

# Segmentação Semi-Automática de Mamografias baseada em Clusterização em Grafo

André de Souza Moreira<sup>1</sup>, Geraldo Braz Júnior<sup>1</sup>, Simara Vieira da Rocha<sup>1</sup>,  
Aristófares Corrêa Silva<sup>1</sup> e Anselmo Cardoso de Paiva<sup>1</sup>

<sup>1</sup>Núcleo de Computação Aplicada - Universidade Federal do Maranhão (UFMA)  
Av. dos Portugueses, s/n – 65085-580 – São Luís, MA - Brasil

andre@nca.ufma.br, {gerald, simara, paiva}@deinf.ufma.br,  
ari@dee.ufma.br

***Abstract.** The use of computer aided diagnosis (CAD) systems has been increasing over recent years. These systems have been widely used to detect cancers at an early stage. This paper presents the use of a graph based image segmentation and geometry measures to discriminate mass regions in mammography images.*

***Resumo.** A utilização de sistemas de auxílio ao diagnóstico (CAD) tem sido crescente ao longo dos últimos anos. Tais sistemas têm sido amplamente utilizados para detecção de neoplasias em fase inicial. Este trabalho apresenta uma metodologia para segmentação de massas em mamografias utilizando algoritmos em grafo e geometria.*

## 1. Introdução

Previsões para o ano de 2012 apontam que o câncer de mama será o tipo de câncer que mais atingirá o sexo feminino [INCA 2012], sendo ele o responsável por 27,9% dos novos casos de câncer no grupo feminino. A melhor forma de combater a doença consiste na sua detecção precoce, elevando consideravelmente as chances de cura.

Diante deste cenário, vários trabalhos têm sido propostos a fim de auxiliar a detecção precoce desta neoplasia. Em [Giuliano et. al. 2008] é utilizado um filtro de suavização anisotrópica via Equações Diferenciais Parciais na fase de pré-processamento para a detecção de regiões suspeitas em mamografias. Em [Zhang 2011] é proposto um algoritmo automático para detecção de tumores em imagens mamográficas através do uso da lógica de fuzzy.

Este trabalho apresenta uma metodologia em desenvolvimento para automatizar e dar suporte ao especialista no processo de detecção de massas em imagens mamográficas. Para atingir tal objetivo, este trabalho propõe a utilização da técnica de segmentação baseada em grafo [Felzenszwalb 2004] a fim de extrair as estruturas do parênquima mamário e, posteriormente, a aplicação de um filtro de geometria sobre os grupos gerados na etapa anterior para redução de regiões que não se enquadram como suspeitas de massas.

## 2. Fundamentação

Esta seção apresenta sucintamente os conceitos a respeito da segmentação de imagens baseada em grafo, assim como do filtro de geometria para redução de resultados falsos positivos.

## 2.1 Segmentação baseada em Grafo

Este trabalho utiliza o algoritmo proposto por [Felzenszwalb 2004] para realizar a segmentação das estruturas internas de uma imagem mamográfica usando Árvore Geradoras Mínimas. O resultado dessa técnica está relacionado com o ajuste de três parâmetros:  $\sigma$ ,  $k$  e  $min$ . O  $\sigma$  controla o grau de suavidade dado para a imagem, o  $k$  é utilizado no cálculo da função limiar  $\tau$  e o  $min$  representa a quantidade mínima de elementos (pixels) presentes em uma região. Seu funcionamento básico consiste em utilizar um predicado ( $D$ ), detalhado na (Eq. 1), para avaliar a existência de uma fronteira entre duas regiões através da comparação das diferenças entre as duas regiões às diferenças internas de cada região.

$$D(C_1, C_2) = f(x) = \begin{cases} \text{verdadeiro}, & \text{Se } Dif(C_1, C_2) > MInt(C_1, C_2) \\ \text{falso}, & \text{Caso contrário} \end{cases} \quad (\text{Eq. 1})$$

onde  $Dif(C_1, C_2)$  representa a diferença entre as duas regiões  $C_1$  e  $C_2$  que equivale ao peso da aresta de menor peso que conecta as duas regiões. A diferença interna mínima,  $MInt$ , é definida como:

$$MInt(C_1, C_2) = \min(Int(C_1) + \tau(C_1), Int(C_2) + \tau(C_2)) \quad (\text{Eq. 2})$$

sendo que a diferença interna de uma região,  $Int(C)$ , é o valor do maior peso presente na árvore geradora mínima desta região. Por fim, a função limiar  $\tau$  controla o grau com que a diferença entre os dois componentes devem ser maior que suas diferenças internas.

## 2.2 Filtro de Geometria

Os grupos gerados pela etapa citada acima são submetidos a uma análise de sua forma geométrica através do índice de circularidade a fim de selecionar as regiões que possuem um aspecto geométrico mais próximo possível de uma circunferência, visto que uma massa possui normalmente um comportamento geométrico circular [Rangayyan et.al. 1997].

## 3. Metodologia

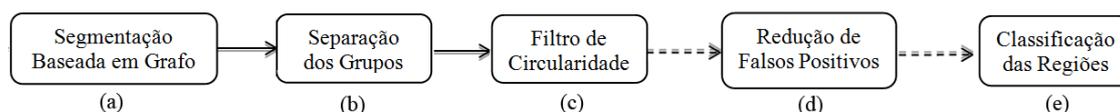


Figura 1. Diagrama esquemático da sequência de atividades.

A metodologia proposta neste trabalho, ainda em desenvolvimento, é apresentada na Figura 1. A primeira etapa, denominada *Segmentação Baseada em Grafo*, utiliza o algoritmo proposto por [Felzenszwalb 2004]. Após a segmentação, é realizada a *Separação dos Grupos* que é a fase responsável pela segregação de cada região gerada na etapa anterior para posteriormente serem utilizados na fase seguinte, *filtro de circularidade*, que utiliza o índice de circularidade para selecionar apenas as regiões que possuem certo grau de similaridade com uma circunferência. Tal índice é calculado através da razão da área da região pela área da menor circunferência que circunda a região.

Por se tratar de um trabalho em andamento, ainda não foram implementadas as duas últimas etapas. A fase de *Redução de Falsos Positivos* é necessária para diminuir a

quantidade de regiões apontadas como suspeitas, mas que na verdade não se enquadram nesta condição. Pretende-se nessa fase realizar inicialmente um tratamento da imagem para retirar o fundo e regiões do músculo peitoral presentes nas mamografias. Em seguida, através de técnicas de descrição e reconhecimento de padrões, reduzir a quantidade de regiões remanescente que não correspondam a regiões de massa. A etapa de *Classificação das Regiões* será responsável por classificar as regiões resultantes de acordo com suas características de malignidade.

#### 4. Resultados Preliminares

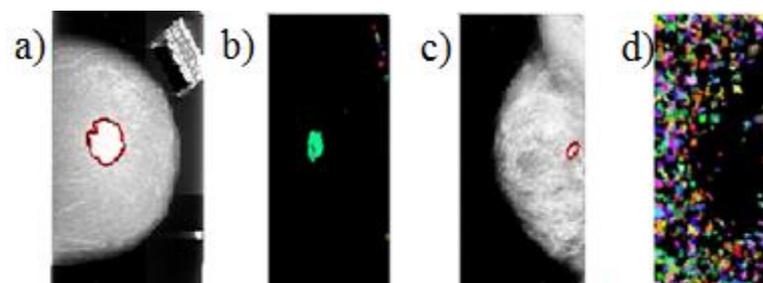
Foram realizados testes em 139 imagens mamográficas obtidas através da base de imagens DDSM [Heath et. al. 2001]. Todas essas imagens coletadas apresentam uma região de massa marcada pelo radiologista, podendo ser regiões benignas ou malignas. Para fins de teste, cada imagem foi submetida a 180 combinações de parâmetros: para o  $\sigma$ , variou-se de 0,2 até 0,5 ao passo 0,1; para o  $k$ , a faixa foi de 200 até 600 ao passo 100; para o  $min$ , variou-se de 150 até 250 ao passo 50 e para o índice de circularidade variou-se de 0,4 até 0,6 em incrementos de 0,1.

O critério utilizado para selecionar a melhor segmentação dentre o conjunto de parâmetros testados foi que a segmentação deveria gerar ao menos um verdadeiro positivo por imagem (VP), garantindo a identificação da massa em todas as imagens, e ao mesmo tempo a quantidade de falsos positivos por imagem (FP) deve ser a mínima possível. A seguir, na Tabela 1, são apresentados os cinco melhores resultados obtidos segundo o critério descrito acima, além da combinação dos parâmetros utilizados e a taxa média de falsos positivos por imagem (Tx. FP médio):

**Tabela 1 – Cinco melhores resultados obtidos da metodologia**

<b>Sigma</b>	<b>K</b>	<b>Min</b>	<b>Circ</b>	<b>TP</b>	<b>FP</b>	<b>Tx. FP médio</b>
0,4	500	200	0,4	139	1806	12,99
0,5	200	150	0,4	149	2000	14,39
0,4	600	150	0,4	148	2181	15,69
0,4	300	250	0,4	151	2241	16,12
0,4	400	200	0,4	147	2294	16,50

No melhor resultado obtido, foi gerado uma taxa média de 12,99 falsos positivos por imagem. De acordo com os parâmetros utilizados, a metodologia pode gerar super-segmentação, como observado nas linhas 2-4 da Tabela 1.



**Figura 2 – Comparação dos resultados obtidos nos testes. (a)-DDSM, cancer\_03, caso: A-1004-1, RIGHT\_CC com marcação do nódulo em vermelho. (b)-Resultado da segmentação de (a) para  $\sigma=0,5$ ,  $k=300$ ,  $min=250$  e  $circ=0,6$ . (c)- DDSM, cancer\_03, caso: A-1000-1, LEFT\_CC com marcação do**

**nódulo em vermelho. (d)-Resultado da segmentação de (c) para  $\sigma=0,2$ ,  
 $k=200$ ,  $\min=150$  e  $\text{circ}=0,4$ .**

Na Figura 2, são apresentadas duas imagens, (a) e (c), que foram utilizadas nos testes e seus respectivos resultados em (b) e (d). Verifica-se que para a imagem (a) a metodologia conseguiu segmentar corretamente a massa, no entanto restaram alguns grupos remanescentes da etiqueta de identificação do paciente. Já para a imagem (c) o resultado, em (d), apresentou grande quantidades de regiões correspondentes ao fundo da imagem devido ao excesso de ruídos presentes no fundo da imagem original.

## **5. Conclusão**

O presente trabalho apresenta resultados promissores nos testes iniciais, encontrando em todas as imagens testadas a região da massa. Todavia, verifica-se que um pequeno conjunto de imagens apresentou uma grande quantidade de falsos positivos, principalmente imagens classificadas como densas.

É possível notar através da Figura 2 que a maioria dessas regiões de falsos positivos está localizada na região de fundo. Por se tratar de um trabalho em andamento, os problemas destacados serão tratados com o uso de técnicas de remoção do fundo e do músculo peitoral na mamografia, além do realce das imagens a serem trabalhadas na etapa seguinte de redução de falsos positivos com o auxílio inclusive de máquinas de reconhecimento de padrões. Logo, acreditamos que a etapa seguinte provocará uma queda brusca de falsos positivos, fazendo desta uma metodologia promissora no sentido de detectar de massas em mamografias.

## **6. Agradecimento**

Agradecemos à CNPQ pelo apoio financeiro concedido por meio de bolsa de iniciação científica que possibilitou o desenvolvimento desta pesquisa.

## **7. Referências**

- Felzenszwalb, Pedro F. (2004). Efficient Graph-Based Image Segmentation. Em: International Journal of Computer Vision, p. 167-181.
- Giuliano, Denise, Barcellos, Celia A., Dias, Walter D. (2008). O Uso de Equações de Difusão no Processo de Detecção de Regiões Suspeitas em Mamografias. Em: WIM – Workshop de Informática Médica, p. 121-130.
- Heath, Michael, Bowyer, Kevin, Kopans, Daniel, Moore Richard and Kegelmeyer, W. Philip (2001). The Digital Database for Screening Mammography. In: Proceedings of the Fifth International Workshop on Digital Mammography, Medical Physics Publishing p. 212-218. ISBN 1-930524-00-5.
- INCA (Instituto Nacional do Câncer) (2012). Disponível em: <http://www.inca.gov.br/estimativa/2012/tabelaestados.asp?UF=BR>. Acessado em: 03 de março de 2012.
- Rangayyan, R.M. and El-Faramawy, N.M. and Desautels, J.E.L. and Alim, O.A. (1997) Measures of acutance and shape for classification of breast tumors. In: Medical Imaging, IEEE Transactions, V. 16 p. 799-810.
- Zhang, L. (2011). A novel automatic tumor detection for breast cancer ultrasound Images. IEEE Transactions, V. 1 p. 401-404.