

# Detecção de Regiões de Suspeitas em Mamogramas.

**Autor: Walter Borges Dias**

**Orientadora: Professora Dra. Célia Zorzo Barcelos**

**Orientadora: Professora Dra. Denise Guliato**

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

walter.dias@gmail.com, {celiazb, guliato}@ufu.br

Nível: Mestrado

Ano de ingresso no programa: 2006

Época esperada de conclusão: 2008

Etapa concluída: Defesa da proposta de dissertação

**Resumo.** *O câncer de mama é uma das doenças que mais causa mortes entre as mulheres. Segundo o Instituto Nacional do Câncer (INCA), no Brasil são registrados cerca de 41 mil novos casos todo ano e mais de nove mil casos de óbitos por consequência da doença. A mamografia é um dos métodos atuais de inquestionável importância no diagnóstico precoce da doença. O presente trabalho propõe um método para detecção de tumores a partir de mamografias digitais baseado em equações diferenciais para suavização do mamograma e utilização das duas vistas da mama para detectar as regiões de interesse e reduzir a ocorrência de falsos-positivos. A avaliação dos resultados será realizada usando a base de dados mamográficas DDSM - Digital Database Screening Mammography. Esta base é pública, contendo imagens de boa qualidade e laudos classificados segundo o sistema BI-RADS.*

**Palavras-Chave.** *Detecção de Lesões, Processamento Digital de Imagens, Mamografia.*

## 1. Introdução e Motivação

O câncer de mama é uma das principais causas de mortes entre as mulheres no mundo. Segundo dados do Instituto Nacional do Câncer (INCA), no Brasil são registrados cerca de 41 mil novos casos todo ano e mais de nove mil casos de óbitos por consequência da doença [Santos et al. 2006]. Um estudo feito em 2007 pela Sociedade Brasileira de Mastologia (SBM), revela um dado preocupante para as mulheres com menos de 40 anos. O estudo mostrou que em 2003 as mulheres nessa faixa etária correspondiam a apenas 5,6% dos casos de câncer de mama, no ano passado esse índice subiu para 16,8% [Barros et al. 2007]. Estes dados mostram que o número de câncer de mama entre as mulheres com idade inferior a 40 anos triplicou. Antes as mulheres dessa faixa etária eram relativamente pouco afetadas pelo câncer de mama, no entanto o estudo da SBM mostra que este grupo pode estar se transformando em um dos alvos primários da doença.

As causas do câncer de mama são pouco conhecidas. Dentre os fatores que podem aumentar as chances de contrair a doença são: primeira menstruação precoce (antes dos 11 anos), menopausa após os 50 anos, primeira gravidez após os 30 anos ou não ter filhos. O aumento do número de câncer de mama também pode estar relacionado com a mudança no estilo de vida das mulheres, sendo que hoje existem mais mulheres fumantes e que fazem uso prolongado de anticoncepcionais. Pelo fato que o agente causador da doença ainda é desconhecimento, torna-se difícil a adoção de programas de prevenção primária, ou seja, medidas tomadas para evitar a surgimento do câncer. Com isso, os esforços para controle da doença estão focados no diagnóstico precoce, pois em sua fase inicial, as células cancerígenas tem maiores chances de serem controladas.

Atualmente não existe uma técnica que seja totalmente eficiente e comprovada para detecção precoce do câncer. A mamografia é um dos métodos atuais de inquestionável importância no diagnóstico precoce da doença é ainda o método mais utilizado para fazer a prevenção, e tem um papel importante nas decisões terapêuticas a serem adotadas em caso de um diagnóstico positivo. Na mamografia digital, a imagem da mama é capturada através de um detector eletrônico de raios X que converte os sinais obtidos para uma foto digital, isto possibilita que a imagem adquirida seja analisada em um monitor de computador. Algumas vantagens da mamografia digital incluem: acesso mais fácil a imagens; diagnóstico assistido por computador; melhores meios de transmissão; recuperação e armazenamento de imagens; uso de uma dose média de radiação mais baixa, sem perda da precisão no diagnóstico [Bloomquist et al. 2005]. Mas apesar de todas essas facilidades, a detecção de lesões em mamografia não é tarefa fácil para os especialistas, isso devido à densidade da mama, que se da pela quantidade de tecido glandular. O tecido glandular aparece como áreas mais claras na mamografia e um nódulo também tem tonalidades claras, dessa maneira a lesão ou o tumor pode ficar oculto ou parcialmente oculto pelo tecido mamário.

No exame mamográfico são obtidas quatro imagens, duas correspondente à mama da direita e duas correspondente à mama da esquerda, tendo assim duas vistas de cada mama (crânio-caudal "CC" e médio oblíquo lateral "MLO"). Para detecção de regiões que possam caracterizar uma possível presença de câncer na mama, os especialistas analisam as duas vistas da mama, o que aumenta a sensibilidade do exame reduzindo também o numero de falso-positivos. Estudos clínicos mostram que a análise das duas vistas mamográficas tornam os diagnósticos mais precisos, quando comparado com o uso de apenas uma vista [Blanks et al. 1999].

As técnicas de processamento digital de imagem são desenvolvidas para aumentar a

qualidade dos mamogramas, realçar aspectos e torná-los mais visíveis, a fim de que possam ser identificados sinais significantes e quantificá-los [Mudigonda et al. 2001]. Atualmente um fator de crescente revolução na análise digital é o desenvolvimento de métodos computacionais que auxiliem na avaliação das informações contidas na imagem. Essa análise digital deve identificar atributos clinicamente relevantes para diagnóstico do câncer de mama. Dentre os métodos computacionais envolvidos nessa análise, podemos destacar: o realce da imagem, descrição e representação das regiões, reconhecimento dessas estruturas, sendo esta última de grande importância em nosso trabalho.

Nosso objetivo é desenvolver um sistema computadorizado para auxílio no diagnóstico ou ("*Computer-Aided Diagnosis*" - CAD) que faça a detecção e reconhecimento de estruturas que sejam clinicamente relevantes, sendo posteriormente apresentadas ao especialista como uma segunda opinião ou uma estratégia de "pre-leitura" da imagem. Os sistemas computadorizados para auxílio no diagnóstico têm mostrado um papel importante no auxílio em laudos médicos. Estudos mostram que os CADs melhoraram na acurácia do diagnóstico, assim como a consistência da interpretação da imagem mamográfica, mediante o uso da resposta do computador como referência.

Na análise digital a maior dificuldade encontrada na segmentação de regiões de interesse são nas mamas de alta densidade, ou seja, mama que possuem uma grande quantidade de tecido fibroglandulares. A mama densa é comum em mulheres mais jovens. Em nossos primeiros estudos realizados foi observado que a densidade da mama é o fator que mais dificulta na detecção de lesões, podendo produzir altos índices de falso-positivos. Isso ocorre porque as lesões podem estar "obscurecidas" pelo tecido denso. Em nosso estudo trabalharemos técnicas capazes detectar regiões suspeitas e reduzir os índices de falso-positivos nas mamas de densidade alta.

Na tentativa de minimizar as dificuldades relacionadas acima, alguns pesquisadores têm proposto métodos para segmentar e modelar a mamografia em regiões. No trabalho de [Miller and Astley 1993] sobre detecção automática de assimetrias, discutem a necessidade da segmentação dos órgãos que compõem a mama, como forma de extrair informações mais próximas às utilizadas pelos radiologistas. [Aylward and Hemminger 1998], propõem um sistema para modelar a imagem do mamograma em cinco componentes distintos: fundo da imagem, tecido adiposo não comprimido, tecido adiposo, tecido denso e músculo peitoral. Essas pesquisas estão focadas na segmentação da mama em partes a fim de facilitar outros algoritmos na detecção de possíveis tumores.

Posteriormente, podemos destacar trabalhos onde o principal objetivo é a detecção de tumores. [Mudigonda et al. 2001] aplicou desenvolveram técnicas para a detecção de lesões, verificando aspectos como tamanho e contraste. [Nishikawa et al. 1993] desenvolveram técnicas para identificar agrupamentos de microcalcificações, usando extração de atributos como textura, distribuição espacial, tamanho e contraste. Logo em seguida esses mesmos autores desenvolveram a transformação área-ponto recursiva e novos métodos para identificação das microcalcificações agrupadas. [Davies and Dance 1990] propuseram algoritmos computacionais baseados em "thresholding" para a detecção de lesões.

Inserido em um projeto de âmbito maior, o módulo a ser desenvolvido neste trabalho "*Detecção de Regiões de Suspeitas em Mamogramas*" fará parte do "Atlas Indexado de Mamografias Digitais" (AMDI). O AMDI é um projeto desenvolvido na Faculdade de Computação da Universidade Federal de Uberlândia pelo grupo de banco de dados. O sistema integra módulos que permite adicionar novos casos na base de dados mamográficas realizados por radiologistas autorizados, também auxilia em pesquisas e atividades de educação sobre câncer de mama,

com uma interface de fácil utilização e flexível toda sendo acessado via web. Ao final desse trabalho, teremos o módulo responsável pela detecção de lesões integrado ao AMDI que auxiliará especialistas e pesquisadores no diagnóstico precoce e detecção do cancer de mama.

## 2. Trabalhos Relacionados

Diversos trabalhos foram desenvolvidos na área de processamento digital de imagem para detecção de regiões suspeitas em mamografias. Os trabalhos aqui relacionados serviram como primeiro impulso para concepção e entendimento do problema.

1. [Mudigonda et al. 2001]. Neste artigo é descrito um método para detecção automática de lesões em mamografias. A técnica utiliza filtros de suavização Gaussiano e sub-amostragem como etapas de pré-processamento. Na detecção das regiões suspeitas, é aplicado uma técnica chamada Iso-Intensity, que analisa a intensidade das regiões suspeitas em vários níveis de intensidade. Em seguida é aplicado princípios Chain-Code para descrever os contornos das regiões. É utilizado um método para redução de falsos-positivos baseado na orientação do fluxo do contorno que classifica de região como falso-positivo ou verdadeira. A base de dados utilizada foi *omni Mammographic Image Analysis Society database* com 56 imagens incluindo 30 casos benigna, 13 casos malignos e 13 casos normais. A avaliação do método apresentou uma sensibilidade de 81% a 2.2 FPs/imagem (Falsos-Positivos por Imagem).
2. [Santos and Schiabel 2002]. Neste trabalho é apresentado uma técnica para detecção de nódulos em imagens mamográficas de mamas densas. Para a segmentação das regiões foi aplicado a transformada de Watershed. As técnicas de pré e pós processamento incluem equalização do histograma da imagem e algoritmos de rotulação de regiões. Os resultados das combinações entre as técnicas citadas no trabalho detectaram aproximadamente 93% dos nódulos existentes na base de dados analisada.
3. [Oliver et al. 2007]. Neste trabalho é apresentado um método para redução de falsos-positivos em imagens mamográficas utilizando o algoritmo (2DPCA) *Two-Dimensional Principal Component Analysis*, recentemente desenvolvido para reconhecimento de faces. A aplicação desse algoritmo para redução de falsos-positivos, a densidade da mama é um fator de peso, ou seja, a partir da classificação da mama a sensibilidade na redução de falsos positivos aumenta. Considerando a densidade da mama, a avaliação do método projetou uma Az de 0.96 melhorando a classificação de [Tourassi et al. 2003] que era de Az 0.87.
4. [Ferrari et al. 2000]. São apresentados neste artigo técnicas para a segmentação de mamogramas em regiões de interesse, sendo estas: borda da mama, músculo peitoral e disco glandular. A detecção da borda da mama é importante para eliminar artefatos presentes na parte externa da mama e delimitar a região a ser processada. O músculo peitoral é detectado pela transformada de Hough, e o disco glandular é segmentado através de método da máxima probabilidade, com o modelo de misturas Gaussianas sendo estimado pelo algoritmo "Expectation-Maximization". As técnicas de segmentação da mama em regiões foi aplicada em 66 imagens da base de dados *mini Mammographic Image Analysis Society database*. A separação de regiões de interesse mostraram ser úteis como etapa de pré-processamento em sistemas CADs para câncer de mama.
5. [Petrick et al. 2005]. Este trabalho trata da importância dos sistemas computadorizados para auxílio no diagnóstico do cancer de mama, e também mostra como a utilização das duas vistas da mama pode reduzir no número de falsos-positivos. São apresentadas

duas técnicas para fazer a combinação entre as duas vistas (Crânio Caudal e Médio Oblíquo Lateral) que baseia-se na localização geométrica da mamografia, características morfológicas e de texturas para fazer a correspondência entre as regiões. O uso de duas vistas para detecção de regiões suspeitas aumentou a sensibilidade de 77% para 91%.

6. [Zheng and Chan 2001]. O artigo apresenta um algoritmo que combina técnicas de inteligência artificial aplicadas à transformada Wavelet para detecção de lesões. Entre as técnicas de inteligência artificial estão incluídas análises de dimensão de fractal e análise da imagem em multi-resoluções. O algoritmo foi aplicada às 322 imagem do *mini Mammographic Image Analysis Society database* e os resultados mostraram uma sensibilidade de 97,3% sob as curvas ROCs e o número de falsos positivos por imagem é de 3,92.

### 3. Metodologia e Estado da Pesquisa

O presente trabalho propõe desenvolver um modelo para detecção de regiões suspeitas. A metodologia baseou-se no método proposto Mudigonda e Rangayyan [Mudigonda et al. 2001]. No método proposto por [Mudigonda et al. 2001] utiliza-se filtros Gaussianos para suavizar a mamografia a ser avaliada, e observa-se apenas uma vista da mama para segmentação de regiões. A abordagem que realizaremos a partir deste projeto, consiste na aplicação de equações diferenciais para suavização do mamograma e utilização das duas vistas da mama para redução de falsos-positivos.

Na imagem mamográfica, as lesões são observadas a partir da características das bordas, e diferentemente do filtro Gaussiano, as equações diferenciais tem uma propriedade importante na suavização de imagens digital. Essa propriedade consiste em suavizar as regiões homogêneas e conservar as informações das regiões de borda. Essa característica nos garantirá a conservação das informações em regiões de transição, sendo as informações de bordas são importantes na etapa de segmentação. Alguns dos primeiros testes mostraram que as equações diferenciais oferecem melhores resultados na segmentação de mamogramas.

A avaliação dos resultados obtidos pelo método de detecção de contorno da mama será realizada a partir de uma base de dados mamográficas, que está sendo construída, as imagens são do DDSM - Digital Database Screening Mammography [Heath et al. 2000]. Essa base de dados DDSM é pública sendo uma das mais completas atualmente, conta com mais de 10480 imagens correspondentes a 2620 casos, todos estão classificados segundo o sistema BI-RADS [Obenauer et al. 2005], e se organizam em quatro tipo de volumes para facilitar na classificação:

- Normal (12 Volumes, 695 casos)
- Câncer (15 Volumes, 914 casos)
- Benigna (16 Volumes, 1011 casos)

A base de dados sendo construída será disponibilizada via Web, contendo imagens num formato bem suportado, e os casos serão associados a um laudo mais completo.

Em nosso trabalho a redução de falsos-positivos que será feita a partir das duas vistas da mama, (crânio-caudal "CC" e médio oblíquo lateral "MLO"). A maioria dos sistemas CAD apresentados na literatura analisam apenas uma vista mamográfica para detecção automática de lesões e microcalcificações [Petrick et al. ] [Mudigonda et al. 2001] estudos indicam que a análise de duas vistas da mama resulta em diagnóstico mais preciso. Poucos trabalhos usam duas vistas mamográficas na detecção de regiões suspeitas [Chan et al. 1999], [Baker et al. 2003] [Good et al. 1999]. Em seguida iniciaremos o desenvolvimento e estudos de novos descritores de características para discriminar se a região detectada é normal ou suspeita.

#### 4. Cronograma do Trabalho até a Defesa

Abaixo estão descritas as fases a serem realizadas durante o curso.

1. Estado da Arte
2. Avaliação do método de detecção de lesões proposto por Mudigonda e Rangayyan [Mudigonda et al. 2001];
3. Aplicação de um filtro de suavização baseado em equações diferenciais;
4. Desenvolvimento de uma nova estratégia para redução de falsos positivos usando informações obtidas a partir da segunda vista mamográfica da mama;
5. Desenvolvimento de um sistema para classificar regiões como normais ou suspeitas e avaliação dos resultados;
6. Avaliação do método proposto usando a base de dados do DDSM;
7. Redação de um artigo a ser submetido à congresso científico;
8. Redação da dissertação e defesa do mestrado.

Meses												
Fase	1	2	3	4	5	6	7	8	9	10	11	12
1	•	•			•	•		•	•	•		
2	•	•	•									
3		•	•	•								
4				•	•	•						
5						•	•	•				
6								•	•			
7								•	•	•		
8									•	•	•	•

Table 1. Cronograma

## References

- Aylward, S. R. and Hemminger, B. M. (1998). Mixture modeling for digital mammogram display and analysis. pages 305–312. Proceedings of the 4th International Workshop on Digital Mammography.
- Baker, J. A., L.Rosen, E., Lo, J. Y., Gimenez, E. I., and Walsh, R. (2003). Computer-aided detection (cad) in screening mammography: Sensitivity of commercial cad systems for detection architectural distortion. pages 1083–1088. American Journal of Roentgenology.
- Barros, A. C. S. D., de Oliveira Ferreira, J. M., and Oliveira, J. F. P. (2007). Estatística do câncer em 2007 no brasil.
- Blanks, R., Wallis, M., and Given-Wilson (1999). Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. pages 152–158. Med. Screen.
- Bloomquist, A. K., Yafee, M. J., and Mawdsley, G. E. (2005). Quality control for digital mammography in the acrin dmist. Med Phys.
- Chan, H.-P., Sahiner, B., Helvie, M. A., Petrick, N., Roubidoux, M. A., Wilson, T. E., Adler, D. D., Paramagul, C., and s. Newman, J. (1999). Improvement of radiologist' characterization of mammographic masses by using computer-aided diagnosis: An roc study. pages 817–827. S. Radiology.
- Davies, D. H. and Dance, D. R. (1990). Automatic computer detection of clustered calcification in digital mammograms. pages 1111–1118. Phys. Med. Biol.
- Ferrari, R. J., Rangayyan, R. M., Desautels, J. E. L., and Frere, A. F. (2000). Segmentação de mamogramas: Identificação da borda da mama, músculo peitoral e disco glandular. 5th International Workshop on Digital Mammography.
- Good, W. F., Zheng, B., Chang, Y.-H., Wang, X.-H., Maitz, G. S., and Gur, D. (1999). Multi-image CAD employing features derived from ipsilateral mammographic views. In Hanson, K. M., editor, *Proc. SPIE Vol. 3661, p. 474-485, Medical Imaging 1999: Image Processing, Kenneth M. Hanson; Ed.*, volume 3661 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 474–485.
- Heath, M., Bowyer, K., Kopans, D., Moore, R., and Jr., P. K. (2000). The digital database for screening mammography. pages 212–218.
- Miller, P. and Astley, S. (1993). Automated detection of mammographic asymmetry using anatomical features. pages 1461–1476. International Journal of Pattern Recognition and Artificial Intelligence.
- Mudigonda, N. R., Rangayyan, R. M., and Desautels, J. E. L. (2001). Detection of breast masses in mammograms by density slicing and texture flow-field analysis. pages 1215–1227. IEEE Trans. Med. Imaging.
- Nishikawa, R. M., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., Metz, C. E., Wu, C. Y., Yin, F.-F., Jiang, Y., Huo, Z., Lu, P., Zhang, W., Ema, T., Bick, U., Papaioannou, J., and Nagel, R. H. (1993). Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. In Acharya, R. S. and Goldgof, D. B., editors, *Proc. SPIE Vol. 1905, p. 422-432, Biomedical Image Processing and Biomedical Visualization, Raj S. Acharya; Dmitry B. Goldgof; Eds.*, volume 1905 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 422–432.

- Obenauer, S., Hermann, K. P., and Grabbe, E. (2005). Applications and literature review of the bi-rads classification. *Eur Radiol*, 15(5):1027–1036.
- Oliver, A., Lladó, X., Martí, J., Martí, R., and Freixenet, J. (2007). False positive reduction in breast mass detection using two-dimensional pca. In *Web site*, pages 154–161. Springer-Verlag Berlin Heidelberg.
- Petrick, N., Chan, H.-P., Sahiner, B., and Wei, D. An adaptive density-weighted contrast enhancement filter for mamographic breast mass detection.
- Petrick, N., Paquerault, S., Chan, H.-P., Sahiner, B., and Helvie, M. A. (2005). Improvement of computerized mass detection on mammograms: Fusion of two-view information. *Med. Phys.*
- Santos, A. M. R., de Oliveira Ferreira, J. M., and Oliveira, J. F. P. (2006). Estimativa 2006 - incidência de câncer no brasil.
- Santos, V. T. and Schiabel, H. (2002). Segmentação de imagens mamográficas para detecção de nódulos em mamas densas. Universidade Federal de São Carlos.
- Tourassi, G., Vargas-Voracek, R., Catarious Jr, D., and Floyd Jr, C. (2003). Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. *Med Phys*, 30(8):2123–30.
- Zheng, L. and Chan, A. K. (2001). An artificial intelligent algorithm for tumor detection in screening mammogram. pages 559–567. IEEE TRANSACTIONS ON MEDICAL IMAGING.