

# Extração de Características de Contornos Côncavos para Diagnóstico de Câncer de Mama

Ricardo Costa da Silva Marques<sup>1</sup>, Marcos Vinicius Benevides<sup>1</sup>,  
Geovane Menezes Ramos Neto<sup>1</sup>, Anselmo Cardoso de Paiva<sup>1</sup>, Geraldo Braz Júnior<sup>1</sup>

<sup>1</sup>Núcleo de Computação Aplicada – Laboratório de Mídias Interativas  
Universidade Federal do Maranhão (UFMA) Campus do Bacanga – São Luís – MA

{rikardomarques, markus, geovane, paiva, geraldo}@nca.ufma.br

**Abstract.** *Breast cancer has the highest rate of occurrence in the female population which accounts for 22% of new cases each year. An approach to early detect such anomaly is the mammography image. However, complex image patterns and the different organization of the breast tissues requires skill and experience by a trained physician to avoid faults in the mammograms interpretation. The main goal of this work utilizes concave geometry for extracting characteristics of mass regions with the purpose of making a diagnosis regarding the pattern of malignancy based on that the shape that circumscribe malignant masses are very irregular.*

**Resumo.** *O câncer de mama tem o maior índice de ocorrência na população feminina que responde por 22% dos novos casos a cada ano. Uma abordagem para detectar precocemente tais anomalias é a imagem de mamografia. No entanto, os padrões de imagem complexos e a diferente organização dos tecidos mamários requer habilidade e experiência de médicos treinados para evitar falhas na interpretação de mamografias. O objetivo principal deste trabalho é utilizar geometria côncava para extrair características de regiões em massa com o objetivo de fazer um diagnóstico quanto ao padrão de malignidade com base em que a forma que circunscreve massas malignas são muito irregular.*

## 1. Introdução

Entre as mulheres, o tipo de câncer mais frequente é o de mama com 57.120 novos casos a cada ano, o que representa 22% de todos os novos casos de câncer no grupo feminino [INCA 2015]. Só para a região Nordeste tem uma estimativa para o ano de 2015 de 10.490 novos casos de câncer de mama, seguindo a média de incidência brasileira.

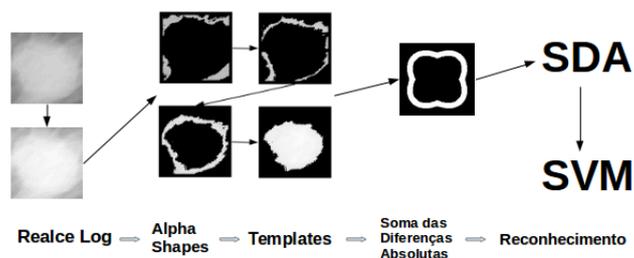
Por sua vez, uma das maneiras para detectar os tumores não palpáveis que causam câncer de mama é justamente a utilização de programas de rastreamento através de mamografias. A partir desse procedimento, foi observada uma redução da taxa de mortalidade associada a essa patologia. Entretanto, a sensibilidade desse exame pode variar, em decorrência da qualidade do mesmo, o que acaba resultando em falhas nos diagnósticos.

Nesse contexto, sistemas de detecção e diagnóstico por computador têm sido propostos com o objetivo de auxiliar o radiologista, indicando áreas suspeitas, bem como anormalidades mascaradas. Estudos demonstram, que o índice de detecção da presença de câncer de mama poderia ser aumentado de 5% a 15% se ferramentas CAD fossem utilizadas [Freer and Ullissey 2001].

Este trabalho apresenta uma metodologia em desenvolvimento para realizar o reconhecimento de regiões de massa através de características de forma. O objetivo principal é reconhecer regiões extraídas de mamografias que representem massas em benignas e malignas, e assim, disponibilizar uma segunda opinião para o especialista.

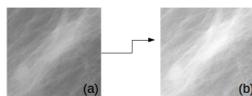
## 2. Metodologia

A metodologia consiste em fazer a análise de concavidade do contorno côncavo extraído de massas, estando subdividida em 5 passos (Figura-1): pré-processamento com realce logarítmico para evidenciar características da imagem, quantização não linear para separar formas contidas de conjuntos de tonalidades distintos e obtenção do Alpha-Shapes de cada nível de quantização, templates de formas para extração de medidas de concavidade através da soma das diferenças absolutas (SDA) entre os templates e os Alpha-Shapes gerados e por fim o uso da Máquina de Vetores de Suporte (SVM) para o reconhecimento de padrões e realização do diagnóstico.



**Figura 1. Etapas da metodologia: A imagem original é realçada, em seguida aplica-se a quantização, assim para cada conjunto de pontos obtém-se um Alpha-Shapes, estes serem comparados a um Template de modo que medidas extraídas dessas comparações sirvam de entrada para a SVM.**

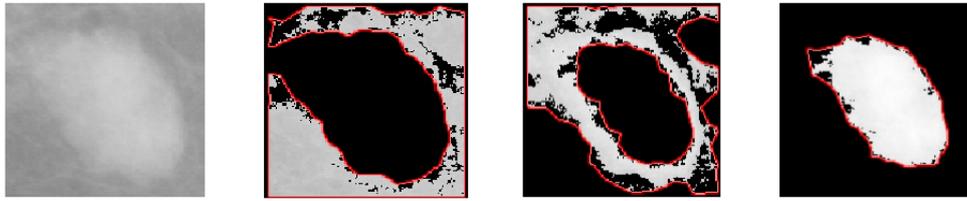
Para pré-processamento, toda imagem é submetida ao Realce Logarítmico [Gonzalez and Woods 2010]. Esta técnica de realce tem como objetivo evidenciar os valores de pixels mais escuros em tons mais claros de cinza, propiciando uma melhor obtenção dos diferentes níveis de quantização (Figura-2).



**Figura 2. (a) Imagem original, (b) Imagem realçada.**

Após o realce logarítmico, cada imagem é quantizada de maneira não linear. Aqui pretende-se dividir a imagem em grupos de tonalidades diferentes, onde inicialmente são determinados todas as tonalidades presentes em uma imagem e em seguida a quantização é realizada sobre este grupo de tonalidades. O resultado, uma nuvem de pontos, é processado pelo Alpha-Shapes para gerar contornos côncavos[Mucke 1994].

Um Alpha-Shapes é uma técnica para definir uma forma a partir de um conjunto qualquer de pontos, sendo que este é capaz de representar contornos côncavos. É interpretado como uma generalização do fecho-convexo. Um exemplo de sua aplicação está em uma imagem de massa quantizada em três níveis apresentado na Figura-3.



**Figura 3. Alpha-Shapes aplicado para três níveis de quantização, onde mais a esquerda estão as faixas de níveis mais baixos e mais a direita a faixa de níveis mais altos.**

Cada sub-imagem resultante do Alpha-Shapes possui 1 ou mais contornos côncavos que serão comparados usando SDA com um template que representa aspectos que desejamos buscar nessa imagem. Os templates utilizados são exemplificados na Figura-4. Tipicamente, uma massa irregular pode possuir uma característica de malignidade, enquanto que uma massa de contorno regular (circular) normalmente é benigna [Duarte 2006]. Nosso método ainda captura o comportamento interno do contorno através da utilização de contornos côncavos de conjunto de tonalidades obtido na quantização não linear.



**Figura 4. Templates utilizados na comparação de formas. O primeiro template tenta capturar regiões circulares, o segundo levemente irregulares e o terceiro muito irregulares.**

Finalmente, o SDA obtém a distância que o template se encontra do caso real. Nossa suposição é encontrar diferenças suficientes para classificação de massas em benignas ou malignas usando Máquina de Vetores de Suporte (MVS) [Vapnik 1998].

Entre os algoritmos que implementam o SVM, foi utilizado a implementação SMO (*Sequential Minimal Optimization*) [Platt et al. 1999] porque este consome menos memória e sua complexidade computacional é menor comparado a outras implementações. O kernel utilizado foi o radial e seus parâmetros foram ajustados usando o pacote Weka [Hall et al. 2009].

### 3. Resultados

A metodologia está sendo validada com o uso de ROIs contendo massas obtidas da base de imagens MIAS [Suckling et al. 1994]. Foram obtidas 67 regiões benignas e 50 malignas.

Foram realizados testes com variação do nível de quantização de 3 a 10. As características extraídas foram testadas separadas por nível de quantização e também unidas em um mesmo volume de treinamento.

Os testes iniciais da metodologia tiveram resultados entre 61% e 67% quando em cada amostragem utilizava-se um nível de quantização por vez, ao passo que ao juntar os níveis em um único parâmetro tivemos um resultado inferior. Esses resultados já eram

esperados já que ao usar templates, ocorrem problemas de posicionamento das formas das massas. Pois estas, podem acabar ficando fora da área de comparação desejada. Portanto, se o template acaba não se posicionando corretamente na ROI não teremos uma comparação bem sucedida. Uma vez que as medidas obtidas geram valores fora dos intervalos esperados pra esse tipo de comparação.

#### 4. Conclusão e Próximos Passos

O trabalho em andamento apresentou uma metodologia para extrair regiões de câncer de mama baseado em características de forma. O resultado encontrado até então demonstra que a abordagem de templates fixos não produz generalização suficiente para ter boa discriminação.

Tendo isto como base, a próxima etapa da metodologia baseia-se na extração de subpadrões de contorno a serem representados com Chain Code [Freeman 1961]. A ideia é quantificar através de histogramas de subpadrões [Berretti et al. 2000] de contorno, como por exemplo, vales e retas. Como essa maneira encontrada para analisar as curvaturas das formas dos subpadrões obtidos, não sofre interferências de translação, escala e rotação. A suposição é diferenciar malignos de benignos baseado na informação de concavidade de suas formas.

#### Referências

- Berretti, S., Del Bimbo, A., and Pala, P. (2000). Retrieval by shape similarity with perceptual distance and effective indexing. *Multimedia, IEEE Transactions on*, 2(4):225–239.
- Duarte, D. L. (2006). *A Mama em Imagens*, volume 1. Guanabara/Koogan, Rio de Janeiro.
- Freeman, H. (1961). On the encoding of arbitrary geometric configurations. *Electronic Computers, IRE Transactions on*, (2):260–268.
- Freer and Ulissey, M. (2001). Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*, 220(3):781–786.
- Gonzalez, R. and Woods, R. (2010). *Processamento Digital de Imagens*. Pearson Prentice Hall, São Paulo, 3 edition.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- INCA (2015). Estimativas 2014-2015: Incidência de Câncer no Brasil. <http://http://www.inca.gov.br/estimativa/2014/index.asp?ID=2/>.
- Mucke, H. E. (1994). Three-dimensional alpha shapes. *ACM Trans. Graph*, 13:43–72.
- Platt, J. et al. (1999). Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods—support vector learning*, 3.
- Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., et al. (1994). The mammographic image analysis society digital mammogram database. In *Exerpta Medica. International Congress Series*, volume 1069, pages 375–378.
- Vapnik, V. (1998). *Statistical Learning Theory*. Wiley New York, New York.