

FILIPPE VERNETTI GIUSTI

**FERRAMENTA PARA AUXÍLIO AO
DIAGNÓSTICO DE IMAGENS
PULMONARES DE TOMOGRAFIA
COMPUTADORIZADA**

Trabalho acadêmico apresentado ao
Bacharelado em Ciência da Computação
da Universidade Federal de Pelotas, como
requisito parcial para a obtenção do grau de
Bacharel em Ciência da Computação

Orientador: Prof. Dr. Lucas Ferrari
Oliveira

Pelotas, 2008

Banca examinadora:

Prof. Dr. Gerson Geraldo Homrich Cavalcheiro

Prof. Dr. Lucas Ferrari Oliveira

Prof. Ms. Ricardo Matsumura de Araújo

*“With great power,
comes great responsibility.”*

— TIO BEN EM HOMEM-ARANHA

AGRADECIMENTOS

Antes de qualquer coisa, agradeço aos meus pais e avós pelo apoio, as vezes até exagerado.

Agradeço ao Lucas, meu orientador, pelo apoio e por ter acreditado no meu trabalho.

Agradeço aos colegas pela imensa troca de experiência e pelos grupos de estudo um dia antes da prova. Amizades que certamente não acabarão junto com o curso.

Por fim, agradeço aos professores que se dedicaram e se empenharam em melhorar o curso de ciência da computação da UFPel, tanto buscando melhorias na organização e infra-estrutura, como aqueles que lecionam com paixão.

DEDICATORIA

Dedico esse trabalho ao meu pai e tutor, Ronaldo, e a minha mãe e protetora, Silvana. Os quais me apoiaram totalmente.

RESUMO

Giusti, Filipe Vernetti. **FERRAMENTA PARA AUXÍLIO AO DIAGNÓSTICO DE IMAGENS PULMONARES DE TOMOGRAFIA COMPUTADORIZADA**. 2008. 50 f. Monografia (Bacharelado em Ciência da Computação). Universidade Federal de Pelotas, Pelotas.

Este trabalho apresenta o resultado do desenvolvimento de uma ferramenta que utiliza técnicas de processamento de imagens e redes neurais artificiais com objetivo de facilitar o diagnóstico de exames de tomografia computadorizada pulmonar. A ferramenta faz isso retornando ao usuário imagens de outros exames de tomografia computadorizada que sejam similares ao exame em que deseja-se realizar o diagnóstico. Sistemas assim são chamados de sistemas de recuperação de imagens baseado em conteúdo. Mas para que possa retornar resultados relevantes, a ferramenta precisa da resposta do usuário sobre quais imagens são realmente similares. Para que possa fazer isso ele utiliza-se de redes neurais artificiais com topologia de pró-alimentação e do algoritmo de retro-propagação para o treinamento, dessa forma o sistema é capaz de aprender a similaridade entre duas imagens. A ferramenta possui uma interface web a qual pode ser instalada em um servidor e acessada de qualquer terminal, facilitando assim a utilização.

Palavras-chave: Processamento de imagens, tomografia computadorizada, redes neurais artificiais, recuperação de imagens baseada em conteúdo.

ABSTRACT

Giusti, Filipe Verneti. **Application for computer aided diagnosis of computed tomography lung images**. 2008. 50 f. Final Dissertation. Universidade Federal de Pelotas, Pelotas.

This paper presents the result of the development of a program that uses techniques of image processing and artificial neural networks to facilitate the diagnosis of lung computed tomography examinations. The system retrieves to the user images from others computed tomography scans that are similar to the exam that you want to do the diagnosis. Systems like this are called content based image retrieval. But to return relevant results, the system needs the user's feedback on which images are similar. In order to do this the program is based on artificial neural networks with feedforward topology and back-propagation algorithm for the training, so the system is able to learn the similarity between two images. The system has a web interface which can be installed on a server and accessed from any terminal, thus facilitating the use.

Keywords: image processing, computed tomography, artificial neural network, content based image retrieval.

LISTA DE FIGURAS

Figura 1	Demonstração de como uma máquina de TC obtém as fatias de imagem do paciente.	18
Figura 2	Forma de movimentação de máquinas de TC de tecnologia antiga.	19
Figura 3	Exemplo de janela.	21
Figura 4	Imagem de TC original e após a aplicação de threshold adaptativo.	23
Figura 5	Exemplos de elementos estruturantes.	25
Figura 6	Imagem de TC após passar por threshold adaptativo e limpeza, submetida a dilatação e erosão utilizando como elemento estruturante um círculo de raio 3 pixels.	26
Figura 7	Imagem de TC após passar por threshold adaptativo, submetida a abertura e fechamento utilizando como elemento estruturante um círculo de raio 6 pixels.	27
Figura 8	Diagrama de um sistema CBIR típico.	32
Figura 9	Uma RNA simples.	33
Figura 10	Funções de ativação comuns em RNAs.	34
Figura 11	Perceptron de múltiplas camadas pró-alimentado simples. . .	35
Figura 12	GUI da ferramenta desenvolvida.	40
Figura 13	GUI da ferramenta desenvolvida com o modo de avaliação ativado.	41
Figura 14	Imagem de TC antes e depois da remoção das partes que não pertencem ao corpo.	43
Figura 15	Imagem de TC antes e depois da eliminação das regiões que não fazem parte do pulmão.	44
Figura 16	Imagem de TC antes e depois da operação morfológica de fechamento utilizando um círculo de raio 6 pixels.	44

Figura 17 Pulmões unidos em 17(a), e depois devidamente separados e
identificados em 17(b) e 17(c). 45

LISTA DE TABELAS

Tabela 1	Valores Hounsfield comuns.	20
Tabela 2	Valores comuns de janela.	21

LISTA DE ABREVIATURAS E SIGLAS

2D	Bidimensional
3D	Tridimensional
ACR	American College of Radiology
CBIR	Recuperação de imagens baseada em conteúdo (content based image retrieval)
DICOM	Digital Imaging and Communications in Medicine
FANN	Biblioteca de redes neurais artificiais rápida (fast artificial neural network library)
IDE	Ambiente de desenvolvimento integrado (integrated development environment)
ITK	Conjunto de ferramentas para segmentação e registro de imagens (insight segmentation and registration toolkit)
NEMA	National Electrical Manufacturers Association
RNA	Rede neural artificial
TC	Tomografia computadorizada
UH	Unidade Hounsfield
VTK	Conjunto de ferramentas para visualização (Visualization Toolkit)

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação	15
1.2	Objetivos	15
1.3	Organização	15
2	PROCESSAMENTO DE IMAGENS	17
2.1	Imagens de tomografia computadorizada	17
2.1.1	Coeficiente de Hounsfield	19
2.1.2	Técnica de Janelas	20
2.1.3	Padrão DICOM	20
2.2	Segmentação	22
2.2.1	Threshold Adaptativo	22
2.3	Morfologia Matemática	23
2.3.1	Erosão e Dilatação	24
2.3.2	Abertura e Fechamento	25
2.4	Extração de características de texturas	26
2.4.1	Matriz de co-ocorrência	28
2.4.2	Características de Haralick	28
3	RECUPERAÇÃO DE IMAGENS BASEADA EM CONTEÚDO	31
3.1	Similaridade	32
3.2	Redes neurais artificiais	33
3.2.1	Perceptron de múltiplas camadas pró-alimentado	34
3.2.2	Algoritmo de retro-propagação	35
3.3	Eficiência de um sistema de CBIR	36
4	FERRAMENTA DESENVOLVIDA	37
4.1	Ambiente de desenvolvimento	37
4.2	Imagens utilizadas	37
4.3	Linguagens de programação	37

4.4	Bibliotecas	38
4.4.1	ITK	38
4.4.2	FANN	39
4.5	Interface Gráfica	39
5	RESULTADOS	42
5.1	Algoritmos	42
5.1.1	Segmentação de imagens de tomografia computadorizada dos pulmões	42
5.1.2	Recuperação de imagens baseada em conteúdo	45
6	CONCLUSÕES	47
6.1	Dificuldades encontradas	47
6.2	Trabalhos Futuros	47
	REFERÊNCIAS	49

1 INTRODUÇÃO

O exame de tomografia computadorizada - TC - resulta em diversas imagens que representam secções do corpo. Elas são geradas através de uma sucessão de raios-x que são posteriormente processados por computador. Essas imagens podem ser reunidas de forma a se obter uma representação tridimensional (3D) do corpo.

Ao analisar um exame de TC para realizar um laudo, o médico faz um julgamento subjetivo com base na sua experiência. Se o médico possuir pouca experiência e não estiver seguro o suficiente para emitir o laudo, terá de realizar uma busca pelas patologias que são mais prováveis e comparar a descrição encontrada com a imagem do paciente, apesar de que às vezes são encontradas descrições com algumas imagens de exemplo, comparar uma descrição textual com uma imagem é um processo muito suscetível à falhas.

O exame de TC gera muitas imagens a serem analisadas, e essa grande quantidade de imagens contribui enormemente para o aumento de falhas humanas, pois analisar diversas imagens é um processo trabalhoso, complexo e tedioso. O grande número possível de combinações de padrões complexos achados nas diversas imagens, a falta de correlação fortemente estabelecida entre os achados radiológicos e patológicos e variações na forma de interpretação e descrição dos achados radiológicos, sem uma definição objetiva, são os principais fatores que resultam em grandes variações entre diagnósticos (UCHIYAMA et al., 2003).

Para tentar acelerar a análise e diminuir o número de laudos incorretos, foi desenvolvida uma ferramenta para auxiliar os médicos na tomada de decisão do laudo. Os resultados alcançados estão expostos nesse trabalho.

A ferramenta é capaz de processar e analisar a imagem do paciente, e então recuperar de uma base de conhecimento, imagens similares, dessa forma possibilitando ao médico a comparação visual com diversas outras imagens de TC, sem a necessidade prévia de um laudo, pois as imagens recuperadas são baseadas em padrões pictóricos da imagem do paciente. Estes padrões

pictóricos são obtidos por meio de extração de características texturais e classificados por uma rede neural.

1.1 Motivação

Tornar a avaliação de imagens de TC menos subjetiva, facilitando o trabalho de médicos menos experientes. Além de fornecer uma forma rápida de comparar exames similares de diferentes pacientes, tornando mais preciso o diagnóstico. Um dos propósitos de se desenvolver uma ferramenta como essa reside no fato de que, reconhecidamente, a avaliação desse tipo de doença é um dos problemas mais difíceis no diagnóstico médico (DOI, 2005), (BICK; DOI, 2006).

É importante evidenciar que a ferramenta desenvolvida é distribuída na forma de software livre, podendo, por isso, ser redistribuída, usada e modificada.

1.2 Objetivos

O objetivo desse trabalho é a realização de um estudo sobre técnicas de segmentação e extração de características de imagens pulmonares de TC e métodos para recuperação de imagens baseado em conteúdo, objetivando o desenvolvimento de um software capaz de auxiliar médicos no diagnóstico desse tipo de imagem.

O software desenvolvido realiza 3 etapas principais:

1. Segmentação das imagens resultantes da TC, visando extrair as regiões de interesse, que são os pulmões, e separá-los. Um dos meios que surgiram para auxiliar na segmentação e registro de imagens médicas foi o framework ITK – Insight Segmentation and Registration Toolkit. Com o qual espera-se conseguir um desenvolvimento mais rápido do software (YOO; ACKERMAN; LORENSEN, 2002).
2. Criação do vetor de características para cada pulmão segmentado na etapa anterior.
3. Comparação da imagem com outras previamente inseridas no software, utilizando uma função de similaridade.

1.3 Organização

Este trabalho, dividido em seis capítulos, apresenta primeiramente os conceitos estudados durante o seu desenvolvimento, seguido da descrição da ferramenta desenvolvida, os resultados obtidos e as conclusões.

No segundo capítulo, é apresentada uma introdução sobre as imagens de TC pulmonar, assim como suas características relevantes, e um estudo sobre as técnicas de processamento de imagens utilizadas para o desenvolvimento da ferramenta.

No capítulo seguinte, são demonstrados os fundamentos e o funcionamento básico de um sistema de recuperação de imagens baseado em conteúdo.

No quarto capítulo, é descrito o software desenvolvido durante este trabalho, bem como as bibliotecas utilizadas.

No quinto capítulo são exibidos os resultados obtidos pela ferramenta desenvolvida e, no último capítulo, são apresentadas as conclusões e as sugestões de trabalhos futuros.

2 PROCESSAMENTO DE IMAGENS

O processamento digital de imagens é um campo cujas aplicações são bastante extensas. Existem métodos de processamento de imagens para interesses muito diferentes, que vão desde a melhoria das informações visuais para a interpretação dos médicos até a computação dos dados da imagem para transmissão, armazenamento ou interpretação autônoma por computador (GONZALEZ; WOODS, 2002).

O processamento digital de imagens médicas é uma área em grande evolução e uma das que mais representa desafios dentro do processamento de imagens. Mas apesar de ser objeto de pesquisas de várias instituições ao redor do mundo, são poucos os sistemas deste tipo utilizados na prática médica. O principal fator é que este tipo de sistema exige alta eficiência, pois deve ser rápido, para que o tempo de resposta ao paciente não seja grande e ter uma baixíssima taxa de erros. Seus resultados são utilizados para tomar decisões em termos de diagnóstico e escolha de tratamento, por isso, alguns tipos de erro são inadmissíveis.

Em geral, os sistemas de processamento de imagens médicas extraem informações que podem ser provenientes de diversos tipos de modalidades de imagens, como Radiografia, Ultra-sonografia e Ressonância Magnética Nuclear, entre outras. Além das técnicas de processamento de imagens, técnicas de inteligência artificial, reconhecimento de padrões, entre outras áreas computacionais, são também aplicadas com o objetivo de melhorar tais imagens e extrair delas informações úteis ao diagnóstico.

2.1 Imagens de tomografia computadorizada

A tomografia era um dos mais importantes métodos de diagnóstico radiológico até a invenção da tomografia computadorizada - TC, na década de 70 do século passado. Sendo um dos primeiros tipos de exame a se beneficiar da popularização da computação.

A utilização de imagens no campo da medicina tem como principal objetivo proporcionar uma avaliação não invasiva dos tecidos e órgãos do corpo humano, tornando possível a verificação de anormalidades causadas por doenças ou acidentes (OLIVEIRA, 1999).

As máquinas de TC possuem uma fonte de raios x e detectores de raios x, os quais ficam em lados opostos um do outro. Para construir as imagens de TC a fonte e os detectores de raios x são movidos para posições pré-determinadas e então obtêm diversas imagens, como demonstrado na fig. 1, as quais representam fatias finas do paciente.

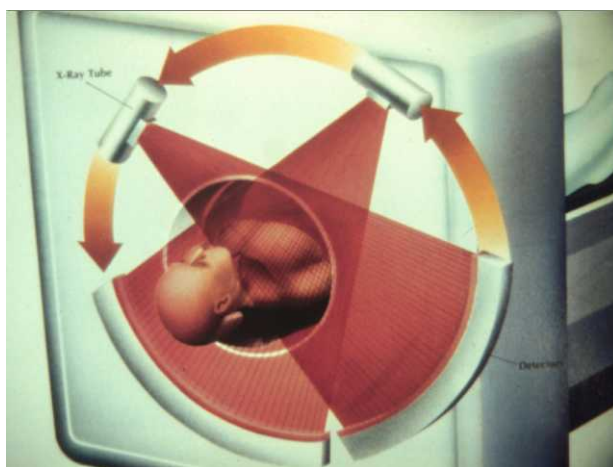


Figura 1: Demonstração de como uma máquina de TC obtém as fatias de imagem do paciente.

Fonte: SPRAWLS 2004

A forma de movimentação depende da tecnologia da máquina utilizada. Nas máquinas mais antigas, eram tiradas várias projeções de uma fatia do paciente, apenas circulando em torno dele, e então depois a "cama" onde o paciente deita-se se move, para outra rodada de projeções, como na fig. 2. Hoje em dia, temos as máquinas helicoidais, os quais descrevem uma hélice em torno do paciente, em vez de uma sucessão de círculos completo. Desta forma é obtida informação de uma forma contínua.

O resultado de cada pixel das fatias obtidas é a soma das energias bloqueadas pelos tecidos do corpo, e sabe-se exatamente em que ângulo cada fatia foi tirada, por isso após a aquisição das seções transversais, é possível a reconstrução da imagem em até 3 dimensões. Para a reconstrução é utilizada uma técnica matemática chamada de projeção retrógrada, ou outras, como a transformada de Fourier.

O campo de visão de uma imagem de TC representa o tamanho máximo do objeto em estudo que ocupa a matriz de pixels, por exemplo, uma matriz 512x512

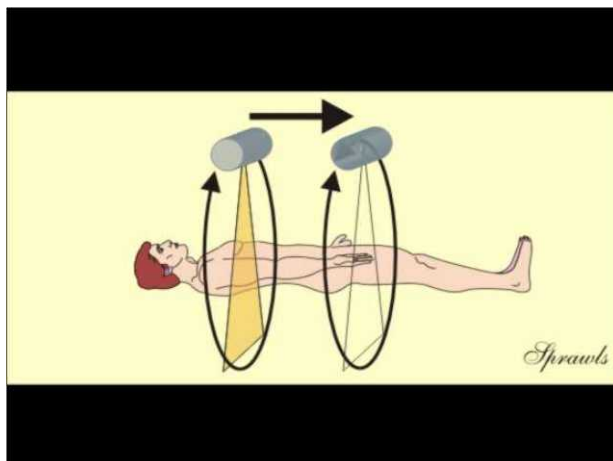


Figura 2: Forma de movimentação de máquinas de TC de tecnologia antiga.

Fonte: SPRAWLS 2004

pixels, se ela possuir um campo de visão de 12cm, cada pixel vai representar cerca de 0,023cm. Assim para o estudo de estruturas delicadas, como o ouvido interno, o campo de visão é pequeno, enquanto que para o estudo do abdômen o campo de visão é maior, em torno de 50cm (se tiver uma matriz de 512x512, então o tamanho da região que cada pixel representa vai ser próximo de 1 mm) (COSTA LEITE; JÚNIOR; OTADUY, 2004).

2.1.1 Coeficiente de Hounsfield

A escala da unidade hounsfield - UH - é uma transformação linear da medida do coeficiente de atenuação linear original, do qual a radiodensidade da água destilada a temperatura e pressão ambientes é definida como 0 UH, enquanto a radiodensidade do ar a temperatura e pressão ambientes é -1000 UH. Para um material X com coeficiente de atenuação μ_X , o correspondente valor em unidades Hounsfield é dado por:

$$\frac{\mu_X - \mu_{H_2O}}{\mu_{H_2O} - \mu_{ar}} \times 1000 \quad (1)$$

onde μ_{H_2O} e μ_{ar} são os coeficientes de atenuação linear da água e do ar, respectivamente, a pressão e temperatura ambientes. Então uma mudança de uma UH representa uma mudança de 0,1% da diferença do coeficiente de atenuação entre a água e o ar, ou aproximadamente 0,1% do coeficiente de atenuação da água, já que o coeficiente de atenuação do ar é aproximadamente 0. Alguns valores Hounsfield conhecidos estão mostrados na tab. 1.

Tabela 1: Valores Hounsfield comuns.

Tecido	Unidade Hounsfield
Ar	-1000
Pulmão	-900 a -400
Gordura	-110 a -65
Água	0
Rim	35
Sangue normal	35 a 55
Sangue coagulado	80
Músculo	40 a 60
Fígado	50 a 85
Ossos	130 a 250

Fonte: OLIVEIRA 1999

2.1.2 Técnica de Janelas

O olho humano tem a capacidade de diferenciar uma escala de cinzas de 10 a 60 tons (a maioria das pessoas distingue 20 tons diferentes), enquanto na tomografia há no mínimo 2000 tons. A técnica de janelas é na verdade uma forma de mostrar apenas uma faixa de tons de cinza que nos interessa, de forma a adaptar a nossa capacidade de visão aos dados obtidos pelo tomógrafo (COSTA LEITE; JÚNIOR; OTADUY, 2004).

Precisamos definir os valores máximo e mínimo da janela, para mapear os valores em UH em níveis de cinza. Todos os valores abaixo do valor mínimo serão mapeados como preto e todos acima do máximo como branco. Para definir estes valores, especificamos o tamanho da janela em UH e o centro dela, como podemos ver na fig. 3.

O uso de diferentes janelas em tomografia permite por exemplo o estudo dos ossos com distinção entre a cortical e a medula óssea ou o estudo de partes moles com a distinção, por exemplo, no cérebro entre a substância branca e a cinzenta. A mesma imagem pode ser mostrada com diferentes ajustes da janela, de modo a mostrar diferentes estruturas de cada vez. Não é possível usar um só ajuste da janela para ver, por exemplo, detalhes ósseos e de tecido adiposo ao mesmo tempo. Alguns valores comuns de janela estão na tab. 2.

2.1.3 Padrão DICOM

A introdução de imagens médicas digitais na década de 70 do século passado e o uso de computadores para processar estas imagens fizeram com que o American College of Radiology (ACR) e a National Electrical Manufacturers Association (NEMA) se juntassem para formar um comitê com o objetivo de criar

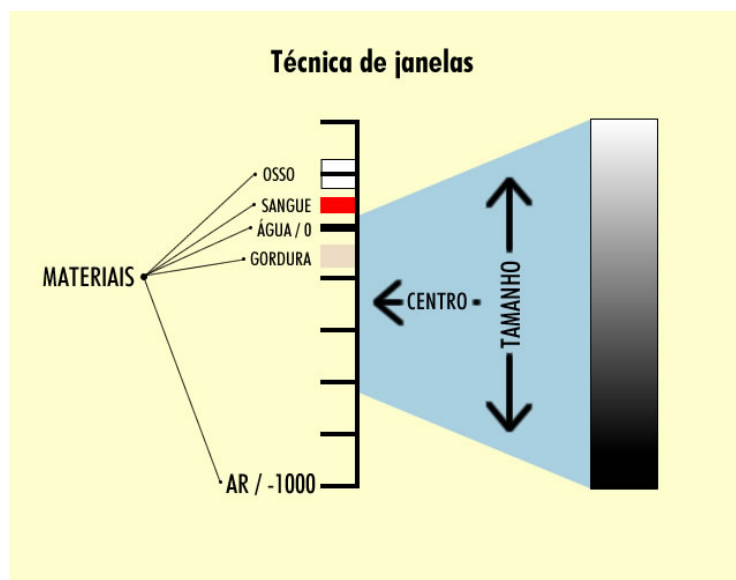


Figura 3: Exemplo de janela.

Tabela 2: Valores comuns de janela.

Tecido a ser destacado	Largura da janela	Centro da janela
Tecido mole (Abdômen)	400	40
Pulmão	1500	-700
Cérebro	100	35
Osso	2000	350
Fígado	2000	-500

Fonte: SHULKIN 2005

um método padrão para a transmissão de imagens médicas e das informações associadas a elas.

Este comitê, constituído em 1983, publicou seu primeiro conjunto de padrões, chamado de ACR-NEMA, em 1985, e o segundo em 1988. Até a publicação dos padrões, a maioria dos equipamentos utilizava formatos proprietários para fazer o armazenamento e comunicação de imagens.

Embora as primeiras versões não tenham obtido êxito total na definição de um padrão comum, a terceira versão, publicada em 1993 sob o nome de DICOM (Digital Imaging and Communications in Medicine), conseguiu estabelecer uma forma padronizada de armazenamento e comunicação de imagens médicas e as correspondentes informações associadas.

Com os melhoramentos promulgados por esta terceira versão, o padrão estava pronto tanto para permitir transferência de imagens médicas em um ambiente com múltiplos fabricantes como também para facilitar o

desenvolvimento e a expansão dos sistemas de armazenamento e de comunicação e a conexão com os sistemas de informação médica (NEMA, 2008).

Nos dias de hoje, a maioria dos fabricantes de equipamentos para aquisição de imagens médicas permite que os arquivos sejam exportados nesse formato, além dos formatos proprietários. Assim como a maioria dos softwares de processamento de imagens médicas também apresenta compatibilidade com esse formato.

2.2 Segmentação

Segmentação é a área do processamento de imagens que trata de isolar as regiões de interesse de uma imagem ou mudar a sua representação para facilitar a sua análise em determinada aplicação. Segmentação de imagens é tipicamente utilizada para localizar objetos e formas (curvas, linhas, etc) em imagens.

A segmentação de imagens não triviais é uma das tarefas mais difíceis no processamento de imagens. A precisão da segmentação determina o sucesso ou falha de um sistema de análise computadorizada (GONZALEZ; WOODS, 2002).

Mas o processo de segmentação é enormemente facilitado quando o domínio das imagens do problema é bem conhecido e restrito. Desse forma permitindo que as técnicas de segmentação possam ser alteradas para trazer resultados mais satisfatórios no domínio do problema.

2.2.1 Threshold Adaptativo

Threshold é um dos métodos mais simples de segmentação de imagens. A partir de uma imagem em tons de cinza, o threshold pode ser usado para torná-la binária.

Durante o processo de threshold, os pixels são divididos em dois grupos, na forma mais simples de threshold, a divisão é feita dependendo apenas se o valor do pixel é maior ou menor que o valor de threshold. Então um grupo é colorido de preto e o outro de branco, tornando a imagem binária. A divisão dos grupos de pixels pode ser feita com base em mais de um valor de threshold, dessa forma os pixels são divididos entre os que estão dentro de um intervalo formado por 2 valores de threshold e os que não estão.

O parâmetro chave que determina a eficiência de um threshold é o valor de threshold usado. Diversos métodos existem para a escolha do valor de threshold, ele pode ser escolhido manualmente, ou por algum algoritmo que o compute,

dependendo da imagem, esse tipo de threshold mais conhecido como threshold automático. Um método simples é escolher o valor da média ou da mediana. Geralmente esse algoritmo só irá atingir bons resultados se a imagem de entrada possuir pouco ruído e for uniforme. Uma abordagem um pouco mais sofisticada seria gerar o histograma da intensidade dos pixels da imagem e escolher como threshold o valor de vale.

A técnica de threshold adaptativo, que pode ser vista na fig. 4, consiste em escolher um valor inicial arbitrário de threshold (T^0), realizar o threshold e calcular o valor médio dos pixels acima (μ_{acima}) e abaixo (μ_{abaixo}) do valor de threshold. Então obter um novo threshold (T^{i+1}), de acordo com a equação 2. Com o novo threshold o processo se inicia novamente, e só para quando atingir a convergência, ou seja, o novo valor for igual ao antigo valor ($T^i = T^{i+1}$). Este algoritmo garante a convergência até um mínimo local.

$$T^{i+1} = \frac{\mu_{acima} + \mu_{abaixo}}{2} \quad (2)$$

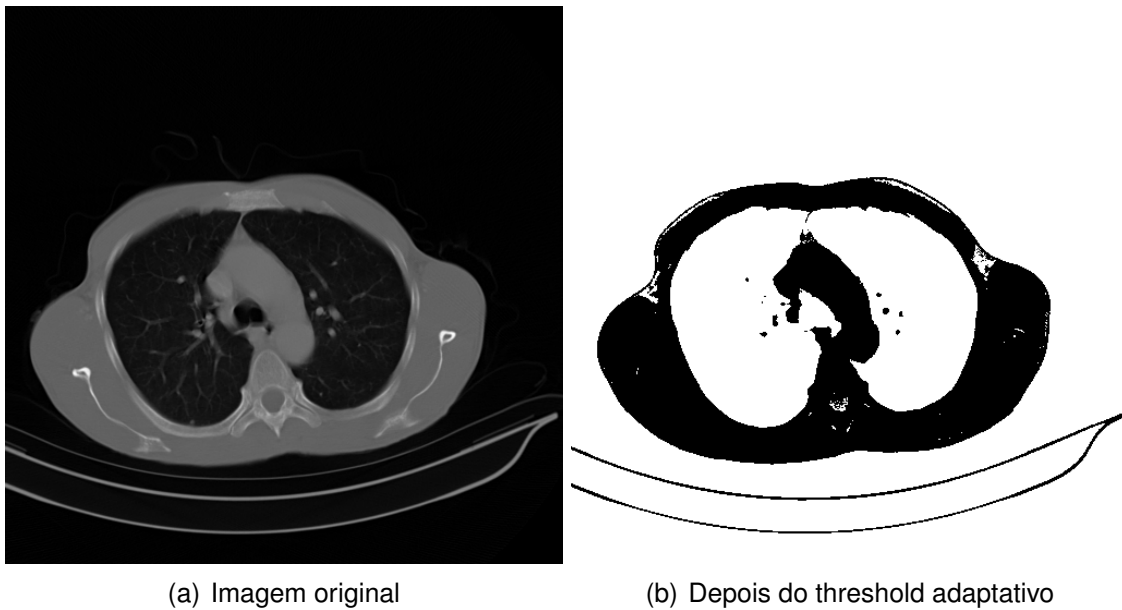


Figura 4: Imagem de TC original e após a aplicação de threshold adaptativo.

2.3 Morfologia Matemática

Morfologia é o estudo da forma. Em processamento de imagens, morfologia matemática é o nome que se dá a um conjunto de métodos, inicialmente desenvolvidos por Georges Matheron e Jean Serra em 1964, que têm em comum o objetivo de estudar a estrutura geométrica de uma imagem.

A linguagem utilizada na morfologia matemática é a teoria dos conjuntos.

Conjuntos na morfologia matemática representam objetos numa imagem. Por exemplo, o conjunto de todos os pixels pretos numa imagem binária é uma descrição morfológica completa da imagem. Em imagens binárias, os conjuntos em questão são membros do espaço dos inteiros bidimensional - $2D - Z^2$, onde cada elemento do conjunto é uma tupla cujas coordenadas são as coordenadas de um pixel preto na imagem. Imagens em tons de cinza podem ser representadas como conjuntos cujos componentes estão em Z^3 . Neste caso, dois componentes de cada elemento do conjunto se referem as coordenadas do pixel, e o terceiro corresponde ao valor do seu nível de cinza. Conjuntos em espaços dimensionais ainda mais elevados podem conter outros atributos da imagem, como cor ou componentes que variam com o tempo (GONZALEZ; WOODS, 2002).

Então temos a morfologia binária que se aplica sobre imagens binárias e a morfologia cinza que se aplica a imagens em níveis de cinza. Uma operação morfológica binária é completamente determinada a partir da vizinhança examinada ao redor do ponto central e do algoritmo utilizado. Já na morfologia cinza, na vizinhança de cada pixel, ou numa parte da vizinhança, é necessário conhecer o valor do pixel mais escuro, e o valor do mais claro. O valor do pixel resultante corresponde a uma combinação particular desses dois. Portanto o tamanho e a forma da vizinhança, as regiões de pesquisa dos valores máximo e mínimo e o algoritmo utilizado, determinam uma operação de morfologia cinza (FACON, 2000).

As operações fundamentais da morfologia matemática são a erosão e a dilatação. Essas operações necessitam do que chamamos de elementos estruturantes. Estes elementos, determinam quais pixels da imagem são retirados (no caso de uma erosão) ou quais são adicionados (no caso de uma dilatação) aos objetos. Elementos estruturantes são simplesmente imagens binárias em que um dos pontos é definido como centro. Como exemplos largamente utilizados podemos citar: um quadrado de 3x3 pixels (fig. 5(a)), uma cruz definida pelo pontos $(-1,0), (0,-1), (0,0), (0,1), (1,0)$ (fig. 5(b)) e um círculo de raio qualquer (fig. 5(c)).

2.3.1 Erosão e Dilatação

Na erosão, o elemento estruturante é sobreposto à imagem em todas as posições possíveis. O pixel que fica na posição da origem do elemento estruturante é modificado da seguinte forma: se o elemento estiver em parte sobre o objeto e em parte sobre o fundo, o pixel que está na posição da origem passa a ser fundo. Ou seja, um pixel só pode permanecer no objeto se quando a origem do elemento estruturante estiver sobre ele, todo o elemento estiver sobre

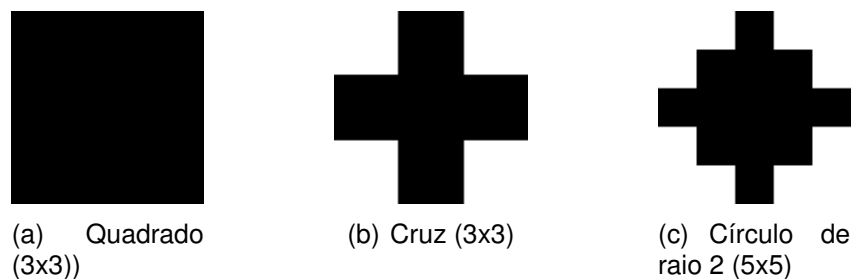


Figura 5: Exemplos de elementos estruturantes.

o objeto, como podemos ver na fig. 6(c).

Efeitos da erosão:

- Diminui as partículas;
- Elimina grãos de tamanho inferior ao tamanho do elemento estruturante;
- Aumenta os buracos;
- Separa grãos próximos.

Na dilatação o que muda é que quando o elemento está em parte sobre o fundo e em parte sobre o objeto, o seu centro passa a ser objeto, como vemos na fig. 6(b). Isto aumenta o tamanho do objeto, como o nome dilatação sugere.

Efeitos da dilatação:

- Aumenta as partículas;
- Preenche pequenos buracos;
- Conecta grãos próximos.

2.3.2 Abertura e Fechamento

Erosão e dilatação podem corrigir defeitos numa imagem com furos, conexões. Porém, o objeto alterado por essas operações não mantém o mesmo tamanho. A erosão reduz e a dilatação aumenta. Com a abertura e o fechamento binários, podemos manter aproximadamente as características de forma e tamanho do objeto.

A abertura é obtida através da erosão, seguida de uma dilatação da imagem resultante com o mesmo elemento estruturante. Esta operação é capaz de eliminar pequenas partículas inferiores em tamanho ao elemento estruturante

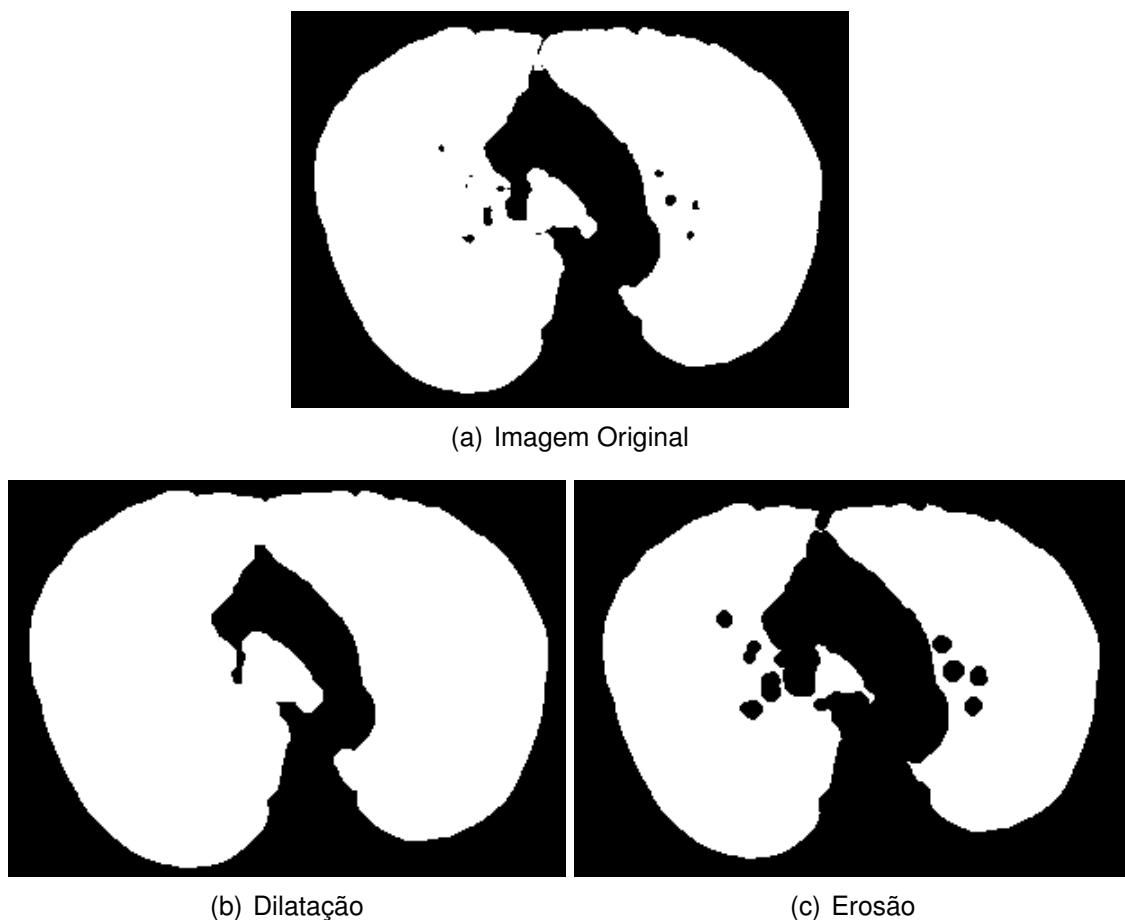


Figura 6: Imagem de TC após passar por threshold adaptativo e limpeza, submetida a dilatação e erosão utilizando como elemento estruturante um círculo de raio 3 pixels.

quase sem modificar o tamanho das outras entidades, além de nivelar os contornos pelo interior, como podemos ver na fig. 7(b).

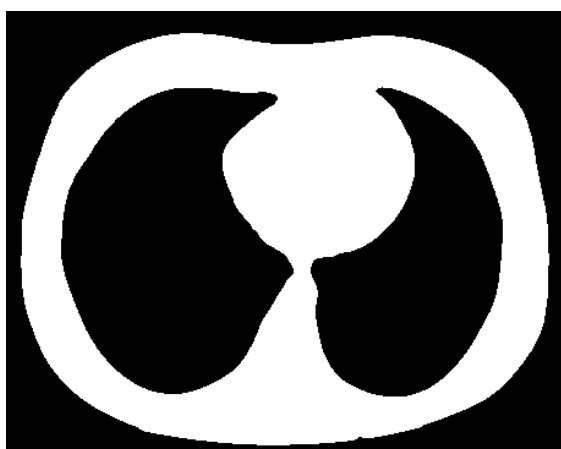
Já o fechamento é o inverso; dilatação, seguida de erosão. Ele conecta partículas próximas, suaviza as bordas pelo exterior e preenche os buracos inferiores em tamanho em relação ao elemento estruturante que estejam no interior das partículas, visto na fig. 7(c).

2.4 Extração de características de texturas

O sistema de visão humano percebe uma cena como sendo variações de intensidade e cor, a imagem toda ou as partes dela que formam certos padrões repetitivos, são chamadas de textura. Numa imagem de textura, um conjunto estatístico ou de atributos, varia lentamente ou se mantém quase periódico. Estes atributos, os quais são repetitivos ou quase, dominam as texturas de uma cena.



(a) Imagem Original



(b) Abertura



(c) Fechamento

Figura 7: Imagem de TC após passar por threshold adaptativo, submetida a abertura e fechamento utilizando como elemento estruturante um círculo de raio 6 pixels.

Uma textura é uma junção de subpadrões que se repetem, que seguem um conjunto de regras de posicionamento bem definido. Estes subpadrões são feitos de unidades mais básicas, chamadas de primitivas. Esta caracterização de texturas é principalmente aplicável a texturas determinísticas, como por exemplo, linhas, tabuleiro de dama, etc. Existem imagens, como imagens de satélite da superfície terrestre, que aparentemente não possuem esses padrões básicos que são repetidos ao longo do padrão mais abrangente (ACHARYA; RAY, 2005).

Uma imagem que seja uma textura, pode então ser toda reduzida num conjunto de características, também chamado de vetor de características. Essa transformação se chama extração de características. Se as características extraídas forem cuidadosamente escolhidas é esperado que o conjunto de características represente toda a informação relevante da textura.

Uma primitiva é um conjunto conectado de pixels, caracterizado por um

conjunto de atributos. A primitiva mais simples é um único pixel com seu atributo de nível de cinza. Um conjunto de pixels conectado que possui o mesmo nível de cinza ou que possui a mesma direção de borda forma uma primitiva. Outros atributos locais podem ser considerados para se definir uma primitiva. Outros atributos incluem a medida de regiões conectadas e a homogeneidade da sua propriedade local. Primitivas podem ser geradas a partir de imagens por vários filtros de vizinhança (ACHARYA; RAY, 2005).

2.4.1 Matriz de co-ocorrência

Matriz de co-ocorrência é um dos métodos de processamento de imagens utilizado para caracterizar texturas. Ele descreve uma imagem, ou uma região de interesse na imagem, em termos da relação entre os valores dos pixels com os valores dos pixels vizinhos.

Uma matriz de co-ocorrência é definida a partir de uma imagem como sendo a distribuição dos valores co-ocorrentes num dado deslocamento. Matematicamente, a matriz de co-ocorrência C é definida a partir de uma imagem I de $N \times M$, parametrizada com um deslocamento de $(\Delta x, \Delta y)$, como podemos ver abaixo:

$$C(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{se } I(p, q) = i \text{ e } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{senão} \end{cases} \quad (3)$$

Para poder ser utilizada, a matriz de co-ocorrência precisa ser normalizada, para isso dividimos cada elemento da matriz pelo número total de pares de pixels.

Percebemos que o deslocamento $(\Delta x, \Delta y)$ faz com que a matriz de co-ocorrência seja sensível a rotação. Dessa forma, uma rotação da imagem diferente de 180 graus irá resultar numa matriz de co-ocorrência diferente para uma mesma imagem (rotacionada). Isso é raramente desejado nas aplicações em que a matriz de co-ocorrência é utilizada, portanto a matriz de co-ocorrência é geralmente utilizada a partir da média em cada valor da matriz para um conjunto de deslocamentos que variam até 180 graus com uma mesma distância entre eles (por exemplo 0, 45, 90 e 135 graus) para não ser influenciada pela rotação.

2.4.2 Características de Haralick

Em HARALICK; SHANMUGAM; DINSTEIN 1973 foi sugerido um total de 14 características que podem ser computadas diretamente da matriz de co-ocorrência, as quais podem ser utilizadas para classificar imagens de textura. Abaixo estão descritas algumas delas.

Considere N como sendo o número de níveis de cinza da matriz de co-ocorrência C .

A Energia, também conhecida como segundo momento angular, avalia a uniformidade textural em uma imagem. A fórmula que descreve a energia pode ser vista na equação 4.

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C(i, j)^2 \quad (4)$$

A Entropia mede a desordem em uma imagem, ou seja, o grau de dispersão de níveis de cinza. Ela é calculada como na equação 5.

$$- \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C(i, j) \log(C(i, j)) \quad (5)$$

O contraste, mede a presença de transição abrupta de níveis de cinza (bordas) na imagem. Ele é definido como:

$$\sum_{x=0}^{N-1} x^2 \left\{ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C(i, j) \right\}, |i - j| = x \quad (6)$$

A homogeneidade local, também conhecido como momento diferencial inverso, como o próprio nome sugere, mede a homogeneidade da imagem. Sua fórmula é:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1 + (i - j)^2} C(i, j) \quad (7)$$

A correlação mede a dependência linear de um nível de cinza em relação aos vizinhos. Ela é calculada como na equação 8.

$$\frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} ij C(i, j) - \mu_i^2}{\sigma_i^2} \quad (8)$$

Onde μ_i representa a média ponderada, como definido na equação 9. Por causa da simetria da matriz de co-ocorrência, não é necessário calcular a média ponderada de μ_j . Já σ_i representa o desvio padrão, como na equação 10, que também se beneficia da simetria.

$$\sum_{j=0}^{N-1} \sum_{i=0}^{N-1} i \cdot C(i, j) \quad (9)$$

$$\sum_{j=0}^{N-1} \sum_{i=0}^{N-1} (i - \mu_i)^2 \cdot C(i, j) \quad (10)$$

Além dessas características, também estão definidas em (HARALICK;

SHANMUGAM; DINSTEIN, 1973): soma dos quadrados, soma da média, soma da variância, soma da entropia, variância diferencial, entropia diferencial, medida de informação da correlação e coeficiente de correlação máxima.

3 RECUPERAÇÃO DE IMAGENS BASEADA EM CONTEÚDO

Os primeiros trabalhos sobre recuperação de imagens são datados do final da década de 1970. Em 1979, uma conferência sobre *Database Techniques for Pictorial Applications* (numa tradução livre "Técnicas de banco de dados para aplicações com imagens") foi realizada em Florência, Itália. Desde então, o potencial de técnicas de gerenciamento de banco de dados de imagens atraiu a atenção dos pesquisadores.

As primeiras técnicas não foram baseadas em características visuais e sim em anotações textuais das imagens. Em outras palavras, as imagens eram primeiro descritas textualmente, para então serem recuperadas usando busca baseada em texto dos sistemas de gerenciamento de banco de dados tradicionais. Através das descrições textuais, as imagens podem ser organizadas por tópicos ou classes hierárquicas para facilitar a busca através de requisições usando expressões lógicas. Entretanto, como a geração automática de descrições para imagens diversas é uma tarefa quase impossível, a maioria dos sistemas de recuperação baseados em imagem requer que as descrições sejam feitas manualmente. Obviamente, fazer anotações manualmente sobre uma imagem é uma tarefa difícil e onerosa para grandes bancos de dados de imagens, além de ser geralmente subjetiva, sensível ao contexto e incompleta (LONG et al., 2003).

O objetivo dos sistemas de recuperação de imagens baseada em conteúdo - CBIR (content based image retrieval) - é buscar em um banco de imagens as imagens mais relevantes relativas a uma requisição visual. Dessa forma eliminando a necessidade de descrições textuais manuais e tornando a busca mais eficiente e menos subjetiva.

Num sistema CBIR típico, o conteúdo visual relevante das imagens no banco de imagens é extraído e armazenado como um vetor de características. Para realizar a recuperação das imagens mais relevantes, o usuário fornece ao sistema uma imagem exemplo ou um esboço do que ele busca. O sistema então realiza a extração das características. A similaridade entre o vetor de

característica da imagem exemplo ou esboço e cada uma das imagens do banco de dados é então calculada e a recuperação das imagens mais similares realizada. Sistemas de CBIR recentes levam em conta a resposta do usuário, sobre quais imagens são relevantes, para modificar o processo de recuperação, de forma a gerar resultados melhores (fig. 8).

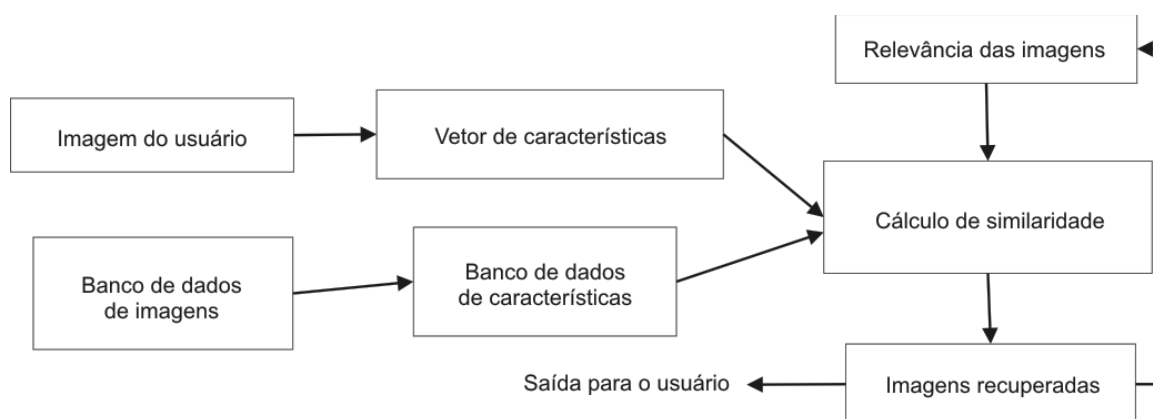


Figura 8: Diagrama de um sistema CBIR típico.

3.1 Similaridade

No dia-a-dia precisamos categorizar diversas coisas no mundo real e da informação, para que, pela comparação com conceitos e situações já existentes, possamos aprender outros. Portanto a capacidade de analisar a similaridade das coisas é importantíssima. Para um sistema de CBIR, é necessário que a medida de similaridade chegue o mais perto possível da percepção do usuário, pois ela irá afetar a eficiência das buscas significativamente.

Várias medidas de similaridade foram desenvolvidas para recuperação de imagens nos últimos anos, a maioria baseada em estimativas empíricas da distribuição de características (LONG et al., 2003). Entretanto essas medidas são estáticas, após escolhidas durante o desenvolvimento do sistema, não são mais alteradas. Dessa forma a utilidade desses sistemas é limitada devido a falta de capacidade de se representar conceitos de alto nível nas imagens utilizando características de baixo nível.

Para contornar esse problema, pesquisadores (EL-NAQA et al., 2004) (LIM et al., 2001) tem usado para a avaliação de similaridade redes neurais artificiais - RNAs, as quais são capazes de aprender através da interação do usuário.

3.2 Redes neurais artificiais

Uma RNA é composta de nodos conectados por arestas direcionadas. Cada aresta tem um peso associado. São esses pesos os responsáveis pelo armazenamento em RNAs, e o aprendizado geralmente acontece com a modificação desses pesos. A modificação de pesos tenta fazer com que o comportamento dos nodos de saída a partir dos nodos de entrada estejam mais próximos do valor esperado.

Cada unidade tem um conjunto de aresta advindas de outros nodos e um conjunto partindo para outros nodos, um nível de ativação e um meio de calcular o nível de ativação, dados as suas entradas e pesos (fig. 9). A idéia é que cada nodo faça um cálculo local baseado das entradas advindas dos nodos vizinhos, mas sem necessidade de qualquer controle global sobre o conjunto de nodos (RUSSELL; NORVIG, 2003).

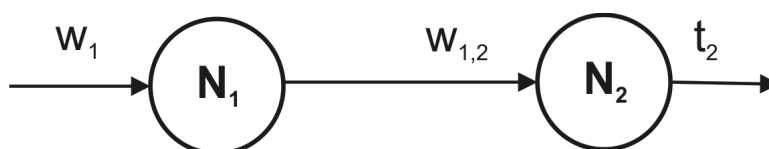


Figura 9: Uma RNA simples.

Para construir uma RNA, primeiro decide-se quantos nodos ela possuirá, de que tipo serão e como serão conectados. Então os pesos são inicializados, e então treinados utilizando um algoritmo de aprendizado a partir de um conjunto de exemplo da tarefa a ser realizada. O uso de exemplo também implica que deve ser decidido como representar os exemplos em termos de entrada e saída da RNA.

Para computar o seu nível de ativação um nodo utiliza duas funções, as quais são chamadas de função de entrada e função de ativação. A função de entrada é linear e usada para computar a soma ponderada dos valores de entrada do nodo. Já a função de ativação é não linear e utilizada para transformar a soma ponderada no valor final que será o valor de ativação desse nodo.

Sendo $w_{i,j}$ o peso da aresta que sai do nodo i até o nodo j e a_i o nível de ativação do nodo i , a função de entrada para j é calculada como na equação 11.

$$\sum_i w_{i,j} \cdot a_i \quad (11)$$

Diferentes modelos de RNAs são obtidos utilizando-se diferentes funções de ativação. As duas escolhas mais comuns são funções lineares e sigmóides, ilustradas na fig. 10. As funções lineares tem um valor de limiar t tal que a

saída da função é 1 quando o valor de entrada é maior que o limiar, e 0 em qualquer outro caso.

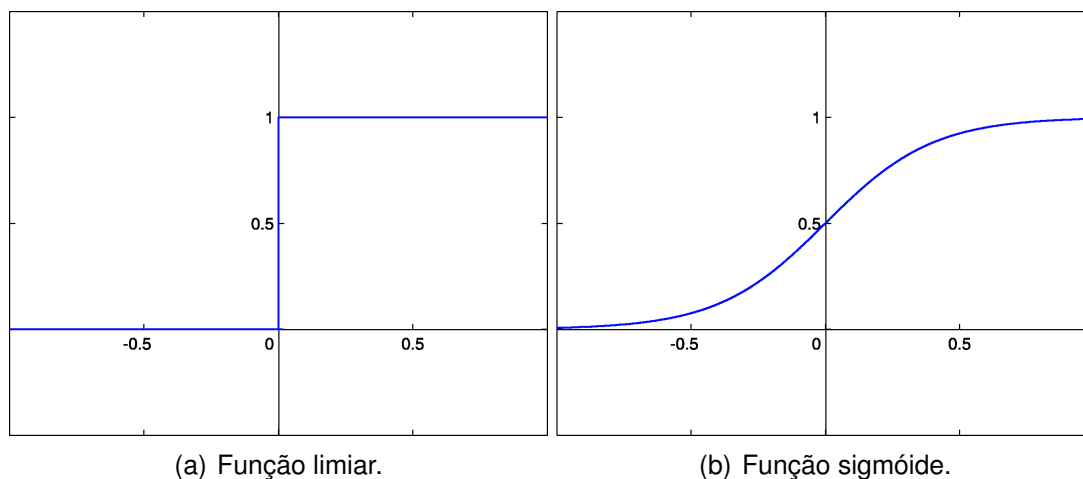


Figura 10: Funções de ativação comuns em RNAs.

Na maioria dos casos, será matematicamente conveniente substituir o valor de limiar com um peso extra de entrada no nodo. Dessa forma podemos utilizar um algoritmo mais simples de aprendizado, pois teremos a necessidade de atualizar apenas os pesos, e não os pesos e os limiares. Esse peso extra que deveria ter um valor de ativação vindo de um nodo ou uma entrada, terá sempre -1 como valor de ativação. Servindo assim como um limiar t , considerando que $w_{0,j} \cdot a_0 = -t$, então todos os nodos poderão ter o limiar fixo em 0.

Existe uma variedade de tipos de estruturas de RNAs, e cada uma resulta em propriedades computacionais muito diferentes. A maior distinção pode ser feita entre as pró-alimentadas e as recorrentes. Nas pró-alimentadas as arestas são unidirecionais e não possuem ciclos, ou seja, são grafos acíclicos direcionados. Numa estrutura recorrente, elas podem formar topologias arbitrárias.

3.2.1 Perceptron de múltiplas camadas pró-alimentado

Um RNA para ser considerada um perceptron de múltiplas camadas deve ter além dos nodos de entrada e saída, nodos intermediários, os quais interagem apenas com outros nodos (SAMARASINGHE, 2006). Podemos ver um perceptron de múltiplas camadas pró-alimentado na fig. 11.

Com uma camada de nodos intermediários (suficientemente grande), é possível representar qualquer função contínua das entradas, já com duas camadas intermediárias é possível representar até funções descontínuas.

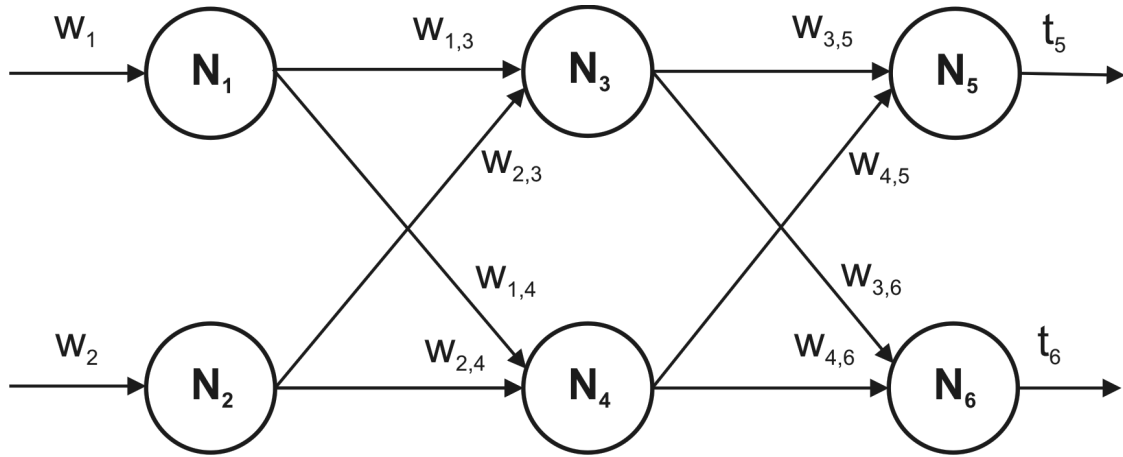


Figura 11: Perceptron de múltiplas camadas pró-alimentado simples.

3.2.2 Algoritmo de retro-propagação

A maneira mais popular de se treinar uma RNA de múltiplas camadas é utilizando o algoritmo de retro-propagação. Ele tenta minimizar o erro entre o valor de saída e o valor esperado, verificando qual é a contribuição de cada peso no erro.

Desta forma, se a RNA calcular um valor correto, os pesos não são alterados. Mas se houver diferença entre o valor esperado e o calculado, os pesos são ajustados de forma a reduzir o erro. Para os nodos de saída utiliza-se a equação 12,

$$w_{i,j} = w_{i,j} + a \cdot a_i \cdot (t_j - o_j) \cdot g'(e_j) \quad (12)$$

onde t_j é a saída do nodo j e o_j a saída esperada do nodo j .

Sendo a uma constante chamada de taxa de aprendizado, e_j o resultado da função de entrada e $g'()$ é a derivada da função de ativação. Será conveniente definir $\Delta_j = (t_j - o_j) \cdot g'(e_j)$. Então a regra de atualização de pesos pode ser escrita como

$$w_{i,j} = w_{i,j} + a \cdot a_i \cdot \Delta_j \quad (13)$$

Para a atualização dos pesos dos nodos de entrada e intermediários nós precisamos definir uma quantidade análoga ao Δ_j dos nodos de saída. Este ponto é onde a retro-propagação do erro ocorre. Cada nodo i é "responsável" por uma fração do erro Δ_j em cada um dos nodos de saída em que ele se conecta. Então o valor de Δ_j é dividido de acordo com a influência da conexão entre o nodo intermediário e o nodo de saída e propagado de volta para que os valores de Δ_i sejam calculados. A regra de propagação é a seguinte

$$\Delta_i = g'(e_i) \cdot \sum_j (w_{i,j} \cdot \Delta_j) \quad (14)$$

Agora a equação de atualização dos pesos dos nodos de entrada e intermediários é igual a dos nodos de saída, devemos apenas tomar cuidado no cálculo dos Δ , que são diferentes.

3.3 Eficiência de um sistema de CBIR

As medidas mais usadas para avaliar a qualidade dos resultados de um sistema de CBIR são a precisão e o *recall*, os quais são calculados pelas equações 15 e 16, respectivamente.

$$precisão = \frac{\text{número de imagens relevantes recuperadas}}{\text{número de imagens recuperadas}} \quad (15)$$

$$precisão = \frac{\text{número de imagens relevantes recuperadas}}{\text{número de imagens relevantes}} \quad (16)$$

É trivial atingir um *recall* de 100% retornando todas as imagens do banco. Por isso o *recall* sozinho não é uma medida boa, deve ser interpretada junto com a precisão do sistema. Geralmente, existe uma relação inversa entre a precisão e o *recall*, onde é possível aumentar um ao custo de reduzir o outro.

4 FERRAMENTA DESENVOLVIDA

Neste trabalho foi desenvolvido uma ferramenta cujo objetivo é auxiliar na interpretação de exames de tomografia computadorizada pulmonares. A ferramenta, através do processamento de imagens digitais e de técnicas de recuperação de imagens baseada em conteúdo, é capaz de oferecer suporte ao médico na realização do diagnóstico do paciente.

Neste capítulo serão demonstrados aspectos da implementação e utilização do protótipo desenvolvido.

4.1 Ambiente de desenvolvimento

O desenvolvimento do trabalho foi realizado utilizando-se um notebook equipado com processador Intel® Core™ 2 Duo T5450 (1.66GHz), 3GB de memória RAM DDR2 e placa de vídeo NVIDIA® GeForce™ 8400M GS com 256MB de memória dedicada.

O sistema operacional de desenvolvimento foi o GNU/Linux da distribuição Debian *testing*, utilizando a IDE NetBeans 6.1 com o auxílio das ferramentas cmake, make e gcc. Todas as ferramentas utilizadas são multiplataforma.

4.2 Imagens utilizadas

Nos diversos experimentos executados durante o desenvolvimento da ferramenta, foram utilizadas imagens médicas de tomografia computadorizada pulmonar cedidas pelo Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

4.3 Linguagens de programação

Para o desenvolvimento da ferramenta foram utilizadas as linguagens de programação C++, PHP e javascript. O C++ é uma linguagem de alto nível de abstração, com sistema de tipos estático e que permite a programação em

múltiplos paradigmas, incluindo programação procedural, genérica e orientada a objetos (STROUSTRUP, 2000). Esta linguagem foi desenvolvida originalmente por Bjarne Stroustrup nos laboratórios Bell como uma melhoria para a linguagem de programação C. Dentre as capacidades do C++, destacam-se a possibilidade de fazer abstração de dados, a sobrecarga de operadores e métodos e tratamento de exceções.

Esta linguagem foi escolhida devido a seu desempenho, suporte a orientação a objetos e velocidade de desenvolvimento, além de ser a linguagem nativa da biblioteca mais utilizadas, a ITK.

Já as linguagens PHP e javascript foram escolhidas para a interface gráfica com o usuário (GUI), pois permitem que a ferramenta esteja disponível pela internet, tornam a centralização do banco de dados muito mais fácil, compartilhando o contínuo treinamento da ferramenta e não exigem que a máquina do usuário tenha grande poder de processamento.

4.4 Bibliotecas

Muitas das funções desempenhadas pela ferramenta desenvolvida incluem o processamento e visualização de imagens e gráficos, e ainda a construção, treinamento e utilização de uma rede neural. O desenvolvimento destas rotinas foi escrito com base em algumas bibliotecas que são referência na área por proverem grande quantidade de algoritmos já codificados, testados e prontos para o uso.

Todas as bibliotecas utilizadas no desenvolvimento da ferramenta são distribuídas livremente, com código-fonte aberto. Além disso todas as bibliotecas são multiplataforma.

4.4.1 ITK

Insight Segmentation and Registration ToolKit (ITK) é uma biblioteca proposta pela Biblioteca Nacional de Medicina dos Estados Unidos da América, voltada para segmentação e registro de imagens (YOO; ACKERMAN; LORENSEN, 2002).

A biblioteca é composta por algoritmos e estruturas de representação de dados com duas finalidades principais: a identificação e classificação de elementos encontrados em uma imagem digital (segmentação) e a tarefa de alinhar imagens ou encontrar correspondências entre dados (registro).

No ITK há um foco em aplicações médicas, embora não haja restrições quanto ao processamento de outros tipos de dados. Essa biblioteca não dá enfoque à parte de visualização deixando a cargo de outras ferramentas que

possam ser utilizadas conjuntamente.

O ITK é mantido, basicamente, por seis instituições: Kitware, GE Corporate R&D, Insightful, University Chapel Hill, University of Utah e University of Pennsylvania, sendo as três primeiras comerciais e as demais são instituições acadêmicas. Outros membros do projeto são Harvard Brigham & Women's Hospital, University of Pittsburgh e Columbia University (KITWARE, 2008).

Quanto à linguagem, ITK também é escrito em C++, mas possui interface para utilização a partir de outras linguagens. É um conjunto de ferramentas que provê uma grande quantidade de algoritmos pré-codificados e testados para fazer registro e segmentação de imagens, bem como rotinas para leitura e decodificação de diversos padrões de arquivos.

O ITK é distribuído livremente e com código-fonte aberto, impondo poucas restrições ao seu uso, modificação e distribuição.

4.4.2 FANN

Fast Artificial Neural Network (FANN) é uma biblioteca de RNAs open source. Ela é simples, bem documentada, versátil, fácil de usar e rápida. A FANN é implementada em C, mas possui versões para C++, Java, Perl, PHP, Python, Ruby, Delphi, Haskell, Mathematica, Matlab, Prolog, Octave, Smalltalk e .NET. Ela foi desenvolvida por Steffen Nissen do departamento de Ciência da Computação da Universidade de Copenhagen, que, logo depois, liberou o código sob a licença GPL.

A escolha dessa biblioteca teve como fator principal a possibilidade de se utilizar as linguagens C++ e PHP. O que não foi encontrado em nenhuma outra biblioteca.

A biblioteca implementa uma RNA de múltiplas camadas totalmente ou esparsamente conectada. Com a FANN, a criação de uma RNA é feita em três níveis. O primeiro é a descrição da rede, o segundo é a criação das conexões da primeira camada, e depois as interconexões das demais camadas. Ela pode trabalhar tanto com números em ponto flutuante quanto números inteiros. Além disso, ela possui um framework para que seja fácil o treinamento a partir de conjuntos de dados.

Ela é flexível, quase todos os valores dos algoritmos de treinamento podem ser alterados, assim como os próprios algoritmos de treinamento.

4.5 Interface Gráfica

A ferramenta foi concebida para ser utilizada em dois modos, o modo normal e o modo de avaliação. No modo normal o usuário irá submeter uma imagem

de TC pela página e escolher se quer ver as imagens mais similares ao pulmão esquerdo ou direito. Então ele poderá ver a segmentação feita nos dois pulmões da imagem enviada, e as 12 imagens mais similares ao pulmão escolhido (fig. 12).

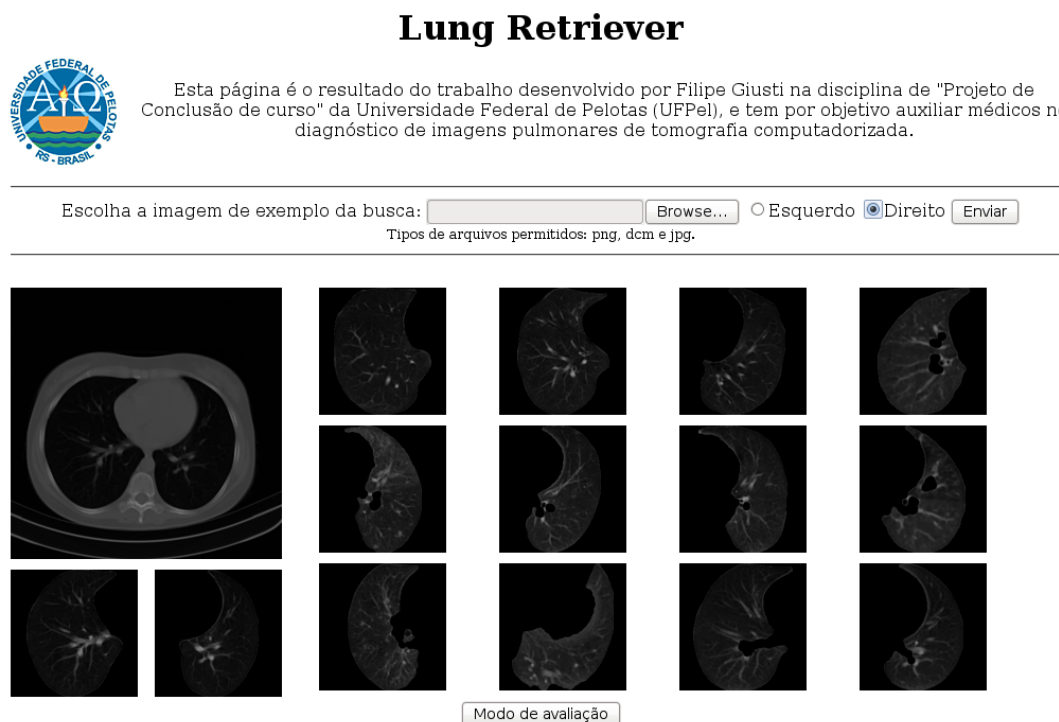



Figura 12: GUI da ferramenta desenvolvida.

Já no modo de avaliação o usuário irá ajudar a aperfeiçoar o sistema. Ele pode avaliar a relevância das imagens e submeter essa avaliação. Assim como também pode avaliar a segmentação feita na imagem enviada, mas que servirá apenas para ajudar a reavaliar o algoritmo de segmentação (fig. 13).

Lung Retriever




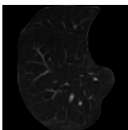
Esta página é o resultado do trabalho desenvolvido por Filipe Giusti na disciplina de "Projeto de Conclusão de curso" da Universidade Federal de Pelotas (UFPel), e tem por objetivo auxiliar médicos no diagnóstico de imagens pulmonares de tomografia computadorizada.

Escolha a imagem de exemplo da busca:

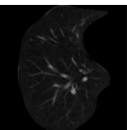
☐ Esquerdo ☒ Direito

Tipos de arquivos permitidos: png, dcm e jpg.

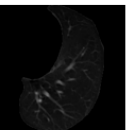




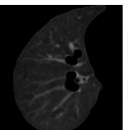
☒☒



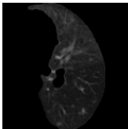
☒☒



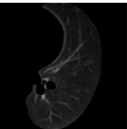
☒☒



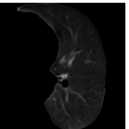
☒☒



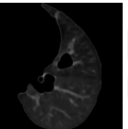
☒☒



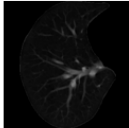
☒☒



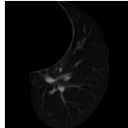
☒☒



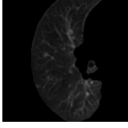
☒☒



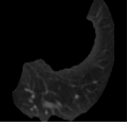
☒☒



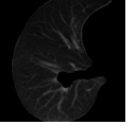
☒☒



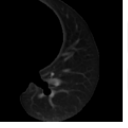
☒☒



☒☒



☒☒



☒☒

Figura 13: GUI da ferramenta desenvolvida com o modo de avaliação ativado.

5 RESULTADOS

Utilizando imagens de exames cedidas pelo Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo foi possível desenvolver algoritmos capazes de segmentar essas imagens e realizar o treinamento da RNA.

Quanto a segmentação dos pulmões foram observados poucos casos em que o resultado não pode ser usada na etapa seguinte, e geralmente pelas imagens estarem em condições extremas.

Na etapa de recuperação das imagens, a RNA conseguiu identificar as imagens utilizadas no conjunto de treinamento, mas por causa do pequeno conjunto de imagens de cada doença (várias possuíam só uma ocorrência) não foi possível a realização da etapa de teste para medir o erro da RNA.

Em ambas as etapas não foi possível ter acesso a um especialista para que uma avaliação do resultado final fosse feita.

5.1 Algoritmos

Grande parte da construção da ferramenta foi constituída pelo desenvolvimento de algoritmos que desempenham as funções propostas. A seguir são apresentados, de modo simplificado, cada um dos algoritmos que obtiveram os melhores resultados.

5.1.1 Segmentação de imagens de tomografia computadorizada dos pulmões

O objetivo deste algoritmo é gerar a partir de uma fatia da imagem de tomografia computadorizada, duas imagens contendo cada uma apenas um dos pulmões com as partes que não pertencem a eles pretas.

São 5 as etapas desse algoritmo:

1. Realizar um threshold adaptativo, como explicado na subseção 2.2.1.
2. Subtrair a região que não pertence ao corpo do paciente.

3. Eliminar pequenas regiões erroneamente marcadas como pulmão pelo threshold.
4. Eliminar buracos dentro do pulmão e juntar pedaços que possam ter ficado separados.
5. Identificar cada um dos pulmões e se necessário separa-los.

Para subtrair a região que não pertence ao corpo do paciente, são removidos todos os pixels que estão conectados e possuem o mesmo nível de cinza dos pixels da borda da imagem. Como demonstrado na fig. 14.

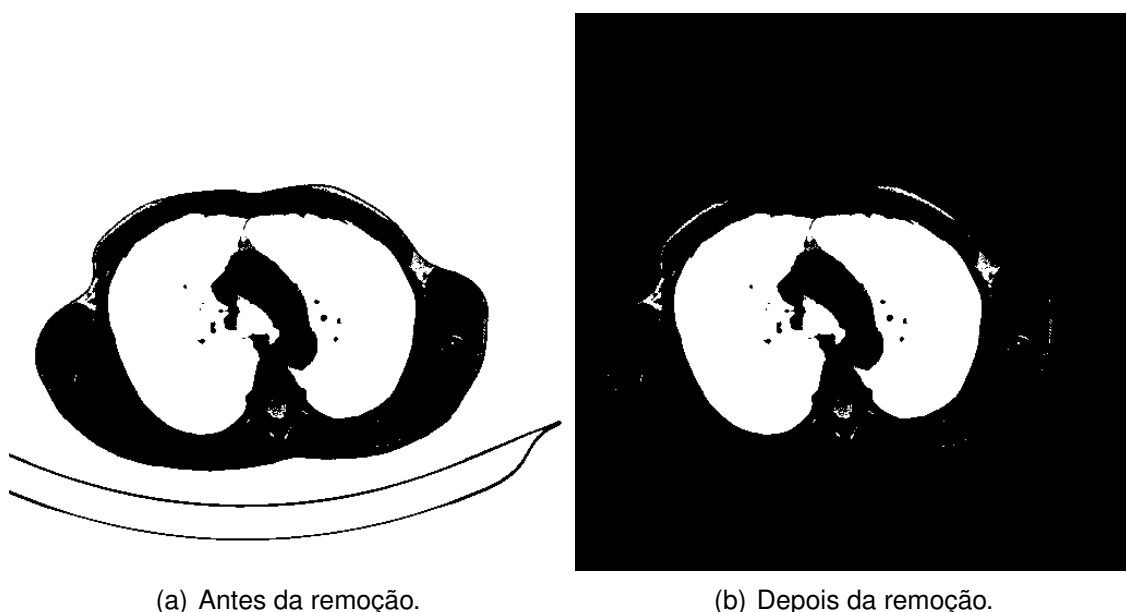


Figura 14: Imagem de TC antes e depois da remoção das partes que não pertencem ao corpo.

A eliminação das regiões que não fazem parte do pulmão é feita analisando-se todos os pixels que representam o pulmão (brancos) na imagem, de maneira que se a soma dos pixels vizinhos que representam o fundo (pretos) exceder em 2 ou mais a soma dos pixels vizinhos que representam o pulmão, então esse pixel se tornará parte do fundo. São considerados pixels vizinhos aqueles que se encontram no quadrado de 5x5 tendo como centro o pixel analisado.

Esse procedimento é realizado até 40 vezes, ou seja, ele será executado uma vez, se algum pixel for alterado, ele será executado novamente, até que nenhum pixel seja alterado ou que ele já tenha se repetido 40 vezes, podemos ver um exemplo na fig. 15.

Para remover os buracos dentro do pulmão e juntar pedaços separados de um mesmo pulmão, é realizado uma operação morfológica de fechamento utilizando

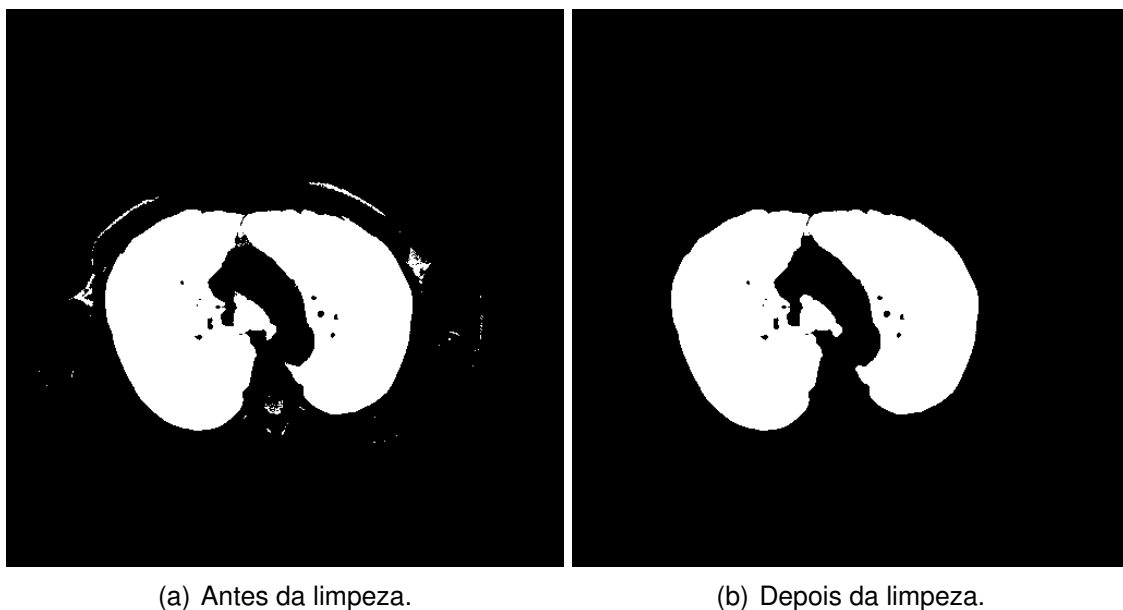


Figura 15: Imagem de TC antes e depois da eliminação das regiões que não fazem parte do pulmão.

como elemento estruturante um círculo de raio 6 pixels, como podemos ver na fig. 16. Essa operação pode causar com que pulmões muito próximos se juntem, tornando a próxima etapa ainda mais importante.

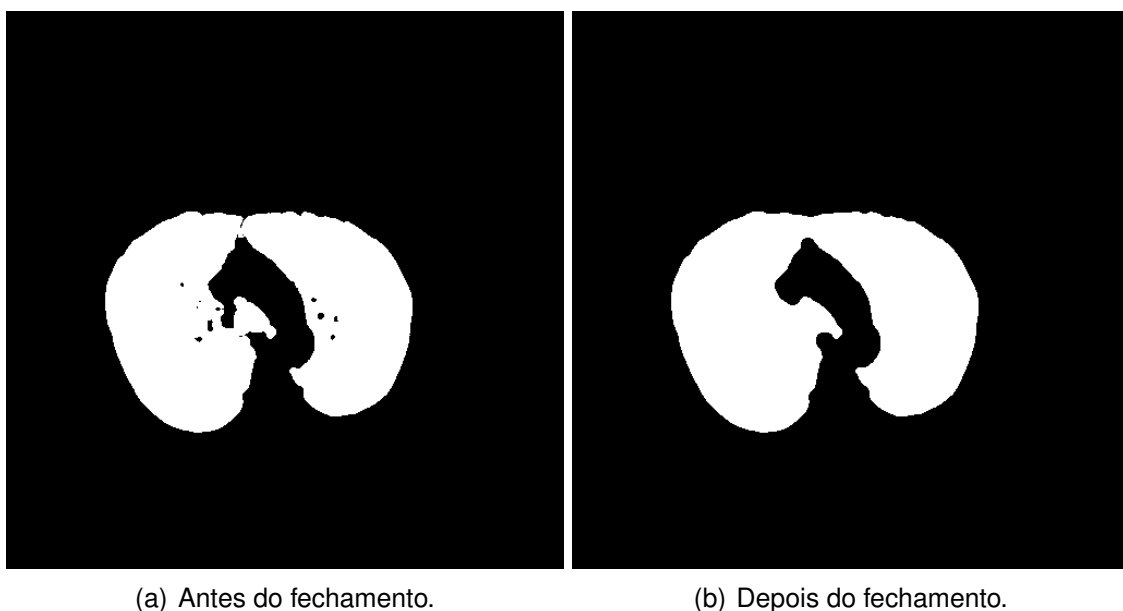


Figura 16: Imagem de TC antes e depois da operação morfológica de fechamento utilizando um círculo de raio 6 pixels.

Para identificar os pulmões primeiro verifica-se quantas regiões conectadas existem, se existir apenas uma, realiza-se uma busca a partir do centro da

imagem pela coluna que possuir menos pixels que representam o pulmão, sendo que a busca para quando a quantidade desses pixels começa a aumentar. Com essa estratégia paramos o algoritmo antes de encontrarmos estruturas pulmonares menores. Podemos ver um exemplo dessa etapa na fig. 17.

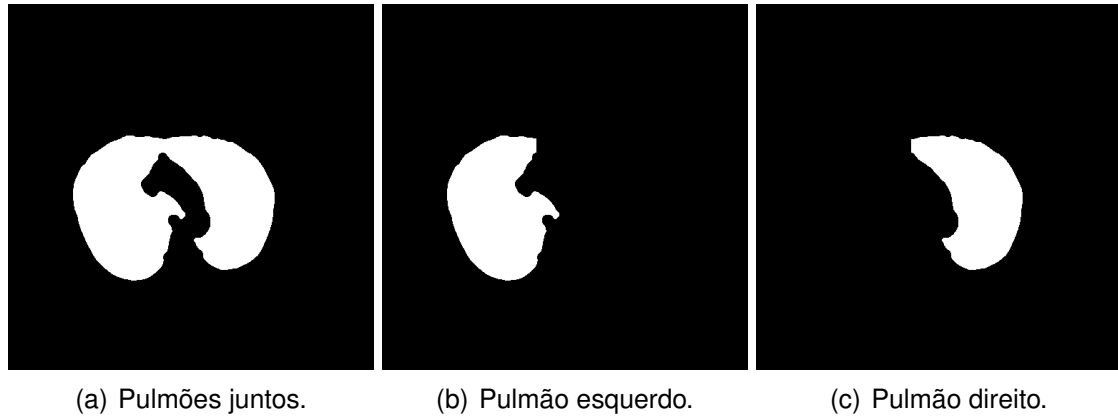


Figura 17: Pulmões unidos em 17(a), e depois devidamente separados e identificados em 17(b) e 17(c).

5.1.2 Recuperação de imagens baseada em conteúdo

Primeiramente foi necessário escolher as características a serem utilizadas para o cálculo de similaridade, as quais foram energia, entropia, correlação, contraste e momento diferencial inverso.

Com isso é preciso que a RNA tenha 10 nodos de entrada, 5 para cada vetor de características. E como a saída da RNA é o valor de similaridade, ela possui apenas um nodo de saída. Para que tenhamos um valor entre 0 e 1 na saída, foi utilizada a seguinte função sigmóide:

$$y = \frac{1}{1 + e^{-x}} \quad (17)$$

A qual é derivada na equação 18, que tem custo computacional baixo, pois como utilizamos para aprendizado o algoritmo de retro-propagação, o valor de y não precisa ser calculado novamente na atualização dos pesos.

$$y \cdot (1 - y) \quad (18)$$

A RNA foi dimensionada com 2 camadas intermediárias, sendo que na primeira camada temos 11 nodos e na segunda camada 13 nodos.

A taxa de aprendizado escolhida para o algoritmo de retro-propagação é 0,2. O valor de o_j , que é o valor de saída esperado, na equação 12 é 1 se o usuário considerar as imagens comparadas similares, 0 se considerar diferentes e 0,5 se

não for considerada igual nem diferente.

O treinamento é realizado com um conjunto de imagens de 20 pacientes. Que é repetido até 1000 vezes ou até que o erro médio do conjunto de imagens seja menor que 0,001. Alguns pares no conjunto de imagens de treinamento foram montados de modo a fazer com que a RNA seja associativa entre duas imagens, ou seja, se a similaridade de uma imagem x com uma imagem y for z , então a similaridade entre a imagem y e a imagem x também deve ser z . Também foram feitos pares no conjunto de imagens de treinamento com a mesma imagem, pois nesse caso o resultado deve ser, idealmente, 1.

6 CONCLUSÕES

Este trabalho propôs o desenvolvimento de uma ferramenta para auxiliar no diagnóstico de imagens pulmonares de tomografia computadorizada, através da extração de características e recuperação de imagens similares ao exame que está sendo diagnosticado.

Atualmente, está ocorrendo uma expansão na aplicação do processamento de imagens à medicina. Os centros de tratamento médico cada vez mais estão utilizando softwares precisos e confiáveis na análise e diagnóstico de diversas doenças. Como a quantidade de dados coletados por um exame de tomografia geralmente é grande, a utilização de técnicas de processamento de imagens para facilitar a visualização e análise automatizada das informações tem se tornado essencial. É importante ressaltar que este trabalho é o início de um trabalho maior, pois outras características podem ser testadas, além de uma necessidade de aumentar a base de dados, para conseguirmos mais casos de todos os tipos utilizados.

6.1 Dificuldades encontradas

Para a segmentação a falta de documentação e alta complexidade do ITK, associado a quantidade de detalhes do padrão DICOM não implementado por muitas ferramentas, até mesmo no ITK, fizeram com que o tempo de aperfeiçoamento do algoritmo fosse diminuído.

Além disso a dificuldade em encontrar especialistas dispostos a ajudar na avaliação do trabalho não permitiu a validação clínica do mesmo.

6.2 Trabalhos Futuros

Este trabalho pode seguir sendo aperfeiçoado mesmo após sua conclusão. As seguintes metas podem ser destacadas:

- Validar a ferramenta com um profissional a área médica;
- Acesso ao laudo das imagens retornadas da busca;
- A portabilidade da ferramenta para outros sistemas operacionais;
- Utilizar outro conjuntos de características, por exemplo, a técnica de wavelets;
- Modularizar a ferramenta de forma a permitir a integração com aplicativos desenvolvidos para segmentação de outros tipos de imagens;
- Indicar nas imagens os possíveis locais de patologias;
- Utilizar outras topologias de RNA.

REFERÊNCIAS

ACHARYA, T.; RAY, A. K. **Image Processing - Principles and Applications**. [S.l.]: Wiley-Interscience, 2005.

BICK, U.; DOI, K. **Tutorial on Computer-aided Diagnosis**. [S.l.: s.n.], 2006.

COSTA LEITE, C. da; JÚNIOR, E. A.; OTADUY, M. G. **Física básica da tomografia computadorizada**. 2004.

DOI, K. Current status and future potential of computer-aided diagnosis in medical imaging. **Br J Radiol**, [S.l.], v.78, p.3–19, 2005.

EL-NAQA, I.; YANG, Y.; GALATSANOS, N. P.; NISHIKAWA, R. M.; WERNICK, M. N. A similarity learning approach to content-based image retrieval: application to digital mammography. **Medical Imaging, IEEE Transactions on**, [S.l.], v.23, n.10, p.1233–1244, 2004.

FACON, J. **Morfologia Matemática**. 2000.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 2.ed. [S.l.]: Prentice Hall, 2002.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural Features for Image Classification. **Systems, Man and Cybernetics, IEEE Transactions on**, [S.l.], v.3, n.6, p.610–621, 1973.

KITWARE. **National Library of Medicine Insight Segmentation and Registration Toolkit Home Page**. [S.l.]: Kitware Inc, 2008.

LIM, J.-H.; WU, J. K.; SINGH, S.; NARASIMHALU, D. Learning Similarity Matching in Multimedia Content-Based Retrieval. **IEEE Trans. on Knowl. and Data Eng.**, Piscataway, NJ, USA, v.13, n.5, p.846–850, 2001.

LONG, F.; ZHANG, H.; SIU, W.-C.; FENG, D. D. **Multimedia Information Retrieval and Management: technological fundamentals and applications**. Berlin: Springer, 2003. (Technological Fundamentals and Applications).

NEMA. **NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION**: dicom homepage. 2008.

OLIVEIRA, K. **Fundamentos de Radiodiagnóstico por Imagem**. Porto Alegre: UFRGS, 1999.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence**: a modern approach. [S.l.]: Pearson Education, 2003.

SAMARASINGHE, S. **Neural Networks for Applied Sciences and Engineering**. Boston, MA, USA: Auerbach Publications, 2006.

SHULKIN, B. L. **Overview of Body CT Scanning in Children**. 2005.

SPRAWLS, P. **Computed Tomography Image Formation**. 2004.

STROUSTRUP, B. **The C++ Programming Language**. 3.ed. [S.l.]: Addison-Wesley Professional, 2000.

UCHIYAMA, Y.; KATSURAGAWA, S.; ABE, H.; SHIRAISHI, J.; LI, F.; LI, Q.; ZHANG, C.-T.; SUZUKI, K.; DOI, K. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. **Medical Physics**, [S.l.], v.30, n.9, p.2440–2454, 2003.

YOO, T. S.; ACKERMAN, M. J.; LORENSEN, W. E. Engineering and algorithm design for an image processing API: a technical report on itk-the insight toolkit. **Proc. of Medicine Meets Virtual Reality**, [S.l.], p.586–592, 2002.