

Identificação de Massas Suspeitas em Mamografias Usando Processamento de Imagens

Pré processamento de imagens e detecção de áreas de interesse

Guilherme Rodrigues Lodron Pires

Departamento de Ciências da Computação (CiC)

Universidade de Brasília (UnB)

Brasília, Brasil

guilodron7@gmail.com

Abstract—Breast cancer is one of the most common kinds of cancer that affects women all around the world. In Brazil, it is the first cause of death by cancer on the feminine population in almost every region. According to the website of the Nacional Institute of Cancer, case rates go up to 18 per 100 thousand women. Having that in mind, early anomaly detection in routine exams is crucial so that the correct treatment is applied. Digital mammography is one of the main methods to identify foreign masses or nodules, although, sometimes the these kind of images are not as clear as expected. To soften this issue there are image processing algorithms that enhance the visibility of these exams to meet medical expectations. This article has the purpose of exploring some image processing methods and its uses and benefits on digital medical images.

Index Terms—image processing, watershed, morphological, report

I. INTRODUÇÃO

O cancer de mama é um dos tipos mais comuns de cancer que afetam mulherem em todo o mundo. No Brasil ele é a primeira causa de morte por cancer na população feminina em quase todas as regiões do país. Segundo o site do instituto nacional do câncer, as taxas de casos no país pode chegar em até 18 para cada 100 mil mulheres.

Tendo em vista esse cenário é importante a detecção de anomalias em exames de rotina o mais cedo possível, para que os devidos tratamentos sejam ministrados. A mamografia digital é um dos principais métodos para identificar massas ou nódulos estranhos, no entanto, muitas vezes a imagem obtida pode não ser tão clara à olho nu. Para amenizar esse problema existem algoritmos de processamento de imagens que ajudam na melhoria da qualidade subjetiva desse tipo de imagem.

Este artigo tem como objetivo abordar alguns métodos de processamento digital de imagens e como podem ser utilizados para o processamento de imagens médicas de mamogramas.

II. TÉCNICAS UTILIZADAS

A. Binarização

Para facilitar operações ou retirar informação indesejada de imagens, uma opção simples é a binarização. Nela é estabelecido um valor limite, chamado threshold, a partir do qual para todos os pixels com valor acima desse marca é

atribuídos o valor 255 (branco) e aos demais é atribuído o valor 0 (preto). Dessa forma resultando em uma imagem composta apenas por preto e branco.

B. Power law ou Correção gama

O realce Power Law ou correção Gamma é uma técnica de processamento de imagens que busca alterar o contraste de uma imagem em partes mais claras ou em partes mais escuras, a depender do valor de gama fornecido para a curva. A equação que descreve essa correção é:

$$s = cr^\gamma \quad (1)$$

Os valores de gama aplicados influenciam diretamente na imagem resultante. Com valores menores do que um aumentando o contraste em áreas escuras e valores maiores que um aumentando o contraste em áreas mais claras.

C. Algoritmos morfológicos

Os algoritmos morfológicos para processamento de imagens utilizam conceitos matemáticos da teoria dos conjuntos para oferecer ferramentas poderosas para problemas de processamento de imagens. Neles são realizadas operações de translação e reflexão sobre um pequeno conjunto ou subimagem, que é denominado elemento estruturante para extrair ou descartar áreas de interesse. Algumas das operações base para esses algoritmos são:

a) *Erosão*: mantém na imagem todos os conjuntos de pixels que têm tamanho suficiente para conter o elemento estruturante utilizado. Com isso o elemento geralmente é afinado ou diminuído. Esse operação é descrita pela seguinte fórmula:

$$A \ominus B = \{z \mid B \subseteq A\} \quad (2)$$

b) *Dilatação*: mantém na imagem todos os deslocamentos de pixels de forma que possuam pelo menos alguma interseção com o elemento estruturante. Ao contrário da erosão o elemento tende a engrossar ou aumentar de tamanho. A fórmula que descreve essa transformação é:

$$A \oplus B = \{z \mid B \cap A \neq \emptyset\} \quad (3)$$

A partir dessas duas operações básicas é possível derivar a operação de fechamento:

c) *Fechamento*: o fechamento geralmente funde as descontinuidades estreitas e alonga os golfos finos, elimina pequenos buracos e preenche as lacunas em um contorno. Essa operação é feita pela dilatação seguida da erosão.

$$A \bullet B = (A \oplus B) \ominus B \quad (4)$$

d) *Abertura*: a abertura geralmente suaviza o contorno de um objeto, rompe os istmos e elimina as saliências finas. Essa operação é feita pela erosão seguida pela dilatação:

$$A \circ B = (A \ominus B) \oplus B \quad (5)$$

As operações citadas acima são feitas em cima de imagens binárias, porém também existem implementações para o seu uso em imagens em escala de cinza.

D. Segmentação da imagem com K-means

Segmentação de imagens é o processo de particionar a imagem em múltiplos segmentos. Seu objetivo é dividir a imagem em partes que sejam mais fáceis ou mais diretas de se analisar. Comumente é utilizado para detectar bordas ou objetos, que é o caso deste artigo.

K-means é um algoritmo de aprendizado não supervisionado utilizado para agrupar dados em um número definido de grupos, chamados clusters, baseado em características em comum. Dado que nosso conjunto de dados são pixels, eles são agrupados de acordo com sua similaridade em intensidade de brilho.

O algoritmo segue os seguintes passos:

- 1 - Escolher o número de clusters;
- 2 - Atribuir de forma aleatória pontos para cada cluster;
- 3 - Calcular o centro dos agrupamentos;
- 4 - Calcular a distância dos pontos até o centro do cluster;
- 5 - De acordo com as distâncias calculadas, reatribuir os pontos para o cluster mais próximo;
- 6 - Calcular o novo centro do cluster;
- 7 - Repetir os passos até que nenhum ponto mude de grupo.

III. IMAGENS UTILIZADAS PARA O PROJETO

Para as análises, foram utilizadas imagens disponíveis na base de dados de mamogramas MIAS (Mammographic Image Analysis Society). Nela as imagens possuem dimensão de 1024 x 1024 pixels e têm um diagnóstico e tipo de anomalia categorizados por profissionais da saúde. As imagens seguem o formato que pode ser observado em 4.

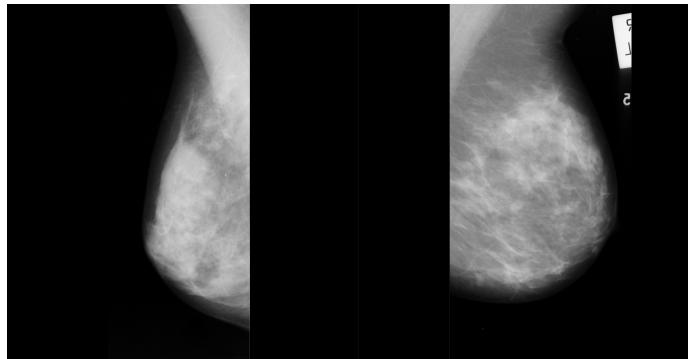


Fig. 1. Imagens extraídas da database MIAS

IV. ALGORITMO PROPOSTO

A. Retirando artefatos

Como pode ser observado na imagem 4, algumas imagens possuem marcações com grande intensidade de brilho, elas serão chamadas de artefatos durante o artigo. Para evitar futuras complicações com a distinção de formas, foi proposto uma algoritmo para remover os artefatos da imagem como descrito a seguir.

A fim de facilitar a detecção da mama e do artefato, primeiro é realizada uma binarização, com limiar escolhido de maneira automática pela biblioteca OpenCV. Em seguida é realizado uma abertura e um fechamento, a fim de remover possível ruído e suavizar contornos. Os resultados obtidos foram:

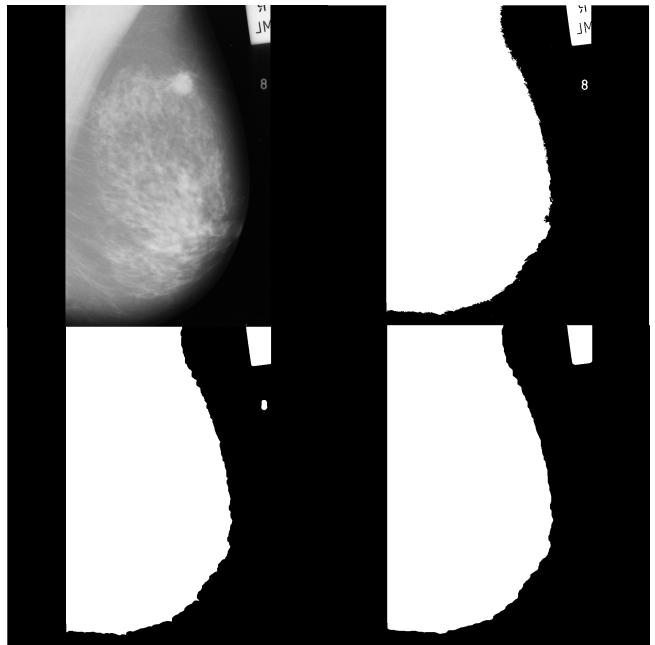


Fig. 2. 1- Imagem original 2- Imagem binarizada 3- Operação de fechamento 4- Operação de abertura

Após essas transformações, a imagem resultante possui apenas as formas da mama e dos possíveis artefatos. Para retirar o artefato, o algoritmo escolhe a maior forma da imagem e descarta as demais. Com isso é realizada uma dilatação, para

recuperar possíveis áreas perdidas no processo de binarização e a imagem resultante é utilizada como máscara para a imagem original, que agora não possui mais artefatos.

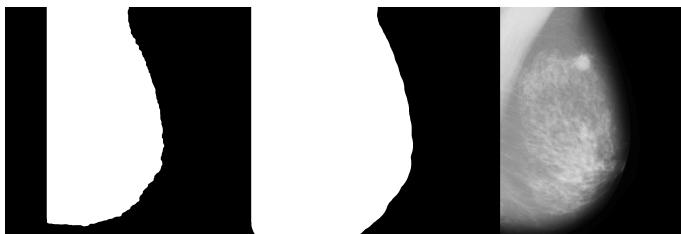


Fig. 3. 1- Maior forma selecionada 2- Dilatação 3- Imagem original após máscara

B. Espelhando imagem

Pelo fato de mamogramas serem realizados nas duas mamas, as imagens da base de dados podem possuir diferentes orientações, como pode ser observado nas imagens em fig 4. Entretanto, para os passos seguintes do algoritmo é necessário que todas as imagens tenham a mesma orientação, no caso, apontando para a direita.

Para detectar se a imagem está na orientação incorreta é obtido o contorno da mama e o mesmo é 'pintado' em branco sobre a imagem. Em seguida o algoritmo faz uma varredura da esquerda para a direita até encontrar o primeiro pixel branco. Nesse momento checa-se se o pixel encontrado faz parte de uma linha reta vertical, caso não faça, a imagem é espelhada.

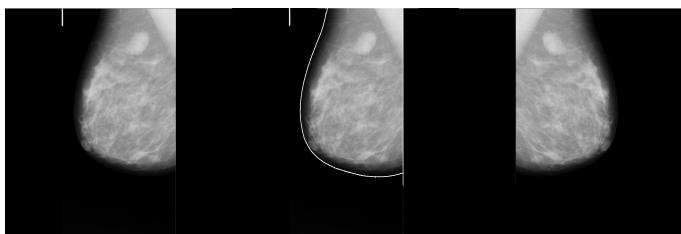


Fig. 4. 1- Imagem inicial 2- Borda encontrada 3- Imagem espelhada

C. Removendo músculo peitoral

O músculo peitoral comumente é observado nas imagens de mamografias, nas imagens utilizadas ele pode ser visto no canto superior direito ou esquerdo. Por ser um músculo, sua densidade é consideravelmente maior que o restante da mama. Então, sua região possui uma intensidade de brilho alta, o que pode dificultar no reconhecimento massas estranhas. Para tratar esse problema o desafio foi retirar o músculo peitoral da imagem sem perder áreas relevantes para o estudo.

Segundo o passo descrito na subseção anterior, é garantido que a imagem terá orientação da esquerda para a direita, dessa forma, é possível afirmar que o músculo peitoral sempre irá se localizar no canto superior esquerdo da imagem. Tendo isso em mente o algoritmo proposto tem como primeiro passo calcular a média de intensidade de um exerto de 50×50 pixels a partir do canto escolhido. Esse valor, com uma margem de 20%, foi utilizado de limiar para uma binarização, que divide

a imagem dando preferência para o destaque do músculo peitoral.

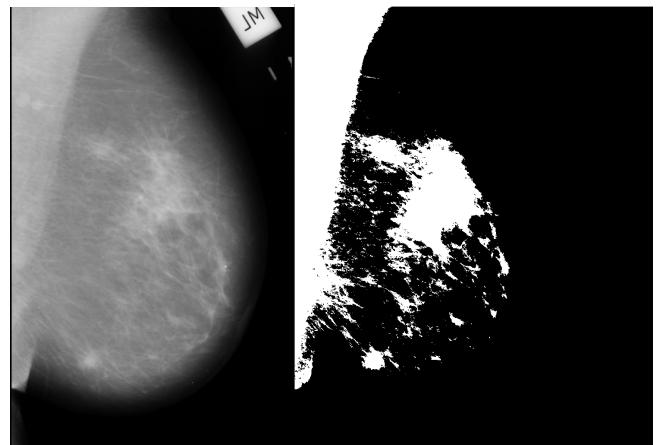


Fig. 5. 1- Imagem inicial 2- Imagem binarizada de acordo com limiar calculado

Como observado na fig 5, partes da mama que sobraram na binarização ainda podem ter conexão com a área de interesse. Para eliminar essa conexão é aplicada uma erosão, e em seguida os menores contornos são eliminados. Ademais é feita uma dilatação no contorno remanescente para recuperar informações perdidas na erosão e por fim o formato encontrado é utilizado de máscara para remover o músculo peitoral da imagem original.



Fig. 6. 1- Erosão 2- Dilatação da área escolhida 3- Imagem original com máscara obtida

V. ENCONTRANDO MASSAS COM K-MEANS

Após a remoção de todos os elementos indesejáveis da imagem, o último passo é identificar massas na imagem. Inicialmente é aplicada uma correção gama com fator de 3.5, para aumentar o contraste nas áreas claras da imagem, que representam as possíveis áreas de interesse.

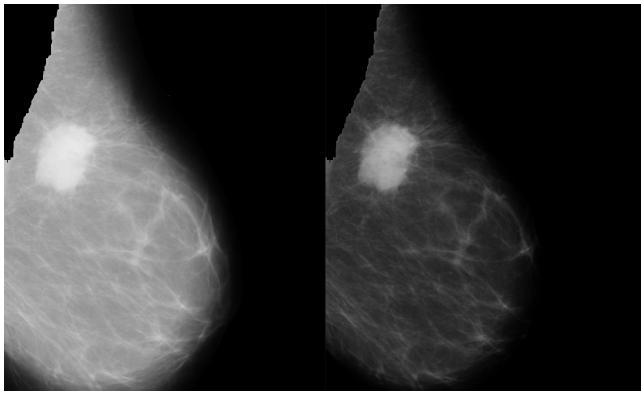


Fig. 7. 1- Imagem inicial 2- Power-law com fator 3.5

Em seguida é executado o método de agrupamento por k-means com o parâmetro de 5 clusters e 15 tentativas. Esse método retorna uma imagem com os grupos encontrados destacados em escalas de cinza. Para selecionar as áreas de interesse, é selecionado o grupo que possuir maior nível de brilho e todos os outros são retirados da imagem.

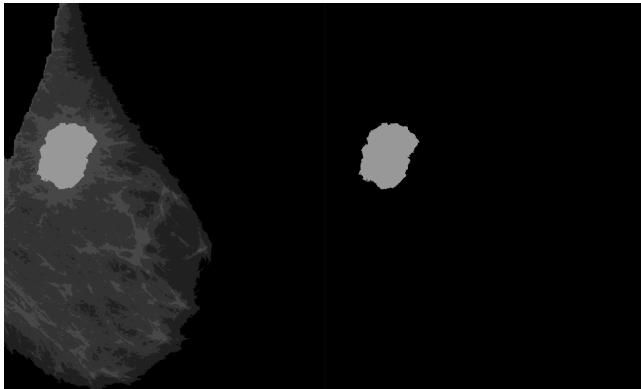


Fig. 8. 1- Clusters encontrados 2- Cluster de maior brilho

Ao final o grupo escolhido é utilizado para contornar as massas encontradas na imagem.

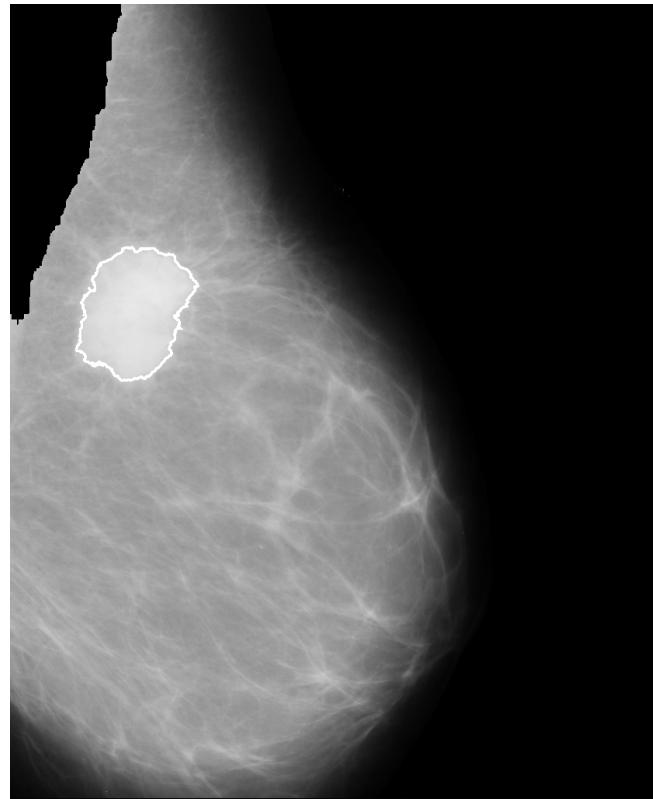


Fig. 9. Resultado final do algoritmo proposto

A. Resultados e análises

O algoritmo proposto foi testado em diversas imagens do dataset MIAS e foram obtidos resultados satisfatórios quanto à lógica de remoção de artefatos e músculo peitoral, chegando a uma acurácia de aproximadamente 60%. Já na segmentação de massas de interesse o algoritmo não se provou tão eficaz à cenários que fugiram muito do ótimo. Isso se dá pela grande variedade na anatomia da mama e na densidade de outros tecidos captados pelo exame. Fato que resultou por vezes na marcação de áreas sem relevância ou no esquecimento de áreas de possível interesse.

Na tabela I estão alguns dos resultados que se comportaram bem ao algoritmo proposto.

VI. CONCLUSÃO

É possível concluir que o projeto teve resultados satisfatórios, uma vez que foi capaz de detectar e realçar anomalias de forma automática em uma grande quantidade de imagens testadas. O algoritmo proposto se adequou bem em cenários diferentes até um certo limite, e por sua vez pode ser passível de melhorias no futuro.

Ao desenvolver o trabalho pode-se perceber os desafios que ainda existem na área de processamento de imagens médicas, e como pequenos detalhes de anatomia podem refletir em uma mudança no resultado final. Os algoritmos desenvolvidos para essa finalidade devem ser duramente testados e criticados antes que possam ter alguma responsabilidade no diagnóstico de pacientes. Dito isso é animador ver o número grande de

TABLE I
RESULTADOS

Imagen	Original	Remoção de Artefato	Espelhamento	Remoção de músculo	K-means	Imagen Final
25						
28						
132						
184						
206						

pesquisas relacionadas à detecção de tumores e a grande gama de ferramentas disponíveis para o processamento desse tipo de imagem.

REFERENCES

- [1] R.C. Gonzalez, R.E. Woods, "Digital Image Processing" 4th edition
- [2] R. Subash Chandra Boss, K. Thangavel, D. Arul Pon Daniel, Automatic Mammogram image Breast Region Extraction and Removal of Pectoral Muscle
- [3] ABHIJITH REDDY BEERAVOLU 1 , SAMI AZAM 1 , MIR-JAM JONKMAN 1 , (Member, IEEE), BHARANIDHARAN SHANMUGAM1 , (Member, IEEE), KRISHNAN KANNOORPATTI 1 , AND ADNAN ANWAR 2, Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN.
- [4] Waleed A. Yousefa, , Ahmed A. Abouelkahireb,1, Deyaaeldeen Almalhallawic,1, Omar S.Marzoukd,1, Sameh K. Mohamede,1 , Waleed A. Mustafaf,1, Omar M. Osamag,1, Ali A. Salehh,1, Naglaa M. Abdelrazeki, Method and System for Image Analysis to Detect Cancer.
- [5] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du, Approaches for automated detection and classification of masses in mammograms.
- [6] Etta D. Pisano, MD, Image Processing Algorithms for Digital Mammography: A Pictorial Essay.
- [7] MIAS database, disponível em <http://peipa.essex.ac.uk/info/mias.html>;
- [8] Mônica de Lourdes Souza Batista1 , Aura Conci1 , Leonardo Motta1 , Sérgio Muinhos Barroso Lima2 , Patrícia Lima Quintão2, Processamento digital de imagens para a detecção e classificação de nódulos em mamografias.
- [9] Documentação OpenCV, disponível em <https://opencv.org/>. Acesso em 18/10/2021