

Processamento automatizado de imagens para o cálculo de dimensão fractal de nódulos mamários malignos e benignos

Leonardo Azeredo¹, Raíza Santana¹ e Válber Laux¹

¹Universidade Federal Rural do Rio de Janeiro
Instituto Multidisciplinar - Departamento de Ciência da Computação
Avenida Governador Roberto da Silveira, Nova Iguaçu – RJ

Abstract

O câncer de mama é, hoje, uma das maiores causas de morte entre as mulheres brasileiras. O diagnóstico precoce, através da mamografia, poderia evitar a morte de muitos dos pacientes que só o descobrem em estado avançado. Esse trabalho propõe uma investigação sobre a possibilidade de um algoritmo automatizado de leitura, tratamento e cálculo da dimensão fractal para imagens de mamografias que auxilia na avaliação de um nódulo mamário detectado como benigno ou maligno. Os testes realizados apontam para uma dificuldade em tornar esse processo automático dadas as características heterogeneidades das imagens obtidas.

Keywords: Câncer de Mama, Dimensão Fractal, Computação Gráfica

1. Introdução

O câncer de mama é o mais comum entre mulheres em todo o mundo. Em 2012, representou 25% do total de casos de câncer no mundo, com aproximadamente 1,7 milhão de casos novos naquele ano [O*13].

Os nódulos são o sintoma mais comum do câncer de mama. No caso dos tumores benignos, geralmente, as bordas são suaves e arredondadas, bem definidas e de baixa densidade; já os tumores malignos, geralmente, apresentam bordas mal definidas e contornos irregulares, além da alta densidade, forma semelhante às formas encontradas nos fractais.

Fractais são objetos geométricos que não respeitam a tradicional geometria euclidiana. [MAN91] Eles são autossimilares e de escala, o que significa um fractal é composto por partes reduzidas com forma semelhantes à dele próprio. Ao aproximarmos tanto quanto se queira o objeto, veremos infinitas repetições de seu padrão. Eles representam muitos objetos da natureza onde a geometria euclidiana falha em representar. Para o caso dos nódulos em questão, os fractais são um bom exemplo de representação mais fiel do objeto.

A mamografia é o exame mais popular e um dos mais eficientes para detecção de câncer de mama aqui no Brasil [MDL03]. Especialistas realizam análises nas imagens obtidas através desse exame, buscando identificar e classificar

tumores. O diagnóstico precoce aumenta significativamente as chances de cura da doença e como, até o presente, não há formas de prevenção para a patologia, essa é uma prática importante.

A mamografia digital é um melhoramento importante na imagiologia mamária, a detecção assistida por computador (CAD) tem ajudado na detecção precoce destacando regiões suspeitas num mamograma, possibilitando uma análise mais efetiva e, por consequência, a redução de mortes causadas pela doença [TCK08].

A fim de otimizar a detecção de áreas suspeitas, o trabalho propõe a aplicação de técnicas de processamento digital de imagem nas mamografias com objetivo permitir o cálculo da dimensão fractal desses nódulos e avaliar, de forma automática, os nódulos cancerígenos.

2. Objetivos

Segundo demonstrado experimentalmente em [ESGJA14], é possível classificar nódulos detectados em imagens mamográficas em grupos de benignos e malignos, dado que a dimensão fractal apresentada pelo segundo grupo é ligeiramente maior do que do primeiro grupo. No entanto, a preparação da imagem para cálculo da dimensão fractal é um desafio, já que cada imagem apresenta diferentes características visuais e nódulos podem estar mais ou menos

visíveis, dependendo da imagem. O presente trabalho consiste em uma proposta de tratar e classificar de forma automática mamografias e dizer se os nódulos identificados têm ou não maiores chances de se tratarem de nódulos malignos.

3. Metodologia

A análise aqui proposta é baseada em um conjunto de mamografias obtidas em [SPD*94]. Inicialmente é realizado o pré-processamento das imagens, que consiste em uma equalização para melhorar a distribuição de cinza nas imagens. Isso faz com que o nódulo fique mais destacado, permitindo a segmentação, que visa retirar a região que não é de interesse, ou seja, a região que não descreve o nódulo em análise. Por fim, a dimensão fractal do nódulo é calculada, utilizando para tal o método *box counting*, que apresenta bons resultados com uma simples implementação, conforme em [LB15].

Em [O*13] estão disponíveis, além das imagens de mamografias, dados sobre as características dos tumores presentes, se houverem. Dentre elas estão as coordenadas do pixel central ao tumor e o raio aproximado da região que o contém. Estes dados são usados para deperminar a região de interesse na imagem, para o cálculo da dimensão fractal.

3.1. Pré-Processamento

Inicialmente, a região de interesse é extraída da imagem original, formando uma nova imagem que contém apenas a área da lesão, identificada em [O*13]. Então, esta nova imagem passa por um processo de equalização de histograma, com o objetivo de realçar a lesão presente e facilitar o processo de binarização necessário para a realização da técnica de cálculo da dimensão fractal por meio de *box-counting*.

A imagem equalizada é então tratada por dois métodos que visam diminuir o ruído natural da imagem de uma mamografia. O primeiro encontra o pixel mais claro na imagem e então elimina todos aqueles que sejam que tenham intensidade menor do que $f * i$, onde f é um parâmetro entre 0 e 1 definido arbitrariamente e i é a intensidade do pixel em questão. O segundo é similar ao anterior, diferindo ao usar sempre a intensidade i do pixel central ao tumor.

3.2. Binarização

Foi desenvolvido um método automatizado para encontrar um limiar de binarização adequado a cada imagem. De forma iterativa, a imagem é binarizada com diferentes limiares em cada iteração, a partir de um limiar baixo inicial. O resultado da binarização é então testado da seguinte forma: os pixels presentes na circunferência da região de interesse são analisados e, quando a razão entre o número de pixels brancos e pretos ultrapassa um valor p definido como aceitável, o limiar é aceito e a binarização pode ser usada.

Espera-se com esse método produzir binarizações adequadas, onde a região do tumor binarizada não ultrapasse a região de interesse definida em [O*13].

3.3. Box-Counting

Apresentado originalmente por [RHO80], foi a técnica pioneira no cálculo de dimensão fractal. É, ainda hoje, um dos métodos mais populares e mais usados para este fim.

Defini-se uma grade de malha com caixas de tamanho r que cobrem a imagem, a dimensão fractal (DF) pode ser definida como

$$DF = - \lim_{r \rightarrow 0} \frac{\log(N(r))}{\log(r)}$$

onde $N(r)$ representa o número de caixas necessárias para cobrir completamente o sinal da imagem.

Para o presente problema, a dimensão fractal é calculada apenas na região de interesse, i.e. a região onde encontra-se o nódulo. Tal informação é fornecida junto da base de imagens em [O*13], sendo composta pelos valores do centro do nódulo e raio da região que o compreende.

O tamanho das caixas para a primeira malha é determinada como 10% da largura da área de interesse. Para uma região de 100 x 100 pixels, por exemplo, a malha seria composta por 100 caixas de tamanho 10 x 10 pixels. Nas próximas iterações, a largura das caixas é dividida pela metade consecutivamente, até atingir o tamanho mínimo de 1 x 1 pixel.

Para a avaliação da quantidade de caixas necessárias para cobrir completamente o nódulo, avaliamos, para cada caixa definida na malha, se esta possui ao menos um pixel branco em seu interior. $N(r)$ é definido, dessa forma, como o somatório dessas caixas citadas. A partir das relações obtidas para cada tamanho de caixa, obtém-se a dimensão fractal do nódulo.

Algoritmo 1: Cálculo da Dimensão Fractal de nódulos mamários

Entrada: Arquivo pgm da mamografia

Saída: Dimensão Fractal

início

 região \leftarrow recorta(mamografia, x, y, raio);

 região \leftarrow pre-processa(região);

 região \leftarrow binariza(região);

$r \leftarrow 10\% * \text{largura}(\text{região})$;

enquanto $r > 1$ **faça**

para cada caixa $c \in$ malha M **faça**

se c contém ao menos 1 px branco **então**

 caixas_necessarias++;

fim

fim

$N(r) \leftarrow$ caixas_necessarias;

$r \leftarrow \frac{r}{2}$;

fim

$\text{dimenso_fractal} = -\lim_{r \rightarrow 0} \frac{\log(N(r))}{\log(r)}$;

retorna dimensão_fractal;

fim

Table 1: Resultados obtidos com pré-processamento simples e o método de binarização

Benignos	DF	Malignos	DF
mdb002.pgm	1.690087	mdb148.pgm	1.846087
mdb005.pgm	1.868483	mdb178.pgm	1.877217
mdb010.pgm	1.85022	mdb186.pgm	1.758245
mdb012.pgm	1.794945	mdb202.pgm	1.880473
mdb015.pgm	1.863722	mdb209.pgm	1.8526
mdb019.pgm	1.9104	mdb213.pgm	1.705969
mdb021.pgm	1.903481	mdb231.pgm	1.734081
mdb025.pgm	1.855956	mdb241.pgm	1.614246
mdb069.pgm	1.906945	mdb249.pgm	1.712463
mdb080.pgm	1.792481	mdb256.pgm	1.731486
mdb083.pgm	1.906098	mdb264.pgm	1.832041
mdb097.pgm	1.791798	mdb265.pgm	1.86794
mdb142.pgm	1.896851	mdb267.pgm	1.872077
		mdb270.pgm	1.823535
		mdb271.pgm	1.900385
		mdb274.pgm	1.812107
média:		média:	1.8013095

4. Experimentos computacionais

O algoritmo proposto foi implementado em linguagem C e compilado em sistema operacional Linux Ubuntu. As imagens de mamografias foram obtidas de [O*13]. Foi utilizado um conjunto de 29 mamografias, divididas em 13 que apresentam tumores benígnos e 16 que apresentam tumores malignos, do mesmo modo como em [ESGJA14], a fim de verificar se os resultados são consistentes com o trabalho citado. Em caso afirmativo, poderíamos afirmar que o pré-processamento automático conseguiu atingir os mesmos resultados do pré-processamento manual de [ESGJA14].

Os métodos automatizados de pré-processamento e binarização foram executados usando parametros de teste iniciais definidos de forma individual e empírica, ao produzirem imagens binarizadas mais claras. São eles $p = 0.3$, para o método que encontra o limiar de binarização dinamicamente, $f = 0.7$, para o método de limpeza baseado no pixel mais claro e $f = 0.8$ para o método baseado no pixel central. A dimensão fractal foi calculada nas imagens produzidas com a binarização após cada um dos métodos de pré-processamento e, então, com a combinação dos três. As tabelas abaixo apresentam os resultados.

As Tabela 1 mostra os resultados obtidos na execução do algoritmo após o pré-processamento simples e utilizando apenas o método de binarização. A Tabela 2 mostra os resultados dos cálculos nas imagens pré-processadas apenas com o primeiro método de limpeza. A Tabela 3 mostra os resultados com o segundo método. A Tabela 4 mostra os resultados quando todos os métodos são usados ao mesmo tempo, com os parâmetros especificados.

Table 2: Resultados nas imagens pré-processadas com o primeiro método de limpeza

Benignos	DF	Malignos	DF
mdb002.pgm	1.752569	mdb148.pgm	1.819398
mdb005.pgm	1.814678	mdb178.pgm	1.871489
mdb010.pgm	1.828479	mdb186.pgm	1.648122
mdb012.pgm	1.78991	mdb202.pgm	1.768241
mdb015.pgm	1.846643	mdb209.pgm	1.819261
mdb019.pgm	1.802071	mdb213.pgm	1.79021
mdb021.pgm	1.826364	mdb231.pgm	1.743071
mdb025.pgm	1.808057	mdb241.pgm	1.614246
mdb069.pgm	1.805844	mdb249.pgm	1.681056
mdb080.pgm	1.792481	mdb256.pgm	1.670914
mdb083.pgm	1.802811	mdb264.pgm	1.823693
mdb097.pgm	1.731762	mdb265.pgm	1.767748
mdb142.pgm	1.83823	mdb267.pgm	1.77444
		mdb270.pgm	1.797931
		mdb271.pgm	1.842176
		mdb274.pgm	1.813007
média:		média:	1.7653126

Table 3: Resultados usando o segundo método de limpeza

Benignos	DF	Malignos	DF
mdb002.pgm	1.792272	mdb148.pgm	1.829652
mdb005.pgm	1.770573	mdb178.pgm	1.843892
mdb010.pgm	1.707519	mdb186.pgm	1.589637
mdb012.pgm	1.771927	mdb202.pgm	1.768241
mdb015.pgm	1.815556	mdb209.pgm	1.867547
mdb019.pgm	1.802071	mdb213.pgm	1.859185
mdb021.pgm	1.800094	mdb231.pgm	1.949146
mdb025.pgm	1.762345	mdb241.pgm	1.899763
mdb069.pgm	1.754368	mdb249.pgm	1.864036
mdb080.pgm	1.792481	mdb256.pgm	1.79928
mdb083.pgm	1.815915	mdb264.pgm	1.812245
mdb097.pgm	1.72569	mdb265.pgm	1.714575
mdb142.pgm	1.83823	mdb267.pgm	1.754372
		mdb270.pgm	1.905501
		mdb271.pgm	1.842176
		mdb274.pgm	1.77528
média:	1.7961776	média:	1.8171586

Table 4: Resultados para a utilização simultânea de todos os métodos de limpeza

Benignos	DF	Malignos	DF
mdb002.pgm	1.690087	mdb148.pgm	1.819398
mdb005.pgm	1.770573	mdb178.pgm	1.843892
mdb010.pgm	1.707519	mdb186.pgm	1.589637
mdb012.pgm	1.78991	mdb202.pgm	1.768241
mdb015.pgm	1.815556	mdb209.pgm	1.819261
mdb019.pgm	1.802071	mdb213.pgm	1.705969
mdb021.pgm	1.800094	mdb231.pgm	1.734081
mdb025.pgm	1.762345	mdb241.pgm	1.614246
mdb069.pgm	1.754368	mdb249.pgm	1.681056
mdb080.pgm	1.792481	mdb256.pgm	1.670914
mdb083.pgm	1.802811	mdb264.pgm	1.812245
mdb097.pgm	1.72569	mdb265.pgm	1.714575
mdb142.pgm	1.83823	mdb267.pgm	1.754372
		mdb270.pgm	1.797931
		mdb271.pgm	1.842176
		mdb274.pgm	1.775283
média:	1.7530623	média:	1.7464548

Como pode-se constatar pelos resultados, apenas o segundo método de limpeza, aplicado individualmente, apresentou um resultado que poderia validar o método automático aqui proposto. Nas outras abordagens os resultados foram opostos ao que se esperaria das dimensões fractais dos nódulos, ou seja, que em média a dimensão dos classificados como benignos fosse menor que a dos malignos.

5. Conclusão e Trabalhos Futuros

O câncer de mama ainda faz milhares de vítimas anualmente no Brasil e milhões no mundo. O diagnóstico precoce poderia salvar a vida de muitos pacientes, que descobrem a doença em estado avançado e têm complicações no tratamento. O uso de ferramentas computacionais que apoiam o profissional médico na avaliação de um nódulo pode ser de grande valia na análise do resultado, uma vez que os nódulos nem sempre são facilmente classificáveis a olho nu. Além disso, os músculos e tecidos presentes na mama podem atrapalhar na visualização do resultado, efeito que pode ser atenuado com o uso de técnicas como a equalização de imagem, apresentada neste trabalho.

Não foi possível, no entanto, encontrar até o presente momento uma configuração válida de parâmetros para o pré-processamento e binarização que pudesse tratar as imagens de forma automática, considerando suas características gráficas distintas entre si. Assim, os testes realizados apresentaram um comportamento insatisfatório, com dimensão fractal muito distante das apresentadas no trabalho de [ESGJA14].

Possíveis explicações para tal fenômeno estão no fato de que o trabalho de [ESGJA14] baseia-se fortemente em tratamentos manuais, dificilmente replicáveis de forma automática. Sendo assim, grande quantidade de ruído se mantém ao redor dos nódulos, afetando a qualidade do cálculo da dimensão fractal. Aprimoramentos na equalização, técnica de cálculo de dimensão fractal mais refinadas ainda poderiam ser aplicados para obtenção de resultados mais precisos.

Uma possível estratégia que pode ser explorada é a experimentação automática de todas as combinações dos parâmetros, a fim de encontrar valores que produzam resultados esperados para as dimensões fractais dos nódulos, de maneira automática, para um grupo selecionado de imagens. A partir daí, verificar se estes permanecem válidos para qualquer outra imagem. Vale notar ainda que não é objetivo garantir que um nódulo é benigno ou maligno, mas sim auxiliar na avaliação feita por um profissional da área.

References

- [ESGJA14] E. S. G. JUNIOR F. B. S. O., AMBRÓSIO P. E.: Utilizando box-counting para calcular a dimensão fractal de nódulos mamários. In *XXIV Congresso Brasileiro de Engenharia Biomédica* (2014), CBEB' 14, pp. 405–408. 1, 3, 4

- [LB15] LOPES R., BETROUNI N.: Fractal and multifractal analysis: A review. *Medical Image Analysis* 13, 4 (2015), 634–649. [2](#)
- [MAN91] MANDELBROT B.: Objectos fractals, ed. *Gradiva, Col. Ciência Aberta* (1991). [1](#)
- [MDL03] MOLINA LUCIANA; DALBEN I., DE LUCA L. A.: Análise das oportunidades de diagnóstico precoce para as neoplasias malignas de mama. *Rev. Assoc. Med. Bras. [online]* 49 (2003), 185–190. [1](#)
- [O*13] ORGANIZATION W. H., ET AL.: International agency for research in cancer. globocan 2012: estimated cancer incidence, mortality and prevalence worldwide 2012, 2013. [1](#), [2](#), [3](#)
- [RHO80] RUSSELL D. A., HANSON J. D., OTT E.: Dimension of strange attractors. *Phys. Rev. Lett.* 45 (1980), 1175–1178. [2](#)
- [SPD*94] SUCKLING J., PARKER J., DANCE D., ASTLEY S., HUTT I., BOGGIS C., RICKETTS I., STAMATAKIS E., CERNEAZ N., KOK S., ET AL.: The mammographic image analysis society digital mammogram database. In *Excerpta Medica. International Congress Series* (1994), vol. 1069, pp. 375–378. [2](#)
- [TCK08] TARTAR M., COMSTOCK C., KIPPER M.: *Breast Cancer Imaging: A Multidisciplinary, Multimodality Approach*. Elsevier Health Sciences, 2008. [1](#)