

# UTILIZAÇÃO DE CONTORNO ATIVO PARA SEGMENTAÇÃO DE NÓDULOS EM IMAGENS DE MAMOGRAFIA

Luiz B. S. e Souza, Michele F. Angelo

Universidade Estadual de Feira de Santana, Feira de Santana, Brasil

e-mail: bernardo.uefs@gmail.com

**Resumo:** O câncer de mama é o segundo tipo mais frequente de câncer do mundo, e a sua identificação precoce apresenta um elevado percentual de cura. Este trabalho tem como proposta o desenvolvimento de um algoritmo capaz de segmentar nódulos em regiões de interesse (RI's) de imagens mamográficas. Para a segmentação dos nódulos foram utilizadas técnicas de pré-processamento e segmentação utilizando contorno ativo. Um conjunto de 180 imagens foi utilizado na fase de avaliação do modelo de contorno ativo através de três abordagens diferentes, tendo a filtragem associada à segmentação apresentado o melhor resultado com a segmentação de 93% dos VP (verdadeiros-positivos) e com 5% de FP (falsos-positivos).

**Palavras-chave:** Segmentação, nódulos mamários, contorno ativo.

**Abstract:** *Breast cancer is the second most common type of cancer in world, and its early detection has a high cure rate. This paper aims at the development of an algorithm capable of segmenting nodes in regions of interest (ROI's) of mammographic images. A set of 180 images was used in the evaluation phase of the active contour through three different approaches models having the filtering associated with the segmentation the best result in the identification of 93% of VP (true positive) and 5% FP (false positive).*

**Keywords:** *Segmentation, breast lumps, active contour.*

## Introdução

O câncer de mama é o segundo tipo mais frequente de câncer do mundo e o mais comum entre a população feminina, afetando aproximadamente uma entre onze mulheres em algum estágio de sua vida (INCA). No Brasil, as taxas de mortalidade por câncer de mama continuam elevadas, muito provavelmente porque a doença ainda é diagnosticada em estágios avançados.

A mamografia é reconhecida como a técnica mais eficaz para a detecção durante a fase inicial do câncer. Uma vez detectada a massa na mamografia, o radiologista recomenda outras investigações, dependendo da probabilidade de malignidade da lesão. Além da complexidade na interpretação das lesões, a imagem mamográfica possui baixo contraste, ruído, entre outros fatores que influenciam na sua qualidade, tais como calibração do equipamento mamográfico, condições de revelação, conservação e manipulação do filme.

Uma maior taxa de acerto no diagnóstico pode ser

obtida com a combinação da interpretação do radiologista e uma análise computacional. Métodos computacionais têm mostrado um grande potencial no auxílio aos radiologistas, fornecendo uma segunda opinião de diagnóstico [1].

Com base nos problemas apresentados para diagnosticar um nódulo, o objetivo deste trabalho foi o desenvolvimento um algoritmo para segmentar nódulo em regiões de interesse (RI) extraídas de imagens mamográficas. Este algoritmo possui dois estágios: o pré-processamento da imagem, utilizando técnicas como filtragem, binarização e equalização, e a segmentação do nódulo utilizando o modelo deformável de contorno ativo (*Snakes*).

## Materiais e métodos

Na metodologia adotada, primeiramente, foi necessária a obtenção de conhecimento sobre a técnica de segmentação utilizando contorno ativo. Para isso foi realizado o levantamento teórico a respeito do assunto, assim como de outras técnicas de processamento digital de imagem.

Este trabalho foi dividido em três etapas: Preparação do ambiente de desenvolvimento, Desenvolvimento do software e Testes e análise dos resultados.

Na primeira etapa foi escolhido o ambiente de trabalho a ser utilizado para a implementação do software. Para isso foram levados em consideração aspectos como: robustez, performance e agilidade no desenvolvimento. Uma das tecnologias que possui essas características é o MATLAB®. Ainda nesta etapa foram coletadas as imagens para a realização dos testes do algoritmo implementado. Foram 180 regiões de interesse (RI's) extraídas de imagens mamográficas, sendo que em 78 delas contém nódulos e 102 não contendo nódulo, todas as imagens pertencem a um banco de imagens mamográficas disponível online [2].

A segunda etapa consistiu no desenvolvimento do algoritmo. Os requisitos destacados no levantamento teórico foram transformados em rotinas para a implementação do algoritmo. Muitas rotinas do MATLAB® já disponibilizava alguns tipos de pré-processamento, interpolação e manipulação de matrizes.

A última etapa consistiu nos testes com as imagens ao aplicar o contorno ativo para a segmentação do nódulo e na análise dos resultados. Para isso foi utilizada a metodologia das curvas ROC (*Response Operating Characteristic*), que utiliza porcentagens de acertos e uma análise

estatística em termos de sensibilidade e especificidade para quantificar os erros em falso-negativo (FN) e falso-positivo (FP) e os acertos em verdadeiro-positivo (VP) e verdadeiro-negativo (VN) [3]. Os verdadeiros positivos (RI's que têm nódulo e nos quais o teste foi positivo), os verdadeiros negativos (RI's que não têm nódulo e nos quais o teste foi negativo), os falsos positivos (RI's que não têm nódulo e nos quais o teste foi positivo) e falsos negativos (RI's que têm nódulo e nos quais o teste foi negativo) são as quatro categorias nos quais os testes foram enquadrados.

A seguir será detalhada a etapa de desenvolvimento do algoritmo, onde são descritos os passos para a implementação e a utilização das técnicas aplicadas neste trabalho.

**Pré-processamento da imagem** - As imagens de mama, no geral, possuem características semelhantes como ruído e baixo contraste, fazendo com que a delimitação do contorno do nódulo seja mais difícil. Para que fosse possível alcançar melhores resultados com o contorno ativo, três técnicas de pré-processamento (filtragem, binarização e equalização) foram utilizadas. As técnicas de pré-processamento foram utilizadas para facilitar a segmentação do nódulo ao executar o algoritmo de contorno ativo.

O filtro de suavização ou passa-baixa foi o escolhido pela sua característica de reduzir o ruído presente nas imagens, impedindo a identificação de falsos contornos. O filtro passa-baixa é um filtro de suavização que remove altas frequências que correspondem às transições abruptas. Neste trabalho, o usuário escolhe o valor do sigma ( $\sigma$ ) que será utilizado para a construção da máscara gaussiana. Foram utilizadas as funções *fspecial* e *filter2* do MATLAB® para a construção da máscara gaussiana (3x3) e filtragem, respectivamente.

A binarização foi a técnica utilizada para destacar o nódulo do fundo da imagem. A função que realiza o processo de binarização é a *im2bw* do MATLAB®. No presente trabalho o valor do limiar de 0.7 foi escolhido manualmente com base em testes realizados, pois ao se utilizar o limiar automático, fornecido pela função *graythresh* do MATLAB® que utiliza o método de *Otsu*, parte do nódulo era perdido ou ofuscado com o fundo da imagem

A equalização foi responsável pelo destaque dos nódulos nas imagens com baixo contraste, devido à equalização do histograma. Nódulos que ficavam mascarados com o tecido puderam ter um maior destaque e visualização. A função *histeq* do MATLAB® foi utilizada para realizar a equalização do histograma.

**Algoritmo de contorno ativo** - A ideia geral do modelo é a utilização de uma curva (*snake*) de minimização de energia para extrair características importantes de uma imagem. Essa energia associada à curva é definida de forma que ela seja mínima quando a curva se encontra sobre uma região com as características desejadas. [4].

A *snake* é uma curva controlada continuamente sob a influência de forças internas e forças externas. As

forças internas atraem a *snake* em direção de características salientes como linhas e contornos. As forças externas são responsáveis por colocar a *snake* próxima ao local de mínima energia [4].

Representando uma *snake* como uma curva parametrizada  $v(s) = (x(s), y(s))$ , pode-se descrever sua energia funcional  $E_{snake}$  como a equação (1) onde  $E_{int}$  representa a energia interna a *snake* e  $E_{ext}$  a energia externa.

$$E_{snake} = \int_0^1 E_{snake}(v(s)) ds \quad (1)$$

$$= \int_0^1 E_{int}(v(s)) + E_{ext}(v(s)) ds$$

A energia interna ( $E_{int}$ ) pode ser escrita como a equação (2).

$$E_{int} = \frac{(\alpha(s)|v'(s)|^2 + \beta(s)|v''(s)|^2)}{2} \quad (2)$$

A energia da curva é composta por uma derivada de primeira ordem controlada por  $\alpha(s)$  e uma derivada de segunda ordem controlada por  $\beta(s)$  como pode ser observado na equação (2). O termo de primeira ordem faz com que a *snake* funcione como uma membrana e o termo de segunda ordem faz com que atue como uma placa fina. Ajustando os pesos de  $\alpha(s)$  e  $\beta(s)$  é possível controlar a relativa importância dos termos de primeira e segunda ordem. Utilizando  $\beta(s)$  com valor zero em um ponto permite a *snake* ser descontínua no termo de segunda ordem e desenvolva um canto.

A energia externa  $E_{ext}$ , como pode ser observada na equação (3) é composta pela combinação de três energias ponderadas por  $w_{line}$ ,  $w_{edge}$  e  $w_{term}$  respectivamente. A depender da ponderação dada as energias diferentes comportamentos da *snake* podem ser criados [4].

$$E_{ext} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \quad (3)$$

A energia  $E_{line}$  ou energia de linha é a energia mais simples e é composta pela própria intensidade dos pixels da imagem. Dependendo do sinal dado a  $w_{line}$ , a *snake* irá ser atraída à linhas mais claras ou mais escuras do contorno mais próximo, ponderações positivas fazem a *snake* ser atraída a linhas escuras e ponderações negativas a linhas claras.

A energia  $E_{edge}$  ou energia de borda se utiliza do gradiente da imagem. Com essa força a *snake* é guiada para locais onde o gradiente é mais alto, ou seja, nos contornos da imagem. Por último a energia  $E_{term}$  a qual permite a curva achar o término de linhas e cantos.

## Resultados

Para a realização dos testes foi desenvolvida uma GUI (*Graphical User Interface*), na qual o usuário realiza todas as ações necessárias à execução do algoritmo. O usuário pode escolher a imagem que quer avaliar e

ponderar as forças (externas e internas) envolvidas no processo e aplicar alguma(s) técnica(s) de pré-processamento disponíveis. Em todos os testes as ponderações utilizadas pelo algoritmo foram as seguintes:  $\sigma = 1$ ,  $w_{\text{line}} = 0$ ,  $w_{\text{edge}} = 1.5$ ,  $w_{\text{term}} = 0$ ,  $\alpha = 0.2$  e  $\beta = 0.9$ . Essas forças foram escolhidas empiricamente, com base em testes feitos nas imagens, observando as influências das forças no comportamento da *snake* e levando em consideração as características comuns entre as imagens, como por exemplo, baixo contraste.

Para a avaliação do algoritmo de contorno ativo foram utilizadas três abordagens, na primeira não é aplicado nenhum tipo de pré-processamento nas imagens, na segunda é aplicada a filtragem e na terceira são aplicadas três técnicas de pré-processamento (filtragem, equalização e binarização) sequencialmente. Para a realização dos testes foram utilizadas 180 RI's, das quais 102 sem nódulo e 78 com a presença de nódulos. Foram utilizadas RI's devido ao fato do modelo de contorno ativo implementado possuir restrições como a limitação da distância na marcação dos pontos ao objeto a ser segmentado e a dificuldade em contornar concavidades mais profundas.

Para determinar o resultado como positivo (nódulo segmentado) ou negativo (nódulo não segmentado) e posteriormente realizar a separação nas categorias (VP, VN, FP e FN) foi adotado o seguinte critério, caso a *snake*, durante o processo iterativo estabilizasse, ou seja, achasse o local de mínima energia, seria considerado como positivo; caso a *snake*, durante o processo iterativo convergisse de tal forma a se tornar um ponto ou traço seria considerado como negativo.

Os resultados obtidos, sem a utilização de técnicas de pré-processamento foram: das 78 RI's com nódulo 60 RI's tiveram o nódulo segmentado, correspondendo a 77% de VP, isso significa dizer que em 23% das imagens que possuíam nódulo o algoritmo não foi capaz de segmentá-los, essa é uma taxa elevada dada a gravidade que essa informação incorreta pode acarretar. Nas imagens normais o percentual de FP foi 22%, conforme apresentado na Tabela 1. A área sob a Curva ROC foi de 0,78 (Figura 1).

Utilizando a filtragem, os resultados obtidos foram 94% de VP e 5% de FP, conforme apresentado na Tabela 1. Os valores encontrados foram significativamente melhores comparados aos encontrados na primeira abordagem devido à eliminação dos ruídos. A área sob a Curva ROC foi de 0,945 (Figura 2).

Na terceira abordagem de avaliação, os resultados obtidos foram 92% de VP e 7% de FP, conforme apresentado na Tabela 1. A área sob a Curva ROC foi de 0,925 (Figura 3).

Tabela 1 – Resultados obtidos com as três abordagens de avaliação: (1) Sem pré-processamento; (2) Filtragem; (3) Filtragem, equalização e binarização.

Abordagem de Avaliação	VP	%	FN	%	VN	%	FP	%
(1)	60	77	18	23	80	78	22	22
(2)	74	94	4	6	97	95	5	5
(3)	72	92	6	8	95	93	7	7

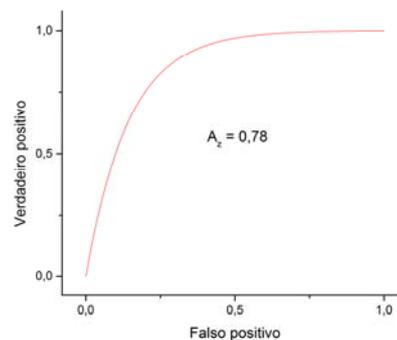


Figura 1 – Curva ROC dos testes realizados com as RI's sem a aplicação de técnicas de pré-processamento

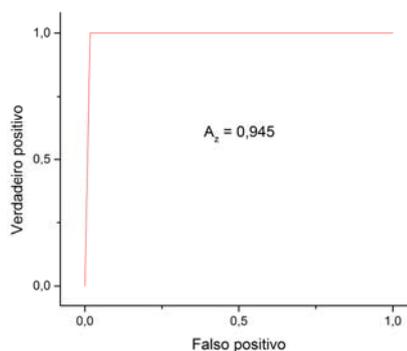


Figura 2 - Curva ROC dos testes realizados com as RI's realizando a filtragem

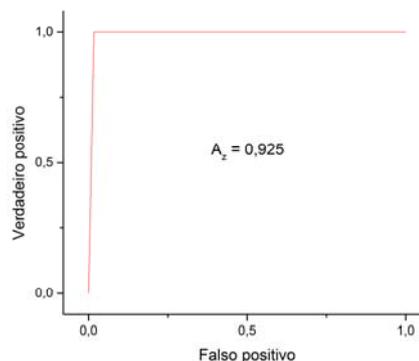


Figura 3 - Curva ROC dos testes realizados com as RI's realizando a filtragem, a equalização e a binarização.

## Discussão

Com base nos resultados obtidos, foi possível comprovar a importância dos métodos de pré-processamento ao se utilizar o contorno ativo como técnica de segmentação de nódulo. Consequentemente, o primeiro teste

aplicado apresentou um resultado bem abaixo em relação aos outros dois realizados.

As imagens utilizadas são ruidosas e possuem baixo contraste causando uma diminuição na eficiência do algoritmo de contorno ativo caso não seja feito um tratamento na imagem. Pelo fato do algoritmo ser sensível a variações abruptas nos valores de intensidade dos pixels, ruídos presentes na imagem podem fazer com que a *snake* pare de convergir em direção as bordas do objeto a ser identificado.

O segundo teste realizado consistiu na aplicação da filtragem nas imagens e apresentou os melhores resultados dentre os três testes realizados. O filtro de suavização eliminou os ruídos que causaram o mau desempenho do algoritmo no primeiro teste, propiciando uma fluidez no processo de convergência da *snake*. O último teste realizado apresentou bons resultados tanto nas RI's com nódulo quanto nas RI's sem nódulo.

Na Figura 4b pode ser observado que o contorno do nódulo não é tão bem definido como o contorno obtido após aplicar um pré-processamento na imagem (Figura 4d). Se tratando de nódulo mamário o formato é um dos fatores analisados para a determinação da sua benignidade ou malignidade, sendo assim o pré-processamento propiciou uma segmentação mais fiel do contorno do nódulo. Esse contorno poderia ser utilizado na etapa de classificação.

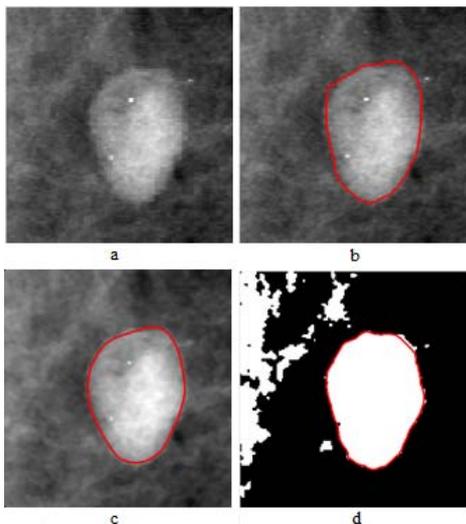


Figura 4 – (a) Imagem original; (b) Contorno obtido sem pré-processamento; (c) Contorno obtido com imagem filtrada; (d) Contorno obtido com filtragem, equalização e binarização

A Figura 5 representa uma RI sem nódulo e que o sistema corretamente não identificou nenhum nódulo após a execução do algoritmo de contorno ativo. Primeiramente foi realizada a marcação dos pontos (Figura 5a) e em seguida foi executado o algoritmo de segmentação (Figura 5b).

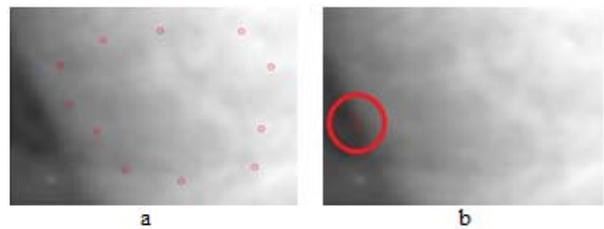


Figura 5 – (a) Marcação dos pontos em uma RI sem nódulo; (b) Contorno final reduzido a um traço após execução do algoritmo.

## Conclusão

O desenvolvimento de sistemas CAD vem crescendo cada dia mais, esses softwares têm se mostrado de grande potencial no auxílio ao radiologista, com o intuito de prover uma segunda opinião ao especialista.

O presente trabalho baseia-se na utilização do modelo de contorno ativo, originalmente proposto por Kass, Witkin e Terzopoulos (1988) [4], para realizar a segmentação de nódulos em RI's extraídas de imagens mamográficas. Como pôde ser observado nos resultados, a utilização do modelo de *snakes*, associado à técnicas de pré-processamento constitui-se em uma ferramenta poderosa para a segmentação de nódulo em RI's.

Apesar dos testes realizados terem apresentado bons resultados para as imagens que tiveram um pré-processamento, o modelo de contorno ativo implementado possui muitas restrições. Devido a esses fatores, como trabalho futuro uma solução seria adicionar a força balão [5], e utilizar a força GVF (*Gradient Vector Flow*) [6] que são variações do modelo onde a distância da marcação dos pontos possui uma limitação muito menor.

## Referências

- [1] Pasquale D, Fantacci ME, Kasae P, Retico A. Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier. *Computers in Biology and Medicine*; 2007. p. 1479–1491.
- [2] Matheus BRN, Schiabel H. Online Mammo-graphic Images Database for Development and Comparison of CAD Schemes. *Journal of Digital Imaging*; 2010. p. 1618-727X. Disponível em: <http://lapimo.sel.eesc.usp.br/bancoweb/>
- [3] Metz CE. Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine*; 1978. p. 283-298.
- [4] Kass M, Witkin A, Terzopoulos D. Snakes: Active Contour Models. *International Journal of Computer Vision*; 1988. p.321-331.
- [5] Yuan Y, Giger ML, Li H, Suzuki k, Sennett C. A dual-stage method for lesion segmentation on digital mammograms. *Medical Physics*. 2007.
- [6] Xu C, Prince JL. Snakes, Shapes, and Gradient Vector Flow. *IEEE Transactions on Image Processing*. v. 7, n. 3, 1998.