

# Segmentação de nódulos em mamografias usando *Fuzzy Connectedness*

G. S. Amaral\*, M. S. Kudo\*, R. S. S. Souza\*, M. C. Moraes\*

\*Departamento de Ciência e Tecnologia/Instituto de Ciência e Tecnologia da UNIFESP, São José dos Campos, Brasil  
e-mail: gamaral@unifesp.br

**Abstract:** *Breast cancer is the most common carcinoma among woman in the world. Caused by the exacerbated growing of mammary tissue, affecting the individual health. There is no cure, but with an early diagnostic there is a chance of a successful treatment. Actually, the recommend examination is the mammography, in other words, a breast X-ray exam. However, the mammography is not easy to analyze when it has a dense breast (high percentage of glandular breast tissue). For having a more precise diagnostic it is already possible to make a computer analysis of the images, and in case the breast has a tumor it is done a segmentation on it. The method is based on the semiautomatic segmentation using morphological operations and Fuzzy Connectedness. The evaluation was made in 10 mammographic images with nodule tumors. Comparatively, with the Golds, the average results of Positive True was  $92 \pm 6$  and Overlap Ratio was  $82 \pm 8$ . These were, as expected, a faster method and with a good performance.*

**Key-words:** *Segmentation, Mammography, Cancer, Fuzzy Connectedness and image processing.*

## Introdução

A taxa de incidência do câncer no mundo atualmente é 10 vezes maior que nos anos 60 e 70. Em 2014 no Brasil, ocorreram 52.120 novos casos e em 2016 surgiram cerca de 57.960 casos de câncer de mama[1].

O carcinoma é um tumor maligno, ou seja, um tecido com alto potencial metastático que pode se desenvolver em qualquer região do corpo [2]. O câncer de mama é o carcinoma encontrado no tecido mamário, podendo estar principalmente na forma de nódulos e calcificações [2]. Há três principais tipos de câncer de mama: o carcinoma ductal in situ (CDSI), carcinoma ductal invasivo (ICD) e carcinoma invasivo lobular (ICL) [3]. Como qualquer câncer, quanto mais cedo for diagnosticado, maior será a chance de se obter êxito no tratamento.

Segundo o Ministério da Saúde, o principal método de diagnóstico capaz de detectar precocemente um nódulo é a mamografia [4]. Estudos mostram que a mamografia em conjunto com os testes clínicos, apresentaram uma redução relativa da mortalidade de 17% das mulheres [5]. Entretanto, deve-se analisar que a mamografia foi um exame criado em 1913. Mesmo com sua evolução ao longo dos anos, continua sendo um exame de Raio-X, ou seja, há a incidência de radiação ionizante sobre o paciente. Ainda que o exame seja digitalizado, há informações subjetivas, uma vez que a diferença de textura entre tecidos saudáveis e patológicos nas imagens são tênues. A partir da análise da imagem de Raio-X da mama, as áreas mais claras podem representar músculos ou glândulas mamárias, enquanto as áreas mais escuras podem ser gordura ou outros tipos de tecidos. Quando a mama é densa, ou seja, apresenta alto percentual de glândulas, a mamografia é mais difícil de ser investigada. Isto deve-se pelo destaque do tecido glandular que pode ocultar massas importantes para um diagnóstico, dificultando o reconhecimento da doença. Assim, é essencial desenvolver e aperfeiçoar métodos e técnicas de processamento de imagens que aumentem a distinção entre o tumor e o restante da imagem. Atualmente diversas pesquisas focam em segmentação de tumores em mamografias. No trabalho de Martins [6] há uma segmentação por crescimento de regiões, porém não apresenta uma avaliação para uma conclusão quantitativa do desempenho do método. O trabalho realizado tem o intuito de segmentar nódulos em mamografias, e comparar o resultado com a segmentação manual de um técnico.

## Materiais e Métodos

Foram utilizadas 10 imagens, com resolução de 1024x1024, do banco de dados mini-MIAS, liberadas para pesquisas científicas. Todas as imagens são de mamografias que apresentam tumores no formato de nódulos em mulheres. A

metodologia é dividida em três partes (Figura 1): Pré-Processamento, Extração de características e



Figura 1: Diagrama em blocos: Resumo da metodologia utilizada para a segmentação do tumor na mama.

Pós-Processamento.

**Pré-Processamento:** Nesta fase a imagem original (Figura 2(A)) é processada com o intuito de realçar o tumor e diminuir a presença das glândulas mamárias. Na imagem de mamografia quando a mama é densa, as glândulas se destacam e borram parcialmente o seio dificultando a análise do tumor. Para fazer a distinção de informações, é utilizado primeiramente um filtro mediana com *kernel* de  $7 \times 7$  (Figura 2(B)), reduzindo os detalhes na região densa da mama e deixando a imagem mais homogênea. Para recuperar as bordas perdidas na filtragem foi implementado um gradiente com máscara Sobel. Essa máscara realça as bordas e para controlar a quantidade de borda inserida na imagem (Figura 2(C)), foi realizado um filtro mediana  $7 \times 7$ . Em seguida, analisa-se o histograma (Figura 2(D)) da imagem (Figura 2(C)), que é um gráfico que relaciona a frequência de ocorrência com valores de intensidade da imagem. Para encontrar a localização do tumor no histograma (Figura 2(C)), estudou-se a intensidade do tumor e constatou-se que este tem, em média de 0,5 a 0,7 do valor máximo de intensidade da imagem. Ou seja, no histograma há um pico de pixels médios onde há a sobreposição informações significativas e insignificativas. Para distinguir essas informações e trabalhar com a região de interesse do histograma, implementou-se um ceifamento (Figura 2(F)). Este ceifamento do histograma consiste em transformar todos pixels maiores que 0.75 em 1. Em seguida foi feita uma transformação de potência na intensidade de cada pixel da imagem mediana ( $I_{mediana}$ ), como descrito

na equação abaixo:

$$I_{pot} = I_{mediana}^{\gamma} \quad (1)$$

Em que  $\gamma$  é o coeficiente da potência e  $I_{pot}$  é a imagem resultante do contraste (Figura 2(C)). Para  $\gamma$  é adotado o valor de 1.2, pois foi o valor que melhor se adaptou na calibração do programa.

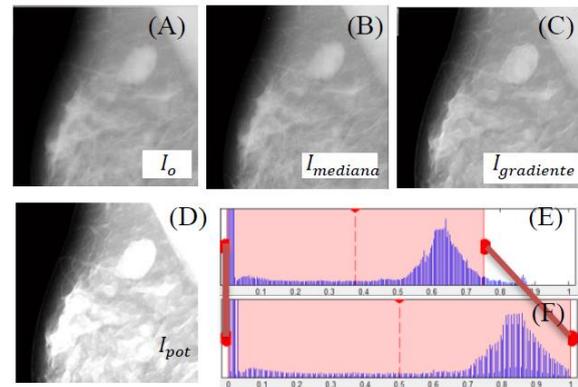


Figura 2: Pré-Processamento de imagens de mamografia. (A) Zoom no nódulo da imagem original. (B) Imagem com filtro mediana. (C) Imagem com gradiente. (D) Imagem com o ajuste no histograma. (E) Histograma da Figura 2(B). (F) Histograma ceifado de 0 a 0.75.

**Extração de características:** A Extração de características consiste na extração do tumor da imagem de mamografia. Para isso, é utilizado o método de segmentação *Fuzzy Connectedness*. Previamente o tumor foi identificado pelo operador, o qual seleciona manualmente uma região dentro deste (Figura 3(A)). Com uma semente no centro da região eleita, a construção da matriz de conectividade é iniciada. Calcula-se uma média e desvio padrão da Intensidade (*MédiaI*, *DesvioPI*) e Homogeneidade (*MédiaH*, *DesvioPH*) desta região, com as equações abaixo:

Intensidade:

$$MédiaI = media\left(\frac{abs(IS + IR)}{2}\right) \quad (2)$$

$$DesvioPI = desviop\left(\frac{abs(IS + IR)}{2}\right) + 0.075 \quad (3)$$

Homogeneidade:

$$MédiaH = media(abs(IS - IR)) \quad (4)$$

$$DesvioPH = desviop(abs(IS - IR)) + 0.075 \quad (5)$$

Nas equações (2), (3), (4) e (5), *media* é a média aritmética, *desviop* é o desvio padrão, *abs* é o absoluto, *IR* é a região selecionada e *Is* é o valor de intensidade do pixel da semente. Com a finalidade

de aumentar a abrangência de *pixels* pertencentes ao tumor, é acrescentado um valor de 0.075 no *DesvioPI* e no *DesvioPH*.

Após o cálculo das características da região selecionada, inicia-se as interações desta com o pixel vizinho. Tais interações objetivam a construção de uma matriz de conectividade. No *Fuzzy Connectedness*, a homogeneidade (*Hom*) e a intensidade (*Int*) são determinadas para cada vizinho (*pix*) a partir da semente (*sem*), conforme as equações a seguir:

$$Hom(pix) = e^{-\frac{1}{2}\left(\frac{abs(Is-Ip)-MédiaH}{DesvioPH}\right)^2} \quad (6)$$

$$Int(pix) = e^{-\frac{1}{2}\left(\frac{0,5(Is+Ip)-MédiaI}{DesvioPI}\right)^2} \quad (7)$$

A partir desses dados calcula-se a afinidade pontual (*AfP*), sendo que *wh* e *wi* são fixos e iguais a 0,5. As equações utilizadas são apresentadas a seguir:

$$AfP = wh.Hom(pix) + wi.Int(pix) \quad (8)$$

$$wh + wi = 1 \quad (9)$$

No qual *Hom* tem o peso de *wh* e *Int* de *wi*. Computou-se então a conectividade a partir da afinidade do caminho com a *Afc*:

$$Afc(pix) = \min(AfP(pix), MédiaH(sem)) \quad (10)$$

$$Conect(pix) = \max(Conect(pix), Afc(pix)) \quad (11)$$

Em que *Conect* é a matriz de conectividade. Como critério de parada, é definido um limiar. Esse processo reduz o custo computacional, uma vez que não há a necessidade de percorrer a imagem inteira. A matriz de conectividade (*Conect*) é construída enquanto os valores de conectividade computados forem maior que o *limiar*. Neste trabalho, propomos um limiar dinâmico de parada, definido por:

$$limiar = \begin{cases} 0,95, & se\ DesvioPI \leq 0,078 \\ 0,85, & se\ 0,078 \leq DesvioPI \leq 0,088 \\ 0,8, & caso\ contrário \end{cases} \quad (12)$$

O *limiar* é calculado a partir dos desvios padrão de *Int* da região selecionada. Assim cada imagem terá seu momento de parada individual, que será consequente de seu desvio padrão. Quando o desvio padrão é alto, há uma maior variação dos pixels na região. Desta forma, o limiar de parada terá que ser menor, pois o valor da conectividade diminuirá mais rápido dentro da região selecionada. E se o desvio padrão for baixo, o limiar será mais alto, já que na região a conectividade não se alterará tão significativamente. Quando concluída a matriz *Conect* é possível visualizar a Figura 3(B) com o

tumor extraído.

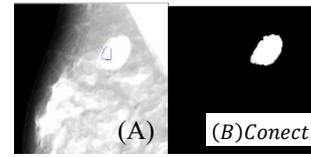


Figura 3: Extração de características. (A) Imagem com região selecionada. (B) Matriz de conectividade resultante do processo de Fuzzy Connectedness.

**Pós-Processamento:** Nesta última etapa, a imagem é binarizada (Figura 4(A)) com um limiar fixo de 0.01 encontrado empiricamente. Em seguida realiza-se um fechamento com elemento estruturante circular de raio 6 (Figura 4(B)) com o intuito de preencher o tumor sem prejudicar as bordas caso haja necessidade. Para uma visualização final, é sobreposto o contorno da segmentação sob a imagem original, como apresentado na Figura 4(C).

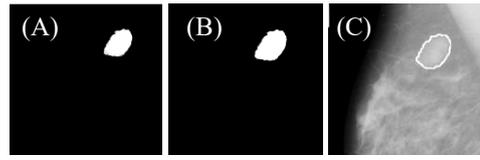


Figura 4: Pós-Processamento. (A) Imagem da matriz de conectividade binarizada. (B) Imagem com fechamento. (C) Sobreposição da borda da Imagem segmentada em vermelho, mais a borda do *Gold* em verde na imagem original.

## Resultados

Os resultados obtidos foram calculados a partir de um limiar de parada da conectividade variável no *Fuzzy Connectedness*. A avaliação do método foi feita através do Verdadeiro Positivo (*VP*), Falso Negativo (*FN*) e Falso Positivo (*FP*), propostos por Udupa [7] e *Overlap Ratio (OR)*. Para o cálculo de tais parâmetros, foi necessário o uso de uma imagem de referência (*Gold*), é aplicado uma segmentação manual do tumor.

O processamento foi realizado no MATLAB (R2012b) 64bits, processador Intel(R) Core(TM) i5-3470 CPU @ 3.20GHz 3.20 GHz e 4GB de memória RAM. O tempo médio foi de  $6,5 \pm 1,4$  segundos.

A avaliação foi realizada analisando a média e desvio padrão do *VP*, *FP*, *FN* e *OR*. Os valores apresentados na Tabela 1, demonstraram acurácias com *VP* de  $92\% \pm 6$ , *FP* de  $12\% \pm 11$ , *FN* de  $8\% \pm 6$  e *OR* a  $82\% \pm 8$ .

Tabela 1: Avaliações do método através de VP, FN, FP e OR.

VP	FP	FN	OR
92 ± 6	12 ± 11	8 ± 6	82 ± 8

## Discussão e Conclusão

A pesquisa foi realizada com o propósito de otimizar processos de segmentação e análise de mamografias, com ênfase em imagens com tumores em forma de nódulos. Outros trabalhos já foram realizados com o mesmo objetivo geral, porém se aprofundando em outras formas de se otimizar a análise das mamografias. Por exemplo uma pesquisa de segmentação de nódulos usando a Transformada Watershed [8] ou automatização da identificação da incidência de mamas esquerdas ou direitas e projeções Crânio Caudal e Médio Lateral Oblíqua [9], são exemplos de trabalhos presentes na literatura.

Como esperado, os resultados da Tabela 1 foram satisfatórios. O alto índice no VP demonstra que a segmentação apresenta alto grau de similaridade da imagem segmentada com o Gold. O OR também apresentou um valor alto de 82%, sendo que leva em consideração o FP, FP e FN, ou seja, é um valor consequente das conquistas e falhas do método. Nos trabalhos de Peyman Rahmati [10] [11], foram segmentados tumores mamários usando contornos ativos. Seus resultados foram de 81% [10] e 87% [11] para o Verdadeiro Positivo, sendo que no método apresentado neste artigo obteve-se o valor de 92%. O curto tempo necessário para o processamento, possibilita que um único técnico consiga analisar mais imagens em um período de tempo.

Este trabalho tem como grande contribuição o auxílio aos técnicos na segmentação de tumores mamários. Além disso, realizou-se a calibração dos parâmetros do Fuzzy Connectedness para mamografia e um novo limiar de parada. Foi realizada a avaliação completa do método utilizando VP, FP, FN e OR. O método desenvolvido pode-se considerar preliminar para avaliação do uso do Fuzzy Connectedness na aplicação, sendo não compatível (por motivos como pequeno número de imagens) comparar com outros artigos.

Para trabalhos futuros, há desejo de automatizar o método, elevar o número de imagens, aumentar a similaridade entre o gold e o tumor segmentado, a fim de aumentar sua acurácia e comparar com outros métodos e artigos.

## Agradecimentos

Instituto de Ciência e Tecnologia da UNIFESP (ICT-UNIFESP), Pilot European Image Processing Archive (PEIPA)

## Referências

- [1] INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Mama. , 2016. Disponível em: <[http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/cancer\\_mama](http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/cancer_mama)>. Acesso em: 02 maio 2016.
- [2] NOGUEIRA SMB, DIÓGENES MAR, SILVA ARV. Autoexame das mamas: as mulheres o conhecem? Rev RENE. 2006;7(1):84-90.
- [3] AMERICAN CANCER SOCIETY. Types of breast cancers., 2016. Disponível em: <<http://www.cancer.org/cancer/breastcancer/detailed-guide/breast-cancer-breast-cancer-types>>. Acesso em: 02 maio 2016.
- [4] TSANTIS S, KAGADIS GC, KATSANOS K, KARNABATIDIS D, BOURANTAS G, NIKIFORIDIS GC. Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography. Med Phys. 2012 Jan; 39(1): 503-13.
- [5] MANDELBLATT JS, CRONIN KA, BAILEY S, BERRY DA, DE KONING HJ, DRAISMA G, ET AL. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Intern Med 2009;151:738-47.
- [6] M. N. MARTINS. Segmentação por crescimento de regiões em imagens de mamografia. CBEB – 2014.
- [7] UDUPA JK, LEBLANC VR, ZHUGE Y, IMIELINSKA C, SCHMIDT H, CURRIE LM, HIRSCH BE, WOODBURN J. A framework for evaluating image segmentation algorithms. Comput Med Imaging Graph 2006, 30(2):75–87.
- [8] SANTOS, V. T. Segmentação de imagens mamográficas para detecção de nódulos em mamas densas. 2002. Dissertação (Mestrado em Engenharia Elétrica) - EESC-USP. São Carlos.
- [9] TEIXEIRA, RITA FILIPA DOS SANTOS. Análise automática de imagens de mamografia:: Distinção entre incidências. 2012. Dissertação (Mestrado em Engenharia Biomédica) - Faculdade de Engenharia da Universidade do Porto. Porto.
- [10] PEYMAN RAHMATI, AHMAD AYATOLLAHI. Maximum Likelihood Active Contours Specialized for Mammography Segmentation. 2009 IEEE.
- [11] PEYMAN RAHMATI , ANDY ADLER, GHASSAN HAMARNEH. Mammography segmentation with maximum likelihood active contours. Medical Image Analysis 16 (2012) 1167–1186.