bandeira Diniz. – São Luís, 2014.

113 f.

Orientador: Aristófanês Corrêa Silva.

Dissertação (Mestrado em Engenharia da Eletricidade) - Universidade Federal do Maranhão, 2014.

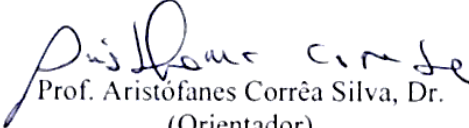
1. Nódulo pulmonar 2. Massa pulmonar 3. Processamento de imagens 4. Análise temporal 5. Índices de similaridade I. Análise temporal de nódulos e massas pulmonares utilizando índices de similaridade.

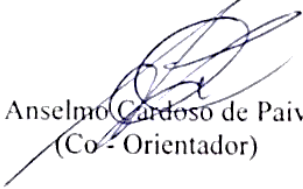
CDU 004.383.5 : 616.006


**ANÁLISE TEMPORAL DE NÓDULOS E MASSAS PULMONARES
UTILIZANDO ÍNDICES DE SIMILARIDADE**

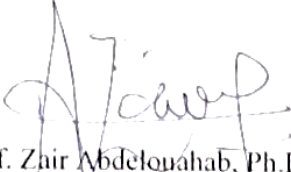
Pedro Henrique Bandeira Diniz

Dissertação aprovada em 03 de janeiro de 2014.


Prof. Aristófanês Corrêa Silva, Dr.
(Orientador)


Prof. Anselmo Cardoso de Paiva, Dr.
(Co - Orientador)


Prof. Rodolfo Acatauassú Nunes, Dr.
(Membro da Banca Examinadora)


Prof. Zair Abdelouahab, Ph.D.
(Membro da Banca Examinadora)

“O homem erudito é um descobridor de fatos que já existem - mas o homem sábio é um criador de valores que não existem e que ele faz existir.” (Albert Einstein)

À minha família, colegas de classe,
amigos e professores.

AGRADECIMENTOS

Deus em primeiro lugar.

Minha família pelo apoio durante todos os anos da minha vida.

Meus amigos, que me acompanharam no decorrer do curso. Em especial, os colegas do mestrado e do NCA que estiveram sempre dispostos a dar ideias e ajudar.

Meus professores, pelo conhecimento que me proporcionaram. Em especial, ao professor Aristófanes, pela orientação nessa dissertação e em diversos outros trabalhos no decorrer da minha vida acadêmica.

RESUMO

Com o advento dos métodos de imagem, o nódulo pulmonar vem se tornando a manifestação mais comum de câncer de pulmão e um dos mais letais de todos os cânceres. Uma forma de pesquisar a malignidade de um nódulo é analisar temporalmente suas medidas de geometria (forma) e textura (tecido). Entre medidas geométricas comumente utilizadas, a taxa de crescimento do nódulo constitui um dos métodos mais precisos não invasivos de aferição da malignidade. Acompanhada a outras medidas de textura obtidas no decorrer do tempo, obtém-se informações valiosas sobre o seu comportamento, de forma que o médico pode usar essas medidas na tomada de decisões. Sabendo disso, é importante a comparação do nódulo em exames extraídos em momentos diferentes. Uma etapa fundamental para essa comparação é verificar a correspondência entre nódulos em exames diferentes, de forma que seja possível determinar se um nódulo X em um exame A é o mesmo nódulo Y em um exame B . Devido a uma série de fatores anátomo-fisiológicos e de aquisição de imagens, um mesmo nódulo pode não estar exatamente na mesma localização em exames diferentes. Para corrigir esse problema, registros de imagem rígidos e deformáveis mostram-se eficientes. Uma vez estabelecida essa correspondência, é possível analisar as mudanças na textura do nódulo através de índices de similaridade. Nesse sentido, o objetivo desse trabalho é apresentar métodos para a análise quantitativa de mudanças de textura em nódulos. Para essa análise serão utilizadas imagens de tomografia computadorizada obtidas em momentos diferentes de um mesmo paciente. Além disso, será apresentado um método para verificar se nódulos encontrados em exames diferentes correspondem ao mesmo nódulo através da aplicação de registros de imagens.

Palavras-chave: Nódulo Pulmonar. Massa pulmonar. Processamento de imagens. Análise Temporal. Índices de Similaridade.

ABSTRACT

With the advent of imaging methods, the pulmonary nodule is becoming the most common manifestation of lung cancer and one of the most lethal of all cancers. Geometry (shape) and texture (tissue) measurements analyzed over time can be used to search the nodule malignancy. Among geometric measures commonly used, the nodule growth rate is one of the most accurate noninvasive methods to evaluate malignancy. Followed by other texture measures achieved over time, it is possible to get valuable information about nodules behavior, so that the doctor can use them to take related decisions. For these reasons, it is important to compare the nodule in exams applied at different moments. A key step for the comparison is to verify the correspondence between nodules of different exams. This correspondence is used to determine if a nodule X in the exam A is the same nodule Y in an exam B . Due to a number of anatomical and physiological factors and image acquisition, the same nodule cannot be in exactly the same location on different exams. To correct this problem, rigid and deformable image registration show up to be efficient. Once established this correspondence, it is possible to analyze the nodule texture changes through similarity indexes. In this sense, the aim of this work is to present methods for quantitative analysis of texture changes in lung nodules. For this analysis, it is used CT scans obtained at different moments from the same patient. Furthermore, it is presented a method to verify nodules found in different exams correspond to the same nodule by applying image registration.

Keywords: Lung nodule. Lung mass. Image processing. Temporal Analysis. Similarity Indexes.

LISTA DE FIGURAS

Figura 1 - Nódulo e outras estruturas do pulmão em uma fatia de tomografia computadorizada.....	26
Figura 2 - Massa pulmonar.....	27
Figura 3 - Tomografia computadorizada.....	27
Figura 4 - Interpolação linear entre dois pontos.....	32
Figura 5 - Processo do registro em um volume de imagens de uma massa.....	34
Figura 6 - Processo de um registro de imagens.....	35
Figura 7 - Registro rígido aplicado em imagens de TC de pulmão.....	37
Figura 8 - Registro <i>B-spline</i> aplicado em imagens de TC de pulmão.....	39
Figura 9 - Exemplo de funcionamento do <i>K-means</i>	42
Figura 10 - Limiarização de Otsu desenvolvida com o ITK e aplicada a uma imagem de massa pulmonar.....	45
Figura 11 - Fluxo da metodologia.....	46
Figura 12 - Massas sintéticas.....	50
Figura 13 - Fluxo do pré-processamento das imagens.....	51
Figura 14 - Exemplo de segmentação do tórax.....	52
Figura 15 - Exemplo de segmentação das estruturas externas do tórax.....	52
Figura 16 - Exemplo de segmentação do parênquima e vias aéreas.....	53
Figura 17 - Reconstrução da borda.....	53
Figura 18 - Funcionamento da remoção da traqueia e brônquios.....	54
Figura 19 - <i>Bounding box</i> de um nódulo.....	55
Figura 20 - Correspondência do nódulo em exames diferentes após o registro.....	56
Figura 21 - Extraindo subvolume contido na <i>bounding box</i> da marcação de uma massa.....	57
Figura 22 - Binarizando nódulo com base na marcação.....	58
Figura 23 - Aplicando a limiarização de Otsu na imagem da massa.....	58
Figura 24 - Criando máscara da massa.....	59
Figura 25 - Melhoramento do contorno da massa.....	59
Figura 26 - Criação de volume auxiliar com maiores dimensões.....	61
Figura 27 - Redesenhando o subvolume da massa sobre o volume auxiliar maior.....	62

Figura 28 - Aplicando o registro de imagens rígido e deformável <i>B-spline</i> nos subvolumes da massa.	63
Figura 29 - Criando volume de diferenças das médias entre dois subvolumes de uma massa.....	63
Figura 30 - Aplicando algoritmo <i>K-means</i> de três classes no volume de diferença das médias.	64
Figura 31 - Identificando os grupos do <i>K-means</i>	65
Figura 32 - Aplicando sucessivos crescimentos de regiões para identificar regiões desconexas.	66
Figura 33 - Binarização da massa.....	67
Figura 34 - Sucessivas extrações de bordas da massa binarizada.	67
Figura 35 - Agrupando bordas extraídas para formar as camadas.....	68
Figura 36 - Formando uma camada com espessura de dois <i>voxels</i>	68
Figura 37 - Aplicando o registro de rígido nos subvolumes da massa.....	69
Figura 38 – Isolando região nova da massa.	70
Figura 39 - Fatias do subvolume da primeira massa no exame mais antigo.	77
Figura 40 - Fatias do subvolume da primeira massa no exame mais recente.....	77
Figura 41 - Fatias do subvolume da segunda massa no exame mais antigo.	80
Figura 42 - Fatias do subvolume da segunda massa no exame mais recente.....	80
Figura 43 - Fatias do subvolume do primeiro nódulo no exame mais antigo.	83
Figura 44 - Fatias do subvolume do primeiro nódulo no exame mais recente.....	84
Figura 45 - Fatias do subvolume da massa no exame mais antigo.....	86
Figura 46 - Fatias do subvolume da massa no exame mais recente.....	86

LISTA DE TABELAS

Tabela 1 - Índices de similaridade. Adaptado de (Meyer, 2002).	44
Tabela 2 - Análise temporal global do par formado por duas cópias da primeira massa sintética.	73
Tabela 3 - Análise temporal global do par formado pela primeira e segunda massas sintéticas.	74
Tabela 4 - Análise temporal por camadas do par formado pela primeira e segunda massas sintéticas	74
Tabela 5 - Análise temporal das regiões de maiores diferenças do par formado pela primeira e segunda massas sintéticas.	75
Tabela 6 - Análise temporal global do par formado pela primeira e terceira massas sintéticas.	75
Tabela 7 - Análise temporal das regiões de maiores diferenças do par formado pela primeira e terceira massas sintéticas.	76
Tabela 8 - Análise temporal global da primeira massa da PLDADR.	78
Tabela 9 - Análise temporal das regiões de maiores diferenças da primeira massa da PLDADR.	78
Tabela 10 - Análise temporal das camadas da primeira massa da PLDADR.	79
Tabela 11 - Análise temporal das regiões iniciais e novas da primeira massa da PLDADR.	79
Tabela 12 - Análise temporal global da primeira massa da PLDADR.	81
Tabela 13 - Análise temporal das regiões de maiores diferenças da segunda massa da PLDADR.	81
Tabela 14 - Análise temporal das camadas da segunda massa da PLDADR.	82
Tabela 15 - Análise temporal das regiões iniciais e novas da segunda massa da PLDADR.	82
Tabela 16 - Análise temporal global da primeiro nódulo da base do <i>Volcano Challenge</i>	84
Tabela 17 - Análise temporal das camadas da primeiro nódulo da base do <i>Volcano Challenge</i>	85
Tabela 18 - Análise temporal das camadas da primeiro nódulo da base do <i>Volcano Challenge</i>	85

Tabela 19 - Análise temporal global da segundo nódulo da base do <i>Volcano Challenge</i>	87
Tabela 20 - Análise temporal das camadas da segundo nódulo da base do <i>Volcano Challenge</i>	87
Tabela 21 - Análise temporal das regiões iniciais e novas do segundo nódulo da base do <i>Volcano Challenge</i>	88

LISTA DE SIGLAS

CAD – *Computer-aided Diagnosis* (Diagnóstico Auxiliada por Computador)

DICOM – *Digital Imaging Communications in Medicine* (Representação Digital para Comunicação em Medicina) **ITK** – *Insight Segmentation and Registration Toolkit*

HU – *Hounsfield Unit* (Unidade de Hounsfield)

ITK – *Insight Segmentation and Registration Toolkit*

NPS – Nódulo Pulmonar Solitário

PLDADR – *Public Lung Database to Address Drug Response*

TC – Tomografia Computadorizada

VOI – Volume de Interesse

SUMÁRIO

1. INTRODUÇÃO	17
1.1. Motivação.....	19
1.2. Objetivos	21
1.3. Trabalhos relacionados.....	21
1.4. Organização do trabalho.....	22
2. FUNDAMENTAÇÃO TEÓRICA	24
2.1. Nódulo pulmonar.....	24
2.1.1. Nódulo pulmonar solitário	24
2.2. Massa pulmonar	26
2.3. Tomografia computadorizada	27
2.4. Padrão DICOM	28
2.5. Diagnóstico Auxiliado por Computador	29
2.6. Interpolação de imagens.....	31
2.6.1. Interpolação Linear	32
2.7. Registro de imagens	33
2.7.1. Registro rígido.....	36
2.7.2. Registro deformável <i>B-spline</i>	37
2.8. Agrupamento.....	40
2.8.1. <i>K-means</i>	40
2.9. Índices de similaridade.....	42
2.10. <i>Insight Segmentation and Registration Toolkit</i>	44
3. METODOLOGIA PROPOSTA.....	46
3.1. Aquisição de imagens.....	47
3.1.1. <i>Public Lung Database to Address Drug Response</i>	47
3.1.2. Base do <i>Volcano Challenge</i>	48
3.1.3. Base de massas sintéticas	49
3.2. Pré-processamento das imagens.....	50
3.2.1. Segmentação do parênquima e vias aéreas.....	51
3.2.2. Remoção da traqueia e brônquios	53
3.3. Correspondência entre nódulos / massas em exames diferentes	54
3.4. Pré-processamento do nódulo / massa.....	57
3.5. Definição das regiões de interesse do nódulo / massa.....	60

3.5.1.	Nódulo / massa inteiro.....	60
3.5.2.	Regiões de maiores mudanças.....	60
3.5.3.	Camadas	66
3.5.4.	Regiões novas.....	69
3.6.	Análise temporal de texturas	70
4.	RESULTADOS.....	72
4.1.1.	Massas sintéticas	72
4.1.2.	Massas da <i>Public Lung Database to Address Drug Response</i>	76
4.1.3.	Nódulos do <i>Volcano Challenge</i>	83
5.	CONCLUSÃO	89
5.1.	Avaliação do trabalho.....	89
5.2.	Trabalhos futuros.....	89
6.	REFERÊNCIAS	91
	APÊNDICE.....	94

1. INTRODUÇÃO

O estilo de vida da humanidade tem, ao longo tempo, trazido diversas mudanças na área saúde. Atualmente, o homem está exposto a uma série de fatores que trazem malefícios à saúde como o aumento do sedentarismo, consumo de bebidas e cigarro, alimentação pobre em nutrientes e a poluição generalizada. Se por um lado surgem avanços na prevenção, detecção e tratamento de doenças, por outro algumas avançam em incidência graças aos hábitos modernos. Entre essas doenças, o câncer se destaca como sendo uma dos mais preocupantes.

Câncer é o nome dado a um conjunto de mais de 100 doenças que têm em comum o crescimento desordenado de células, que invadem tecidos e órgãos. Dividindo-se rapidamente, estas células tendem a ser muito agressivas e incontroláveis, determinando a formação de tumores malignos, que podem espalhar-se para outras regiões do corpo. As causas de câncer são variadas, podendo ser externas ou internas ao organismo, estando inter-relacionadas. As causas externas referem-se ao meio ambiente e aos hábitos ou costumes próprios de uma sociedade. As causas internas são, na maioria das vezes, geneticamente pré-determinadas, e estão ligadas à capacidade do organismo de se defender das agressões externas (INCA, 2013).

Entre os cânceres, o de pulmão é conhecido por ser um dos mais letal. É o mais comum de todos os tumores malignos, apresentando aumento de 2% por ano na sua incidência mundial (INCA, 2013).

Há três fatores principais para o surgimento do câncer de pulmão: tabagismo, poluição atmosférica e componente hereditário. Em relação a primeira causa, estima-se que em 90% dos casos diagnosticados, o câncer de pulmão está associado ao consumo de derivados de tabaco (INCA, 2013). Evidências na literatura também mostram que pessoas que têm câncer de pulmão apresentam risco aumentado para o aparecimento de outros cânceres de pulmão e que irmãos, irmãs e filhos de pessoas que tiveram câncer de pulmão apresentam risco levemente aumentado para o desenvolvimento desse câncer. Mesmo esses sendo os dois principais fatores, é difícil estabelecer o quanto desse maior risco decorre de fatores hereditários e o quanto é por conta do hábito de fumar.

Com o advento dos modernos métodos de imagem, o nódulo pulmonar é a apresentação mais comum de câncer de pulmão. Como todo tipo de câncer, dependendo se a detecção é precoce ou não, pode ser difícil de diagnosticar. Nódulos grandes, por exemplo, geralmente definidos como sendo maiores do que 1 cm de diâmetro, podem ser diagnosticados por biópsia na agulha. Por outro lado, nódulos menores trazem mais dificuldade para uma biópsia na agulha e o diagnóstico passa a exigir um método mais invasivo. Ressalta-se pois, que a maioria dos nódulos pulmonares pequenos são de natureza benigna.

Quando um nódulo é de natureza indeterminada, cabe ao médico especialista optar ou não pela remoção cirúrgica. Em alguns casos de diagnóstico indefinido ou, principalmente, em casos com forte suspeita de benignidade, o nódulo passa a ser acompanhado para observar seu comportamento no decorrer do tempo. Nesse caso são feitos exames periódicos e comparações das informações desses exames para verificação de eventuais modificações.

O acompanhamento de uma doença no decorrer do tempo, através da comparação de exames periódicos, é chamada de análise temporal. Reeves *et al* (2006) afirma que muitos trabalhos destacam a análise temporal como a melhor forma não invasiva de diagnóstico de nódulos pulmonares. A taxa de crescimento do nódulo, por exemplo, pode ser utilizada como um dos métodos mais precisos não invasivos de determinação da malignidade. Somada a ela, outras medidas geométricas e de textura extraídas no decorrer do tempo podem oferecer informações importantes sobre o comportamento do nódulo. Essas medidas podem ser usadas pelo médico na tomada de decisões mais coerentes. Por essas razões, confirma-se a importância da análise temporal de nódulos pulmonares.

A análise temporal é feita através da comparação de informações extraídas das imagens de exames. O exame mais utilizado em casos de câncer de pulmão é a tomografia computadorizada (ou TC) de tórax.

Para que a análise temporal seja efetiva, é necessário que haja qualidade na aquisição das imagens. Nas últimas décadas muitos avanços foram feitos na área de aquisição de imagens médicas. Grande investimento acadêmico e industrial tem sido utilizado para o desenvolvimento de novos sistemas de aquisição, que são cada vez mais rápidos e oferecem maior resolução e qualidade de imagem.

Acompanhando o crescimento tecnológico da aquisição de imagens, a área de processamento de imagens médicas também tem evoluído. O processamento dessas imagens permite diversas aplicações práticas, como por exemplo: detecção de patologias, diagnóstico, planejamento cirúrgico, etc.

No entanto, no que diz respeito à comparação de imagens, ainda existe obstáculos. Considerando, por exemplo, o acompanhamento de um nódulo de pulmão no decorrer do tempo, pode haver dificuldade em localizá-lo em exames diferentes, porque as coordenadas dele no primeiro exame podem ser diferentes em relação ao segundo exame. De fato, isso pode ocorrer por diferenças anatômicas do paciente no decorrer do tempo, mas também pode ser por erro de aquisição, por conta de mau posicionamento ou efeitos da respiração do paciente.

Outro fator importante que pode atrapalhar a comparação de imagens médicas é a mudança do equipamento de aquisição de imagens. Tomógrafos diferentes podem gerar volume de imagens com dimensões e espaçamentos variados. Dessa forma, a localização e a comparação do mesmo nódulo em aparelhos diferentes passam a ser dificultadas.

Este trabalho apresenta uma metodologia de análise temporal de nódulos e massas pulmonares baseado em textura (tecido) utilizando volumes de imagens de TC. Primeiramente, a metodologia busca estabelecer uma correspondência entre a localização de nódulos / massas em exames diferentes. Essa correspondência é obtida utilizando principalmente técnicas de pré-processamento e registros de imagens. Uma vez estabelecida essa correspondência, o subvolume de imagens do nódulo / massa será também pré-processado para eliminar componentes indesejados. Em seguida, serão definidas as regiões de interesse do nódulo / massa. Essas regiões serão, por fim, utilizadas na análise temporal, através da aplicação de índices de similaridade.

1.1. Motivação

O câncer do pulmão é o tipo de câncer que mais causa mortes no mundo. A mais recente estimativa mundial apontou uma incidência de 1,61 milhão de casos novos de câncer do pulmão para o ano de 2008, representando 12,7% de todos os novos casos de câncer. Estimam-se 17.210 casos novos de câncer de pulmão em homens e 10.110 em

mulheres, no Brasil, no ano de 2012. Esses valores correspondem a um risco estimado de 18 casos novos a cada 100 mil homens e 10 a cada 100 mil mulheres. (INCA, 2013)

No Brasil, foi responsável por 20.622 mortes em 2008, sendo o tipo que mais fez vítimas. Altamente letal, 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. No fim do século XX, o câncer de pulmão se tornou uma das principais causas de morte evitáveis. (INCA, 2013)

Sabendo disso, é importante o acompanhamento de nódulos e massas pulmonares quando existe alguma dúvida na sua natureza. Em especial, o acompanhamento nódulos de diagnóstico indeterminado pode evitar procedimentos cirúrgicos desnecessários, uma vez que a análise temporal realizada em exames de momentos diferentes é eficaz na caracterização do nódulo.

A análise da mudança de tecido (ou análise da mudança de textura) dos nódulos no tempo ainda é pouco explorada, porém não menos importante. Apenas as mudanças geométricas (ou de formato) podem não ser o suficiente para avaliar o comportamento do nódulo. Essa mudança de textura no nódulo pode indicar a sua natureza em casos em que, por exemplo, houve pouca mudança geométrica. Somado a isso, os médicos em alguns casos tem dificuldade em avaliar por si só a textura e por isso o auxílio de técnicas computacionais pode fazer a diferença.

Além disso, para verificar o comportamento do nódulo de forma não invasiva, o médico precisa estar apoiado de ferramentas computacionais. É comum, no entanto, que as ferramentas usadas sejam apenas visuais e de medição manual. Atualmente, existe uma carência de ferramentas computacionais mais robustas para auxiliar na detecção e acompanhamento dos nódulos pulmonares, em especial no Brasil.

Por essas razões, é necessário o desenvolvimento de técnicas computacionais que auxiliem aos profissionais da saúde no acompanhamento dos nódulos e massas pulmonares e os forneçam informações relevantes sobre mudanças de textura globais e locais no decorrer do tempo.

1.2. Objetivos

Desenvolver uma metodologia computacional que analise as mudanças no decorrer do tempo em um mesmo nódulo (ou massa) pulmonar. Nesse processo serão usadas imagens de TC extraídas de um paciente em momentos diferentes.

Esse trabalho não tem como foco o diagnóstico do nódulo pulmonar. Ao invés disso, busca-se oferecer ao médico informações úteis sobre mudanças nos padrões de textura do nódulo / massa no decorrer do tempo de avaliação do paciente. Espera-se que essas informações possam não só auxiliar o médico especialista no diagnóstico como também na decisão do melhor tratamento baseado no comportamento do nódulo / massa.

Para concretizar este objetivo geral, alguns objetivos específicos devem ser alcançados:

- Desenvolver uma técnica que permita verificar se dois nódulos / massas detectados cada um em um exame de TC de um mesmo paciente são de fato o mesmo nódulo / massa e não nódulos / massas diferentes.
- Analisar as mudanças de padrões de textura globais e locais nos nódulos / massas no decorrer do tempo utilizando índices de similaridade.

1.3. Trabalhos relacionados

Alguns trabalhos na literatura propõem formas de análise temporal de nódulos pulmonares.

Ko *et al* (2001) utiliza técnicas de detecção de nódulos pulmonares e análise temporal. Nele demonstra-se a importância da correspondência entre nódulos de exames diferentes para que seja possível a análise temporal e propõe o uso de registro de imagens como solução. A análise temporal baseia-se principalmente na mudança de volume do nódulo no decorrer do tempo.

Reeves *et al* (2006) também propõe técnicas de detecção e análise temporal de nódulo pulmonares. Ele destaca o tempo de duplicação como uma das melhores medidas não invasivas de determinação da malignidade dos nódulos pulmonares e, dessa forma, fundamental na análise temporal. Nele utiliza-se o registro de imagens para correspondência dos nódulos e como forma de aperfeiçoar a medição dos nódulos.

Sluimer *et al* (2006) destaca a importância do registro de imagens aplicado a imagens de tomografia computadorizada de pulmão e quais suas vantagens. Entre elas, a combinação de exames feitos em pacientes no decorrer do tempo para acompanhar a reação a um tratamento e/ou detectar e quantificar mudanças num intervalo de tempo, em especial, em nódulos pulmonares. Dentre as técnicas recentes, ele cita o uso de registros rígido seguido do deformável *B-spline*.

Zheng *et al* (2007) utiliza o registro de imagens de exames diferentes para melhorar a precisão da segmentação dos nódulos. É proposto o uso do registro B-spline nas imagens de TC para aproximar os exames em forma, mas para evitar deformações dos nódulos, utiliza-se um registro rígido local.

Gu *et al* (2011) compara algoritmos rígidos e deformáveis em eficiência e tempo computacional em imagens. O principal objetivo de seu estudo é verificar quais registros eram melhores para o casamento de nódulos pulmonares em imagens de TC. Ele afirma que, em testes, os registros rígidos tem menos custo computacional, mas em precisão, os melhores são os deformáveis e destaca que os mais utilizados na literatura atual são o *B-spline*, *Demon* e variantes desses.

Os trabalhos atuais mostram os avanços da análise temporal dos nódulos pulmonares e como as técnicas de registro podem ajudar na precisão da avaliação dos médicos. No entanto, eles focam principalmente na análise temporal geométrica desses nódulos, mas sabe-se também que a análise de textura é importante. Assim sendo, uma importante contribuição desse trabalho será a análise conjunta de diferentes técnicas de geometria e textura dos nódulos pulmonares ao longo do tempo.

1.4. Organização do trabalho

Este trabalho apresenta a seguinte organização:

No Capítulo 2, Fundamentação Teórica, segue informações importantes para o contexto e entendimento do trabalho, tais como Padrão DICOM e Registro de Imagens.

No Capítulo 3, Metodologia, será explicada a metodologia utilizada como ponto de partida para o desenvolvimento desse trabalho.

No Capítulo 4, Resultados, mostra-se a ferramenta desenvolvida, através dos resultados alcançados na aplicação da metodologia.

No Capítulo 5, Conclusão, apresenta-se a conclusão do trabalho. Nela está contida uma retrospectiva do que foi falado como todo e uma avaliação dos resultados obtidos.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, serão abordados os conceitos necessários para entender o funcionamento da metodologia.

2.1. Nódulo pulmonar

Um nódulo pulmonar é uma formação esférica de tecido circuncidada em sua maior parte pelo parênquima pulmonar. Na sua maioria, os nódulos são benignos, e não requerem nenhuma ação médica.

Um nódulo benigno pode ser definido como uma formação circunscrita ao pulmão que não invade outros tecidos e não origina metástase. Quando aumentam de volume, esses nódulos tendem a crescer mais lentamente do que os malignos e são menos susceptíveis de causar problemas compressivos. Em alguns casos, no entanto, os nódulos benignos podem interferir no funcionamento do órgão afetado ou se tornarem malignos. Por essa razão, até os nódulos benignos necessitam de alguma forma de acompanhamento.

Tendões e músculos são ambos propensos ao desenvolvimento de nódulos em resposta às lesões. Nódulos podem também aparecer na pele ou nas camadas da pele, caso em que, eles são geralmente notáveis por serem visíveis. Outros órgãos e outras partes do corpo podem também desenvolver nódulos, como mama, pulmão, próstata, entre outros. (WisegEEK, 2008).

A fase inicial do desenvolvimento do nódulo pulmonar é assintomática (não apresenta sintomas) e é formada por uma área não identificável por exames de imagem devido as suas dimensões que são pequenas demais para a resolução do tomógrafo. Com o passar do tempo, ocorre o desenvolvimento do nódulo e, no caso de um nódulo maligno, pode haver o espalhamento de suas células alteradas pela corrente sanguínea, desenvolvendo o processo de metástase.

2.1.1. Nódulo pulmonar solitário

O nódulo pulmonar solitário (NPS) é uma opacidade pulmonar arredondada, circunscrita, envolta por parênquima pulmonar normal, com diâmetro menor ou igual a 3 centímetros e não associada à pneumonia ou atelectasia. Reforçando esse conceito, Tan

et al. (2003) afirma que um nódulo pulmonar solitário é uma lesão esférica ou oval, menor ou igual a 3 centímetros de diâmetro rodeado pelo parênquima pulmonar.

Dentre as várias lesões que se apresentam como NPS, as de origem benigna são as mais frequentes. Pereira (2009) explica que 60% dos nódulos pulmonares solitários são benignos.

A tuberculose é uma importante causa de nódulos pulmonares em países subdesenvolvidos e em desenvolvimento. Já entre as causas malignas, o carcinoma broncogênico é a mais comum e o principal foco de preocupação do médico ao deparar-se com a presença de um nódulo pulmonar solitário (Zamboni, 2000). Dentre outras causas benignas destacam-se as infecções por fungos, os hamartomas e os tumores benignos. Os hamartomas são aglomerações de alguns tecidos normais, mas que não deveriam estar presentes naquele órgão naquela proporção. As células são normais, mas são mal organizadas. No pulmão o tipo mais frequente envolve um agrupamento nodular de cartilagem e gordura.

A preocupação da detecção de um nódulo pulmonar reside na possibilidade do mesmo ser um câncer (30%) ou um câncer proveniente de outro lugar que se disseminou para o pulmão (10%). (Pereira, 2009).

Dois padrões de margens de um nódulo são relativamente específicos. Um deles é o sinal corona radiata, que consiste em fios muito finos lineares que se estendem de 4 a 5 mm para fora do nódulo, originalmente descritos em tomógrafos simples, pois eles têm uma aparência espiculada em radiografias simples. Ao passo que uma borda mais lisa é mais sugestiva de um diagnóstico benigno (David *et al.*, 2003).

As principais características do nódulo pulmonar solitário que devem ser observadas em um exame de imagem são: tamanho, bordas, calcificação, densidade e crescimento (Zamboni, 2000). Quanto ao tamanho, é importante notar que as lesões maiores apresentam maior risco de malignidade. A maioria dos nódulos com diâmetro maior que 2 centímetros são lesões malignas, enquanto apenas metade das lesões de diâmetro menor que 2 centímetros são malignas. Já em relação às bordas, lesões benignas geralmente apresentam bordas regulares ou lobuladas, enquanto as lesões malignas em geral apresentam bordas irregulares ou espiculadas (Barcellos, 2002).

O nódulo é frequentemente descoberto por acaso, em uma radiografia ou tomografia de tórax feita por outras razões. Entre esses dois tipos de exame, a tomografia computadorizada visualiza melhor todas as características importantes do nódulo, distingue melhor o nódulo de alterações da parede torácica, pode identificar lesões adicionais, analisa melhor os ápices, regiões peri-hilares e ângulos costofrênicos e permite melhor estudo do mediastino (Zamboni, 2011). A Figura 1 exemplifica algumas estruturas do pulmão (incluindo um nódulo) numa imagem de tomografia computadorizada correspondente a uma fatia do pulmão.

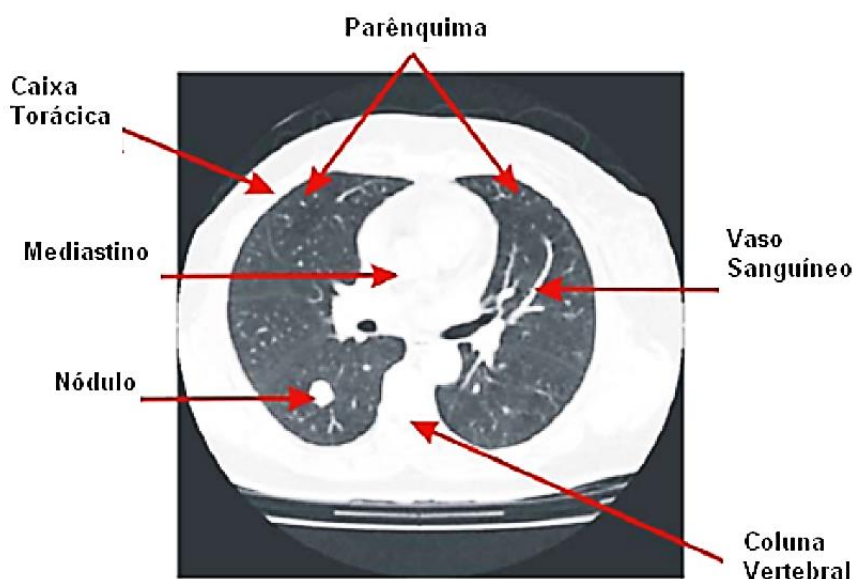


Figura 1 - Nódulo e outras estruturas do pulmão em uma fatia de tomografia computadorizada. Fonte: (Netto, 2010)

2.2. Massa pulmonar

A massa pulmonar é uma lesão no pulmão com as mesmas características de nódulo pulmonar, porém com dimensões maiores. Assim como o nódulo pulmonar, a massa pulmonar é uma lesão não associada à pneumonia ou atelectasia, mas são maiores que três centímetros (Zamboni, 2000). Pereira (2009) explica que quando a lesão tem mais de 3 cm de diâmetro é chamada de massa; na presença de massa a chance de câncer é alta.

A massa pulmonar geralmente é resultado de câncer avançado. Devido às suas dimensões, é mais fácil de ser detectado por exames de TC e seus sintomas são mais visíveis. A Figura 2 mostra um exemplo de massa pulmonar.

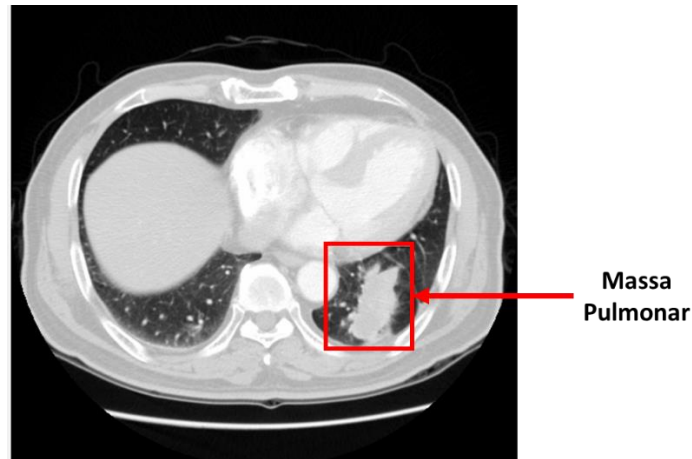


Figura 2 - Massa pulmonar.

2.3. Tomografia computadorizada

A tomografia computadorizada (TC) foi desenvolvida por Godfrey Hounsfield e Allan Cormack em 1972 e o seu funcionamento consiste basicamente em uma máquina que emite Raios X e vai girando em volta do corpo do paciente e à medida que gira emite Raios X em 360°, ou seja, fazendo uma circunferência completa em torno do paciente (Figura 3).

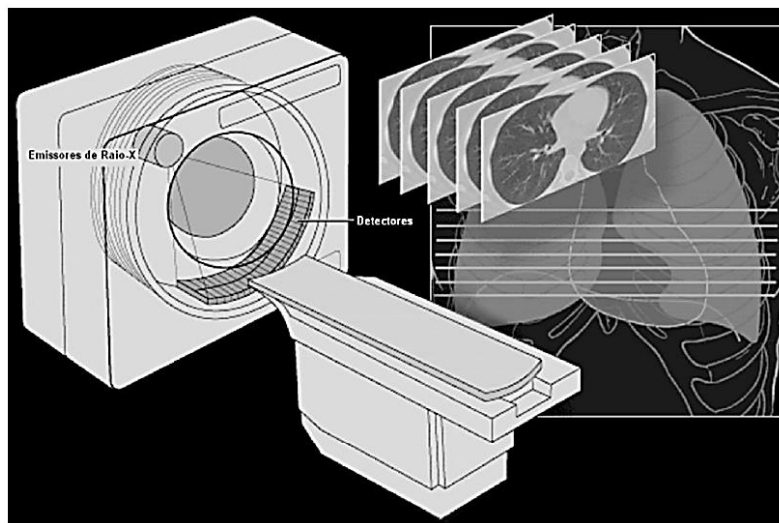


Figura 3 - Tomografia computadorizada. Durante uma sessão de tomografia computadorizada, um fino Raio X gira em torno de uma área do corpo gerando uma imagem tridimensional das estruturas internas. Fonte: (Daros, 2006)

A tomografia computadorizada é um exame que permite a obtenção de imagens de cortes do corpo do paciente, sendo bastante utilizada como um exame médico de diagnóstico por imagem. Comparada à radiografia tradicional, a TC apresenta maior precisão e sensibilidade, além de não apresentar uma imagem com sobreposição de tecidos (Santos, 2011).

Com a TC é possível uma visualização tridimensional dos órgãos em que ela foi realizada, em especial no tórax. A TC proporciona uma maior flexibilidade nos estudos das imagens com maior número de detalhes e com maior número de níveis de cinza se comparado com um exame de raios X. Essa seria então a grande vantagem do uso da TC.

A TC é excelente para detecção de alterações agudas ou crônicas no parênquima pulmonar, como as decorrentes de enfisema ou fibrose (doença pulmonar obstrutiva crônica, que causa a perda de capacidade respiratória e uma oxigenação insuficiente). É também usada na região do tórax para diagnóstico de várias outras infecções.

O resultado visual da TC é monocromático, ou seja, são mostrados apenas os vários níveis de cinza, indo do totalmente preto ao branco, mesmo assim são muito mais numerosas que as variações de níveis de cinza dos raios X convencionais.

2.4. Padrão DICOM

A palavra DICOM é uma abreviação de *Digital Imaging and Communications in Medicine* (ou comunicação de imagens digitais em medicina).

O padrão DICOM foi criado pelo ACR (*American College of Radiology*) e pelo NEMA (*National Electrical Manufacturers Association*), em um comitê em 1983. Ele surgiu com o objetivo de padronizar o formato das imagens e suas informações, usado entre os diferentes fabricantes de tomografia computadorizada, de ressonância magnética nuclear, ultra-som e etc. Com essa padronização seria possível a troca de informações entre médicos e especialistas que se encontram em localidades distintas, principalmente com o uso de PACS (*Picture Archiving and Communications Systems*), que propiciam a criação de bases de dados (NEMA, 2009).

O padrão DICOM internamente é um conjunto de normas para tratamento, armazenamento e transmissão de informação médica num formato eletrônico, estruturando um protocolo. Em outras palavras, o padrão DICOM é uma série de regras que permite que imagens médicas e informações associadas sejam trocadas entre equipamentos de diagnóstico geradores de imagens, computadores e hospitais. O padrão estabelece uma linguagem comum entre os equipamentos de marcas diferentes, que geralmente não são compatíveis, e entre equipamentos de imagem e computadores, estejam estes em hospitais, clínicas ou laboratórios.

O padrão DICOM deve especificar as seguintes informações: o conjunto de protocolos a serem seguidos pelas aplicações e equipamentos, a sintaxe e semântica dos comandos do protocolo e das informações associados, e os requisitos mínimos de conformidade. Apesar disso, o padrão não precisa especificar os detalhes de uma implementação. Ele também não garante a interoperabilidade entre equipamentos e não especifica procedimentos para teste de validação.

Entre características importantes do padrão DICOM, podemos destacar:

- É aplicável em ambiente de redes;
- Especifica como os equipamentos que usam o padrão DICOM devem reagir a comandos e dados realizados;
- Especifica níveis de conformidade para as aplicações;
- Está estruturado em documentos de múltiplas partes de forma a facilitar a evolução do protocolo;
- Introduce o conceito de objetivos da informação, não somente para imagens e gráficos, mas para estudos e relatórios;
- Utiliza uma técnica de identificação única dos objetos de informação.

O arquivo DICOM ao contrário da maioria dos outros padrões para imagens médicas é composto por um cabeçalho e um corpo. O cabeçalho contém informações úteis na interpretação do conteúdo do corpo do arquivo. Nele pode-se encontrar o tamanho da imagem, o nível de compressão, a data do exame, etc. o corpo carrega a imagem propriamente dita, que pode ser armazenado tanto em formato não processado quanto compactado em formato JPEG. (Sousa, 2007)

2.5. Diagnóstico Auxiliado por Computador

Diagnóstico auxiliado por computador, também conhecido pela sigla CAD (de *Computer-Aided Diagnosis*) é definido como um diagnóstico realizado pelo especialista, utilizando o resultado de análises quantitativas automatizadas de imagens médicas como uma segunda opinião para tomada de decisões diagnósticas (Seixas, 2005). Em outras palavras, os sistemas CAD tem como objetivo auxiliar o médico especialistas nas tarefas de detecção e diagnóstico através de ferramentas computacionais aplicadas a imagens médicas.

As imagens médicas utilizadas pelo sistema CAD podem ser obtidas de exames das mais diferentes modalidades dependendo do escopo do sistema, tais como tomografia computadorizada, mamografias ou ultra-sonografias. Alguns sistemas podem combinar imagens de diferentes modalidades com o objetivo para obter o melhor visualização. Outros podem combinar imagens para acompanhamento de mudanças, no caso de aquisições feitas em momentos diferentes. Essas imagens também podem ser de qualquer órgão do corpo, como pulmão, rim, mama, etc.

O auxílio obtido por esses sistemas também é variado. Os sistemas CAD mais completos podem conter ferramentas de organização das imagens médicas, ferramentas de visualização, edição e medição de artefatos da imagem e até mesmo oferecer uma segunda opinião. Essa segunda opinião é obtida de ferramentas automáticas de detecção e diagnóstico.

Recentes pesquisas têm demonstrado as vantagens do uso de sistemas CAD no auxílio ao médico especialista na tarefa de detecção e diagnóstico de nódulo pulmonar. Foi demonstrado que a revisão de um diagnóstico feito por dois especialistas eleva significativamente as chances de acerto. Um sistema CAD funcionaria como o segundo especialista (Seixas, 2005). Experimentos mostram que o especialista ou sistema CAD sozinho apresenta menor desempenho para detecção de nódulos pulmonares do que o médico com o auxílio pelo sistema CAD. (Diniz, 2011)

Os sistemas CAD são especialmente usados quando as tarefas de detecção e diagnóstico do nódulo pulmonar são realizadas contínua e repetidamente pelo médico especialista. Nesses casos, o principal interesse em seu uso está na facilidade com que evidenciam achados radiológicos suspeitos e indicam possíveis diagnósticos, em um tempo menor que o especialista sozinho. Isso ocorre porque um sistema computacional não perde eficiência com o tempo ou não “cansa”, não tem fadiga visual, não deixa passar informações despercebidas por distração. Ao mesmo tempo, ele evita que o especialista passe por esses problemas.

Por essas razões, o desenvolvimento de sistemas CAD para auxílio ao médico especialista vem crescendo a cada dia. A necessidade de se obter um diagnóstico preciso de doenças para possibilitar e a identificação rápida para possibilitar também um

tratamento mais preciso e rápido tem estimulado muitos pesquisadores a buscarem novas ferramentas de apoio e formas de detecção e diagnóstico com auxílio do computador.

No caso do Brasil, o desenvolvimento de um sistema de diagnóstico auxiliado por computador trará muitos benefícios, tais como, a redução do custo do sistema de saúde pela melhoria do desempenho do diagnóstico, nivelamento da qualidade de atendimento e redução da demanda de exames adicionais realizados pelas clínicas médicas. (SEIXAS, 2005).

2.6. Interpolação de imagens

Interpolação é um método que permite construir um novo conjunto de dados a partir de um conjunto discreto de dados pontuais conhecidos. Em engenharia e ciência, dispõe-se habitualmente de dados pontuais, obtidos a partir de uma amostragem ou experimento. Através da interpolação pode-se construir uma função que se "encaixe" nestes dados pontuais.

Existem várias outras definições para interpolação. Uma delas define a interpolação como uma estimativa informada do desconhecido (Watson, 1992). Thévenaz *et al* (2000) a define, de forma menos concisa, como a recuperação baseada em modelos de dados contínuos a partir de dados discretos dentro de um domínio conhecido. Essa última definição é preferível por permitir uma distinção mais clara entre a interpolação e extrapolação. Thévenaz *et al* (2000) afirma ainda que a interpolação considera apenas a existência de um intervalo conhecido, onde se aplica o modelo e que os dados contínuos deterministicamente recuperados são inteiramente descritos pelos dados discretos, enquanto que a extrapolação permite a utilização do modelo de fora do intervalo conhecido, com o pressuposto de que o modelo é "bom" perto amostras de dados e, possivelmente, menos bom em outro lugar.

As três hipóteses mais importante para a interpolação são:

1. Os dados subjacentes são continuamente definidos;
2. Tendo em conta as amostras de dados, é possível calcular um valor de dados da função contínua subjacente;
3. A avaliação da função contínua subjacente nos pontos de amostragem produz o mesmo valor que os próprios dados.

A interpolação é bastante utilizada em processamento de imagens. De fato, os métodos de interpolação de imagem são tão antigos quanto a computação gráfica e o processamento de imagens. Nos primeiros anos, os algoritmos simples, como a interpolação do vizinho mais próximo e a interpolação linear, foram utilizados para a corrigir problemas de reamostragens. Como resultado da teoria da informação introduzida pelo Shannon, no final dos anos 1940, a função de sincronismo foi aceita como a função de interpolação. A partir daí muitas outras surgiram para utilizando princípios matemáticos mais elaborados como os métodos de polinômios de Taylor ou Lagrange. Finalmente, devido à sua eficiência numérica, diferentes famílias de funções *spline* têm sido utilizados para interpolação. (Lehmann, 1999)

2.6.1. Interpolação Linear

A interpolação linear é comumente usada por ser simples de entender e implementar e por ser eficiente em diversos problemas. Ela é ideal quando há poucos pontos a serem estimados.

Em seu funcionamento, a interpolação linear traça um segmento de reta entre dois pontos desejados e calcula a sua equação. Através dessa equação é possível estimar valores entre os pontos extremos do segmento, baseando-se na sua localização. Essa operação pode ser vista na Figura 4.

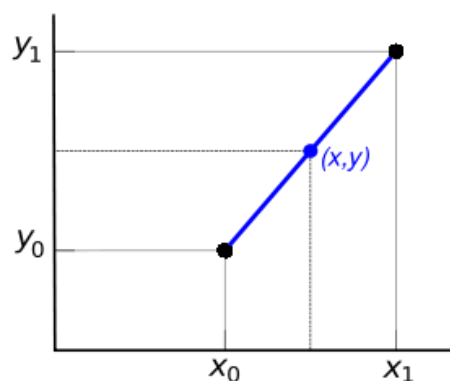


Figura 4 - Interpolação linear entre dois pontos

Em relação as imagens, a interpolação linear traça um segmento de reta entre cada par de *pixels* (ou *voxels*) vizinhos. A interpolação utiliza esse segmento para estimar valores entre os *pixels* vizinhos, que são os pontos extremos do segmento. Essa estimativa é feita baseando-se na distância entre os pontos extremos e seus valores. Um *pixel*

interpolado encontrado exatamente no centro dos seguimento de reta, por exemplo, possui valor igual a média aritmética dos *pixels* dos pontos extremos.

Matematicamente, a interpolação linear pode ser definida da seguinte forma: sendo um intervalo $[x_0, x_1]$ e os valores f_0 e f_1 para os pontos extremos da função $f(x)$, obtém-se o polinômio interpolador definido na equação 1.

$$f(x) = f_0 + \left[\frac{x-x_0}{x_1-x_0} \right] (f_1 - f_0) \quad (1)$$

A interpolação linear será utilizada na metodologia para interpolar os nódulos / massas pulmonares para que possam ter espaçamentos entre fatias iguais e também estimar *voxels* suficientes para que possam ser melhor analisados.

2.7. Registro de imagens

O registro de imagens tem sido de grande importância em diversas áreas, em especial na área médica. Ele tem sido empregado em segmentação, diagnóstico e também na análise temporal de doenças.

A literatura emprega diversas definições ao registro de imagens. Brown (1992) define o registro de imagens como o processo de transformação de diferentes conjuntos de dados para um mesmo sistema de coordenadas. Os dados podem ser várias fotografias, dados de diferentes sensores, de diferentes épocas, ou a partir de diferentes pontos de vista.

Crum (2004) por sua vez, explica o registro de imagens como um processo para a determinação da correspondência entre as características de imagens recolhidas em momentos diferentes ou utilizando diferentes métodos de imagem. As correspondências podem ser usadas para alterar a aparência - por rotação, tradução, etc alongamento - de uma imagem, de forma que mais se assemelhe a outra a fim de que o par possa ser diretamente comparado, combinado ou analisado. O uso mais intuitivo de inscrição é para corrigir posições diferentes de pacientes entre as varreduras. Registro de imagem não tem uma finalidade em si, mas agrega valor às imagens, por exemplo, permitindo que imagens estruturais (TC, RM, ultra-som) e funcional (PET, SPECT, ressonância magnética funcional (fMRI)) possam ser vistas e analisadas no mesmo sistema de coordenadas, e facilita novas utilizações de imagens, por exemplo, para monitorar e quantificar a

progressão da doença ao longo do tempo em indivíduo ou para construir modelos estatísticos de variação estrutural em uma população.

Uma definição mais direta é a utilizada por Parraga (2008) que conceitua o registro de imagens como um processo que busca encontrar a transformação $T(x, y, z)$ que melhor mapeia uma imagem $I(x, y, z)$ em outra. Essa melhor transformação é buscada de acordo com um critério de similaridade. A Figura 5 ilustra o processo do registro. Na Figura 5(a) mostra a imagem que será a referência para a Figura 5(b) de forma que essa possa ser deformada para assemelhar-se a 5(a). O registro busca encontrar a correspondência entre as imagens. Como resultado temos a Figura 5(c), que é a 5(b) deformada para se assemelhar a 5(a).

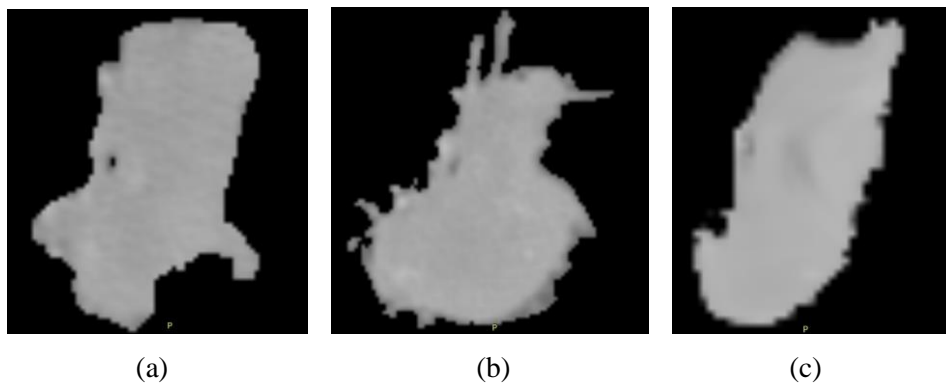


Figura 5 - Processo do registro em um volume de imagens de uma massa. (a) é a imagem de referência. (b) é a imagem a ser transformada para se assemelhar a imagem (a). Em (c) é visto o resultado da transformação.

Em outras palavras, o registro de imagens trata-se de um processo que busca uma série de operações e as aplica numa imagem (chamada de imagem móvel) de forma que ela se assemelhe o máximo possível de outra (chamada de imagem fixa). Essas operações formam a transformação da imagem $T(x, y, z)$. Denota-se a imagem fixa de $F(x, y, z)$ e a móvel de $M(x, y, z)$. Uma vez que o registro encontra $T(x, y, z)$, ela é usada sobre a $M(x, y, z)$ para gerar a imagem registrada $M'(x', y', z')$, como mostrado na equação 1.

$$M'(x', y', z') = M(x - T_x(x, y, z), y - T_y(x, y, z), z - T_z(x, y, z)) \quad (1)$$

A transformação usada no registro pode ser classificada em global ou local. A global é aquela em que, uma vez aplicada, todos os pixels da imagem são deslocados em uma direção. A local gera deslocamentos diferentes para cada pixel ocasionando deformações locais na imagem.

Historicamente, os registros dividem-se em dois grupos: registros rígidos e registros deformáveis (ou não rígidos). Os registros rígidos consideram que a imagem móvel é um objeto que precisa ser rotacionado e/ou transladado para que haja uma correspondência satisfatória com a imagem fixa. Ou seja, esse tipo de registro usa transformação global. Os registros deformáveis, por outro lado, obtêm essa correspondência através de algumas deformações localizadas. Esse por sua vez usa transformação local.

O registro de imagens é composto pelos seguintes elementos: duas imagens, sendo uma fixa e uma móvel, uma medida de similaridade, uma função de transformação, de um otimizador e um interpolador.

A medida de similaridade é uma medida numérica responsável por dizer o quanto a imagem móvel está similar a fixa. A função de transformação é a função que mapeia a imagem móvel na imagem fixa. Essa função de transformação é calculada pelo otimizador, otimizando certa medida de similaridade. Dessa forma, o registro pode ser visto com um problema de otimização, no qual um otimizador estima a função transformação que melhor mapeia a imagem fixa da imagem móvel, de acordo com um valor obtido pela medida de similaridade. Esse processo pode ser visto na Figura 6.

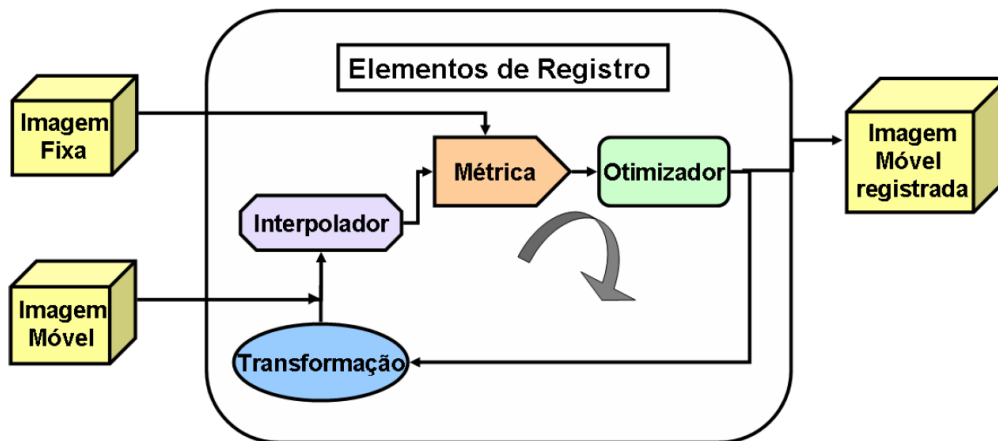


Figura 6 - Processo de um registro de imagens. Fonte: (Parraga, 2008).

O registro de imagens médicas, por sua vez, é uma técnica de processamento de imagens que pode ser definido como processo de alinhamento de imagens para que características correspondentes sejam facilmente relacionadas (HAJNAL, 2001), sendo que, estas imagens podem ser adquiridas por sensores diferentes ou em instantes diferentes.

O registro é um processo essencial em várias áreas de aplicação ligadas a imagens médicas. Uma utilização bastante comum é a realização de comparação e análise de imagens de pacientes adquiridas em momentos diferentes (análise temporal). Nesses casos, o registro auxilia o especialista no acompanhamento da evolução da patologia do paciente. Por exemplo, acompanhar o crescimento de um tumor ou verificar imagens adquiridas antes e após a utilização de contraste.

Outras aplicações muito comuns são o procedimento cirúrgico guiado por imagem (GOSHTASBY, 2005), no qual o registro de imagens pré-operatórias e imagens obtidas durante a cirurgia é de fundamental importância para a localização da patologia e a junção de tipos diferentes de exames permitindo agrupar informações complementares em uma mesma imagem.

A seguir será apresentada a metodologia usada na implementação dos registros. Na metodologia, serão explicados dois tipos de algoritmos de registro em três dimensões, que são comumente utilizados em imagens médicas. São eles: registro rígido e registros deformável *B-spline*. Parraga (2008) apresenta uma boa explicação sobre registros de imagem, em especial estes utilizados neste trabalho.

2.7.1. Registro rígido

O registro rígido é o tipo de registro mais simples. A sua aplicação resume-se apenas a aplicações de operações e translações, de forma que a imagem móvel não é deformada, e sim posta no mesmo sistema de coordenadas da imagem fixa.

O registro rígido, por si só, não é o mais adequado para a maioria dos problemas de desalinhamento de imagens. Isso ocorre porque ele não consegue corrigir erros locais que apenas seriam resolvidos por meio de deformações. No entanto, ele é comumente usado como pré-processamento de registros deformáveis, possibilitando que os mesmos sejam executados com maior eficiência e menor custo computacional. Isso acontece porque, uma vez que a imagem móvel está no mesmo sistema de coordenadas da fixa, as deformações passam a ser mais locais e menos frequentes.

Mais formalmente, a transformação de coordenadas de imagem móvel, $M(x)$, nas coordenadas da imagem fixa $F(x)$ é dada pela Equação 1.

$$x' = T(x) = R_x(\Theta_x)R_y(\Theta_y)R_z(\Theta_z)x + t \quad (1)$$

onde R_x , R_y , e R_z são as matrizes de rotação em torno dos eixos x , y e z , respectivamente, considerando um ângulo Θ e t é o vetor translação. As matrizes de rotação são representadas pelas Equações 2, 3 e 4.

$$R_x(\Theta_x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \Theta_x & -\sin \Theta_x \\ 0 & \sin \Theta_x & \cos \Theta_x \end{pmatrix} \quad (2)$$

$$R_y(\Theta_y) = \begin{pmatrix} \cos \Theta_y & 0 & \sin \Theta_y \\ 0 & 1 & 0 \\ -\sin \Theta_y & 0 & \cos \Theta_y \end{pmatrix} \quad (3)$$

$$R_z(\Theta_z) = \begin{pmatrix} \cos \Theta_z & -\sin \Theta_z & 0 \\ \sin \Theta_z & \cos \Theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

A Figura 7 mostra um exemplo de registro rígido aplicado a imagens médicas.

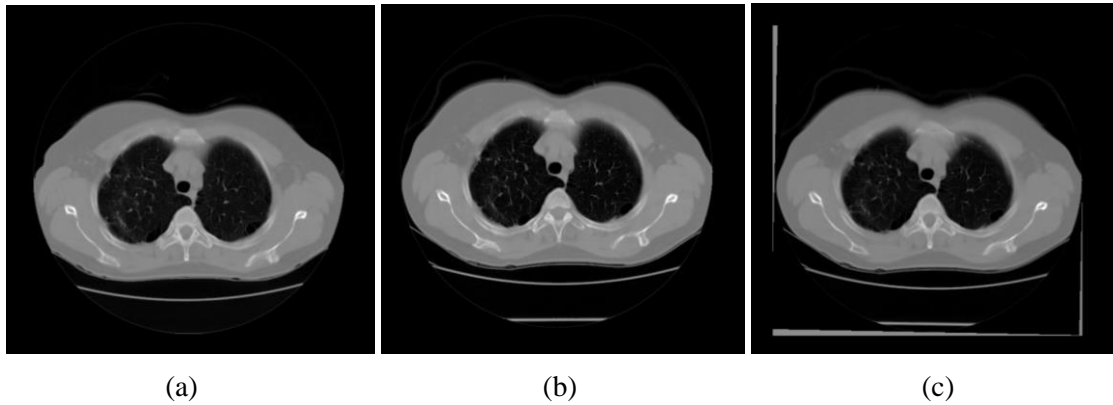


Figura 7 - Registro rígido aplicado em imagens de TC de pulmão. (a) é a imagem fixa, (b) é a imagem móvel e (c) é imagem resultante. (b) foi levemente rotacionada para se aproximar ao formato de (a).

2.7.2. Registro deformável *B-spline*

O Registro *B-spline* é um dos tipos de registro deformáveis mais utilizados. O seu modelo de transformação é descrito por uma *B-spline* de terceira ordem. É um registro do tipo paramétrico e retorna os parâmetros do modelo representando o campo de deformação.

A transformação de imagem móvel em um ponto qualquer $x = [x, y, z]/t$ num registro *B-spline* pode ser representada pela Equação 5.

$$T(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 \beta_l(u)\beta_m(v)\beta_n(w) \Phi_{i+l,j+m,k+n} \quad (5)$$

onde as funções β são polinômios *B-spline* de terceira ordem, representadas pelas Equações 6, 7, 8 e 9:

$$\beta_0(t) = (-t^3 + 3t^2 - 3t + 1)/6 \quad (6)$$

$$\beta_1(t) = (3t^3 - 6t^2 + 4)/6 \quad (7)$$

$$\beta_2(t) = (-3t^3 + 3t^2 + 3t + 1)/6 \quad (8)$$

$$\beta_3(t) = t^3/6 \quad (9)$$

O parâmetro $\Phi_{i,j,k}$ é um conjunto de pontos de controle. Esses pontos são coeficientes que representam o campo de deformação. Eles formam uma malha de pontos de tamanho n_x, n_y, n_z , que são distribuídos sobre a imagem móvel, com espaçamentos $\delta_x, \delta_y, \delta_z$.

Para uma *B-spline* de terceira ordem, quatro funções *B-spline* são necessárias para um determinado ponto x . Como o registro é tridimensional, leva-se em consideração coordenadas x, y e z , o que implica que 43 *B-splines* são importantes para cada ponto. Dessa forma, a transformação *B-spline* é aplicada para cada *voxel* da imagem móvel nas posições em torno dos $4 \times 4 \times 4$ pontos de controle da malha $\Phi_{i,j,k}$. Os índices i, j e k representam os índices dos pontos de controle da malha. São eles vistos nas Equações 10, 11 e 12.

$$i = \left\lfloor \frac{x}{\delta_x} \right\rfloor - 1 \quad (10)$$

$$j = \left\lfloor \frac{y}{\delta_y} \right\rfloor - 1 \quad (11)$$

$$k = \left\lfloor \frac{z}{\delta_z} \right\rfloor - 1 \quad (12)$$

onde u, v e w são as posições relativas de x, y e z dos pontos da malha no espaço 3D. São eles vistos nas Equações 13, 14 e 15.

$$u = \frac{x}{\delta_x} - \left\lfloor \frac{x}{\delta_x} \right\rfloor \quad (13)$$

$$w = \frac{y}{\delta_y} - \left\lfloor \frac{y}{\delta_y} \right\rfloor \quad (14)$$

$$v = \frac{z}{\delta_z} - \left\lfloor \frac{z}{\delta_z} \right\rfloor \quad (15)$$

O registro *B-spline* é um problema de otimização, que busca encontrar os parâmetros $\Phi_{i,j,k}$ que maximizam a semelhança entre as imagens, ou seja, maximizar a

medida de similaridade em busca dos parâmetros que melhor representam a transformação $T(x, y, z)$. O processo do registro é representado resumidamente pelas seguintes etapas:

1. Escolher o tamanho da malha de pontos de controle (quantidade de pontos);
2. Inicializar os valores do conjunto de parâmetros $\Phi_{i,j,k}$;
3. Calcular a medida de similaridade para $\Phi_{i,j,k}$;
4. Atualizar o conjunto de parâmetros $\Phi_{i,j,k}$;
5. Aplicar a transformação $T(x, y, z)$;
6. Repetir o passo 3 até algum critério de parada ser satisfeito.

Uma característica importante do registro *B-spline* é que suas deformações são locais, o que impede alterações desnecessárias em outras estruturas. Isso ocorre porque alterações nos pontos de controle afetam somente a vizinhança local da malha. No entanto, algumas consequências sobre a parametrização da malha devem ser levadas em consideração. Uma malha muito grande (ou seja, com muitos pontos de controle) define deformações mais locais na imagem, uma vez que o espaçamento entre os pontos será menor, mas uma malha menor (menos pontos de controle) resulta em deformações mais globais.

A Figura 8 mostra a aplicação do registro deformável *B-spline*.

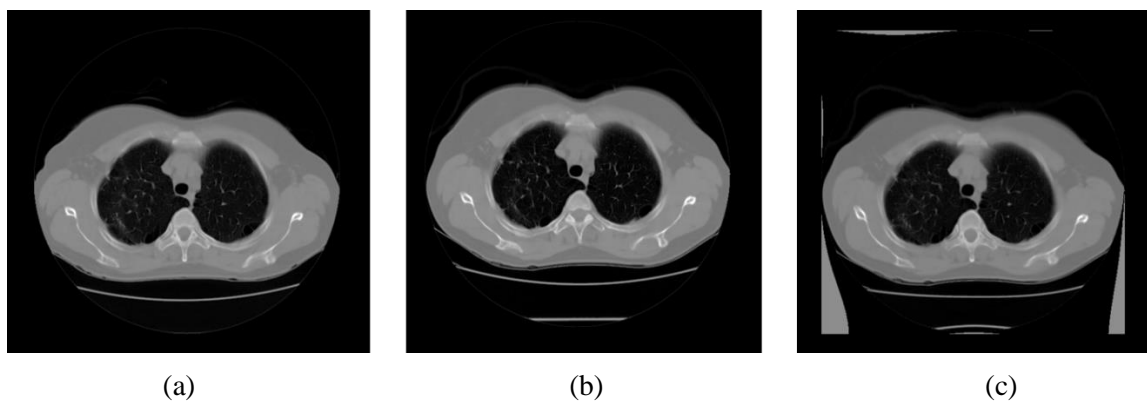


Figura 8 - Registro *B-spline* aplicado em imagens de TC de pulmão. (a) é a imagem fixa, (b) é a imagem móvel e (c) é imagem resultante.

Na metodologia, os registros rígido e *B-spline* são usados em conjunto para aproximar em formato e localização os volumes de imagens de pulmão e também subvolumes de nódulos / massas.

2.8. Agrupamento

Agrupamento, também conhecido como aglomeração ou clusterização, é o processo pelo qual indivíduos de determinada amostra são divididos em grupos (*clusters*) obedecendo a um ou mais critérios que calculem a similaridade entre esses indivíduos. Um agrupamento efetivo deve criar os grupos de forma que os indivíduos de um mesmo grupo sejam os mais semelhantes possíveis e os indivíduos de grupos diferentes sejam os mais diferentes possíveis. Em outras palavras, um agrupamento busca maximizar a distância interclasse (distância entre grupos) e minimizar a distância intraclasse (distância entre indivíduos do mesmo grupo).

O Agrupamento pode ser encontrado sob diferentes nomes em diferentes contextos, tais como aprendizagem não supervisionada e aprender sem um professor (em reconhecimento de padrões), taxonomia numérica (em biologia, ecologia), tipologia (em ciências sociais), e partição (em teoria dos grafos). O exemplo a seguir é inspirado pela biologia: (Theodoridis et al, 2009)

2.8.1. *K-means*

O *K-means* é um algoritmo de agrupamento bastante conhecido, principalmente devido a sua eficiência. Ele é simples de utilizar, uma vez que exige apenas uma quantidade de grupos como parâmetro.

O algoritmo *K-means* tenta de encontrar os "blocos" naturais nos dados. O usuário define o número desejado de grupos e, em seguida, o *K-means* rapidamente encontra um bom posicionamento para os centros de fragmentação, onde "bom" significa que os centros do conjunto tendem a acabar localizados no meio dos grupos naturais de dados. É uma das técnicas mais utilizadas de agrupamento e tem grandes semelhanças com o algoritmo de maximização da expectativa de mistura Gaussiana, bem como algumas semelhanças com o algoritmo de *Mean-shift*. O *K-means* é um algoritmo iterativo e é também conhecido como o algoritmo de Lloyd ou (equivalente) "iteração Voronoi". O algoritmo é executado como se segue: (Bradski et al, 2008)

1. Toma-se como entrada um conjunto de dados e um número de grupos k , escolhido pelo usuário.
2. Aleatoriamente, escolhe-se os centros dos grupos;

3. Associe-se cada ponto de dados com seu centro de grupo mais próximo;
4. Move-se os centros de grupos para os centroides de seus pontos de dados;
5. Repete-se o passo 3 até que haja uma convergência (centroide não mova mais).

A Figura 9 mostra o funcionamento do *K-means*. Neste caso ele leva apenas duas iterações para convergir. Em casos reais, no entanto, ele pode tanto convergir rápido ou levar um grande número de iterações.

O *K-means* é um algoritmo de agrupamento extremamente eficaz, mas tem três problemas (Bradski et al, 2008).

1. Não garante encontrar a melhor solução possível para localizar os centros de grupo. No entanto, ele sempre converge para uma solução (isto é, as iterações não irão continuar indefinidamente).
2. Não diz quantos centros de grupo são necessários usar. Se fossem escolhidos dois ou quatro conjuntos para o exemplo da Figura 9, os resultados seriam diferentes e, talvez, não intuitivos.
3. Presume que a covariância no espaço ou não importa ou já foi normalizada.

Apesar desses problemas encontrados no *K-means*, existem formas de melhorar o seu desempenho. No *K-means*, cada grupo possui os seus pontos de dados e um centro. A partir destes, é possível calcular a variância desses pontos. O melhor agrupamento minimiza a variação sem causar muita complexidade (grupos demais). Dessa forma, uma boa solução seria executar o *K-means* sucessivas vezes, com quantidades diferentes de classes e com centros iniciais em locais diferentes a cada iteração, até que se encontre o resultado desejado.

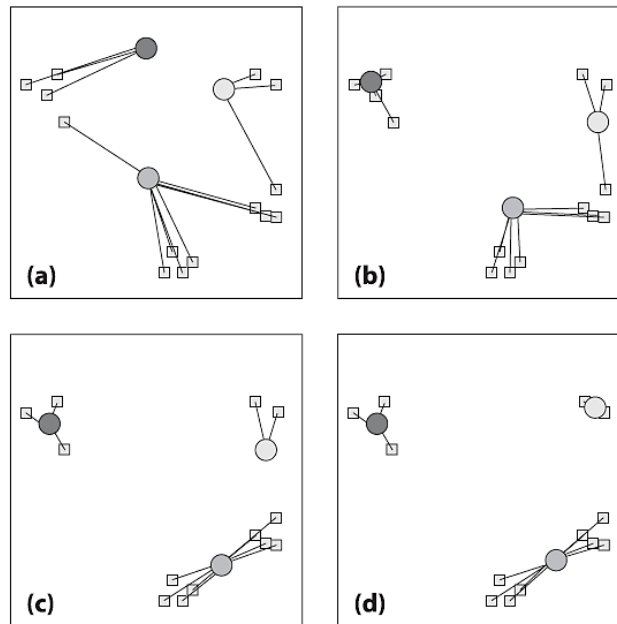


Figura 9 - Exemplo de funcionamento do *K-means*. (a) Os centros dos grupos são colocados de forma aleatória e cada ponto de dados é então atribuído ao seu centro mais próximo; (b) os centros de grupos são movidos para o centroide dos seus pontos; (c) os pontos de dados são novamente atribuídos ao seu centro de grupos mais próximo; (d) Os centros de grupo são novamente mudou-se para o centroide de seus pontos. Fonte: livro Learning OpenCV (Bradski et al, 2008)

O *K-means* será utilizado na metodologia para ajudar a agrupar regiões de de textura semelhantes e diferentes entre nódulos / massas.

2.9. Índices de similaridade

Os índices de similaridade são equações utilizadas para medir semelhança entre grupos. Os parâmetros dessas fórmulas são, em geral, o número de tipos de indivíduos presentes e ausentes em cada grupo e a quantidade de indivíduos de cada tipo. O resultado da aplicação dessas fórmulas é um valor numérico que representa um grau de semelhança. O domínio do valor resultante da fórmula varia de índice para índice, mas comumente apresenta valores entre 0 e 1.

Os índices de similaridade são comumente usados em biologia para fazer comparações entre grupos de espécies. Na biologia, muitas vezes é desejável fazer comparações entre amostras de fauna ou florais tomadas em momentos diferentes, em lugares diferentes, ou por diferentes técnicas (Meyer, 2002). Essas comparações são feitas baseando-se nas características individuais e a abundância de indivíduos das amostras. A necessidade de quantificar essas comparações tornou popular o uso dos índices de

similaridade. Apesar da popularização, um grande número dos índices utilizados atualmente foram criados no século passado (Meyer, 2002).

Alguns destes índices apenas levam em consideração a presença ou ausência de espécies das amostras, enquanto outros incorporam informação sobre a abundância relativa da espécie. O índice preferível em um determinado caso depende das situações e dos tipos de dados disponíveis. Huhta (1979) testou uma longa série de índices de similaridade com dados reais e descobriu que os resultados dependem em grande parte do índice escolhido. Isso sugere que em muitos casos há um perigo de se escolher um índice e este não ser capaz de provar o que se quer demonstrar. Assim, uma avaliação cuidadosa de os vários índices é essencial.

Além dos índices de similaridade, também existem os índices de diversidade. Conceitualmente, a diferença entre os dois é que o primeiro busca medir o grau de semelhança entre grupos e o segundo busca medir o grau de diferença. Matematicamente, um índice de similaridade apresenta alto valor quando há uma forte semelhança entre os grupos, enquanto um índice de diversidade apresenta alto valor quando há baixa semelhança entre os grupos. Apesar de buscarem resultados distintos, geralmente é fácil transformar um índice de similaridade em um de diversidade e vice-versa. Um índice de similaridade s com domínio $[0, 1]$, por exemplo, pode ser transformado em um índice de diversidade d através da operação complementar $d = (1 - s)$.

São muitos os índices de similaridade encontrados na literatura. Nesta dissertação será utilizado um número de índices considerados apropriados para o problema, descritos em (Meyer, 2002). Esses índices de similaridade são:

- Jaccard
- Anderberg
- Czekanowsky
- Kulczynski 1
- Kulczynski 2
- Ochiai

Considerando dois grupos, chamados grupo 1 e grupo 2, no qual deseja-se medir a similaridade, descreve-se as seguintes variáveis:

- a : tipos de indivíduos que pertencem tanto ao grupo 1 quanto grupo 2;
- b : tipos de indivíduos que pertencem grupo 1 e não pertencem ao grupo 2;
- c : tipos de indivíduos que pertencem grupo 2 e não pertencem ao grupo 1.

Tendo conhecimento dessas variáveis, na Tabela 1, são listadas as fórmulas dos índices de similaridade utilizados.

Tabela 1 - Índices de similaridade. Adaptado de (Meyer, 2002).

Índice de similaridade	Equação	Intervalo de Ocorrência
Jaccard	$\frac{a}{a + b + c}$	[0, 1]
Anderberg	$\frac{a}{a + 2(b + c)}$	[0, 1]
Czekanowsky	$\frac{2a}{2a + b + c}$	[0, 1]
Kulczynski 1	$\frac{a}{b + c}$	$[0, +\infty)$
Kulczynski 2	$\frac{a}{2} \left(\frac{1}{a + b} + \frac{1}{a + c} \right)$	[0, 1]
Ochiai	$\frac{a}{\sqrt{(a + b)(a + c)}}$	[0, 1]

Na metodologia, os índices de similaridade serão usados na análise temporal de nódulos / massas baseado em textura. Através dele será possível definir se há semelhanças ou diferenças nos padrões de textura locais e globais nos nódulo / massas no decorrer do tempo.

2.10. Insight Segmentation and Registration Toolkit

O *Insight Segmentation and Registration Toolkit* (ou ITK) é um *framework* de desenvolvimento multiplataforma de aplicações de código aberto amplamente utilizado para o desenvolvimento de programas de segmentação de imagem e de registro de imagens. Ele também pode é definido como um kit de ferramentas flexível e multiplataforma, ou seja, funciona muito bem independente do sistema operacional do computador.

O ITK foi desenvolvido com financiamento da *National Library of Medicine* como um recurso aberto de algoritmos para analisar as imagens do *Visible Human Project*.

O ITK é um projeto de código aberto, ou seja, os desenvolvedores de todo o mundo podem usar, depurar, manter e estender o software. ITK usa um modelo de desenvolvimento de software de programação extrema. As principais características da programação extrema são comunicação e testes. A comunicação entre os membros da comunidade ITK é o que ajuda a gerenciar a rápida evolução do software. O teste é o que mantém a estabilidade do software.

Ele fornece algoritmos de segmentação e de registro em duas, três e mais dimensões. Os algoritmos providos por ele são os mais diversos, tanto na área de processamento de imagens como reconhecimento de padrões. O ITK também é reconhecido por ser utilizado com imagens médicas. A Figura 10 mostra a aplicação da limiarização de Otsu numa massa pulmonar.

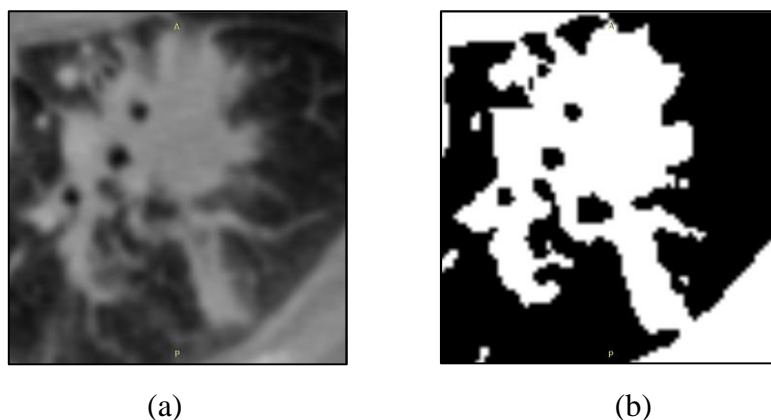


Figura 10 - Limiarização de Otsu desenvolvida com o ITK e aplicada a uma imagem de massa pulmonar. (a) é a imagem original e (b) é a limiarizada.

No caso desta dissertação, é conveniente que ele tenha sido implementado em C++, embora também seja compatível a outras linguagens com Java, Python e Tcl. Além disso, o ITK tem seu código-fonte aberto e gratuito, e é constantemente utilizado, testado e modificado pela comunidade online em favor de melhorias e detecção de bugs, o que garante a sua qualidade, em especial em projetos médicos. Nesta dissertação, o ITK é essencial pela disponibilidade de suas ferramentas de processamento de imagens e reconhecimento de padrões. A maioria dos algoritmos utilizados nesse trabalho, como registro, interpolação e agrupamento foram desenvolvidos com auxílio do ITK.

3. METODOLOGIA PROPOSTA

Nesse capítulo será apresentado a metodologia proposta nesta dissertação. A metodologia possui em cinco etapas bem distintas: aquisição de imagens, pré-processamento das imagens, localização dos nódulos em exames diferentes, pré-processamento do nódulo / massa e, finalmente a análise temporal. A metodologia pode ser resumida no seguinte fluxo (Figura 11):

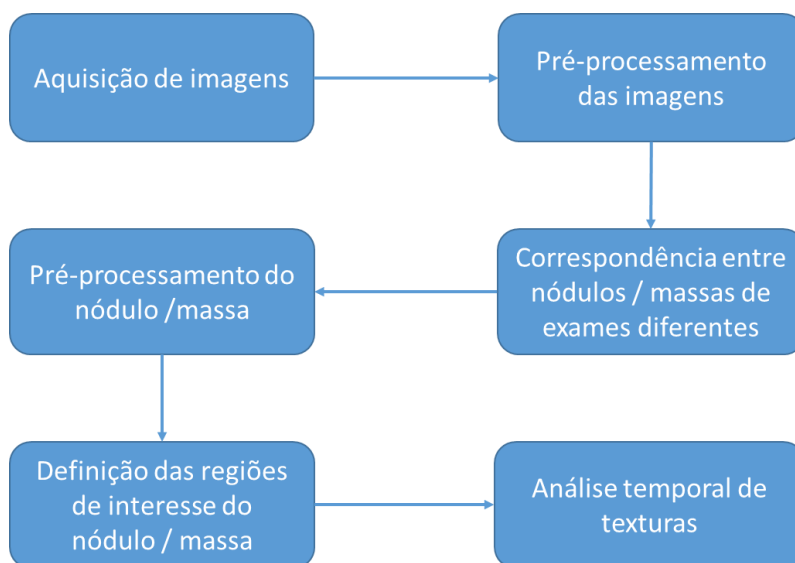


Figura 11 - Fluxo da metodologia

A primeira etapa especifica o tipo de imagens utilizadas e de que fontes elas serão extraídas. Na segunda etapa, os volumes de imagens serão melhorados. Isso quer dizer que serão removidos elementos desnecessários das imagens, como o fundo e a caixa torácica que podem atrapalhar as etapas subsequentes. A terceira etapa mostra as técnicas que serão usadas para determinar uma correspondência entre os nódulos / massas localizados em imagens de exames realizados em momentos diferentes em um mesmo paciente. Na quarta etapa, as imagens dos subvolumes dos nódulos / massas são melhoradas para seja possível analisá-los. Na quinta etapa serão definidas as regiões de interesse do nódulo / massa que se deseja analisar. Essa análise é feita na sexta etapa. Através da análise temporal feita na quinta etapa pretende-se obter informações sobre as mudanças de tecido (textura) do nódulo / massa.

3.1. Aquisição de imagens

As imagens utilizadas nesta metodologia são imagens DICOM de tomografia computadorizada (TC) de tórax. Essas imagens serão obtidas de bases de imagem.

As bases de imagem são conjuntos de imagens e suas informações, de mesma modalidade ou modalidades diferentes, mas agrupadas para um mesmo fim. Elas podem vir acompanhadas de ferramentas de visualização e de banco de dados para busca, tanto das imagens quanto de suas informações.

Em imagens médicas, essas imagens tendem a buscar algum tipo de padronização de aquisição, de forma que possam ser usadas em conjunto, embora na prática, nem sempre isso ocorra. Geralmente as imagens da base são usadas em conjunto na comparação de resultados de estudos aplicados. Muitas bases de imagens são criadas principalmente para a aplicação de técnicas computacionais.

No caso das áreas de processamento de imagens e reconhecimento de padrões, as bases de imagens são utilizadas para a validação de algoritmos, comumente de aprendizagem e detecção de artefatos da imagem.

Para este trabalho as bases de imagens utilizadas devem conter correspondência temporal. O objetivo é obter imagens de exames de um mesmo paciente tiradas em momentos diferentes, de modo que se possa analisar as mudanças ocorridas nos nódulos pulmonares. As bases de imagem utilizadas serão a *Public Lung Database to Address Drug Response* (PLDADR) e base *Volcano Challenge*. Além dessas duas, foi desenvolvida uma base com massas sintéticas. Essas “massas” são volumes de imagens criados para validar a eficiência da análise temporal. A seguir serão explicadas as características dessas bases e como serão utilizadas.

3.1.1. *Public Lung Database to Address Drug Response*

A *Public Lung Database to Address Drug Response* (PLDADR) foi criada pela Universidade Cornell com ajuda de doações da Fundação de Prevenção ao Câncer (*Prevent Cancer Foundation* ou PRF) e trabalhando em conjunto com o Instituto Nacional do Câncer americano (*National Cancer Institute* ou NCI). O objetivo de sua criação é acelerar o progresso no desenvolvimento de técnicas de monitoramento

quantitativo de doenças assistidas por computador. Uma característica importante dessa base é que as lesões acompanhadas pelos exames encontram-se em tratamento para avaliar a resposta a medicamentos.

As imagens desta base são de TC de tórax e estão no formato DICOM. Acumulação de dados para esse banco de dados está em andamento. O banco de dados atual contém um número limitado de exames de imagem TC anotadas que destacam muitas das questões-chave para medir grandes lesões no pulmão. Todas as imagens estão disponíveis gratuitamente e podem ser baixadas por meio de registro de usuário.

A base apresenta tanto imagens de exames sequenciais, que são úteis a essa dissertação, como exames únicos. Os exames sequenciais, no entanto, são em grande maioria para o acompanhamento de massas pulmonares. No total, a PLDADR possui exames sequenciais de 24 lesões, sendo que cada lesão pode ter sido acompanhada por dois a quatro exames. As imagens possuem espessuras de fatias variadas, com maioria variando em torno de 2,5 mm.

O site desta base fornece um conjunto de ferramentas de visualização de imagens para ambas as imagens de TC e suas anotações. Ele também inclui as ferramentas e informações de medições da lesão e a análise de crescimento. Esses recursos estão disponíveis em um protótipo de software online que utiliza banco de dados.

3.1.2. Base do *Volcano Challenge*

O *Volcano Challenge* é uma competição online cujo objetivo é reunir equipes interessadas em a análise de mudança tridimensional de nódulos pulmonares em imagens de TC. O foco do desafio não é diretamente na própria segmentação, mas a mudança no tamanho da lesão gravado em duas imagens separadas pelo tempo. Para possibilitar esse desafio, é oferecida uma base de imagens temporais de exames de TC. Embora o objetivo dessa dissertação não seja análise de mudanças geométricas, que é o foco desse desafio, a sua base de imagens apresenta uma boa quantidade de exames temporais além de informações e ferramentas de visualização online úteis.

Os dados de imagem disponibilizados foram adquiridos para ser futuramente incluídos na PLDADR (Subseção 3.1.2) e foram fornecidos pela Faculdade de Medicina Weill Cornell.

Todas as imagens para este desafio são disponibilizadas em formato DICOM com todas as informações do paciente removidas. As datas originais foram removidas a partir de pesquisas e substituído com datas correspondentes a um intervalo de tempo de 100 dias entre cada par de exames.

Os exames originais foram cortados na direção axial para conter apenas as fatias do nódulo e mais cinco fatias acima e cinco fatias abaixo, caso estas fatias extras estivessem disponíveis. Cada um desses exames contém um conjunto de fatias suficientemente grande para possibilitar a visão pulmão e também, no mínimo, englobar todas as fatias do nódulo.

Os casos selecionados contem pelo menos um nódulo de consistência sólida, que está presente em, pelo menos, dois exames de TC. Apenas nódulos visíveis em pelo menos três fatias de ambos os exames foram incluídos.

O conjunto de dados de avaliação consiste de 49 nódulos divididos em três categorias. A primeira categoria consiste de 27 nódulos visíveis em exames de TC de 1,25 mm de espessura fatia, com pouca mudança de tamanho observado, e um intervalo de diâmetro entre 4 mm e 24 mm. Estes casos abrangem os tamanhos de maior interesse para a medida de crescimento de nódulos e representam exames de boa qualidade. A segunda categoria de nódulos incluiu 13 nódulos cujos exames de cada nódulo tem espessuras diferentes, para analisar o efeito da espessura de corte sobre o desempenho. Os nódulos variavam em tamanho entre 8mm e 30 mm. A terceira categoria consiste em um adicional de 9 nódulos visíveis em exames de TC de 1,25 milímetros de espessura, mas com uma grande mudança de tamanho. Estes nódulos variam em tamanho entre 5 mm e 14 mm.

3.1.3. Base de massas sintéticas

A base de massas sintéticas é simplesmente uma pequena base composta de três pequenos volumes de imagens, gerados artificialmente e que contém cada um uma estrutura esférica com características de textura singulares. Essas estruturas são chamadas de massas sintéticas porque irão ser usados como massas. No caso dessa dissertação, eles serão usados em pares para a aplicação da análise temporal de textura.

Essa base foi criada para melhor análise da metodologia desta dissertação. Por apresentar estruturas artificiais e com texturas diferentes entre si, é mais fácil controlar e explicar os resultados obtidos pela análise temporal.

As massas sintéticas pode ser vistos na Figura 12. A primeira massa sintética (Figura 12a) possui valores de *voxel* variando em ordem crescente do centro do nódulo a até a sua borda em intervalos regulares. A segunda massa sintética (Figura 12b) também apresenta seus valores variando do centro para a borda, porém de forma decrescente, fazendo com que o centro apresente valor máximo (250 HU) e a borda valor mínimo (0 HU). A última massa sintética (Figura 12c) tem comportamento crescente valores de *voxel* variando em ordem crescente do centro da massa a até a sua borda em intervalos regulares, assim como o primeiro. No entanto, seus valores variam de 100 HU até 350 HU.

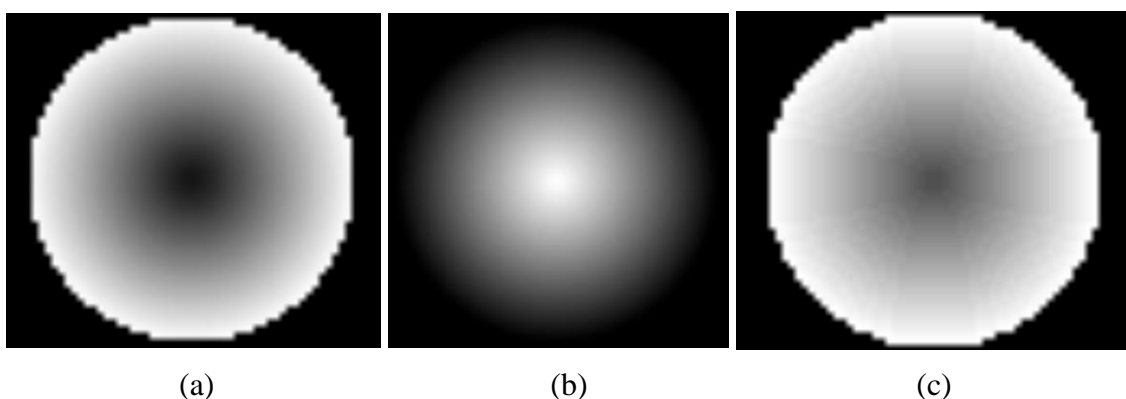


Figura 12 - Massas sintéticas. (a) possui valores de *voxels* de 0 HU a 250 HU, variando de forma crescente do centro para a borda. (b) apresenta os mesmos valores de (a), mas variando de forma crescente da borda para o centro. (c) apresenta valores entre 100 HU e 250 HU, variando de forma crescente do centro para a borda.

Todos os três volumes de imagens apresentam espessura de 1 mm. O diâmetro de cada nódulo é de 50 mm.

Na próxima seção, os volumes de imagens obtidos das bases PLDADR e da Base Volcano serão pré-processados. O objetivo é eliminar elementos indesejados, como traqueia e brônquios e mantendo apenas o parênquima.

3.2. Pré-processamento das imagens

A metodologia usada é baseada na dissertação de (Santos, 2011). O pré-processamento é responsável por eliminar elementos desnecessários das imagens para melhorá-la para a etapa subsequente. Após o pré-processamento, espera-se que apenas o

parênquima das imagens seja mantido. Ele pode ser descrita pelo fluxo apresentado na Figura 13.

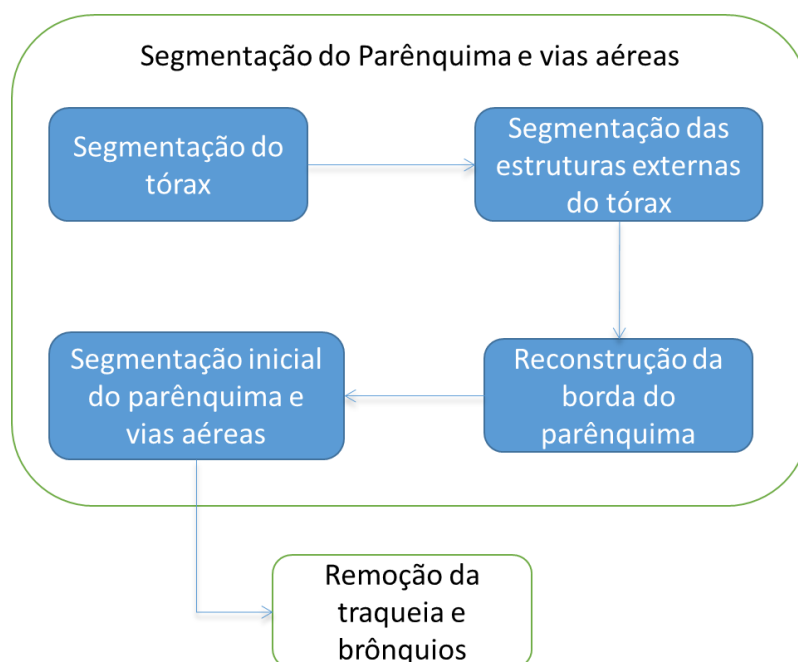


Figura 13 - Fluxo do pré-processamento das imagens

O pré-processamento é dividido em duas etapas: segmentação do parênquima e vias aéreas e remoção da traqueia e brônquios. A primeira etapa resulta em imagens contendo apenas o parênquima, traqueia e brônquios. A segunda remove a traqueia e os brônquios.

3.2.1. Segmentação do parênquima e vias aéreas

A segmentação do parênquima, por sua vez se divide nas seguintes subetapas: segmentação do tórax, segmentação das estruturas externas do tórax, segmentação inicial do parênquima e vias aéreas, reconstrução da borda do parênquima. Elas serão mais bem explicadas seguir.

Primeiramente, é feita segmentação do tórax através de uma limiarização do volume de imagens de TC. O limiar utilizado é -500 HU. Tudo abaixo desse limiar pode ser considerado o fundo e algumas regiões vazias do pulmão. O fundo será pintado de branco e o restante da imagem de preto. O resultado é visto na Figura 14.

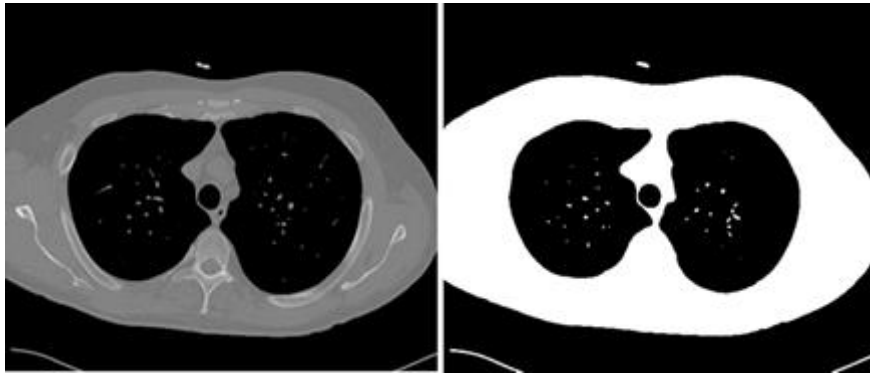


Figura 14 - Exemplo de segmentação do tórax. É aplicada uma limiarização, cujo limiar é -500 HU. Tudo abaixo desse limiar é considerado fundo e pintado de preto. O restante é pintado de branco

A segunda subetapa elimina as estruturas externas à caixa torácica. Essa eliminação é feita através da aplicação de vários algoritmos de crescimentos de região com sementes nas bordas da imagem resultante da subetapa anterior, segmentação do tórax. O resultado é visto na Figura 15.



Figura 15 - Exemplo de segmentação das estruturas externas do tórax. Aplica-se que algoritmos de crescimento de região 3D, com sementes nos quatro cantos do volume.

Na terceira subetapa, é obtido apenas o parênquima e vias aéreas. Para isso, basta utilizar a operação *AND* nas imagens resultantes das subetapas anteriores. Em seguida é feito mais um crescimento de região com sementes nas bordas da imagem e depois uma operação de inversão de valores de *voxels*. O resultado é visto na Figura 16.



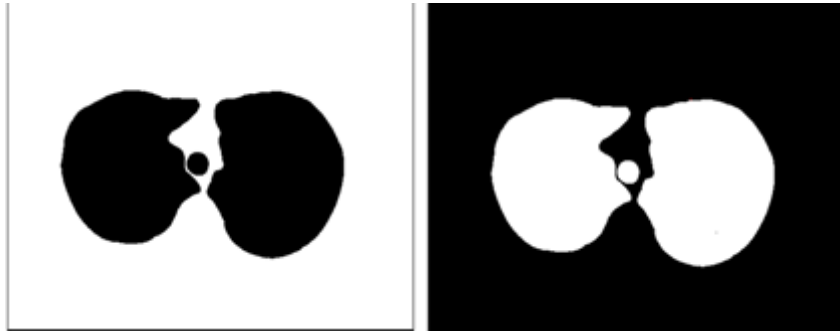


Figura 16 - Exemplo de segmentação do parênquima e vias aéreas. Primeiro usa-se a operação *AND* nas imagens resultantes das subetapas anteriores. Em seguida usa-se sucessivos crescimentos de região com sementes nos cantos do volume e por fim inverte-se os valores de *voxels* do volume.

Por fim, na quarta subetapa, é corrigida uma deficiência das subetapas anteriores. Essa deficiência é a eliminação de regiões do pulmão que são úteis e ocorre quando os crescimentos de região ultrapassam a região esperada e avançam além da borda do pulmão. Para corrigir isso, utiliza-se a operação morfológica de fechamento. Essa operação é ilustrada na Figura 17.

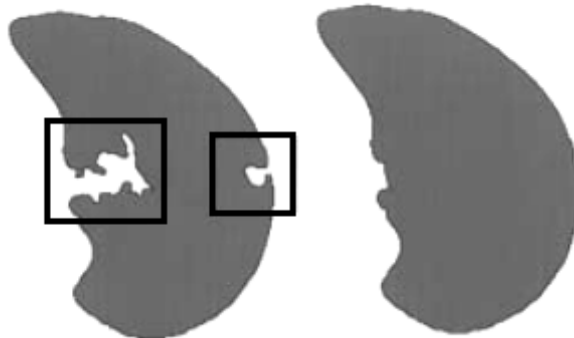


Figura 17 - Reconstrução da borda. Utiliza-se uma operação morfológica de fechamento.

3.2.2. Remoção da traqueia e brônquios

Essa etapa é responsável pela remoção da traqueia e brônquios, com o objetivo de obter-se o volume de imagens final pré-processado contendo apenas um pulmão (esquerdo ou direito) ou os dois.

O funcionamento dessa etapa é representado pela Figura 18. Primeiramente divide-se a imagem em dois setores, um de 40% (lado do menor pulmão) e outro de 60% (lado do maior pulmão). Em seguida, para cada setor, busca-se a linha de maior somatório de valor de *voxels*, ou seja, a linha que mais contem *voxels* brancos. O primeiro ponto branco de cada linha será usado como semente do algoritmo de crescimento de região. Se

for desejado os dois pulmões, utiliza-se as sementes das duas linhas. Se for desejado apenas um, utiliza-se apenas a semente da correspondente a pulmão desejado.

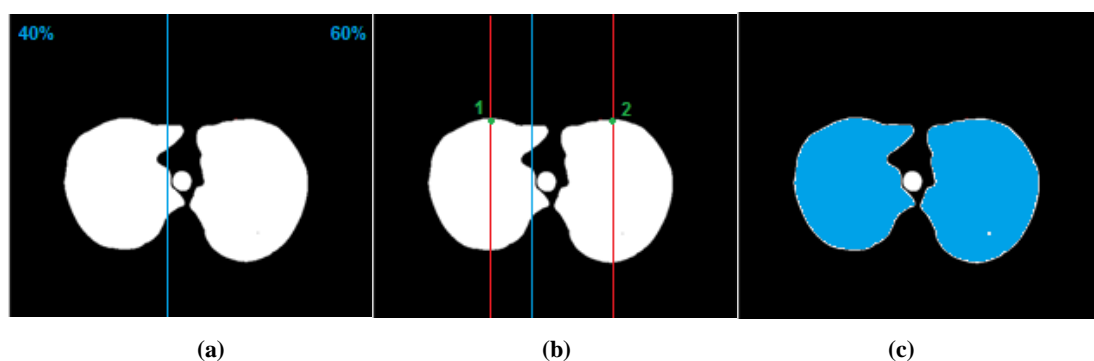


Figura 18 - Funcionamento da remoção da traqueia e brônquios. Em (a), divide-se a imagem em dois setores. Em (b), busca-se em cada setor a linha que o somatório de valor de *voxels* é mais alto e define o primeiro pixel branco de cada linha como uma semente. Em (c), aplica-se o crescimento de região com a semente.

3.3. Correspondência entre nódulos / massas em exames diferentes

Essa etapa é responsável por determinar uma correspondência entre os nódulos / massas localizados em volumes imagens diferentes. Esses volumes são obtidos de exames realizados em momentos diferentes em um mesmo paciente. Em outras palavras, busca-se verificar se o nódulo / massa *A* no exame *X* é o mesmo nódulo / massa *B* no exame *Y*. É após a determinação dessa correspondência que se pode determinar o quanto um mesmo nódulo / massa mudou do exame mais recente até o mais atual nas etapas posteriores.

Gerar essa correspondência, no entanto, nem sempre é uma tarefa fácil. Essa etapa é não trivial e necessária porque o nódulo / massa pode não estar na mesma localização nos exames *X* e *Y*. Essa diferença na localização pode acontecer por diversas circunstâncias. Primeiramente, embora seja seguido um protocolo nos exames de TC, pode haver diferenças de posicionamento e respiração do paciente nos exames. Alguns outros casos também podem ocorrer mudanças consideráveis na estrutura e massa corporal do paciente entre o intervalo de tempo dos exames. A diferença de localização pode atrapalhar principalmente a identificação de nódulos por que são menores que massas e, dessa forma, são mais difíceis de estabelecer uma correspondência quando essas mudanças ocorrem.

Entre as diversas técnicas utilizadas para tentar solucionar esses impasses, uma das utilizadas é o registro de imagens (Seção 2.7). Os registros utilizados são o rígido 3D

e o deformável B-spline 3D, nessa ordem. Esses dois registros utilizarão como parâmetro os volumes de imagens dos exames de TC pré-processados.

O pré-processamento usado na etapa anterior (Seção 3.2) deve manter apenas o pulmão que contém o nódulo / massa em que se deseja estabelecer a correspondência. A eliminação dos demais elementos é feita para aumentar a precisão dos registros de imagens. Entende-se que o nódulo / massa não se deslocará para fora do espaço desse pulmão, então não há necessidade de levar em consideração elementos que não precisam ser aproximados. A utilização desses elementos desnecessários poderia diminuir a precisão do registro porque o alinhamento deles poderia atrapalhar no alinhamento do pulmão onde se encontra o nódulo / massa, que é o órgão de interesse.

O volume fixo pré-processado será o do exame mais antigo e o volume móvel pré-processado será o do exame mais recente. Inicialmente, é aplicado o registro rígido 3D no volume móvel para que ele seja posto no mesmo sistema de coordenadas do volume fixo. Em seguida utiliza-se o registro *B-spline* 3D para corrigir problemas de deformação. O resultado da aplicação dos dois registros será o volume de imagens registrado. Esse volume, explicando de forma sucinta, é o volume móvel com formato e localização aproximados aos do volume fixo. Com essa aproximação, o nódulo / massa também estará em coordenadas parecidas no volume fixo e no registrado.

A marcação utilizada nesse dissertação foi feita manualmente através das informações encontradas nas bases de imagens. Uma vez que se sabe a localização do nódulo / massa no volume fixo e registrado, então calcula-se a *bounding box* 3D de suas marcações. A *bounding box* pode ser definida como o menor quadrilátero envolvente de um objeto específico. A representação de uma *bounding box* pode ser vista na Figura 19.

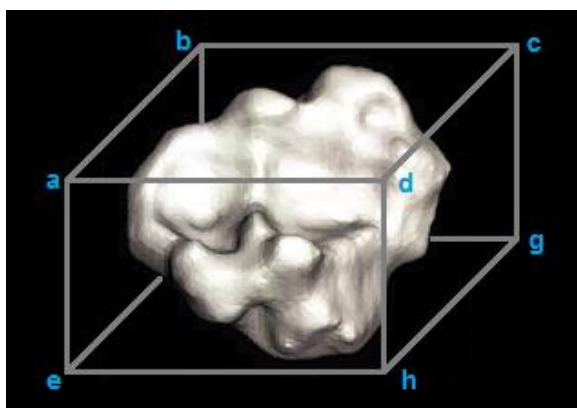


Figura 19 - *Bounding box* de um nódulo.

O cálculo dessa *bounding box* é simples e envolve a determinação de dois pontos, que sozinhos definem o quadrilátero. Esses pontos são, respectivamente, o ponto de menores coordenadas (x, y, z) do objeto e o ponto de maiores coordenadas (x, y, z) do objeto. Na Figura 19, esses pontos são, respectivamente, o ponto b e o ponto h . O demais pontos do quadrilátero podem ser calculados a partir desses dois.

Uma vez calculada a *bounding boxes* da marcação nos volumes fixo e móvel, busca-se não verificar se essas marcações possuem localização próximas. A definição de localização próximas nesse caso, é quando há sobreposição das áreas das *bounding boxes*. Essas operações são ilustradas na Figura 20.

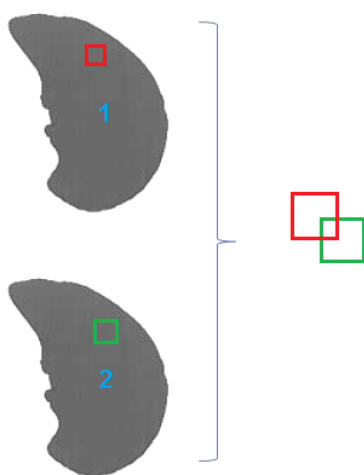


Figura 20 - Correspondência do nódulo em exames diferentes após o registro.

Se houver sobreposição das *bounding boxes*, ou seja, há voxels dentro da *bounding box* 1 cujas coordenadas também se encontram na *bounding box* 2, então assume-se que o nódulo / massa da *bounding box* 1 é o mesmo nódulo / massa da *bounding box* 2.

Esse procedimento busca estabelecer uma correspondência entre nódulos / massas de exames diferentes, para que a análise temporal seja aplicada sobre os nódulos / massas corretos, evitando que o nódulo / massa em um primeiro exame não seja confundido com outro num segundo exame, ou simplesmente não seja achado. No entanto, esse método serve apenas para estabelecer essa correspondência. O nódulo / massa do exame registrado não será utilizado porque provavelmente foi deformado pelo registro *B-spline*. Ao invés disso, nas próximas etapas será usado o nódulo / massa do volume móvel em que se comprovou ser o mesmo do volume fixo por intermédio do exame registrado.

A próxima etapa será responsável por pré-processar o nódulo / massa. Esse pré-processamento permite que a análise das mudanças da textura do nódulo / massas sejam mais precisas, evitando comparar elementos externos, como fundo, vasos ou próprio parênquima.

3.4. Pré-processamento do nódulo / massa

Essa etapa é responsável por preparar o nódulo para a extração de medidas usadas na análise temporal. Essa preparação é necessária porque a marcação do nódulo / massa nem sempre é precisa, seja ela feita manualmente por especialistas ou utilizando técnicas de detecção automática. Essa imprecisão pode atrapalhar extração de medidas na próxima etapa, uma vez que a marcação pode incluir *voxels* de partes da imagem que não pertencem ao nódulo / massa. As operações a seguir são utilizadas para melhorar o contorno do nódulo / massa.

Primeiramente é preciso extrair o menor subvolume do volume de imagens do pulmão que contenha o nódulo. Esse subvolume possui a mesma localização e dimensões da *bounding box* da marcação do nódulo / massa. Dessa forma, para obter esse subvolume é necessário calcular essa *bounding box*.

No caso do nódulo / massa, o objeto considerado é sua marcação. Uma vez determinada a sua *bounding box*, pode-se usá-la para extrair o subvolume do nódulo / massa. A extração do subvolume através da *bounding box* da marcação pode ser ilustrada pela Figura 21.

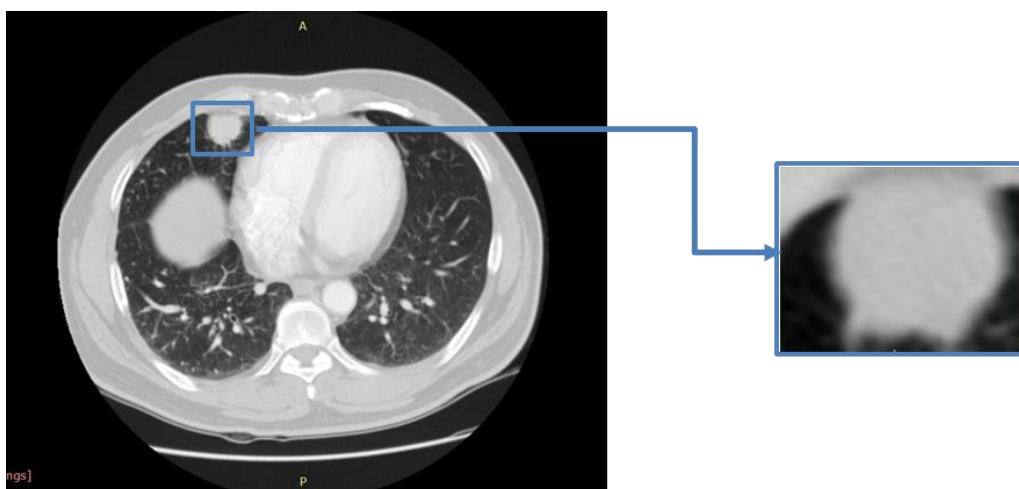


Figura 21 - Extraindo subvolume contido na *bounding box* da marcação de uma massa.

Após extrair o subvolume da *bounding box* do nódulo / massa é feita uma binarização dos *voxels* que estão dentro e fora da marcação, como visto na Figura 22. A parte branca da imagem corresponde ao nódulo / massa enquanto a preta corresponde todas as demais estruturas ou fundo.

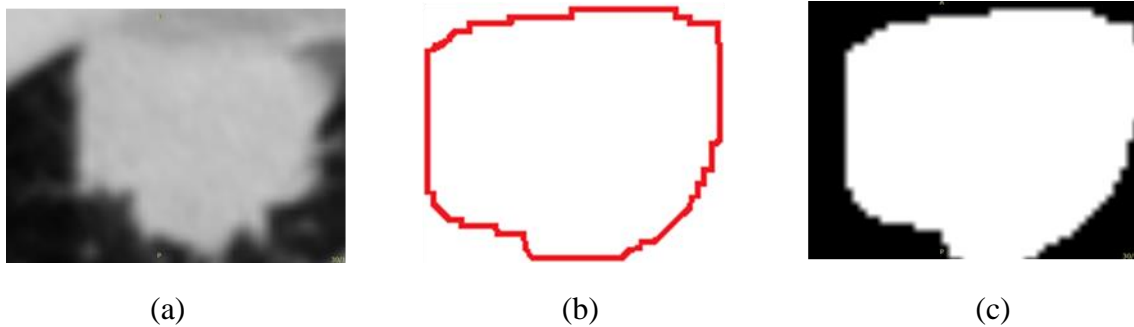


Figura 22 - Binarizando nódulo com base na marcação. (a) corresponde a imagem extraída da *bounding box* da marcação e (b) corresponde a marcação. Em (c), encontra-se o resultado da binarização de (a) com base em (b), onde a parte branca é a região interna a marcação.

É comum que a marcação do nódulo ou da massa incluía *voxels* de outras estruturas ou do fundo. Quando isso ocorre, esses *voxels* desnecessários podem atrapalhar a análise do nódulo / massa. Em alguns casos a marcação é feita assim pelo próprio especialista, devido a forma complexa do nódulo / massa. Em outros casos, a imperfeição é originada de algoritmos de detecção que geram a marcação automaticamente, mas com imprecisão. Para auxiliar na correção do contorno da marcação, utiliza-se uma série de algoritmos. O primeiro desses algoritmos a ser utilizados é a limiarização de Otsu. A Figura 23 mostra o resultado dessa limiarização.



Figura 23 - Aplicando a limiarização de Otsu na imagem da massa.

Na Figura 23, observa-se que a limiarização de Otsu é eficaz para separar o fundo (região preta) das demais estruturas (região branca). No entanto, estruturas que não pertencem ao nódulo / massa, como a parede torácica também precisam ser removidas da imagem.

Em seguida, para obter um contorno mais preciso, utilizou-se a operação morfológica *AND* entre a imagem resultante da binarização feita usando a marcação e a imagem resultante da limiarização de Otsu. Isso pode ser visto na Figura 24.

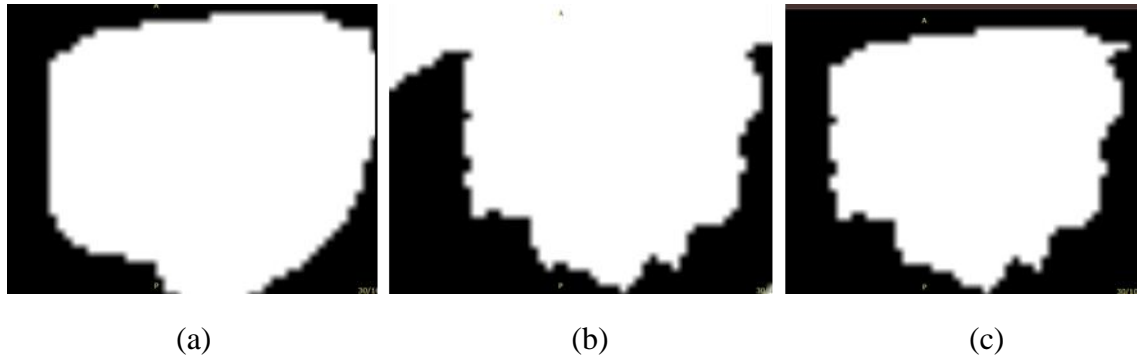


Figura 24 - Criando máscara da massa. (a) corresponde a imagem binarizada da massa com base na marcação e (b) corresponde a imagem da limiarização de Otsu. Em (c), encontra-se o resultado da operação morfológica *AND* entre (a) e (b).

Por fim, a imagem resultante da operação morfológica *AND* é utilizada como máscara sobre o subvolume da *bounding box* da marcação (Figura 25). Através da aplicação da máscara, elimina-se valores de *voxels* excessivos e indesejados e mantém-se aqueles que pertencem ao nódulo / massa.

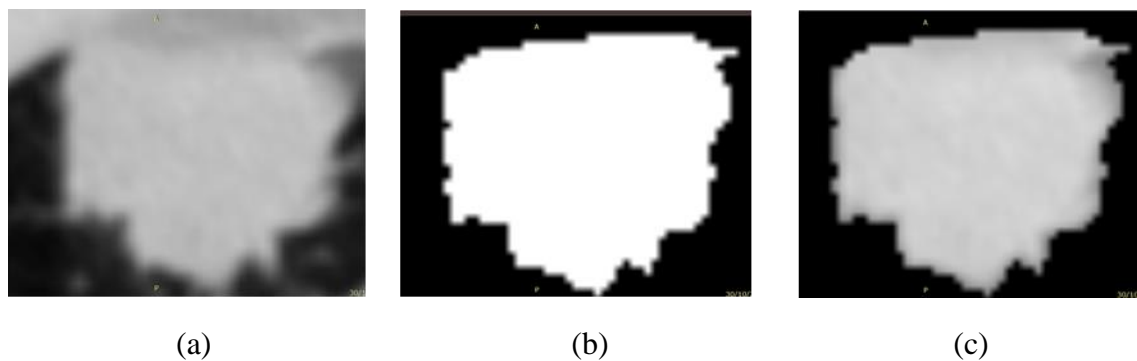


Figura 25 - Melhoramento do contorno da massa. (a) corresponde a imagem extraída da *bounding box* da marcação e (b) corresponde a máscara que será usada para eliminar *voxels* indesejados. Em (c) o nódulo pré-processado

Essa etapa foi responsável por pré-processar o nódulo / massa de forma a aprimorar a precisão da marcação. Através desse aprimoramento, os *voxels* que não pertencem ao nódulo / massa são eliminados.

A próxima seção, Definição das regiões de interesse do nódulo, mostrará as técnicas utilizadas para obter as regiões de interesse do nódulo / massa pré-processados. São essas regiões que serão utilizadas para análise temporal de textura.

3.5. Definição das regiões de interesse do nódulo / massa

Essa etapa será responsável por definir as regiões do nódulo / massa onde serão aplicadas a análise temporal. Como a análise temporal deve avaliar o nódulo / massa em tempos diferentes, as regiões de interesse serão definidas em dois subvolumes de nódulo / massa pré-processados de um mesmo paciente. Para uma definição mais precisa dessas regiões, os subvolumes serão pré-processados, como explicado na Seção 3.4.

O nódulo / massa é abordado de quatro formas diferentes para a definição dessas regiões. Primeiro, ele pode ser abordado como um todo. Dessa forma, serão extraídos todos os seus voxels no tempo um e no tempo dois e eles serão confrontados para averiguar se houve mudanças gerais. A segunda abordagem busca encontrar as regiões onde houveram maiores mudanças entre o tempo 1 e o tempo 2. A terceira forma de abordar o nódulo / massa, é como um objeto que possui “camadas”, de forma que apenas os voxels de cada camada no tempo 1 será avaliada com a sua respectiva camada no tempo 2. A seguir serão explicadas melhor essas abordagens. A última abordagem busca encontrar a região que cresceu no nódulo ou massa no decorrer do tempo, ou seja, a região que se encontra no exame mais recente, mas não se encontra no mais antigo. Essa última abordagem, no entanto, só é possível se houve aumento do volume do nódulo no decorrer do tempo.

3.5.1. Nódulo / massa inteiro

Quando o nódulo / massa é usado como um todo, é simples obter seus voxels para posterior análise. Como pré-processamento remove o fundo e demais objetos desnecessários do subvolume, é possível ignorar todos os voxels de valores preto (que correspondem ao fundo) e buscar apenas os demais voxels. Isso deve ser feito nos dois subvolumes do nódulo / massa. Todos esses voxels, então, serão utilizados na análise temporal.

3.5.2. Regiões de maiores mudanças

Essa abordagem busca encontrar as regiões de onde há maiores diferenças de textura nos dois subvolumes do nódulo / massa. As operações vistas no pré-processamento do nódulo anteriores devem ser aplicadas nos dois subvolumes correspondentes obtidos dos dois exames.

Para que essas regiões de maiores mudanças sejam encontradas, é necessário que haja uma correspondência entre os *voxels* dos dois subvolumes. Nesse caso, essa correspondência é obtida espacialmente, ou seja, os *voxels* de mesma coordenada nos dois subvolumes serão *voxels* correspondentes. No entanto, sabe-se que o nódulo / massa tem formatos diferentes nos dois subvolumes e isso impede que essa correspondência entre os *voxels* seja satisfatória. Para resolver isso, o nódulo / massa devem ter formato aproximado.

Para que os subvolumes possam ter o formato aproximado, eles serão registrados. Mas para que o registro seja possível, é necessário que os subvolumes fixo e móvel possuam as mesmas dimensões x , y e z . Isso deve ser feito sem deformar as imagens do nódulo e mantendo-as concêntricas. Esses pré-requisitos permitem maior eficiência do registro de imagens. Uma das formas de adequar as dimensões das imagens é criando uma terceira imagem auxiliar, na qual os subvolumes dos nódulo / massas possam ser redesenhados. Essa imagem auxiliar terá todos os *voxels* com valor de fundo e com dimensões x , y , z iguais as maiores dimensões entre as os subvolumes fixo e móvel. Essa imagem auxiliar pode ser vista na Figura 26.

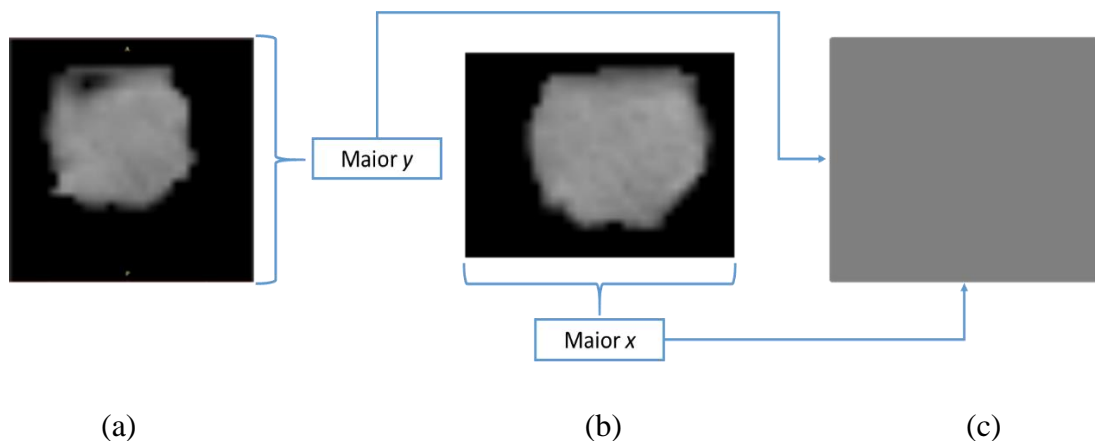


Figura 26 - Criação de volume auxiliar com maiores dimensões. (a) corresponde ao subvolume da massa mais antigo e (b) corresponde ao subvolume da massa mais recente. Em (c), é construído um volume auxiliar de valores de *voxels* de fundo e com dimensões iguais as maiores dimensões encontradas em (a) e (b).

Uma vez criado esse volume auxiliar, os dois subvolumes do nódulo / massa podem ser “redesenhados” sobre ele de forma concêntrica. A Figura 27 mostra como ficaria o um subvolume massa redesenhado sobre o volume auxiliar. Após o redesenho dos dois subvolumes do nódulo / massa, eles passarão a ter as mesmas dimensões.

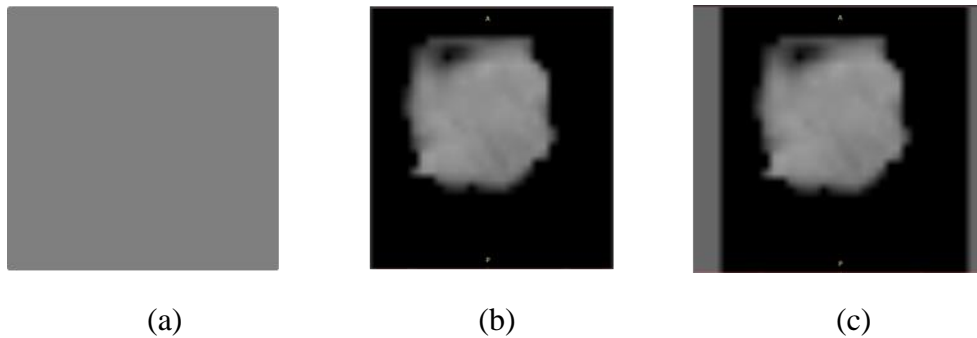


Figura 27 - Redesenhando o subvolume da massa sobre o volume auxiliar maior. (a) ao volume auxiliar. (b) corresponde ao subvolume da massa. (c) é resultado do redesenho de (b) sobre (a).

Antes do registro, mais uma etapa ainda precisa ser aplicada: a interpolação. A interpolação dos volumes de imagens é importante por dois motivos. Primeiramente, é possível que os subvolumes do nódulo / massa não possuam a mesma quantidade de fatias, mesmo com a adequação das dimensões das imagens. Isso ocorre devido a resoluções diferentes entre os tomógrafos usados na aquisição das imagens. Com resoluções diferentes, há espaçamentos entre fatias diferentes. A interpolação, nesse caso, será usada para adequar o espaçamento entre os volumes. O outro motivo é que, em geral, o espaçamento entre nódulos / massas é muito grande e, assim, há poucas fatias a serem analisadas e, conseqüentemente, poucos *voxels*. Nesse caso, a interpolação estimará mais *voxels* para facilitar a análise. Apesar desses *voxels* novos nem sempre possuírem os valores reais para as suas localizações, seus valores tendem a ser próximo ao valor real.

A interpolação utilizada é a linear (Subseção 2.6.1.). Apesar da simplicidade, ela é suficiente pois o espaçamento dos volumes de imagem das bases já é pequeno, de forma que não será preciso estimar novas fatias muito distantes das já existentes. Devido a essa aproximação entre as fatias, os valores estimados serão bem próximos do real, não sendo, assim, necessário usar interpolações mais sofisticadas. A interpolação linear estimará fatias com 1 mm de espessura para massas e 0.5 mm para nódulos.

Uma vez que os subvolumes estão com as mesmas dimensões e interpolados, então pode-se aplicar os registros de imagens. Os registros utilizados serão o rígido seguido do *B-spline*. A Figura 28 mostra o resultado da aplicação dos registros.

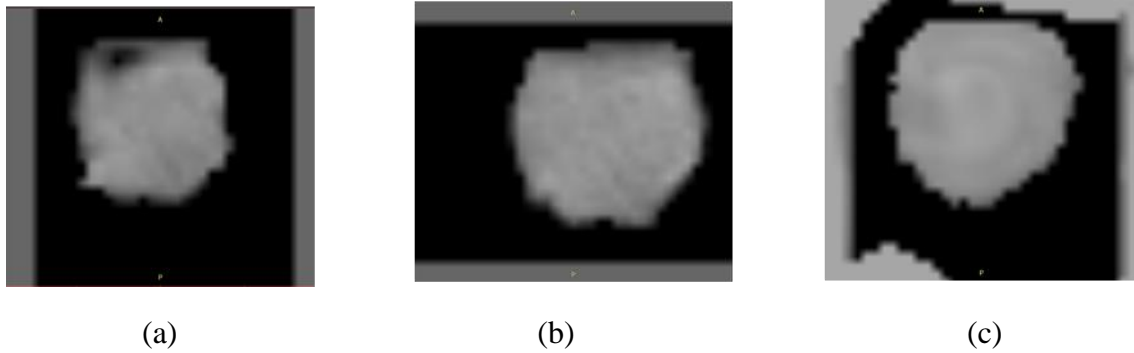


Figura 28 - Aplicando o registro de imagens rígido e deformável *B-spline* nos subvolumes da massa. (a) corresponde ao subvolume fixo (b) corresponde ao subvolume móvel. Em (c), encontra-se o subvolume resultante da combinação dos registros rígido e deformável *B-spline* usando (a) e (b).

Após as operações anteriores, os subvolumes possuirão dimensão, espaçamento (1 mm) e formato do nódulo / massa semelhantes. Os subvolumes agora estão prontos para a identificação de regiões de diferenças. Essa identificação é feita aplicando uma operação que verifica as diferenças locais dos volumes de imagens. A determinação das diferenças locais é obtida primeiramente através do módulo da diferença das médias de regiões cúbicas de dimensões de $3 \times 3 \times 3$ *voxels*, que sejam comum a dois nódulos / massas. A Figura 29 mostra a aplicação dessa operação. Essa operação é aplicada sucessivas vezes ao longo das regiões comuns entre os nódulos e o resultado é aplicado ao volume resultante. As operação não são mais aplicadas após todas as regiões do nódulo tenham sido varridas. Essas regiões comuns não incluem o *voxels* do fundo da imagem por ser desnecessário.

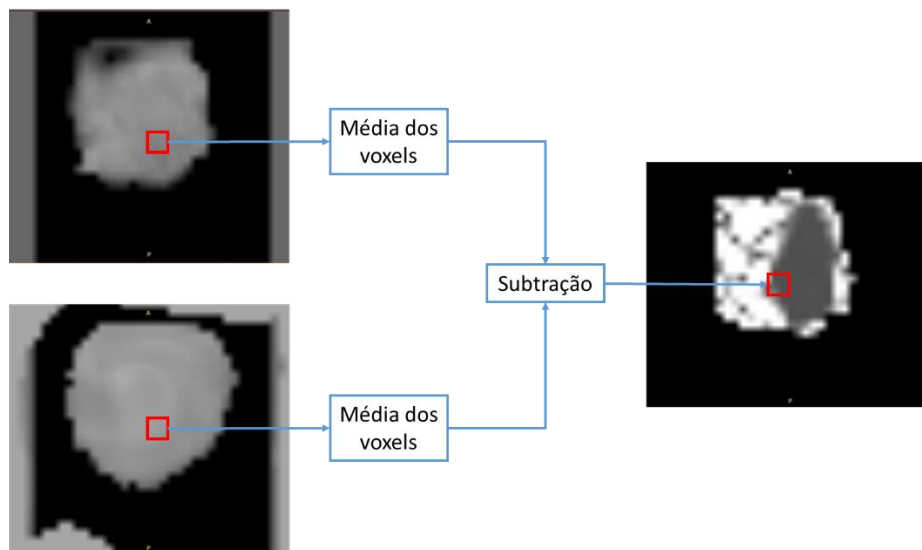


Figura 29 - Criando volume de diferenças das médias entre dois subvolumes de uma massa. Primeiramente busca-se todas as regiões $3 \times 3 \times 3$ dos dois subvolumes que não tem *voxels* de fundo. Depois aplica-se o módulo da diferença das médias das regiões correspondentes num novo volume.

Uma vez criado o volume da diferenças das médias, é necessário focar nas regiões que apresentam as maiores diferenças. Nesse sentido, é preciso remover as regiões de fundo e baixa diferença, que não são importantes, e manter apenas as regiões de alta diferença. Uma boa solução para essa separação é a aplicação do algoritmo de agrupamento *K-means*. Para esse problema, o *K-means* deve ser configurado para encontrar três classes, que correspondem os três tipos de regiões citadas. O próprio *K-means* é responsável por definir o que são regiões de fundo, baixa e alta diferença.

O resultado do agrupamento é vista na Figura 30.

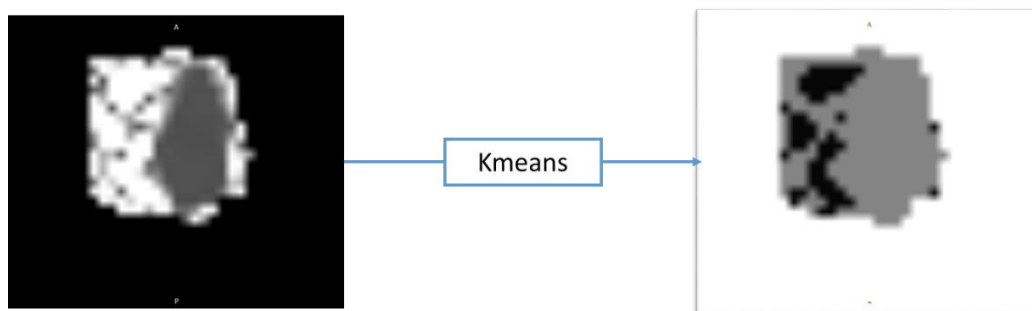


Figura 30 - Aplicando algoritmo *K-means* de três classes no volume de diferença das médias.

O *K-means* obtém as três regiões de agrupamento (grupos) desejadas. Os grupos podem ser distinguidos por suas cores. Nesse caso, temos as cores branco, cinza e preto, representando cada uma das classe de agrupamento. No entanto, o *K-means* sozinho não é capaz de identificar cada grupo como fundo, região de baixa diferença ou região de alta diferença. Sem essa identificação, não é possível isolar o grupo que representa as região de alta diferença, que é desejada para a análise temporal. Para obtermos essa identificação, segue-se o seguinte processo:

1. Primeiramente, obtém-se as médias aritméticas de valores de *voxels* de cada região.
2. Em seguida, ordena-se essas médias em ordem crescente.
3. Identifica-se, então, os grupos pelas suas médias de valores de *voxel* correspondentes, ou seja, identifica-se o fundo como o grupo de menor média, a região de baixa diferença como grupo de média intermediária e a região de alta diferença como o grupo de maior média.
4. Por fim, cria-se um novo volume apenas com a região de alta diferença

Isso pode ser visualizado na Figura 31.

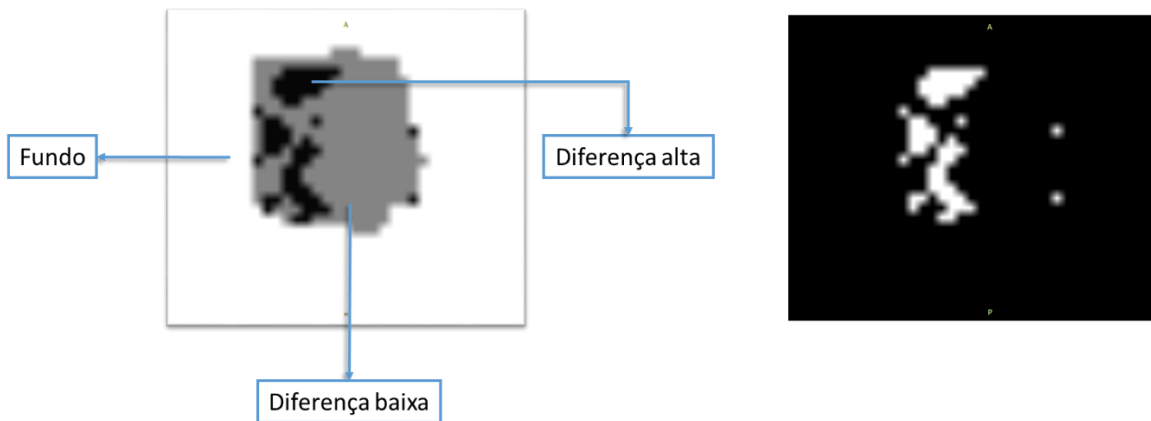


Figura 31 - Identificando os grupos do *K-means*. Para cada grupo, tira-se as médias aritméticas dos *voxels* do volume de diferença das médias (Figura 29) e usa-as para identificar os grupos.

Como visto na Figura 31, é possível identificar as região de alta diferença do nódulo / massa. No entanto, percebe-se que a mesma possui várias sub-regiões desconexas. Para a análise temporal, é interessante avaliar cada uma dessas sub-regiões separadamente. Dessa forma foi utilizado o algoritmo de crescimento de região para separar essas sub-regiões. Ele foi aplicado da seguinte forma:

1. Percorre-se a imagem em busca de *voxels* brancos (branco é a cor da região de alta diferença);
2. Uma vez encontrado um *voxel* branco, utiliza-o como semente e aplica-se o algoritmo de crescimento de região;
3. A sub-região desconexa definida pelo algoritmo é separada para análise posterior;
4. Pinta-se a sub-região desconexa definida pelo algoritmo de preto, para evitar que o ele seja aplicado novamente em outro de seus *voxels*.
5. Continua-se o processo a partir do passo 1 até pintar toda imagem de preto.

Esse processo pode ser melhor visualizado na Figura 32.

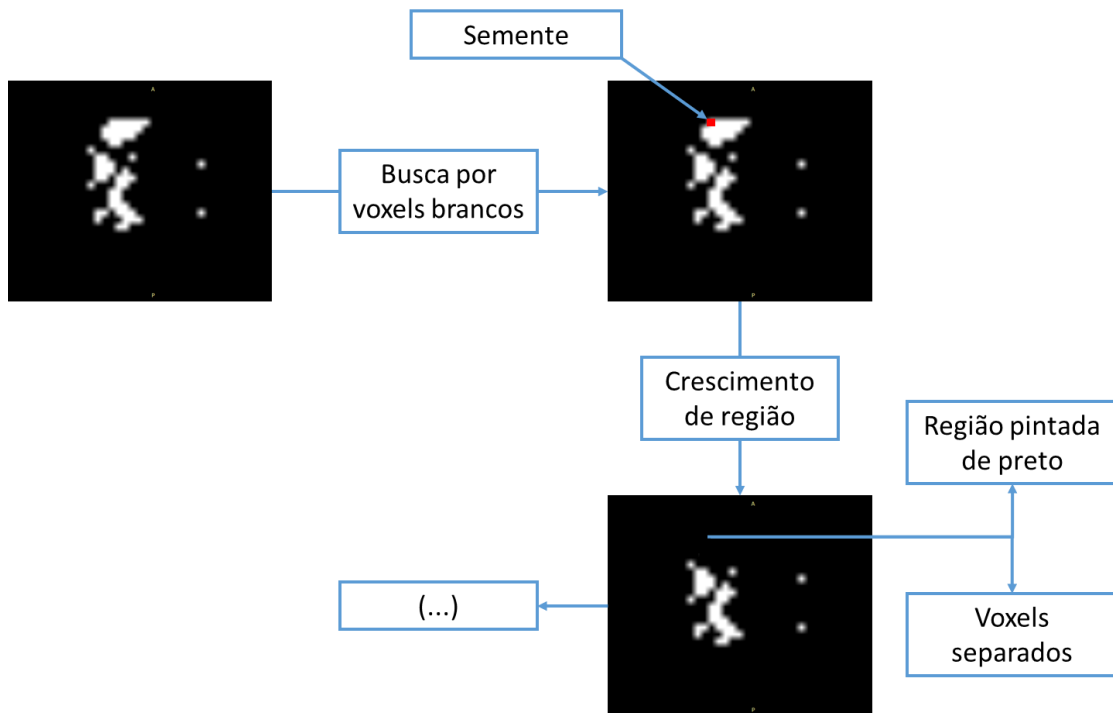


Figura 32 - Aplicando sucessivos crescimentos de regiões para identificar regiões desconexas.

Uma vez que as regiões são separadas, é possível analisá-las. No entanto, nem todas as regiões podem ser consideradas relevantes, como é o caso de regiões muito pequenas. Dessa forma, optou-se por manter apenas as regiões com uma quantidade de relevante de *voxels*. Essa quantidade foi definida como 20 ou mais *voxels*.

3.5.3. Camadas

As mudanças de textura do nódulo podem ser estudadas pelas suas camadas. Considerando que um nódulo possui 3 camadas, em seus dois subvolumes, uma mais interna, outra mais externa e outra intermediária, é possível comparar as mudanças em cada uma dessas camadas no decorrer do tempo. As técnicas aplicadas a seguir mostrarão como extrair essas camadas.

Primeiramente, usa-se cada um dos dois subvolumes pré-processados e em seguida os interpola para obterem espaçamento de 1 mm caso seja um subvolume de massa e 0.5 mm caso o subvolume seja de um nódulo, tal como é feito na abordagem de regiões de maiores mudanças (Subseção 3.5.3). Em seguida essa imagem deverá ser binarizada. A binarização considera uma região preta como o fundo e a uma branca como a região do nódulo interpolado. Isso pode ser visto na Figura 33.

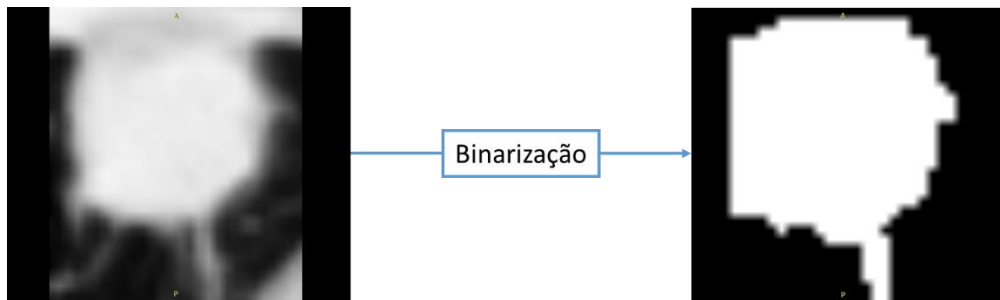


Figura 33 – Binarização da massa.

Após a binarização será aplicada uma operação para extração das bordas do volume. Para que isso seja feito a imagem deve ser varrida em busca de pixels brancos. Uma vez encontrado, a sua vizinhança imediata é analisada. Se houver algum valor preto em volta, esse *voxel* será considerado como borda. Ao final do processo, todos os voxels considerados como borda serão armazenados e pintados de preto na imagem. Esse processo é repetido até que não sobrem mais *voxels* brancos. Esse processo pode ser visto na Figura 34.

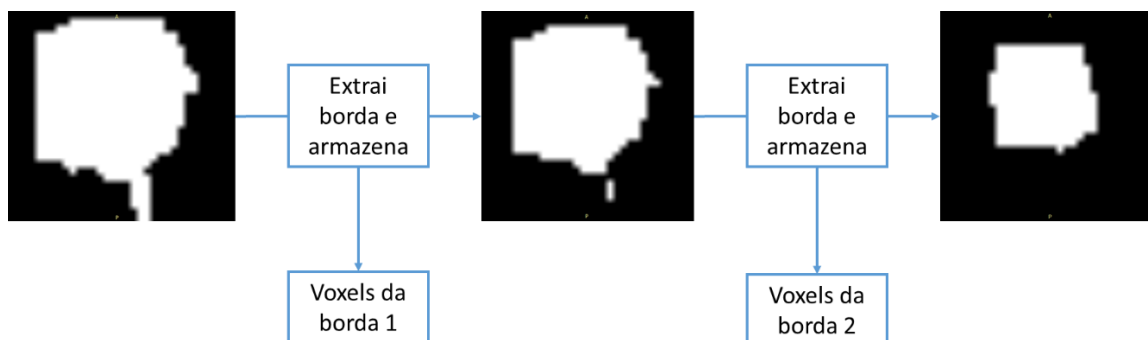


Figura 34 – Sucessivas extrações de bordas da massa binarizada.

A identificação e remoção das bordas de forma iterativa não só possibilita encontrar os *voxels* de cada borda a cada iteração como também permite armazenar esses *voxels* de forma hierárquica. Pode-se considerar, por exemplo, que os *voxels* de menor hierarquia serão os obtidos na primeira extração de borda e os *voxels* de maior hierarquia serão aqueles obtidos na última extração de borda.

O objetivo de organizar esses *voxels* por hierarquia é para que seja possível definir as “camadas” do nódulo / massa. Uma camada será considerada um subconjunto da hierarquia de *voxels*. Se for desejado que uma camada tenha espessura de um *voxel*, os *voxels* de menor hierarquia serão classificados como a primeira camada. No entanto, a espessura de um *voxel* pode não ser interessante para análise, por ser muito fina e também

porque, em casos de grandes massas, pode gerar muitas camadas. Se for desejado que uma camada tenha dois *voxels* de espessura, então a primeira camada será composta pelos *voxels* da primeira e segunda hierarquia. A Figura 35 mostra como ficaria essa composição.

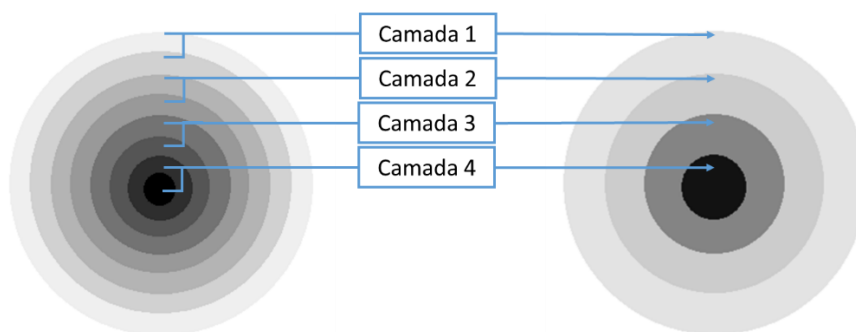


Figura 35 – Agrupando bordas extraídas para formar as camadas.

Retornando a Figura 34, pode-se compor os *voxels* da borda 1 com a borda 2 e obter-se primeira camada da imagem. O resultado pode ser visto na Figura 36.

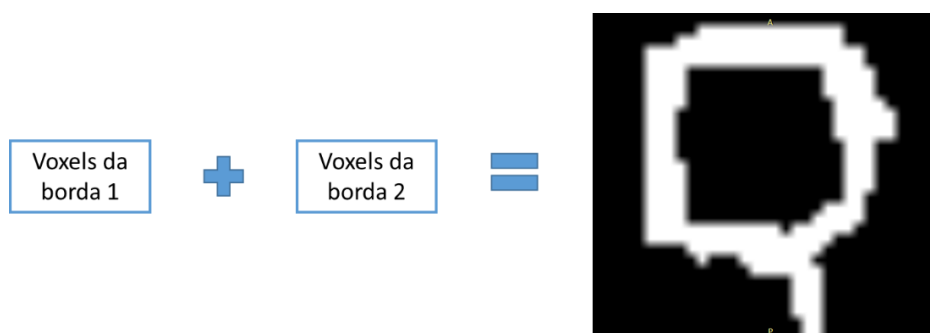


Figura 36 – Formando uma camada com espessura de dois *voxels*.

A Figura 36 mostra uma imagem binária. Cada uma das camadas binárias geradas pelo processo pode ser aplicada como máscara na imagem original e assim obter os valores de *voxels* da imagem original. Uma vez obtido esses valores de *voxels*, para cada camada de cada subvolume, estes podem ser usadas na análise temporal.

A próxima seção, Análise temporal de texturas, utilizará os *voxels* das regiões de interesse obtidas nessa etapa para analisar as mudanças de textura do nódulo / massa. No caso, serão utilizados, por vez, dois subvolumes, um de cada exame obtido de um paciente. A análise temporal buscará as mudanças de textura ocorridas no decorrer do tempo no nódulo / massa utilizando índices de similaridade.

3.5.4. Regiões novas

Essa última abordagem busca isolar as regiões novas do nódulo / massa, que se desenvolveram no decorrer do tempo. Essa abordagem é interessante porque, uma vez isolada essa nova região, é possível confrontá-la com a região antiga e verificar se o tecido que surgiu entre os exames sofreu mudanças. Como já citado, essa abordagem só é possível se houve aumento do volume do nódulo / massa no decorrer do tempo e por isso não pode ser aplicada em casos onde o nódulo / massa regrediu.

Para isolar as regiões novas do nódulo / massa, utilizasse o nódulo pré-processado, com dimensões ajustadas e interpolado, tal como é feito na subseção 3.5.3. Em seguida, aplica-se apenas o registro rígido considerando o subvolume mais antigo como fixo e o mais recente como móvel. O registro rígido é usado para que os subvolumes possam encontrar-se centralizados. Esse registro pode ser visto na Figura 37.

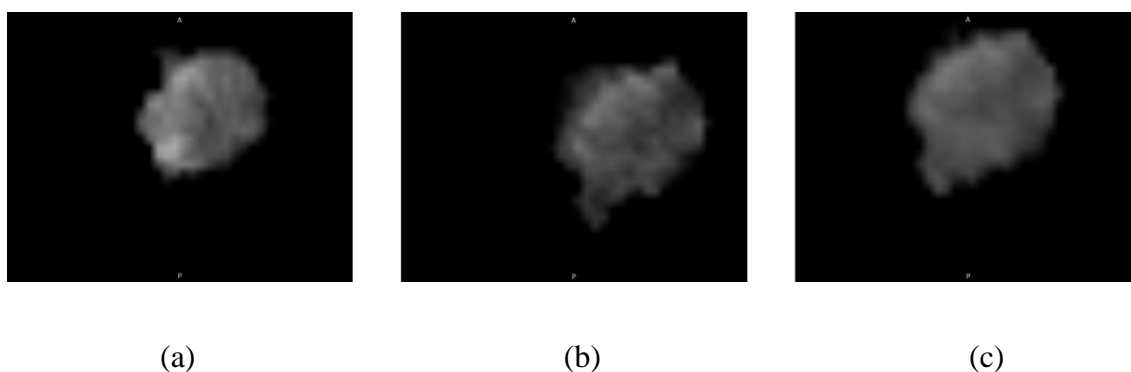


Figura 37 - Aplicando o registro de rígido nos subvolumes da massa. (a) corresponde ao subvolume fixo (b) corresponde ao subvolume móvel. Em (c), encontra-se o subvolume resultante da combinação dos registros rígido usando (a) e (b).

Uma vez que os subvolumes encontram-se centralizados, então busca-se os *voxels* do nódulo / massa que estão presentes no subvolume do exame mais recente e não estão presentes no subvolume do exame mais antigo. Esses *voxels* são, de fato, das regiões que cresceram no nódulo / massa no decorrer do tempo entre os dois exames, pois no exame mais antigo, esses *voxels* correspondiam ao fundo. Isso pode ser melhor visualizado na Figura 38.

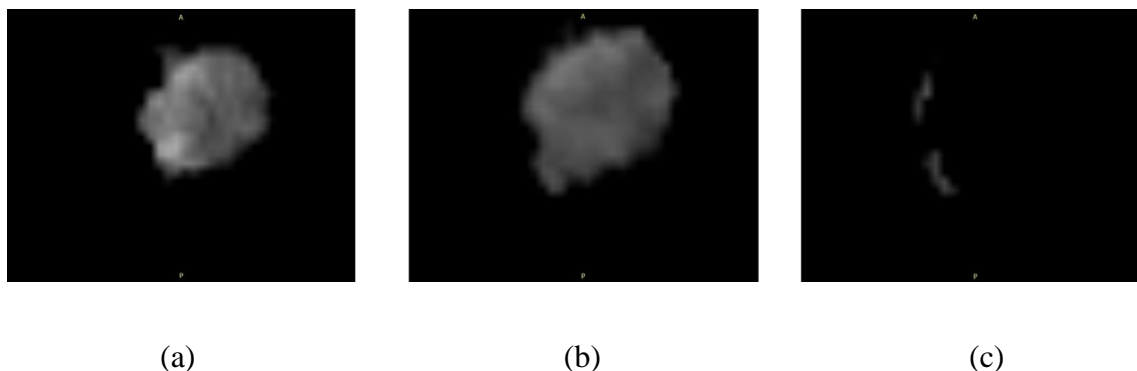


Figura 38 – Isolando região nova da massa. (a) corresponde ao subvolume mais antigo e (b) corresponde ao subvolume mais recente. Em (c), encontra-se a região que se desenvolveu no intervalo entre os exames.

3.6. Análise temporal de texturas

A etapa da análise temporal é responsável por analisar mudanças de tecido (textura) em nódulos / massas ocorridas no decorrer do tempo. Ela será aplicada em dois subvolumes de um nódulo / massa por vez, um obtido no exame de TC mais antigo e outro obtido no exame de TC mais atual. Antes da análise, esses subvolumes são pré-processados, como visto na Seção 3.4, para eliminar regiões indesejadas para a análise e em seguida são obtidas as regiões de interesse através das abordagens apresentadas na Seção 3.5.

As medidas de mudanças de textura são obtidas através da análise do padrão de textura do nódulo / massa que é diretamente relacionada aos seus valores de *voxel*. Nessa etapa, são calculados índices de similaridade (Seção 2.9). Os índices de similaridade utilizados neste trabalho são os descritos em (Meyer, 2002). No trabalho em questão eles serão utilizados para comparar a similaridade de texturas do nódulo / massa em exames diferentes. Considerando as equação as fórmulas dos índices de similaridade da Tabela 1 (Seção 2.8), as variáveis *a*, *b* e *c* serão redefinidas como:

- *a*: valores de *voxels* que pertencem tanto ao nódulo / massa no exame 1 quanto ao nódulo / massa no exame 2;
- *b*: valores de *voxels* que pertencem ao nódulo / massa no exame 1 mas não pertencem ao nódulo / massa no exame 2;
- *c*: valores de *voxels* que pertencem ao nódulo / massa no exame 2 mas não pertencem ao nódulo / massa no exame 1;

Considerando essas variáveis e as fórmulas dos índices de similaridade, o próximo passo é obter o conjunto de valores de *voxel* obtidos na etapa anterior (Seção 3.5). Como foi visto, os valores de *voxels* obtidos podem ser de todos os *voxels* dos dois subvolumes, apenas das regiões de alta diferença entre os dois subvolumes, de suas camadas, ou entre a região inteira do nódulo / massa no primeiro exame e a nova região que se desenvolveu no intervalo entre os exames.

No primeiro caso, os índices de similaridade calcularão a similaridade da textura geral do nódulo / massa, obtendo uma medida de mudança dele como um todo. No segundo caso, a similaridade calculada será apenas das regiões de maiores diferenças e por essa razão, tendem a ter a menor similaridade entre as regiões do nódulo. No terceiro caso, é verificado a similaridade do nódulo via suas camadas, obtendo assim valores que indiquem quais camadas sofreram maiores mudanças de textura. No último caso, busca-se verificar se a região que se desenvolveu no tempo tem textura diferente da região inicial do nódulo / massa.

Os resultados da análise de textura geral, utilizando todos os valores dos *voxels* do nódulo / massa, podem ser utilizados como referência para as análise de textura locais. Dessa forma é possível apontar quais regiões mudaram mais em relação as mudanças gerais do nódulo / massa e ter uma ideia de quão acentuada foi a mudança quantitativamente.

O próximo capítulo, Resultados, apresentará os resultados obtidos com aplicação de toda a metodologia. Será dado destaque a técnica de estabelecimento de correspondência entre subvolumes dos nódulos / massas e análise dos valores obtidos pelo índices de similaridade na análise temporal.

4. RESULTADOS

Neste capítulo, serão apresentados os resultados obtidos através da aplicação da metodologia (Capítulo 3) nas bases de imagens. As bases de imagens utilizadas são as vistas na Seção 3.1. Primeiramente, para melhor análise das técnicas, foi utilizada a base de massas sintéticas (Subseção 3.1.3). Os resultados em casos reais foram obtidos através da aplicação das técnicas na PLRADR (Subseção 3.1.1.) e na base do *Volcano Challenge* (Subseção 3.1.2).

Por questão de simplificação, não serão mostrados os resultados obtidos em todos os exames das bases, já que há uma grande quantidade de informações. Então, para a melhor exposição dos resultados, foram utilizados os três massas sintéticas, da base de massas sintéticas, e escolhidos dois nódulos da base do *Volcano Challenge* e duas massas da PLRADR. Nesses casos, foram aplicadas todas as etapas da metodologia. Será mostrado, portanto, os valores obtidos pela etapa final, análise temporal de textura (Seção 3.6). Considerando que há quatro abordagens para a definição das regiões de interesse do nódulo / massa (Seção 3.5), os resultados serão divididos para cada abordagem nesses nódulos / massas, quando possível aplicar cada abordagem. Resultados de mais casos dessas bases podem ser encontrados no Apêndice.

A seguir, será mostrado os resultados obtidos na análise temporal de textura aplicada aos casos mencionados em cada uma das abordagens de definição das regiões de interesse. No caso da PLDADR e a base do *Volcano Challenge*, os nódulos / massas serão identificados por meio de uma chave. Essa chave é um conjunto de caracteres e números que serve para encontrar essas lesões nas bases. Cada chave corresponde a um paciente. No final, será feita uma discussão dos valores obtidos nessas bases.

4.1.1. Massas sintéticas

Na base de massas sintéticas (Subseção 3.1.3), a análise temporal será aplicada para entender melhor o funcionamento dos índices de similaridade e para mostrar a sua eficiência. Serão mostrados os resultados obtidos em três combinações de pares de massas dessa base, utilizando as três primeiras abordagens definição das regiões de interesse da massa.

Para essa base, só é necessária a aplicação das etapas das Seções 3.5 e 3.6. As demais etapas, como o pré-processamento do nódulo (Seção 3.4), são desnecessárias porque os volumes de imagens das massas sintéticas já foram criados com fundo possuindo valor de *voxel* diferenciado e que será ignorado na análise temporal.

Inicialmente, o primeiro nódulo (Figura 12a) fará par com ele mesmo na análise temporal. O objetivo é mostrar como os índices se comportarão numa estrutura idêntica. As Tabelas 2 mostram os resultados obtidos utilizando a massa inteira como região de interesse. Como pode ser visto nela, os índices apresentaram valor máximo, uma vez que as massas sintéticas são idênticas.

Tabela 2 - Análise temporal global do par formado por duas cópias da primeira massa sintética.

Índice de similaridade	Valores
Anderberg	1.0
Jaccard	1.0
Kulczynski 1	1.0
Kulczynski 2	+ ∞
Ochiai	1.0
Czekanowsky	1.0

A próxima abordagem é dividir o nódulo em camadas. Foram obtidas oito camadas com espessura de dois *voxels*. Todas as camadas apresentaram mesmo valor para os índices. Esses valores são os mesmos mostrados na Tabela 2. Como os nódulos são idênticos, a terceira abordagem de definição de região de interesse não encontrou nenhuma região de alta diferença.

Em seguida a análise temporal foi utilizada com o par formado pela primeira massa sintética e segunda massa sintética (Figura 12b). Os resultados utilizando todos os *voxels* como região de interesse dos dois nódulos são vistos na Tabela 3. Como pode ser visto nessa tabela, os índices também apresentaram valor máximo. Esses valores são esperados, já que as duas massas sintéticas apresentam, globalmente, os mesmos valores de *voxels*, embora em distribuição diferente.

Tabela 3 - Análise temporal global do par formado pela primeira e segunda massas sintéticas.

Índice de similaridade	Valores
Anderberg	1.0
Jaccard	1.0
Kulczynski 1	1.0
Kulczynski 2	+ ∞
Ochiai	1.0
Czekanowsky	1.0



Na Tabela 4, os índices foram aplicados dessa vez nas oito camadas obtidas de cada nódulo. As camadas 1, 2, 3, 6, 7 e 8 apresentaram valores 0 (mínimo) pra cada índice e as camadas 5 e 6 valores diferentes de 0. Isso acontece por que a distribuição dos valores desses nódulos é invertida (o primeiro tem distribuição crescente a medida que afasta do centro e o segundo crescente a medida que se aproxima do centro). Dessa forma, os valores de *voxels* iguais nos dois nódulos só aparecem nas camadas intermediárias (camadas 4 e 5), principalmente na camada mais central, que é a camada 4.

Tabela 4 - Análise temporal por camadas do par formado pela primeira e segunda massas sintéticas

Índice de similaridade	Valores por camada							
	1	2	3	4	5	6	7	8
Anderberg	0.0	0.0	0.0	0.0909091	0.428571	0.0	0.0	0.0
Jaccard	0.0	0.0	0.0	0.166667	0.6	0.0	0.0	0.0
Kulczynski 1	0.0	0.0	0.0	0.2	1.5	0.0	0.0	0.0
Kulczynski 2	0.0	0.0	0.0	0.285714	0.75	0.0	0.0	0.0
Ochiai	0.0	0.0	0.0	0.285714	0.75	0.0	0.0	0.0
Czekanowsky	0.0	0.0	0.0	0.285714	0.75	0.0	0.0	0.0

A terceira abordagem encontrou duas regiões de alta diferença. Essas regiões são formadas pelas áreas mais externas e internas dos nódulos, respectivamente. Como explicado anteriormente, isso acontece pois os *voxels* só tem valores iguais nas áreas mais intermediárias dos nódulos. Os resultados da análise de textura podem ser vistos na Tabela 5.

Tabela 5 - Análise temporal das regiões de maiores diferenças do par formado pela primeira e segunda massas sintéticas.

Índice de similaridade		
Anderberg	0.0	0.0
Jaccard	0.0	0.0
Kulczynski 1	0.0	0.0
Kulczynski 2	0.0	0.0
Ochiai	0.0	0.0
Czekanowsky	0.0	0.0

Finalmente, a análise temporal é aplicada utilizando a primeira e terceira massas sintéticas (Figura 12c). Os resultados da análise global são mostrados na Tabela 6. As medidas de similaridades globais nesse par de nódulo são consequência da presença e ausência de certos valores de *voxels*. Como os valores de *voxels* do primeiro nódulo variam entre 0 HU e 250 HU e os do terceiro variam entre 100 HU e 350 HU, então, por interseção de conjuntos, sabe-se que os valores entre 0 HU e 99 HU e entre 251 HU a 350 HU não serão comuns aos dois nódulos.

Tabela 6 - Análise temporal global do par formado pela primeira e terceira massas sintéticas.

Índice de similaridade	Valores
Anderberg	0.285714
Jaccard	0.444444
Kulczynski 1	0.8
Kulczynski 2	0.615385
Ochiai	0.615385
Czekanowsky	0.615385

Em seguida, é analisada textura das camadas. Os resultados das análises apresentaram valor zero para todas as camadas. Isso acontece porque um *voxel* em uma

certa posição no primeiro nódulo sempre terá 100 HU a menos que outro *voxel* na mesma posição do terceiro nódulo. Como a camada tem espessura de apenas dois *voxels*, é impossível haver valores de *voxels* em camadas de mesma posição nesses nódulos.

Por fim, os resultados das regiões de maiores diferenças. Como todos os *voxels* de mesma localização no primeiro e terceiro nódulos tem 100 HU de diferença, então haverá uma única região de alta diferença que abrangerá todo o nódulo. Nesse caso, a análise temporal dessa região apresentará valores semelhantes aos vistos na análise global. Esses resultados podem ser vistos na Tabela 7.

Tabela 7 - Análise temporal das regiões de maiores diferenças do par formado pela primeira e terceira massas sintéticas.

Índice de similaridade	Valores
Anderberg	0.272727
Jaccard	0.6
Kulczynski 1	0.428571
Kulczynski 2	0.75
Ochiai	0.6
Czekanowsky	0.6

4.1.2. Massas da *Public Lung Database to Address Drug Response*

O primeira massa a ser utilizada é do paciente identificado pela chave SS0024. Seus dois exames de TC possuem intervalo de tempo de 125 dias. No primeiro exame, a massa tem volume 10939.55 mm³ e no segundo exame tem volume 12831.17 mm³, resultando em um tempo de dobra de 542.72 dias. A Figura 39 e 40 mostram as fatias dos subvolumes mais antigo e recente, respectivamente, obtidos de seus dois exames de TC. O quadrado vermelho nas imagens mostra a localização das massas

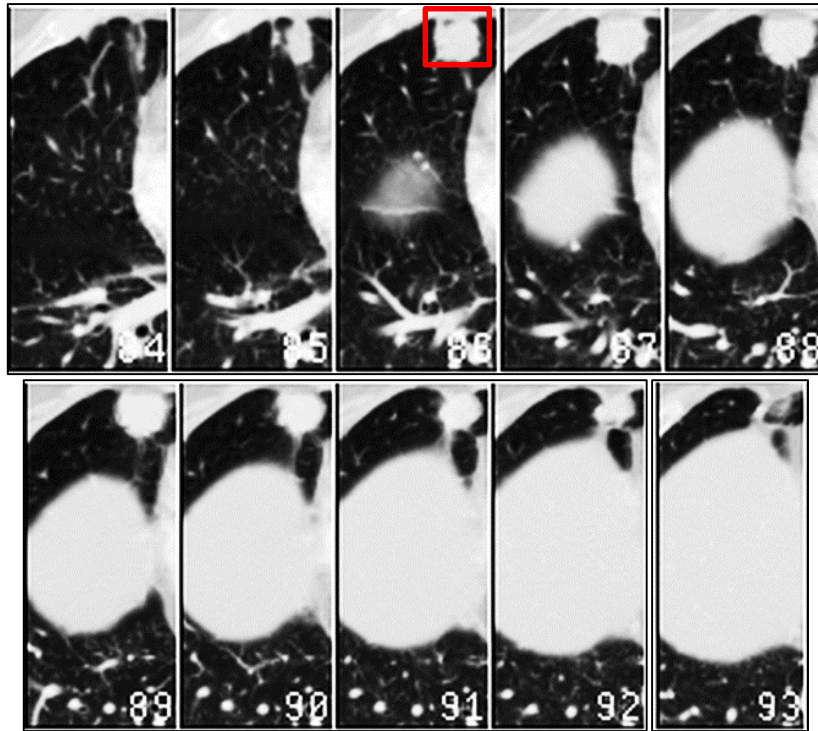


Figura 39 - Fatias do subvolume da primeira massa no exame mais antigo.

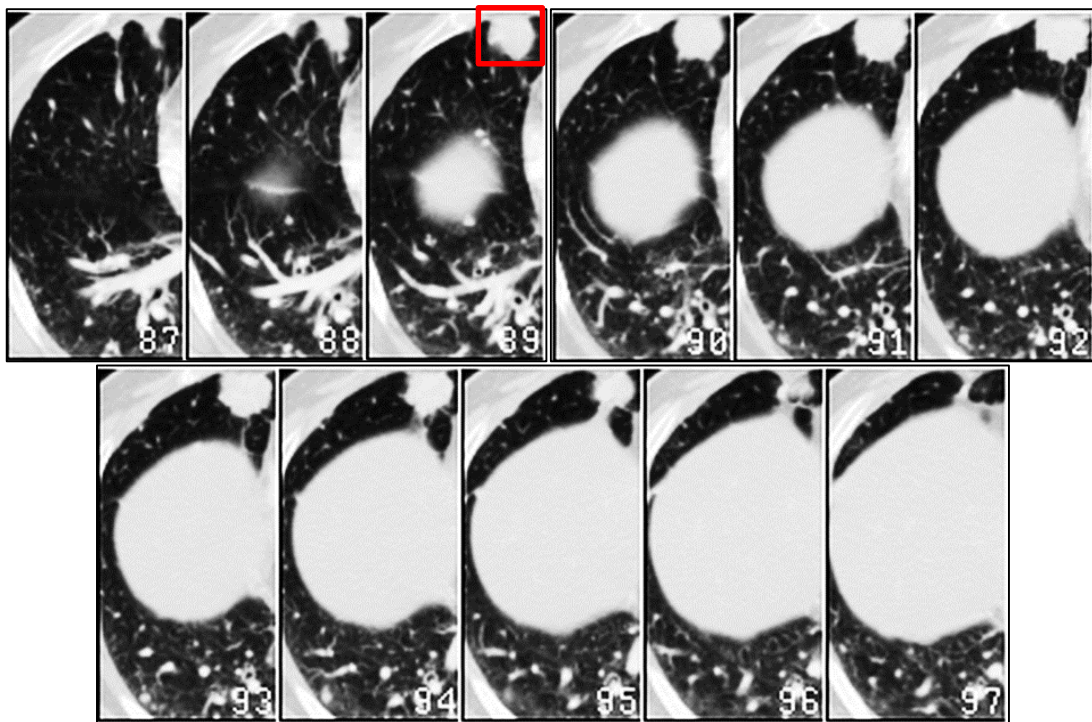


Figura 40 - Fatias do subvolume da primeira massa no exame mais recente.


Primeiramente, foi feita a análise de texturas global dessas massas. Os resultados obtidos pelos índices de similaridade encontra-se na Tabela 8.

Tabela 8 - Análise temporal global da primeira massa da PLDADR.

Índice de similaridade	Valores
Anderberg	0.739007
Jaccard	0.849918
Kulczynski 1	5.66304
Kulczynski 2	0.919432
Ochiai	0.919151
Czekanowsky	0.918871



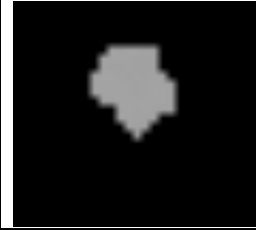



Em seguida a análise temporal foi feita localmente, utilizando a abordagem das regiões de maiores diferenças e em seguida na abordagem de camadas. Os resultados podem ser vistos nas Tabelas 9 e 10.

Tabela 9 - Análise temporal das regiões de maiores diferenças da primeira massa da PLDADR.

Índice de similaridade	
Anderberg	0.0
Jaccard	0.0
Kulczynski 1	0.0
Kulczynski 2	0.0
Ochiai	0.0
Czekanowsky	0.0

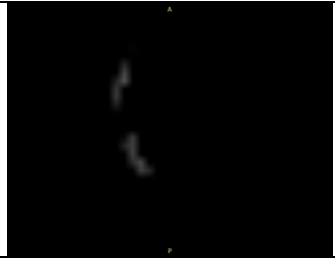
Na Tabela 9, as imagens encontradas na primeira linha são as regiões desconexas de maiores diferenças entre os dois nódulos. Na Tabela 10, as imagens encontradas nas duas primeiras linhas são as camadas da massa no exame mais antigo e mais recente, respectivamente. Essas camadas estão ordenadas da esquerda para a direita, de forma que as camadas mais à esquerda são as mais externas da massa.

Tabela 10 - Análise temporal das camadas da primeira massa da PLDADR.

Índice de similaridade			
			
Anderberg	0.749271	0.4	0.486726
Jaccard	0.856667	0.571429	0.654762
Kulczynski 1	5.97674	1.33333	1.89655
Kulczynski 2	0.923161	0.772268	0.793379
Ochiai	0.922981	0.749433	0.792372
Czekanowsky	0.922801	0.727273	0.791367

Por fim, a análise temporal foi aplicada entre toda a região da massa no primeiro exame e a região nova que se desenvolveu no intervalo entre os exames. Os resultados podem ser vistos na Tabela 11.

Tabela 11 - Análise temporal das regiões iniciais e novas da primeira massa da PLDADR.

Índice de similaridade	
Anderberg	0.384434
Jaccard	0.555366
Kulczynski 1	1.24904
Kulczynski 2	0.771515
Ochiai	0.742268
Czekanowsky	0.714129

A segunda massa é do paciente identificado pela chave SS0010. Ela possui intervalo de tempo entre os exames de 203 dias. No primeiro exame, a massa tem volume 29843.87 mm^3 e no segundo exame tem volume 36552.93 mm^3 , resultando em um tempo de dobra de 693.96 dias. A Figura 41 e 42 mostra os subvolumes da massa no exame mais antigo e recente, respectivamente.

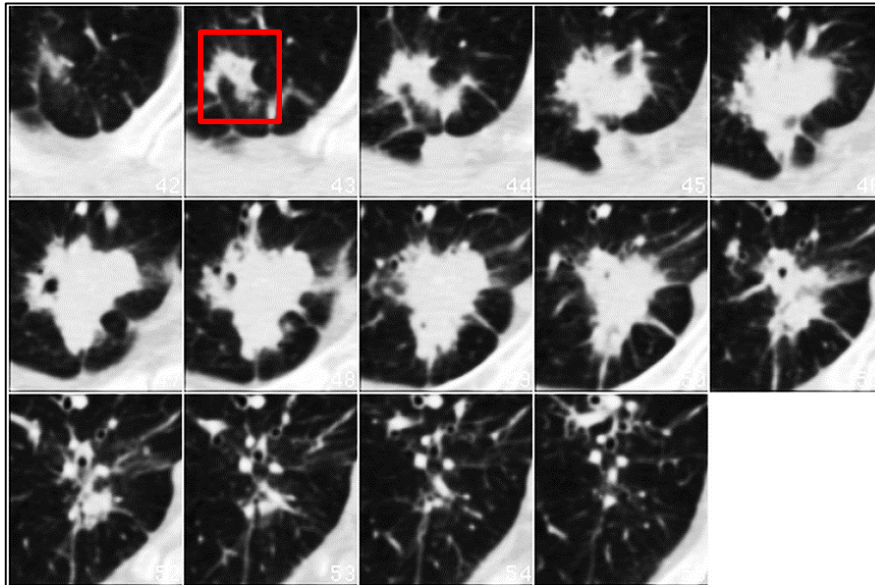


Figura 41 - Fatias do subvolume da segunda massa no exame mais antigo.

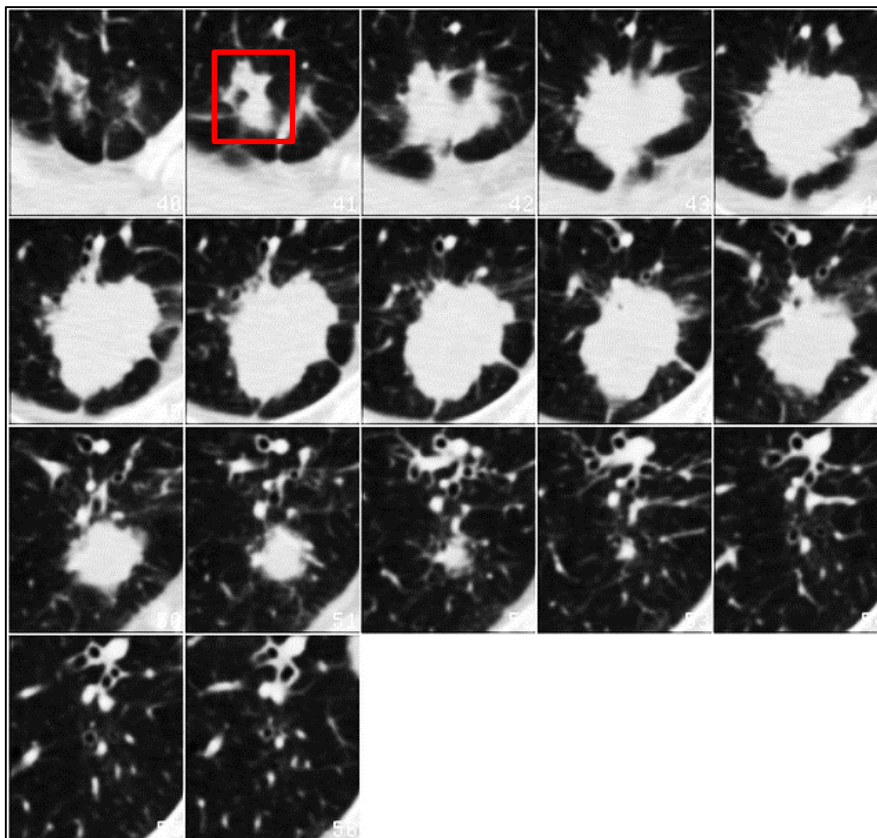


Figura 42 - Fatias do subvolume da segunda massa no exame mais recente.

As tabelas mostram os resultados da análise temporal de forma semelhante a primeira massa. A Tabela 12 é da análise temporal de textura global da massa. A Tabela 13 é das regiões de maiores diferenças, a Tabela 14 é das camadas e a Tabela 15 é da análise entre regiões iniciais e novas. Na tabela 13 são mostradas apenas as quatro maiores regiões desconexas de alta diferença por simplificação. No total essa massa tem cinco regiões desconexas de alta diferenças.

Tabela 12 - Análise temporal global da primeira massa da PLDADR.

Índice de similaridade	Valores
Anderberg	0.808181
Jaccard	0.893916
Kulczynski 1	8.42647
Kulczynski 2	0.944564
Ochiai	0.944275
Czekanowsky	0.943987

Tabela 13 - Análise temporal das regiões de maiores diferenças da segunda massa da PLDADR.



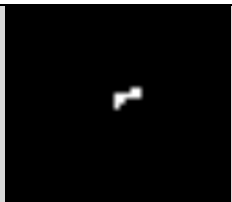

Índice de similaridade	Valores por região desconexa			
				
Anderberg	0.00787402	0.0	0.0	0.0
Jaccard	0.015625	0.0	0.0	0.0
Kulczynski 1	0.015873	0.0	0.0	0.0
Kulczynski 2	0.0590909	0.0	0.0	0.0
Ochiai	0.0426401	0.0	0.0	0.0
Czekanowsky	0.0307692	0.0	0.0	0.0

Tabela 14 - Análise temporal das camadas da segunda massa da PLDADR.



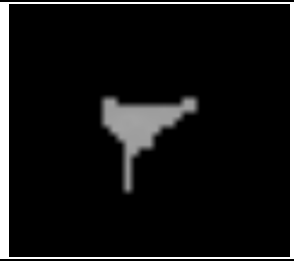




Índice de similaridade			
			
Anderberg	0.794944	0.474903	0.568345
Jaccard	0.885759	0.643979	0.724771
Kulczynski 1	7.75342	1.80882	2.63333
Kulczynski 2	0.940212	0.786631	0.852094
Ochiai	0.939816	0.785034	0.84624
Czekanowsky	0.939419	0.78344	0.840426

Tabela 15 - Análise temporal das regiões iniciais e novas da segunda massa da PLDADR.

Índice de similaridade	
Anderberg	0.673582
Jaccard	0.804959
Kulczynski 1	4.12712
Kulczynski 2	0.898318
Ochiai	0.895124
Czekanowsky	0.891941

Como pode ser visto nas fatias dos subvolumes (Figuras 39, 40, 41 e 42), há pouca mudança na textura da primeira e da segunda massa. Esses comportamento é refletido nos resultados dos índices de similaridade tanto na análise global das texturas quanto na

análise das camadas. A análise das regiões desconexas mostrou baixa similaridade, como esperado, o que mostra que a abordagem das regiões de maiores diferenças encontra as regiões de menores similaridade. As regiões encontradas se concentram nas áreas mais externas das massas. A análise entre regiões iniciais e novas mostra que as regiões novas que se desenvolveram no intervalo de tempo entre os exames apresentam alguma mudança na textura em relação a análise global. Isso demonstra que as mudanças de textura tendem a se concentrar mais em tecidos novos.

4.1.3. Nódulos do *Volcano Challenge*

Foram selecionados dois nódulos na base do *Volcano Challenge*. Os procedimentos aplicados foram os mesmos das massas da PLDADR. Os resultados serão mostrados também como os anteriores.

O primeiro nódulo escolhido foi do paciente identificado pela chave SC0002 e foi acompanhado por dois exames de intervalo de tempo de 100 dias entre os dois exames. Trata-se de um nódulo pequeno de formato circular. No primeiro exame, a massa tem volume 1816.16 mm^3 e no segundo exame tem volume 1638.03 mm^3 , resultando em um tempo de dobra de -671.48 dias. A Figura 43 e 44 mostram os subvolumes obtidos dos exames mais antigo e recente, respectivamente, desse nódulo.

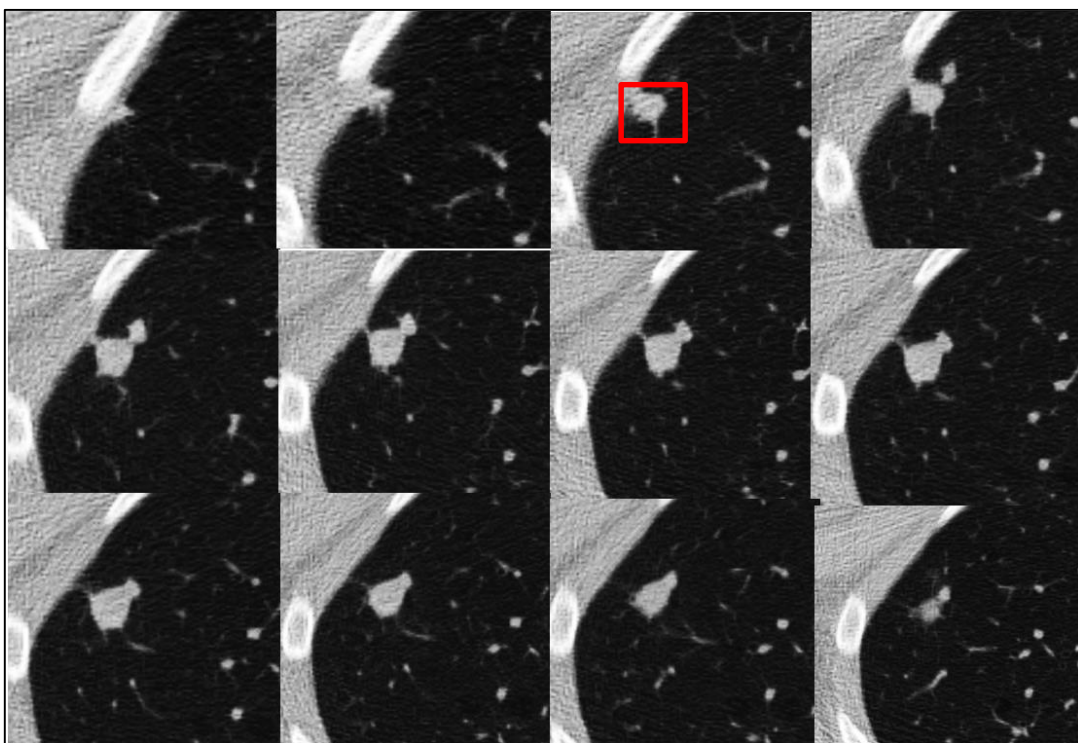


Figura 43 - Fatias do subvolume do primeiro nódulo no exame mais antigo.

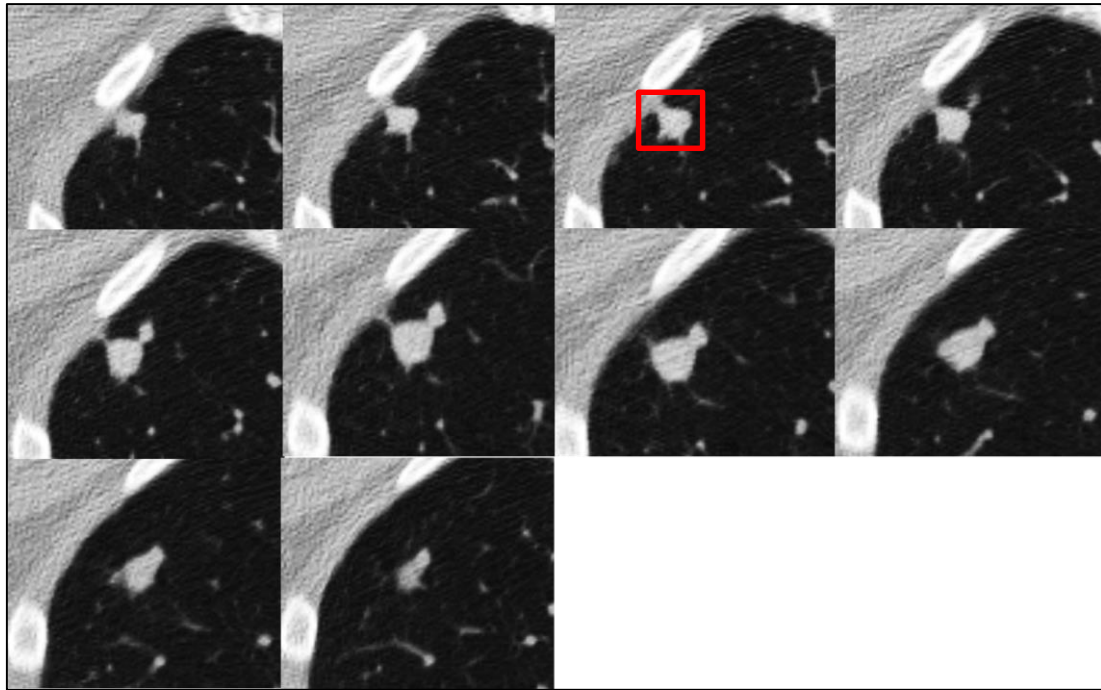


Figura 44 - Fatias do subvolume do primeiro nódulo no exame mais recente.

As Tabelas 16 e 17 mostram os resultados da análise temporal aplicada a todas os valores de *voxels* do nódulo e por camadas, respectivamente. A Tabela 18 mostra o resultado para as regiões desconexas de alta diferença. Como não houve crescimento do nódulo, então não há tabela para a análise entre regiões iniciais e novas. De fato, houve uma redução no volume do nódulo.

Tabela 16 - Análise temporal global da primeiro nódulo da base do *Volcano Challenge*.

Índice de similaridade	Valores
Anderberg	0.70524
Jaccard	0.905396
Kulczynski 1	0.827145
Kulczynski 2	4.78519
Ochiai	0.905432
Czekanowsky	0.905414

Tabela 17 - Análise temporal das camadas da primeiro nódulo da base do *Volcano Challenge*.








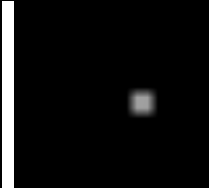
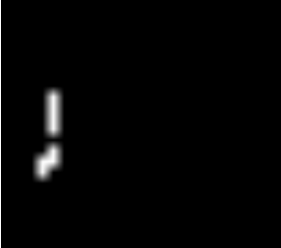
Índice de similaridade				
				
Anderberg	0.716667	0.50499	0.297405	0.0526316
Jaccard	0.834951	0.671088	0.458462	0.1
Kulczynski 1	5.05882	2.04032	0.846591	0.111111
Kulczynski 2	0.911866	0.803693	0.630589	0.385965
Ochiai	0.910959	0.803434	0.62964	0.264906
Czekanowsky	0.910053	0.803175	0.628692	0.181818

Tabela 18 - Análise temporal das camadas da primeiro nódulo da base do *Volcano Challenge*.

Índice de similaridade	Valores por região desconexa
	
Anderberg	0.027027
Jaccard	0.0526316
Kulczynski 1	0.0555556
Kulczynski 2	0.104167
Ochiai	0.102062
Czekanowsky	0.1

O segundo nódulo escolhido foi do paciente identificado pela chave SC0011 acompanhado por dois exames de intervalo de tempo de 100 dias entre os dois exames. Trata-se de um nódulo grande de formato circular. No primeiro exame, a massa tem

volume 1675.21 mm^3 e no segundo exame tem volume 2387.58 mm^3 , resultando em um tempo de dobra de 195.614 dias. A Figura 45 e 46 mostram os subvolumes obtidos dos exames mais antigo e recente, respectivamente, desse nódulo.

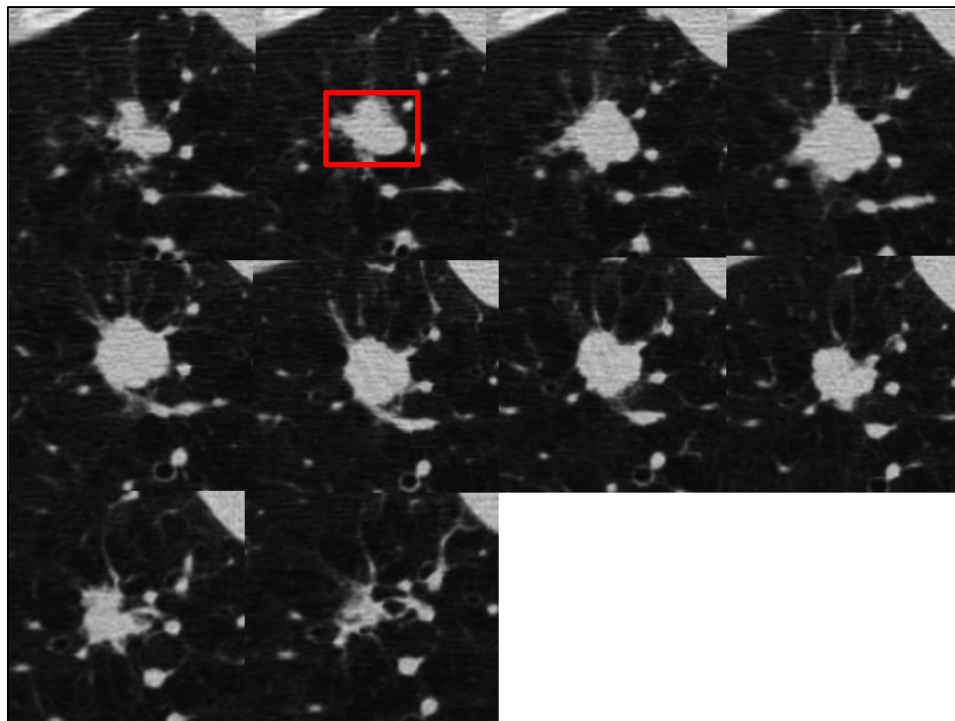


Figura 45 - Fatias do subvolume da massa no exame mais antigo.

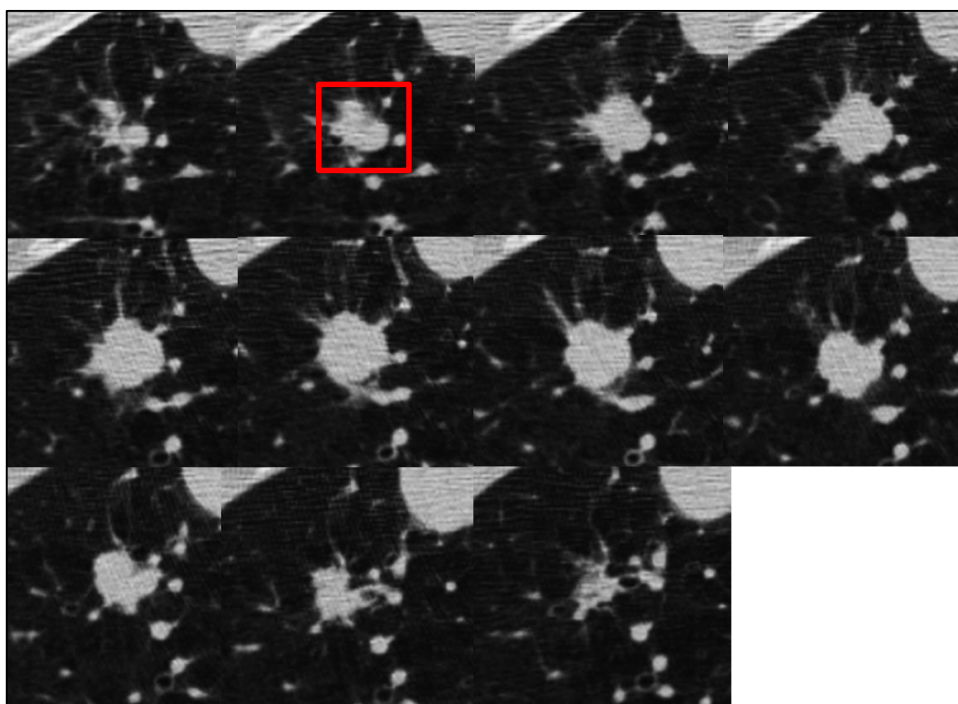


Figura 46 - Fatias do subvolume da massa no exame mais recente.

As Tabelas 19 e 20 mostram os resultados da análise temporal aplicada a todos os valores de *voxels* do nódulo e por camadas, respectivamente. Não houve regiões desconexas de alta diferença com tamanho relevante nesse nódulo. A tabela 21 mostra a análise entre regiões iniciais e novas do nódulo.

Tabela 19 - Análise temporal global da segundo nódulo da base do *Volcano Challenge*.

Índice de similaridade	Valores
Anderberg	0.79726
Jaccard	0.887195
Kulczynski 1	7.86486
Kulczynski 2	0.942746
Ochiai	0.941485
Czekanowsky	0.940226

Tabela 20 - Análise temporal das camadas da segundo nódulo da base do *Volcano Challenge*.




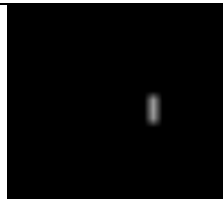





Índice de similaridade				
				
Anderberg	0.797183	0.616766	0.556391	0.0201342
Jaccard	0.887147	0.762963	0.714976	0.0394737
Kulczynski 1	7.86111	3.21875	2.50847	0.0410959
Kulczynski 2	0.942699	0.871702	0.852811	0.235119
Ochiai	0.941448	0.868619	0.843253	0.133631
Czekanowsky	0.940199	0.865546	0.833803	0.0759494

Tabela 21 - Análise temporal das regiões iniciais e novas do segundo nódulo da base do *Volcano Challenge*.

Índice de similaridade	
Anderberg	0.506953
Jaccard	0.672819
Kulczynski 1	2.05641
Kulczynski 2	0.830592
Ochiai	0.817398
Czekanowsky	0.804413

As texturas do primeiro e segundo nódulos sofreu pouca mudança na textura, como pode ser visto nas fatias dos subvolumes (Figuras 43, 44, 45 e 46). Nesses casos, os índices de similaridade também refletiram esse comportamento, tanto na análise global das texturas quanto na análise das camadas. A análise das regiões desconexas também mostrou baixa similaridade nas regiões encontradas. Essas regiões se encontram mais externamente nos nódulos. No caso do segundo nódulo, a região que se desenvolveu no intervalo entre os exames tem textura levemente diferente a região inicial, se comparado a análise global.

5. CONCLUSÃO

Esta dissertação apresentou uma metodologia de análise temporal de nódulos de pulmão. Essa metodologia mostra-se importante uma vez que possibilita fazer uma correspondência entre nódulos em exames diferentes e fornecer ao usuário informações quantitativas sobre as mudanças de textura utilizando índices de similaridade. Essas informações podem ser obtidas do nódulo como um todo, através de suas camadas ou localmente. Com essas medições, será possível um melhor acompanhamento dos nódulos, a fim de determinar o seu comportamento no decorrer do tempo. Dessa forma, pressupõe-se que o médico terá melhor conhecimento sobre as mudanças de textura (tecido) do nódulo e possa usá-las para assistir diagnóstico ou acompanhamento terapêutico.

5.1. Avaliação do trabalho

A metodologia possui bons resultados na determinação de correspondência do nódulo em exames diferentes, obtendo acerto entre grandes massas até pequenos nódulos. Os resultados mostraram também que o pré-processamento do volume de imagens do pulmão possibilitou melhor eficiência no uso do registro, através da exclusão de elementos desnecessários e conseqüentemente possibilitou os resultados obtidos da determinação de correspondências.

O pré-processamento do nódulo possibilitou uma melhor marcação do nódulo, corrigindo erros que são comuns em marcações manuais devido a pressa ou desatenção de especialistas. A correção da marcação possibilitou também melhor precisão na obtenção dos valores dos *voxels* e conseqüentemente uma melhor aplicação da análise temporal.

A utilização dos índices de similaridade mostrou bons resultados para determinar mudanças de textura, tanto gerais quanto locais.

5.2. Trabalhos futuros

Apesar dos resultados obtidos, ainda há um bom número de mudanças de podem ser feitas para maior eficiência da metodologia. A seguir serão apresentados os problemas e sugestões de como eles podem ser corrigidos em trabalhos futuros:

1. As bases de imagens utilizadas contem nódulos / massas que apresentaram poucas mudanças de textura no decorrer do tempo. Isso acontece principalmente porque essas lesões estão em tratamento. Futuramente busca-se validar a metodologia em outras bases públicas. Há também disponível uma outra base do Hospital Pedro Ernesto, Rio de Janeiro – RJ.
2. Os registros de imagens utilizados na Seção 3.3 e na Subseção 3.5.2 demandam muito tempo, principalmente o deformável *B-spline*. Existem outros registros disponíveis na literatura que podem ser pesquisados e utilizados para reduzir o tempo.
3. Os registros de imagens utilizados na Subseção 3.5.2 demandam muito tempo, principalmente o deformável *B-spline*, quando utilizando grandes massas. A solução possível é a mesma do problema 3.
4. Os índices de similaridade utilizados consideram a existência ou ausência de cada valor de *voxels*, mas não levam em consideração a quantidade de *voxels* de cada valor. Levar em consideração essa quantidade podem melhorar os resultados. Existem outros índices de similaridade com essa característica que podem ser utilizados.

6. REFERÊNCIAS

BRADSKI, G.; KAEHLER, A. *Learning OpenCV*. 2008. 1ª Edição. O'Reilly. ISBN: 978-0-596-51613-0.

BROWN, L. G. *A survey of image registration techniques*. 1992. ACM Computing Surveys (CSUR) archive. 24 (4). 325 – 376.

CRUM, W. R.; HARTKENS T.; HILL, D. L. G. *Non-rigid image registration: theory and practice*. 2004. The British Journal of Radiology. 77. 140 – 153.

DINIZ, P. H. B.; SOUSA, H. P.; CAMPOS, T. L.; MOTTA, S. A. C. S.; SILVA, A. C.; PAIVA, A. C. *SAIM: Software para Auxiliar Médicos na Visualização, Detecção e Diagnóstico de Nódulo Pulmonar*. 2011. SIBGRAPI 2011. Maceió.

GOSHTASBY, A. A. *2-D and 3-D image registration for medical, remote sensing, and industrial applications*. 2007. Morgan Kaufmann.

GU, S.; WILSON, D.; TAN, J; PU, J. *Pulmonary nodule registration: Rigid or nonrigid?* 2011. Med Phys. Volume 38(7). Páginas 4406-4414.

HAJNAL, J. V.; HILL, D. L. G.; HAWKES, D. J. *Medical Image Registration*. 2001. 1ª Edição. CRC Press.

HUHTA, V. *Evaluation of different similarity indices as measures of succession in arthropod communities of the forest after clearcutting*. 1979. Oecologia (Berl). Volume 41. Páginas 11-23.

IBÁÑEZ, L; SCHROEDER, W; NG, L; CATES, J. *The ITK Software Guide Second Edition Updated for ITK version 2.4*. 2005.

Instituto Nacional do Câncer – INCA. *O que é câncer?* Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/cancer/site/oquee>. Acessado em: 22/10/2013.

KO, P. J.; BETKE, M. *Chest CT: Automated Nodule Detection and Assessment of Change over Time - Preliminary Experience 1*. 2001; 218:267–273.

LEHMANN, T. M.; GONNER, C.; SPITZER, K. *Survey: interpolation methods in medical image processing*. 1999. IEEE Transactions on Medical Imaging. Volume 18. Número 11. Páginas 1049-1075.

MEYER, A. S. *Comparação de Coeficientes de Similaridade usados em Análises de Agrupamento com Dados de Marcadores Moleculares Dominantes*. 2002. Dissertação. Universidade de São Paulo.

National Electrical Manufactures Association – NEMA. *Digital Imaging and Communications in Medicine (DICOM)*. 2009. PS 3-1, 3, 5, 6 e 10. Estados Unidos.

NETTO, S. M. B. *Segmentação de Automática de Nódulos Pulmonares usando Growing Neural Gas e Máquina de Vetores de Suporte*. 2010. Dissertação (Mestrado em Engenharia Elétrica). UFMA.

PARRAGA, A. *Atlas Anatômico da Região da Cabeça e do Pescoço: em Direção à Radioterapia Adaptativa*. 2008. Tese (Doutorado em Engenharia Elétrica). Universidade Federal do Rio Grande do Sul.

PREIM, B.; BARTZ, D. *Visualization in Medicine: Theory, Algorithms and Applications*. 2007. Morgan Kaufmann.

REEVES, A. P.; CHAN, A. B.; YANKELEVITZ, D. F.; HENSCHKE, C. I.; KRESSLER, B.; KOSTIS, W. J. *On measuring the change in size of pulmonary nodules*. 2006. IEEE Trans Med Imaging. 25:435–450.

SANTOS, A. M. *Detecção De Nódulos Pulmonares Pequenos usando Modelo de Mistura Gaussiana e Matriz Hessiana*. 2011. Dissertação (Mestrado em Engenharia Elétrica). UFMA.

SEIXAS, F. L. *Diagnóstico Auxiliado por Computador*. 2005. Relatório Técnico do Laboratório MídiaCom. Departamento de Engenharia de Telecomunicações da UFF. Niterói. RJ.

SLUIMER, I; SCHILHAM, A; PROKOP, M; van GINNEKEN, B. *Computer Analysis of Computed Tomography Scans of the Lung: A Survey*. 2006. IEEE Transactions on Medical Imaging. Volume 25(4). Páginas 385-405.

THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 2009. 4ª edição. Elsevier. ISBN: 978-1-59749-272-0.

WATSON, D. F. *Contouring: A Guide to the Analysis and Display of Spatial Data*. 1992. New York: Pergamon Press.

WOLDA, H. *Similarity Indices, Sample Size and Diversity*. 1981. Oecologia (Berl). Volume 50. Páginas 296-302.

ZHENG, Y; STEINER, K; BAUER, T; YU, J; SHEN, D; KAMBHAMETTU, C. *Lung Nodule Growth Analysis from 3D CT Data with a Coupled Segmentation and Registration Framework*. 2007. IEEE. Páginas 1-8.

APÊNDICE

Esta seção apresenta os resultados obtidos pelos índices de similaridade aplicados a 20 massas da PLDADR (duas para cada paciente) e 20 nódulos da base do Volcano *Challenge* (duas para cada paciente). Os exames escolhidos foram aqueles que apresentaram melhor qualidade. Os nódulos / massas são agrupados pela chave identificadora do paciente, para que seja possível identificá-los em sua base. Para cada paciente há cinco tabelas sendo a primeira para as informações sobre o nódulo / massa, a segunda para a análise global das texturas (Subseção 3.5.1), a terceira para a análise das camadas (Subseção 3.5.2), a quarta para a análise das regiões desconexas de maiores diferenças (Subseção 3.5.3) e a quinta para análise entre regiões iniciais e novas. Em casos onde houve mais de quatro regiões desconexas relevantes, foram apresentadas apenas as quatro primeiras encontradas. Caso não haja regiões desconexas relevantes, não haverá uma tabela para essa abordagem. Caso o nódulo ou massa não tenha aumentado de volume no decorrer do tempo, também não haverá tabela para a última abordagem. O valores obtidos pelos índices de similaridade nessas bases é visto a seguir.

1) Massas da PLDADR

a) Chave: SS0001.

Informações	
Intervalo entre exames (em dias)	45
Volume exame mais antigo (em mm ³)	70275.88
Volume exame mais recente (em mm ³)	36335.55
Tempo de dobra (em dias)	-47.27

Índice de similaridade	Valores
Anderberg	0.776824
Jaccard	0.874396
Kulczynski 1	6.96154
Kulczynski 2	0.93353
Ochiai	0.93326
Czekanowsky	0.93299

Índice de similaridade	Valores por camada				
	1	2	3	4	5
Anderberg	0.793103	0.535168	0.42963	0.394636	0.270588

Jaccard	0.884615	0.697211	0.601036	0.565934	0.425926
Kulczynski 1	7.66667	2.30263	1.50649	1.3038	0.741935
Kulczynski 2	0.938989	0.824212	0.79781	0.782967	0.712963
Ochiai	0.938882	0.822903	0.773953	0.752286	0.65263
Czekanowsky	0.938776	0.821596	0.750809	0.722807	0.597403

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.134328	0.0	0.0243902	0.129032
Jaccard	0.236842	0.0	0.047619	0.228571
Kulczynski 1	0.310345	0.0	0.05	0.296296
Kulczynski 2	0.391667	0.0	0.192982	0.382222
Ochiai	0.387298	0.0	0.132453	0.377124
Czekanowsky	0.382979	0.0	0.0909091	0.372093

b) Chave: SS0002

Informações	
Intervalo entre exames (em dias)	91
Volume exame mais antigo (em mm ³)	312863.60
Volume exame mais recente (em mm ³)	730849.74
Tempo de dobra (em dias)	74.46

Índice de similaridade	Valores
Anderberg	0.366224
Jaccard	0.536111
Kulczynski 1	1.15569
Kulczynski 2	0.768056
Ochiai	0.732196
Czekanowsky	0.698011

Índice de similaridade	Valores por camada									
	1	2	3	4	5	6	7	8	9	10
Anderberg	0.373	0.292	0.481	0.635	0.7	0.762	0.709	0.458	0.406	0.259
Jaccard	0.543	0.452	0.650	0.777	0.823	0.865	0.830	0.628	0.578	0.412371
Kulczynski 1	1.190	0.827	1.858	3.485	4.666	6.437	4.888	1.690	1.372	0.701754
Kulczynski 2	0.771	0.720	0.820	0.880	0.903	0.930	0.915	0.814	0.789	0.706186
Ochiai	0.737	0.670	0.803	0.877	0.903	0.929	0.911	0.792	0.760	0.642161
Czekanowsky	0.704	0.623	0.788	0.874	0.903	0.927	0.907	0.771	0.732	0.583942

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0410959	0.0181818	0.170213	0.189024
Jaccard	0.0789474	0.0357143	0.290909	0.317949
Kulczynski 1	0.0857143	0.037037	0.410256	0.466165
Kulczynski 2	0.176724	0.0782726	0.478114	0.500691
Ochiai	0.160817	0.0734718	0.464207	0.491507
Czekanowsky	0.146341	0.0689655	0.450704	0.48249

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.393197
Jaccard	0.564453
Kulczynski 1	1.29596
Kulczynski 2	0.781639
Ochiai	0.751019
Czekanowsky	0.721598

c) Chave: SS0003

Informações	
Intervalo entre exames (em dias)	183
Volume exame mais antigo (em mm ³)	11382.45
Volume exame mais recente (em mm ³)	8071.59
Tempo de dobra (em dias)	-369.44

Índice de similaridade	Valores
Anderberg	0.389982
Jaccard	0.561133
Kulczynski 1	1.27859
Kulczynski 2	0.720823
Ochiai	0.71985
Czekanowsky	0.718879

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.3875	0.51073	0.351852
Jaccard	0.558559	0.676136	0.520548

Kulczynski 1	1.26531	2.08772	1.08571
Kulczynski 2	0.718834	0.806789	0.691475
Ochiai	0.717798	0.806784	0.688072
Czekanowsky	0.716763	0.80678	0.684685

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0410959	0.0	0.0	0.0
Jaccard	0.0789474	0.0	0.0	0.0
Kulczynski 1	0.0857143	0.0	0.0	0.0
Kulczynski 2	0.157692	0.0	0.0	0.0
Ochiai	0.151911	0.0	0.0	0.0
Czekanowsky	0.146341	0.0	0.0	0.0

d) Chave: SS0010

Informações	
Intervalo entre exames (em dias)	203
Volume exame mais antigo (em mm ³)	29843.87
Volume exame mais recente (em mm ³)	36552.93
Tempo de dobra (em dias)	693.96

Índice de similaridade	Valores
Anderberg	0.808181
Jaccard	0.893916
Kulczynski 1	8.42647
Kulczynski 2	0.944564
Ochiai	0.944275
Czekanowsky	0.943987

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.794944	0.474903	0.568345
Jaccard	0.885759	0.643979	0.724771
Kulczynski 1	7.75342	1.80882	2.63333
Kulczynski 2	0.940212	0.786631	0.852094
Ochiai	0.939816	0.785034	0.84624
Czekanowsky	0.939419	0.78344	0.840426

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.00787402	0.0	0.0	0.0
Jaccard	0.015625	0.0	0.0	0.0
Kulczynski 1	0.015873	0.0	0.0	0.0
Kulczynski 2	0.0590909	0.0	0.0	0.0
Ochiai	0.0426401	0.0	0.0	0.0
Czekanowsky	0.0307692	0.0	0.0	0.0

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.384434
Jaccard	0.555366
Kulczynski 1	1.24904
Kulczynski 2	0.771515
Ochiai	0.742268
Czekanowsky	0.714129

e) Chave: SS0012

Informações	
Intervalo entre exames (em dias)	37
Volume exame mais antigo (em mm ³)	845460.47
Volume exame mais recente (em mm ³)	1481180.56
Tempo de dobra (em dias)	45.48

Índice de similaridade	Valores
Anderberg	0.483146
Jaccard	0.651515
Kulczynski 1	1.86957
Kulczynski 2	0.821769
Ochiai	0.805213
Czekanowsky	0.788991

Índice de similaridade	Valores por camada			
	1	2	3	4
Anderberg	0.544304	0.219876	0.178571	0.127273
Jaccard	0.704918	0.360489	0.30303	0.225806
Kulczynski 1	2.38889	0.563694	0.434783	0.291667
Kulczynski 2	0.848996	0.566639	0.647025	0.612903

Ochiai	0.837887	0.547982	0.548582	0.475191
Czekanowsky	0.826923	0.52994	0.465116	0.368421

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0	0.0	0.0	0.0
Jaccard	0.0	0.0	0.0	0.0
Kulczynski 1	0.0	0.0	0.0	0.0
Kulczynski 2	0.0	0.0	0.0	0.0
Ochiai	0.0	0.0	0.0	0.0
Czekanowsky	0.0	0.0	0.0	0.0

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.00235571
Jaccard	0.00470035
Kulczynski 1	0.00472255
Kulczynski 2	0.402353
Ochiai	0.0613572
Czekanowsky	0.00935673

f) Chave: SS0016

Informações	
Intervalo entre exames (em dias)	138
Volume exame mais antigo (em mm ³)	63285.95
Volume exame mais recente (em mm ³)	41643.66
Tempo de dobra (em dias)	-229.02

Índice de similaridade	Valores
Anderberg	0.534466
Jaccard	0.696615
Kulczynski 1	2.29614
Kulczynski 2	0.824691
Ochiai	0.822935
Czekanowsky	0.821182

Índice de similaridade	Valores por camada			
	1	2	3	4
Anderberg	0.534466	0.489796	0.730769	0.468927
Jaccard	0.696615	0.657534	0.844444	0.638462
Kulczynski 1	2.29614	1.92	5.42857	1.76596
Kulczynski 2	0.824691	0.793876	0.921025	0.779497
Ochiai	0.822935	0.793632	0.91834	0.77942
Czekanowsky	0.821182	0.793388	0.915663	0.779343

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.137255	0.0144928	0.025641	0.146154
Jaccard	0.241379	0.0285714	0.05	0.255034
Kulczynski 1	0.318182	0.0294118	0.0526316	0.342342
Kulczynski 2	0.390093	0.0602007	0.147498	0.407361
Ochiai	0.38949	0.0578315	0.118522	0.406889
Czekanowsky	0.388889	0.0555556	0.0952381	0.406417

g) Chave: SS0021

Informações	
Intervalo entre exames (em dias)	141
Volume exame mais antigo (em mm ³)	71264.00
Volume exame mais recente (em mm ³)	111060.73
Tempo de dobra (em dias)	220.21

Índice de similaridade	Valores
Anderberg	0.62766
Jaccard	0.771242
Kulczynski 1	3.37143
Kulczynski 2	0.885621
Ochiai	0.878204
Czekanowsky	0.870849

Índice de similaridade	Valores por camada			
	1	2	3	4
Anderberg	0.62766	0.423341	0.30137	0.319838
Jaccard	0.771242	0.594855	0.463158	0.484663
Kulczynski 1	3.37143	1.46825	0.862745	0.940476

Kulczynski 2	0.885621	0.766929	0.645148	0.720238
Ochiai	0.878204	0.756376	0.639092	0.685739
Czekanowsky	0.870849	0.745968	0.633094	0.652893

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0155039	0.0133333	0.011236	0.0
Jaccard	0.0305344	0.0263158	0.0222222	0.0
Kulczynski 1	0.0314961	0.027027	0.0227273	0.0
Kulczynski 2	0.0878334	0.0633117	0.0721569	0.0
Ochiai	0.0721453	0.0569803	0.0560112	0.0
Czekanowsky	0.0592593	0.0512821	0.0434783	0.0

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.661364
Jaccard	0.79617
Kulczynski 1	3.90604
Kulczynski 2	0.89571
Ochiai	0.891103
Czekanowsky	0.886519

h) Chave: SS0022

Informações	
Intervalo entre exames (em dias)	223
Volume exame mais antigo (em mm ³)	24308.44
Volume exame mais recente (em mm ³)	32214.25
Tempo de dobra (em dias)	548.93

Índice de similaridade	Valores
Anderberg	0.815231
Jaccard	0.898212
Kulczynski 1	8.82432
Kulczynski 2	0.946767
Ochiai	0.946572
Czekanowsky	0.946377

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.815231	0.350427	0.372549
Jaccard	0.898212	0.518987	0.542857
Kulczynski 1	8.82432	1.07895	1.1875
Kulczynski 2	0.946767	0.722679	0.746997
Ochiai	0.946572	0.702731	0.725027
Czekanowsky	0.946377	0.683333	0.703704

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0185185	0.0377358	0.0	0.0117647
Jaccard	0.0363636	0.0727273	0.0	0.0232558
Kulczynski 1	0.0377358	0.0784314	0.0	0.0238095
Kulczynski 2	0.0904762	0.155263	0.0	0.0606061
Ochiai	0.0796819	0.145095	0.0	0.0524864
Czekanowsky	0.0701754	0.135593	0.0	0.0454545

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.673279
Jaccard	0.804742
Kulczynski 1	4.12143
Kulczynski 2	0.898784
Ochiai	0.895289
Czekanowsky	0.891808

i) Chave: SS0023

Informações	
Intervalo entre exames (em dias)	84
Volume exame mais antigo (em mm ³)	13666.62
Volume exame mais recente (em mm ³)	13015.02
Tempo de dobra (em dias)	-1190.99

Índice de similaridade	Valores
Anderberg	0.568224
Jaccard	0.724672
Kulczynski 1	2.63203

Kulczynski 2	0.840572
Ochiai	0.840466
Czekanowsky	0.840359

Índice de similaridade	Valores por camada	
	1	2
Anderberg	0.56676	0.208556
Jaccard	0.72348	0.345133
Kulczynski 1	2.61638	0.527027
Kulczynski 2	0.839789	0.538542
Ochiai	0.839673	0.525697
Czekanowsky	0.839557	0.513158

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0107527	0.0	0.0	0.00671141
Jaccard	0.0212766	0.0	0.0	0.0133333
Kulczynski 1	0.0217391	0.0	0.0	0.0135135
Kulczynski 2	0.042328	0.0	0.0	0.131944
Ochiai	0.0419961	0.0	0.0	0.0589256
Czekanowsky	0.0416667	0.0	0.0	0.0263158

j) Chave: SS0024

Informações	
Intervalo entre exames (em dias)	125
Volume exame mais antigo (em mm ³)	10939.55
Volume exame mais recente (em mm ³)	12831.17
Tempo de dobra (em dias)	542.72

Índice de similaridade	Valores
Anderberg	0.739007
Jaccard	0.849918
Kulczynski 1	5.66304
Kulczynski 2	0.919432
Ochiai	0.919151
Czekanowsky	0.918871

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.749271	0.4	0.486726
Jaccard	0.856667	0.571429	0.654762
Kulczynski 1	5.97674	1.33333	1.89655
Kulczynski 2	0.923161	0.772268	0.793379
Ochiai	0.922981	0.749433	0.792372
Czekanowsky	0.922801	0.727273	0.791367

Índice de similaridade	Valores por região desconexa
	1
Anderberg	0.0
Jaccard	0.0
Kulczynski 1	0.0
Kulczynski 2	0.0
Ochiai	0.0
Czekanowsky	0.0

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.384434
Jaccard	0.555366
Kulczynski 1	1.24904
Kulczynski 2	0.771515
Ochiai	0.742268
Czekanowsky	0.714129

2) Nódulos do *Volcano Challenge*

a) Chave: SC0001

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	6984.48
Volume exame mais recente (em mm ³)	239.286
Tempo de dobra (em dias)	-20.5451

Índice de similaridade	Valores
Anderberg	0.485714

Jaccard	0.653846
Kulczynski 1	1.88889
Kulczynski 2	0.826923
Ochiai	0.808608
Czekanowsky	0.790698

Índice de similaridade	Valores por camada	
	1	2
Anderberg	0.497076	0.205298
Jaccard	0.664063	0.340659
Kulczynski 1	1.97674	0.516667
Kulczynski 2	0.832031	0.67033
Ochiai	0.8149	0.58366
Czekanowsky	0.798122	0.508197

Índice de similaridade	Valores por região desconexa			
	1	2	3	4
Anderberg	0.0	0.0	0.00917431	0.0070922
Jaccard	0.0	0.0	0.0181818	0.0140845
Kulczynski 1	0.0	0.0	0.0185185	0.0142857
Kulczynski 2	0.0	0.0	0.0608696	0.0421053
Ochiai	0.0	0.0	0.0466252	0.0341993
Czekanowsky	0.0	0.0	0.0357143	0.0277778

b) Chave: SC0002

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	1816.16
Volume exame mais recente (em mm ³)	1638.03
Tempo de dobra (em dias)	-671.48

Índice de similaridade	Valores
Anderberg	0.70524
Jaccard	0.905396
Kulczynski 1	0.827145
Kulczynski 2	4.78519
Ochiai	0.905432
Czekanowsky	0.905414

Índice de similaridade	Valores por camada			
	1	2	3	4
Anderberg	0.716667	0.50499	0.297405	0.0526316
Jaccard	0.834951	0.671088	0.458462	0.1
Kulczynski 1	5.05882	2.04032	0.846591	0.111111
Kulczynski 2	0.911866	0.803693	0.630589	0.385965
Ochiai	0.910959	0.803434	0.62964	0.264906
Czekanowsky	0.910053	0.803175	0.628692	0.181818

Índice de similaridade	Valores por região desconexa
	1
Anderberg	0.027027
Jaccard	0.0526316
Kulczynski 1	0.0555556
Kulczynski 2	0.104167
Ochiai	0.102062
Czekanowsky	0.1

c) Chave: SC0007

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	274.235
Volume exame mais recente (em mm ³)	179.614
Tempo de dobra (em dias)	-163.797

Índice de similaridade	Valores
Anderberg	0.308861
Jaccard	0.471954
Kulczynski 1	0.893773
Kulczynski 2	0.654693
Ochiai	0.647942
Czekanowsky	0.641262

Índice de similaridade	Valores por camada	
	1	2
Anderberg	0.308311	0.062069

Jaccard	0.471311	0.116883
Kulczynski 1	0.891473	0.132353
Kulczynski 2	0.65233	0.283516
Ochiai	0.646473	0.243599
Czekanowsky	0.640669	0.209302

d) Chave: SC009

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	5013.42
Volume exame mais recente (em mm ³)	4226.99
Tempo de dobra (em dias)	-406.23

Índice de similaridade	Valores
Anderberg	0.608377
Jaccard	0.75651
Kulczynski 1	3.10695
Kulczynski 2	0.86303
Ochiai	0.862204
Czekanowsky	0.861379

Índice de similaridade	Valores por camada				
	1	2	3	4	5
Anderberg	0.656943	0.429936	0.409949	0.501299	0.101449
Jaccard	0.792958	0.601336	0.581509	0.66782	0.184211
Kulczynski 1	3.82993	1.50838	1.38953	2.01042	0.225806
Kulczynski 2	0.884722	0.799494	0.78938	0.824699	0.359435
Ochiai	0.884623	0.77489	0.761904	0.812677	0.334401
Czekanowsky	0.884525	0.751043	0.735385	0.80083	0.311111

Índice de similaridade	Valores por região desconexa
	1
Anderberg	0.04
Jaccard	0.0769231
Kulczynski 1	0.0833333
Kulczynski 2	0.143182
Ochiai	0.143019
Czekanowsky	0.142857

e) Chave: SC0010

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	2015.27
Volume exame mais recente (em mm ³)	2076.8
Tempo de dobra (em dias)	2304.73

Índice de similaridade	Valores
Anderberg	0.829971
Jaccard	0.907087
Kulczynski 1	9.76271
Kulczynski 2	0.951987
Ochiai	0.951633
Czekanowsky	0.95128

Índice de similaridade	Valores por camada		
	1	2	2
Anderberg	0.778761	0.373083	0.0530973
Jaccard	0.875622	0.543424	0.10084
Kulczynski 1	7.04	1.19022	0.11215
Kulczynski 2	0.935302	0.720639	0.236089
Ochiai	0.934494	0.712362	0.207974
Czekanowsky	0.933687	0.70418	0.183206

Índice de similaridade	Valores por região desconexa	
	1	2
Anderberg	0.0	0.0123457
Jaccard	0.0	0.0243902
Kulczynski 1	0.0	0.025
Kulczynski 2	0.0	0.0535714
Ochiai	0.0	0.0505076
Czekanowsky	0.0	0.047619

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.093779
Jaccard	0.171477

Kulczynski 1	0.206967
Kulczynski 2	0.585739
Ochiai	0.414098
Czekanowsky	0.292754

f) Chave: SC0011

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	1675.21
Volume exame mais recente (em mm ³)	2387.58
Tempo de dobra (em dias)	195.614

Índice de similaridade	Valores
Anderberg	0.79726
Jaccard	0.887195
Kulczynski 1	7.86486
Kulczynski 2	0.942746
Ochiai	0.941485
Czekanowsky	0.940226

Índice de similaridade	Valores por camada			
	1	2	2	4
Anderberg	0.797183	0.616766	0.556391	0.0201342
Jaccard	0.887147	0.762963	0.714976	0.0394737
Kulczynski 1	7.86111	3.21875	2.50847	0.0410959
Kulczynski 2	0.942699	0.871702	0.852811	0.235119
Ochiai	0.941448	0.868619	0.843253	0.133631
Czekanowsky	0.940199	0.865546	0.833803	0.0759494

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.506953
Jaccard	0.672819
Kulczynski 1	2.05641
Kulczynski 2	0.830592
Ochiai	0.817398
Czekanowsky	0.804413

g) Chave: SC00020

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	1449.58
Volume exame mais recente (em mm ³)	1438.14
Tempo de dobra (em dias)	-8745.16

Índice de similaridade	Valores
Anderberg	0.818908
Jaccard	0.900439
Kulczynski 1	9.04412
Kulczynski 2	0.94902
Ochiai	0.948316
Czekanowsky	0.947612

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.828986	0.507614	0.431193
Jaccard	0.906498	0.673401	0.602564
Kulczynski 1	9.69491	2.06186	1.51613
Kulczynski 2	0.951588	0.82679	0.769785
Ochiai	0.951272	0.815736	0.760841
Czekanowsky	0.950956	0.804829	0.752

Índice de similaridade	Valores por região desconexa			
	1	2	2	4
Anderberg	0.047619	0.0410959	0.011236	0.0638298
Jaccard	0.0909091	0.0789474	0.0222222	0.12
Kulczynski 1	0.1	0.0857143	0.0227273	0.136364
Kulczynski 2	0.166667	0.146429	0.0435606	0.216774
Ochiai	0.166667	0.146385	0.0435194	0.215526
Czekanowsky	0.166667	0.146341	0.0434783	0.214286

h) Chave: SC0028

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	943.039

Volume exame mais recente (em mm ³)	471.519
Tempo de dobra (em dias)	-100

Índice de similaridade	Valores
Anderberg	0.76
Jaccard	0.863636
Kulczynski 1	6.33333
Kulczynski 2	0.929314
Ochiai	0.928071
Czekanowsky	0.926829

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.706539	0.622642	0.109375
Jaccard	0.90593	0.868421	0.329412
Kulczynski 1	0.828037	0.767442	0.197183
Kulczynski 2	4.81522	3.3	0.245614
Ochiai	0.909827	0.868421	0.598592
Czekanowsky	0.907876	0.868421	0.444053

i) Chave: SC0037

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	732.422
Volume exame mais recente (em mm ³)	1009.77
Tempo de dobra (em dias)	215.855

Índice de similaridade	Valores
Anderberg	0.74746
Jaccard	0.855482
Kulczynski 1	5.91954
Kulczynski 2	0.924871
Ochiai	0.923491
Czekanowsky	0.922113

Índice de similaridade	Valores por camada		
	1	2	3
Anderberg	0.735119	0.5338	0.201646
Jaccard	0.847341	0.696049	0.335616
Kulczynski 1	5.55056	2.29	0.505155
Kulczynski 2	0.921597	0.822066	0.539465
Ochiai	0.919478	0.821427	0.520688
Czekanowsky	0.917363	0.820789	0.502564

Índice de similaridade	Valores entre regiões iniciais e novas
Anderberg	0.536754
Jaccard	0.698556
Kulczynski 1	2.31737
Kulczynski 2	0.835
Ochiai	0.828741
Czekanowsky	0.822529

j) Chave: SC0040

Informações	
Intervalo entre exames (em dias)	100
Volume exame mais antigo (em mm ³)	720.112
Volume exame mais recente (em mm ³)	557.709
Tempo de dobra (em dias)	-271.216

Índice de similaridade	Valores
Anderberg	0.57673
Jaccard	0.731552
Kulczynski 1	2.72512
Kulczynski 2	0.844971
Ochiai	0.844969
Czekanowsky	0.844967

Índice de similaridade	Valores por camada	
	1	2
Anderberg	0.573276	0.290598
Jaccard	0.728767	0.450331
Kulczynski 1	2.68687	0.819277

Kulczynski 2	0.843464	0.621471
Ochiai	0.843285	0.621238
Czekanowsky	0.843106	0.621005

Índice de similaridade	Valores por região desconexa
	1
Anderberg	0.0285714
Jaccard	0.0555556
Kulczynski 1	0.0588235
Kulczynski 2	0.110145
Ochiai	0.107676
Czekanowsky	0.105263