

DETECÇÃO AUTOMÁTICA DE TUMORES EM MAMOGRAFIAS UTILIZANDO TÉCNICAS DE PROCESSAMENTO DIGITAL DE IMAGEM

T. C. Silva^{1,2}, C. F. Boos², D. S. Medeiros¹, E. M. Lobato¹ e F. M. de Azevedo²

¹ Instituto Federal de Santa Catarina, São José, Brasil

² Instituto de Engenharia Biomédica da UFSC, Florianópolis, Brasil
E-mail: thaysechristinesilva@gmail.com

Resumo: O câncer de mama é o segundo tipo de câncer mais frequente no mundo e o primeiro entre as mulheres. Os melhores exames para diagnóstico prematuro deste tipo de câncer são primeiramente a mamografia e depois o autoexame. O presente trabalho tem como objetivo propor uma metodologia que, usando técnicas de processamento digital de imagens, possa auxiliar no diagnóstico desses tumores em imagens de mamografias. A metodologia proposta consiste na aplicação de filtros, modificações de contraste e realização de sucessivas decomposições de uma imagem mamográfica de forma a permitir uma melhor visualização e localização de possíveis tumores. Espera-se que este processamento permita que possíveis nódulos de câncer no mamograma sejam ressaltados, auxiliando o médico a realizar um diagnóstico mais preciso e rápido, possibilitando um melhor prognóstico ao paciente e encaminhamento ao tratamento adequado. A metodologia proposta mostra-se promissora considerando os resultados obtidos, quais sejam sensibilidade e especificidade de 86% e 75%, respectivamente, e 84% de eficiência na detecção dos tumores.

Palavras-chave: Mamografia, Processamento Digital de Imagem, Tumores.

Abstract: *Breast cancer is the second most common cancer in the world and first among women. The best tests for early diagnosis of this cancer are first mammography and then self-examination. The present study aims to propose a methodology that, using techniques of digital image processing, can facilitate the diagnosis of tumors in mammography images. The proposed methodology consists of applying filters, contrast changes and performing successive decompositions of a mammographic image to allow a better visualization and localization of potential tumors. It is expected that this process allows potential cancer nodules in the mammogram to be highlighted, helping the physician to perform a more accurate and rapid diagnosis, enabling better patient prognosis and referral to appropriate treatment. The proposed methodology shows itself promising considering the results obtained, namely, sensitivity and specificity of 86% and 75%, respectively, and 84% efficiency in the detection of tumors.*

Keywords: *Mammography, Digital Image Processing, Tumors.*

Introdução

A área de processamento digital de imagens tem apresentado um expressivo desenvolvimento nas últimas décadas. Nesta área vem sendo desenvolvidos diversos estudos, a fim de auxiliar na resolução de problemas dos mais diferentes domínios do conhecimento humano [1].

Para auxiliar no diagnóstico e tratamento das mais diversas doenças, tem sido cada vez mais presente o uso de computadores ou microcontroladores. Estes dispositivos vêm munidos de sofisticados algoritmos especializados em processamento de sinais e/ou imagens, facilitando o diagnóstico e tratamento de determinada enfermidade. Particularmente, técnicas de processamento de imagens são utilizadas para auxiliar o diagnóstico de diversas doenças, através da análise computacional das imagens, podendo ser utilizado paralelamente técnicas de inteligência artificial para identificação de padrões. Estas técnicas oferecem um diagnóstico automatizado ao profissional médico, que terá a mais nobre missão de analisar e repassar as informações ao paciente.

Para o diagnóstico do câncer de mama, a mamografia é um exame amplamente utilizado, tendo uma grande importância para detecção e identificação precoce de possíveis tumores, permitindo que seja feito o tratamento prévio do paciente melhorando, portanto, o prognóstico da doença [2].

Este trabalho tem como objetivo principal a proposição de uma metodologia para a detecção de tumores em mamografias baseada nas metodologias de Borges [3], Nishikawa *et al.* [4] e Dengler, Behrens e Desaga [5]. Para tanto foi implementado, em MatLab®, um script que, a partir das imagens fornecidas pelo mamógrafo, detecte a ocorrência de nódulos que dão origem ao câncer de mama.

Materiais e métodos

A metodologia proposta, conforme ilustra a Figura 1, foi dividida em diversas etapas de processamento digital de imagens que estão detalhadas nos itens a seguir.

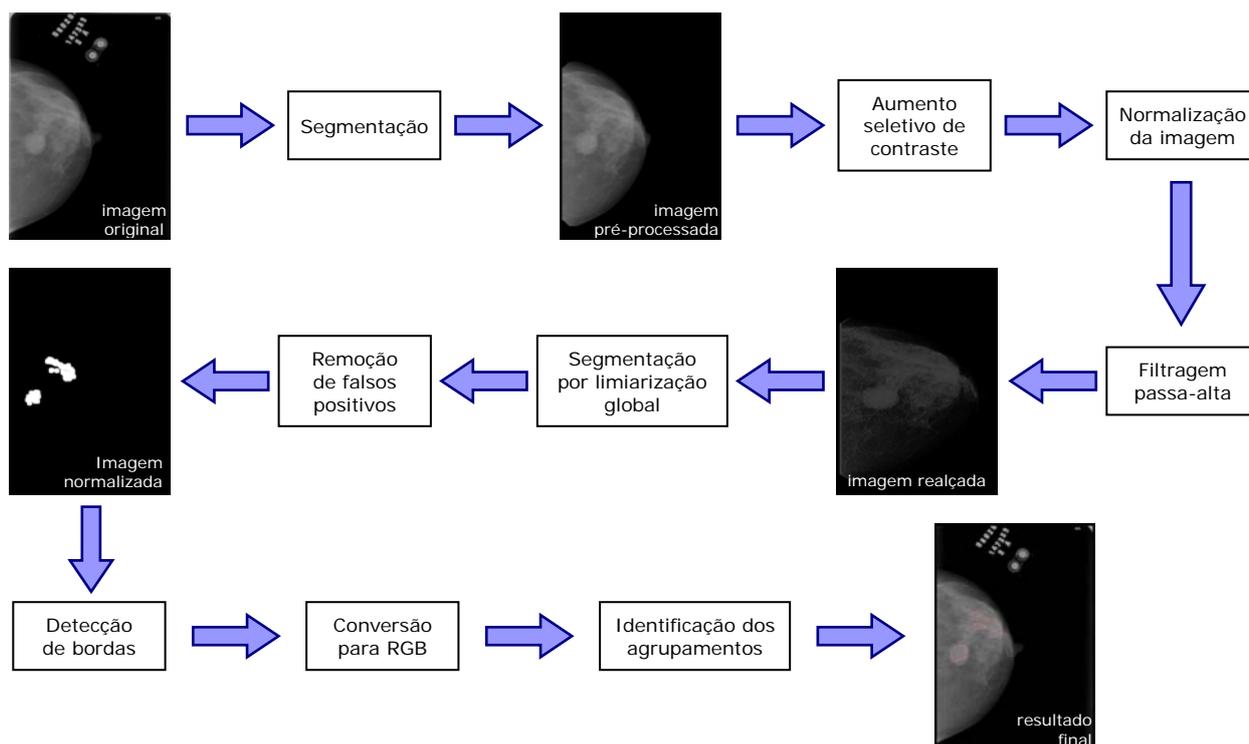


Figura 1: Diagrama de desenvolvimento da proposta, na qual a imagem original passa pelos processos de segmentação, aumento seletivo de contraste, normalização, filtragem para gerar uma imagem realçada. Esta imagem é segmentada, são detectados e removidos possíveis falsos positivos, e é feita a detecção das bordas do tumor. Após finalização da detecção é feito o processamento da imagem para visualização do resultado (localização do tumor).

Banco de imagens – As imagens utilizadas para a implementação da metodologia são pertencentes ao banco de dados [6] do Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO) da Universidade de São Paulo (USP). Este banco de dados contém imagens mamográficas, e que são fornecidas aos pesquisadores que possuem acesso restrito ao sistema.

Nas imagens fornecidas consta um breve laudo sobre o nódulo com características que permitem a identificação da localização do tumor.

Segmentação – Esta etapa é dividida em alguns métodos utilizados para a segmentação da imagem mamográfica. A segmentação da mama se faz necessária para separar a mama do fundo da imagem. É importante então para determinar a extensão da mama, separando-a dos possíveis ruídos que possam existir na imagem. Esses ruídos podem interferir negativamente no resultado da identificação.

Pré-processamento – Esta subetapa consiste em reduzir o tamanho da imagem original e inserir uma borda, com o objetivo de melhorar o desempenho do algoritmo, devido ao fato da dimensão para varredura de informações da imagem ser menor [7]. É importante salientar que esta redução não pode permitir a perda de parte das características da imagem, que pode ocasionar um falso diagnóstico.

É necessária a inserção de uma borda preta ao redor do mamograma, com o objetivo de retirar a curvatura

existente nesta borda, além de algum possível ruído que possa estar presente na mesma.

Limiarização Global – Para fazer a separação de onde termina o objeto e onde começa o fundo, é necessária a criação de um limiar que irá indicar essa região de separação. Para realizar a limiarização faz-se uma varredura pixel a pixel da imagem e verificando-se o valor dos pixels. Os pixels que possuem valores maiores que o limiar determinado recebem o valor um e os valores abaixo, recebem o valor zero. Neste estudo foi utilizado um limiar com valor 76.

Filtro Morfológico de Abertura – Após a limiarização é necessário separar a mama do que é considerado o fundo da imagem. Para tanto, foi utilizado um filtro morfológico de formato redondo e abertura 25 [9], pois o formato arredondado do filtro se adequa ao formato e tamanho da mama nas imagens.

Retirada do Fundo do Mamograma – Para retirar o fundo do mamograma é necessário realizar uma multiplicação ponto a ponto da imagem que havia sido reduzida e inserida a borda, com a imagem resultante do filtro morfológico de abertura. O resultado desta multiplicação é uma imagem sem ruído na região externa ao tecido mamário contendo, da imagem original, somente a mama. Todos os pixels que formam o fundo da imagem são pixels de preenchimento na cor preta.

Aumento Seletivo de Contraste – O aumento seletivo de contraste consiste em diferenciar os objetos existentes na imagem. A aplicação desta técnica visa

melhor o contraste da imagem através da equalização de histograma. Após a equalização o nódulo tende a ficar mais claro do que o tecido mamário resultante.

Normalização – A normalização de imagens permite a preparação das imagens para a execução de outras tarefas. Este processo tem como objetivo reescalonar a imagem, para que a mesma esteja com os pixels dentro do limite de [0,255].

Filtragem Passa-Alta – Como as bordas e outras mudanças abruptas de intensidade estão associadas às componentes de alta frequência, o aguçamento de imagens pode ser obtido no domínio da frequência pela filtragem passa-alta, que atenua as componentes de baixa frequência sem afetar as informações de alta frequência [8].

Com a passagem do filtro passa-alta as transições entre diferentes regiões da mama são realçadas, destacando os nódulos do fundo do tecido mamário.

O filtro utilizado para realçar esses nódulos foi um filtro linear no qual a soma de seus coeficientes é diferente de zero. Utilizando esse filtro a informação dos pixels é mantida, pois cada pixel alterado (quando o somatório dos coeficientes é entre zero e um) ou mantido original (quando o somatório dos coeficientes é igual a um). Os coeficientes utilizados foram os mesmos do filtro de [3]. O filtro está descrito na Equação 2.

$$f = [-0,65 \quad -0,40 \quad 3,00 \quad -1,00] \quad (2)$$

Segmentação via Limiarização Global – Para esta segunda limiarização, o procedimento utilizado foi o mesmo utilizado na seção Limiarização Global. Porém é utilizado um limiar de 220, obtido através testes realizados em várias imagens. Este limiar permite selecionar o agrupamento de nódulos do mamograma.

Como resultado desta limiarização percebe-se a acentuação, em branco, dos pontos onde existem agrupamentos de nódulos, porém ainda existe muito ruído resultante do tecido mamário que será removido nas próximas etapas.

Remoção de Falsos Positivos – A remoção de falsos positivos é uma etapa fundamental para a detecção dos agrupamentos de nódulos, pois nem todos os pontos existentes nas imagens são considerados nódulos reais, o que pode gerar um diagnóstico incorreto aos pacientes.

Para realizar a detecção de quais pontos podem realmente ser considerados verdadeiros ou não, foi utilizado um filtro morfológico de abertura, utilizando um círculo de valor seis como elemento estruturante. Com a utilização deste filtro, somente os possíveis nódulos com área menor que o limiar permaneceram na imagem.

Detecção de Bordas – Esta etapa consiste em detectar as bordas dos nódulos indicados após o processo de remoção dos falso-positivos. A borda deve ser identificada, pois na imagem final é esperado apenas um contorno do nódulo, para que indique onde ele se localiza e qual é o seu tamanho. Caso não fosse feita a detecção das bordas dos nódulos, ficaria na imagem

apenas um borrão preto no lugar onde deveria conter a imagem do nódulo.

A fim de fazer a detecção das regiões consideradas bordas dos nódulos, foi necessário utilizar detectores de bordas já existentes. Então, para realizar a detecção da borda, foi utilizado o detector de Canny, que recebe uma imagem em tons de cinza ou binária, e retorna uma imagem binária do mesmo tamanho da imagem de entrada, inserindo “1” onde é borda e “0” para o restante da imagem.

Conversão para RGB – Para sinalizar a localização dos agrupamentos de nódulos, é necessário transformar a imagem vinda do filtro de Canny em uma imagem colorida, para que realmente essa sinalização apareça na imagem original. Então foi feita a conversão das imagens, pré-processada e somente com as bordas, que originalmente possuem duas camadas (tons de cinza) para uma imagem de três camadas (colorida, RGB).

Identificação dos Agrupamentos – A transformação automática da imagem contendo o contorno dos agrupamentos, resulta em uma imagem branca com o contorno em verde, isso indica que possui informação nos três planos dessa imagem, não sendo interessante na hora de obter a imagem final. Para retirar as informações sobressalentes foi necessário inserir os valores da imagem em RGB com agrupamentos, na primeira camada, e nas demais camadas foram inseridos zeros, mantendo o tamanho original da imagem.

Possuindo agora duas imagens RGB, original e com contorno de agrupamentos, é necessário somá-las, para resultar em uma única imagem com as características da imagem original e o contorno dos nódulos.

Avaliação da metodologia – A avaliação da metodologia proposta utilizou 45 imagens com laudos feitos por especialista, das quais 43 apresentavam laudo constatando a presença de tumores e duas são imagens normais, ou seja, não apresentavam tumores.

Foram realizadas duas análises: uma da imagem como um todo e outra da detecção de cada tumor. Para análise da imagem completa foi comparado o laudo do especialista e o resultado do sistema, ou seja, se a metodologia proposta identificou algum tumor na imagem. A segunda análise, da detecção de tumores, foi realizada comparando-se todos os tumores identificados pelo especialista e aqueles identificados pelo sistema. As imagens normais foram consideradas como duas ausências de tumores. A partir das comparações foram identificados valores de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN) das imagens, foi calculada a taxa de falsos tumores por imagem e a eficiência do sistema.

Resultados

Os resultados obtidos com as duas análises podem ser observados na Tabela 1. Após realização da análise das imagens o sistema obteve 84,44% de eficiência, corretamente identificando tumores em 38 das 45 imagens, classificando três imagens como normais e não

identificando nenhum tumor quatro imagens cujo laudo apresentava presença de tumores.

Tabela 1: Resultados das análises comparativas.

Análise	VP	VN	FP	FN	Total
imagem	38	3	0	4	45 imagens
detecção dos tumores	101	3	8	15	116 tumores

Com a análise da detecção dos tumores, obteve-se uma taxa de 0,1778 falsos positivos por imagem e eficiência na detecção dos tumores de 85,83%. A taxa de acerto dos tumores foi de 87,07%.

Os dois falsos positivos encontrados na detecção dos tumores aconteceram na mesma imagem e os três verdadeiros negativos identificados na segunda linha da Tabela 1 referem-se a não identificação de tumores nas três imagens normais utilizadas nos testes.

Conclusão

Neste trabalho foi desenvolvido e implementado um script cuja função é identificar e localizar tumores em imagens mamográficas.

Para verificar o código foram feitos testes analisando várias imagens resultando em uma eficiência de 84,44% considerando a comparação com o laudo do especialista e uma taxa de acerto de 85,83% na detecção dos tumores. Além disso, foi obtida uma taxa de falsos positivos por imagem de 17,78%.

Os valores de eficiência e taxa de acerto dos tumores podem ser considerados relativamente bons, no entanto, seria necessária a realização de testes com um maior número de imagens para que uma conclusão definitiva sobre a eficiência do sistema possa ser feita.

Um aspecto que poderia aumentar a eficiência da metodologia e diminuir a taxa de falsos positivos por imagem é a resolução das imagens. Contudo, ao utilizar imagens com maior resolução, seria necessária uma reestruturação de algumas partes do sistema para adequação a nova quantidade de pixels. Imagens com maior resolução apresentariam características mais bem definidas o que poderia trazer uma maior eficácia ao script e proporcionar um diagnóstico mais preciso.

Referências

- [1] Solomon C, Breckon T. Fundamentals of digital Image Processing: A Practical Approach with Examples in Matlab. West Sussex: John Wiley & Sons; 2011.
- [2] O'Malley MS, Earp JA, Hawley ST, Schell MJ, Mathews HF, Mitchell J. The association of race/ethnicity, socioeconomic status, and physician recommendation for mammography: who gets the message about breast cancer screening? American Journal of Public Health. 200; 91(1): 49-54.

- [3] Borges VRP, Guliato D. Detecção de Agrupamentos de Microcalcificações em Imagens Mamográficas Digitais Utilizando Etapas de Segmentação da Mama e Realce. Horizonte Científico (Uberlândia). 2010; 4:1-29.
- [4] Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA. Computer-aided detection of clustered microcalcifications: An improved method for grouping detected signals. Medical Physics. 1993; 20(6):1661-1666.
- [5] Dengler J, Behrens S, Desaga JF. Segmentation of Microcalcifications In Mammograms, IEEE Transactions on Medical Imaging. 1993; 12(4):634-642.
- [6] Matheus BRN. BancoWeb: base de imagens mamográficas para auxílio em avaliações de esquemas CAD. [dissertação]. São Carlos: Universidade de São Paulo; 2010.
- [7] Marques Filho ON, Vieira H. Processamento Digital de Imagem. Rio de Janeiro: Brasport; 1999.
- [8] Gonzalez RC, Woods RE. Processamento de Imagens Digitais. 3ª ed. São Paulo: Pearson; 2011.
- [9] Nunes FLS, Schiabel H. Detecção de agrupamentos de microcalcificações mamárias através de processamento de imagens mamográficas com detecção semi-automática de regiões de interesse. Revista Brasileira de Engenharia Biomédica (Impresso). 2001; 16(13):109-121.