

Explorando a Ripplet-II para Identificação de Lesões em Imagens Histológicas de Mama

Wagner L. Moreira Junior*, Marcelo Z. do Nascimento**, Leandro Alves Neves***

*Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, Santo André, SP

**Faculdade de Computação – FACOM - Universidade Federal de Uberlândia, Uberlândia, MG

*** DCCE, Universidade do Estado de São Paulo (UNESP), São José do Rio Preto, SP

Email: wmoreira@gmail.com, marcelo.zanchetta@gmail.com, neves.leandro@gmail.com

Abstract – *In this study, it was evaluated the Ripplet-II transform to identify textural characteristics of histological images of breast lesions, in comparison to wavelet, which proves not to be very efficient for the identification of image contours. The ReliefF and Random Forest (RaF) algorithms were applied for selection and classification of the characteristics. Rate of 0,923 for area under the ROC curve was the most significant result obtained with the proposed approach.*

Palavras-chave: Breast Cancer, Histological Imaging, Ripplet, Texture Analysis.

Introdução

O câncer de mama é um dos tipos mais comuns de câncer entre as mulheres acima de 35 anos, respondendo por cerca de 28% dos casos novos a cada ano. Em 2016, um relatório apresentado pelo INCA previa um número estatístico de cerca de 57.960 novos casos de câncer de mama no Brasil [1].

Diante dos recentes avanços na geração de bancos de dados de tecidos, e da catalogação de estudos histológicos digitalizados, agora é possível utilizar padrões de tecido histológico com análise de imagem assistida por computador para facilitar a classificação da doença [2]. A extração de dados texturais vem sendo utilizada na análise e interpretação de imagens médicas para determinação de padrões na classificação de um determinado tipo de imagem [3]. Essa etapa tem contribuído para o desenvolvimento de sistemas computacionais de apoio ao diagnóstico (*Computer Aided Diagnosis – CAD*) para diversos tipos de câncer.

Uma variedade de métodos de processamento de imagens é utilizada na obtenção de informações de textura. Uma representação esparsa e eficiente das imagens é altamente necessária no reconhecimento de padrões, numa visão computacional e no processamento de imagens. Embora vários métodos, como a transformada Wavelet e a transformada Ridgelet, tenham sido propostos na literatura, eles não conseguem resolver as descontinuidades bidimensionais (2D) ao longo de qualquer curva em uma imagem [4]. A Ripplet-II, apresentada por Xu e Wu em 2010, pretende resolver essa questão.

Juntamente com os coeficientes Ripplet-II, outras características estatísticas de textura podem ser derivadas. Uma das abordagens utilizadas para adquirir informações sobre transições de níveis de intensidade entre dois pixels é a obtida através da construção da matriz de co-ocorrência [5].

Este trabalho visa explorar um novo método para obtenção de descritores de textura em imagens histológicas de câncer de mama, utilizando a capacidade da transformada Ripplet-II em identificar e distinguir as massas benignas ou malignas de maneira mais assertiva.

Materiais e Métodos

A Figura 1 demonstra a sequência utilizada para realizar os experimentos propostos.

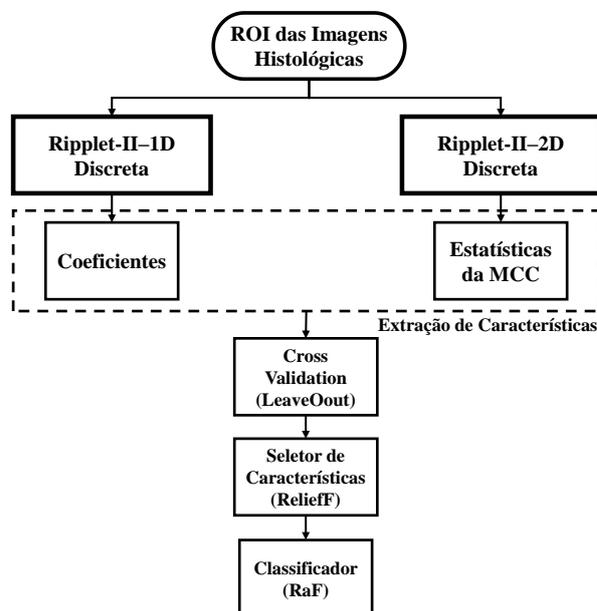


Figura 1. Etapas do experimento.

Para este estudo foram utilizadas 52 imagens histológicas digitalizadas, de câncer de mama, sendo 26 de lesões malignas e 26 de lesões benignas, extraídas de regiões de interesse (ROI) do banco de imagens do Instituto Nacional do Câncer e Instituto Nacional do Envelhecimento nos Estados Unidos, as

quais se encontram disponíveis para download em [6]. As imagens microscópicas foram adquiridas por meio de um microscópio de luz (Zeiss Axioscope), com objetiva de 20x e câmera digital colorida (AXio Cam MR5) acoplada. As imagens foram selecionadas em regiões de interesse no formato *tif*, padrão de cor RGB, resolução de 1388x1040 pixels, com quantização de 24 bits. As ROI's foram recortadas com tamanho 512x512 pixels, onde o centro da imagem corresponde ao centro da anormalidade presente. Esse tamanho foi selecionado devido aos resultados promissores em diferentes estudos presentes na literatura [7], [8].

Inicialmente foram extraídas, para cada imagem histológica, as transformadas Discretas Ripplet-II 1D e Ortogonal 2D. Para isso, utilizou-se três funções wavelet-mãe para efeito de comparação de resultados: Daubechies4, Symlet8 e a Biorthogonal1.3 [9-10].

Em seguida, a partir dos coeficientes 2D foram calculadas as medidas estatísticas da matriz de co-ocorrência. Em uma terceira etapa, combinando-se as medidas estatísticas com os coeficientes das transformadas 1D, foram geradas as amostras de treino e de teste, para cada grupo de lesão, através do método de *cross validation LeaveOneout*. Para eliminar ruídos e características não relevantes, foi aplicado o seletor *ReliefF*. Por fim, o classificador *Random Forest* foi executado para avaliar as combinações propostas. O método proposto foi desenvolvido em linguagem MATLAB e a plataforma WEKA (*Waikato Environment for Knowledge Analysis*) [11]. A seguir são detalhadas as etapas do experimento:

A. Extração da Ripplet-II Discreta

Para extração dos coeficientes Ripplet-II, as imagens passaram por um filtro para transformação das mesmas em escala de níveis de cinza.

A Ripplet-II se baseia na transformada generalizada de Radon [12], [13]. A transformada generalizada de Radon converte curvas em pontos. Ela cria picos localizados nos parâmetros da curva correspondentes. Intuitivamente, a transformada Ripplet-II consiste em duas etapas: 1) usar a transformada generalizada de Radon para converter singularidades ao longo de curvas em singularidades de pontos no domínio generalizado de Radon; 2) usar a transformada wavelet para resolver singularidades pontuais no domínio de Radon[4].

A transformada de Radon clássica é definida no espaço 2D como a integral de uma função 2D de entrada sobre linhas retas. Para uma função integral 2D $f(x,y)$, onde $(x,y) \in \mathbb{R}^2$, a transformada clássica de Radon $f(x,y)$ é definida por:

$$R(r, \theta) = \iint f(x,y) \delta(x \cos \theta + y \sin \theta - r) dx dy \quad (1)$$

Pode-se converter $f(x,y)$ para as coordenadas polares $f(\rho, \phi)$, e a Eq.(1) ficaria assim:

$$R(r, \theta) = \iint f(\rho, \phi) \delta(\rho \cos(\phi - \theta) - r) \rho d\rho d\phi \quad (2)$$

Para estender a transformada clássica de Radon, os pesquisadores propuseram a transformada generalizada de Radon, que é baseada em uma integral ao longo de uma família de curvas [13], [14]. Num sistema polar com coordenadas (ρ, ϕ) , uma curva pode ser determinada por:

$$\rho^{1/d} \cos\left(\frac{1}{d}(\phi - \theta)\right) = r^{1/d} \quad (3)$$

Onde r e θ são fixos, e d denota uma graduação. Para $d=1$ e $d=2$, a Eq. (3) representa uma linha reta e uma parábola. Quando $0 < d < 1$ ou $-1 < d < 0$, as curvas se cruzam pelo menos uma vez. Com isso, tem-se uma única curva. Caso contrário, as curvas não se cruzam, o que levará a situações complicadas. Então, considera-se somente $|d| \geq 1$. Refere-se à $d > 0$ como "curvas positivas" e $d < 0$ como "curvas negativas".

A transformada generalizada de Radon ao longo de curvas pode ser definida nas coordenadas polares (ρ, ϕ) por:

$$GR_d(r, \theta) = \iint \rho f(\rho, \phi) \delta(r - \rho \cos^d((\phi - \theta)/d)) d\rho d\phi \quad (4)$$

Segundo [4], se a entrada da transformada Ripplet-II é uma imagem digital, deve se utilizar a transformada Ripplet-II Discreta. A transformada Ripplet-II Discreta da função f pode ser obtida calculando-se primeiro a discreta GRT (DGRT) de f , e depois extraíndo-se a Wavelet Discreta 1D (DWT) da DGRT de f como:

$$f(\rho, \phi) \xrightarrow{DGRT} GR_d[f](r, \theta) \xrightarrow{1D-DWT} R_f(a, b, d, \theta) \quad (5)$$

Se for aplicada a transformada wavelet 2D aos coeficientes generalizados de Radon, a transformada wavelet adicional ao longo do ângulo θ mantém o potencial de melhorar a dispersão dos coeficientes da transformada. Os autores deram a esta nova transformada o nome de Ripplet-II Ortogonal. De maneira similar a Ripplet-II, a sua versão Ortogonal pode ser obtida por:

$$f(\rho, \phi) \xrightarrow{DGRT} GR_d[f](r, \theta) \xrightarrow{1D-DWT} R_f(a, b_1, b_2, d) \quad (6)$$

Para este estudo, três níveis de decomposição foram escolhidos para as transformadas 1D, obtendo-se um total de 479 coeficientes das sub-imagens, e 265.470 coeficientes das sub-imagens de detalhe de 2 níveis de decomposição para as transformadas 2D.

B. Estatísticas da Matriz de Co-ocorrência

Uma das técnicas muito utilizada na extração de características de textura é através do método da matriz de co-ocorrência (MCC) [14]. Uma matriz de co-ocorrência definida por P de uma imagem pode ser associada a uma direção (ângulo) θ e uma distância d entre os pixels de valor, $P(i, j, d, \theta)$ armazenado em uma linha i e coluna j . Essa matriz consiste no número de vezes em que o pixel de valor i possui vizinhos de valor j à esquerda ou à direita e a uma distância $d = 1$ entre os pixels vizinhos. Foram utilizadas quatro direções, 0° , 45° , 90° e 135° .

Cada elemento da matriz de co-ocorrência representa a frequência com que um pixel com nível de cinza i e outro com nível de cinza j ocorrem na imagem, separados por uma distância d , na direção θ , ou separados entre si de Δx colunas e Δy linhas [15].

Com o objetivo de descrever as propriedades contidas na textura de uma imagem, foram calculados o *contraste*, *correlação*, *energia* e *homogeneidade*.

C. Seleção e Classificação de Características

Com intuito de eliminar ruídos das imagens e características não relevantes para a etapa de classificação, foi utilizado o seletor de características *ReliefF* (uma extensão do *Relief*), um algoritmo com abordagem de ponderação de características, que busca vários vizinhos mais próximos para tornar o processo de seleção mais robusto a ruído, e lida com várias classes [16]. Este seletor fornece como resultado um ordenamento (ranking) das características, das mais relevantes para as menos relevantes. Neste estudo o seletor *ReliefF* selecionou apenas 12,5% das características mais relevantes, após a etapa de rank.

Para demonstrar a eficácia da técnica utilizada, é necessário utilizar-se um algoritmo de classificação. Neste estudo foi utilizado o classificador *Random Forest*. Ele tem sua definição baseada em um conceito de árvores de regressão, induzido por um método de amostragem com repetição, *bootstrap*, de um conjunto de dados de treinamento, usando descritores aleatórios, selecionados no processo de indução de árvore. Este classificador vem sendo explorado em imagens de mama, e tem se destacado conforme demonstram os estudos de [11], [17], [18]. Nesta etapa, foi utilizado o método de validação cruzada *LeaveOneout*.

Avaliações quantitativas foram realizadas por meio das métricas de *sensibilidade*, *especificidade* e *área sob a curva ROC* (*Receiver Operating Characteristic*).

Resultados

A Tabela 1 mostra o resultado das métricas de *sensibilidade* e *especificidade* da avaliação do descritor baseado em Ripplet-II. Com objetivo de uma análise comparativa, a transformada wavelet com funções Daubechies4 (DB4), Biortogonal1.3 (BIOR1.3) e Symelet 8 (SYM8) foram empregadas nesse estudo.

Características	Com Seletor (ReliefF)			
	Sens.	Esp.	Média (Acurácia)	Qtd. Caract.
Coef Ripplet-II + MCC (BIOR1.3)	80,77%	65,38%	73,08%	120
Coef Ripplet-II + MCC (DB4)	86,54%	76,92%	81,73%	70
Coef Ripplet-II + MCC (SYM8)	88,46%	61,54%	75,00%	197

Tabela 1: métricas de *sensibilidade* e *especificidade* dos descritores da Ripplet-II com as medidas estatísticas da matriz de co-ocorrência (MCC).

A Tabela 2 mostra os resultados da abordagem proposta para avaliação do descritor Ripplet-II, com ou sem o uso do seletor *ReliefF*. Nessa tabela também é apresentado os resultados do uso da transformada wavelet como descritor de características.

Discussão e Conclusão

Os resultados mostram que a combinação entre as informações Ripplet-II 1D com a função mãe DB4 e as medidas estatísticas obtiveram o melhor desempenho nos valores das médias entre *sensibilidade* e *especificidade* (ver Tabela 1).

Os resultados de *área sob a curva ROC* (Tabela 2) demonstram que as informações obtidas sob as subbandas Ripplet-II foram superiores aos dados obtidos com a transformada wavelet. Neste experimento pode-se notar que o classificador teve seu melhor desempenho quando as estatísticas da matriz de co-ocorrência estão presentes, obtendo-se **0,923 de área sob a curva ROC**, utilizando-se a DB4 como função wavelet-mãe.

Pode-se observar, diante dos resultados obtidos, que a Ripplet-II apresenta melhor desempenho que a Wavelet, quando analisadas as imagens histológicas

de lesões de mama. Em trabalhos futuros, pretende-se aplicar esta mesma metodologia com uma base de imagens radiológicas de câncer de mama, e explorar outros seletores e classificadores de características.

Características	Sem Seletor		Com Seletor (ReliefF)	
	Área ROC	Qtd. Caract.	Área ROC	Qtd. Caract.
Coef. Wavelet (BIOR1.3)	0,481	457	0,510	380
Coef. Ripplet-II (BIOR1.3)	0,519	457	0,663	380
Coef. Wavelet (DB4)	0,539	462	0,635	300
Coef. Ripplet-II (DB4)	0,596	462	0,663	300
Coef. Wavelet (SYM8)	0,587	479	0,664	30
Coef. Ripplet-II (SYM8)	0,654	479	0,673	470
Coef Wavelet + MCC (BIOR1.3)	0,692	553	0,731	120
Coef Ripplet-II + MCC (BIOR1.3)	0,712	553	0,788	120
Coef Wavelet + MCC (DB4)	0,740	558	0,779	70
Coef Ripplet-II + MCC (DB4)	0,827	558	0,923	70
Coef Wavelet + MCC (SYM8)	0,750	575	0,769	30
Coef Ripplet-II + MCC (SYM8)	0,817	575	0,885	197

Tabela 2: Resultado da Classificação, área sob a curva ROC das características combinadas com os coeficientes da Ripplet-II 1D e wavelet 1D.

Referências

- [1] INCA (2017). Estimativa 2016: Incidência de câncer no Brasil. Disponível em: <https://goo.gl/8b3TP6>, acesso em: 23/07/2017.
- [2] Gurcan, et al., Histopathological image analysis: a review. *IEEE Rev Biomed Eng*, 2, 147–171, 2010.
- [3] X. Liu, X. You, and Y. Cheung, Texture image retrieval using non-separable wavelets and local binary patterns, *Computational Intelligence and Security, CIS09. International Conference on*, vol. 1, pp. 287–291, 2009.
- [4] J. Xu, D. Wu, Ripplet-II transform for feature extraction, *Proc. SPIE 7744, Visual Communications and Image Processing, 77441R-77441R-10*, 2010.
- [5] H. Pedrini e W. Schwartz, W., *Análise de imagens digitais: princípios algoritmos e aplicações*. São Paulo: Thomson Learning, 2008.
- [6] L. Shamir, et al., A proposed benchmark suite for biological image analysis, *Medical & biological engineering & computing*, 46 (9), 943–947, 2008.
- [7] R. D. Dantas, et al., Fusion of two-view information: SVD based modeling for computerized classification of breast lesions on mammograms in Mammography, *Intech*, pp. 261–278, 2011.
- [8] R. S. Jacomini, et al., Classificação de lesões benigna e maligna combinando informações morfológicas e textura obtidas em duas visões mamográficas. *XXIII Congresso Brasileiro de Engenharia Biomédica*. 2012.
- [9] G. R. Bahendwar, Y. Sinha, Efficient Algorithm For Denoising Of Medical Images Using Discrete Wavelet Transforms *Faculty of Engineering and Technology of Shri Shankaracharya Technical Campus Bhilai, 17th Int. Conf. Math. Methods, Comput. Tech. Intell. Syst. (MAMECTIS 15) Tenerife*, no. April 2016, pp. 17–22, 2015.
- [10] P. M. Prasad, et al., Performance Analysis of Orthogonal and Biorthogonal Wavelets for Edge Detection of X-ray Images, *Procedia Comput.Sci.*, vol.87, pp.116-121, 2016.
- [11] L. Vibha, et al., Classification of mammograms using decision trees, in *Database Engineering and Applications Symposium, IDEAS'06. 10th International IEEE*, pp. 263-266, 2006.
- [12] A. Cormack, The Radon transform on a family of curves in the plane (I), *Proceedings of the American Mathematical Soc.* 83 (2), 325-330, 1981.
- [13] A. Cormack, The Radon transform on a family of curves in the plane (II), *Proceedings of the American Mathematical Soc.* 86 (2), 293-298, 1982.
- [14] R. M Haralick; K Shanmugam; Its'hak Dinstein, Textural Features for Image Classification, *IEEE Transactions on Systems, Man, and Cybernetics. SMC-3 (6): 610–621*, 1973.
- [15] E. Azevedo, A. Conci, *Computação Gráfica: Teoria e Prática*, volume 1. [S.l.]: Elsevier, Rio de Janeiro, Brasil, 11 edition, 2008.
- [16] I. Kononenko, Estimating attributes: Analysis and extension of RELIEF. *Proceedings of the European Conference on Machine Learning (pp. 171{182)*. Catania, Italy: Springer-Verlag, 1994.
- [17] G. B. Junior, et al., Classification of breast tissues using moran's index and geary's coefficient as texture signatures and SVM, *Computers in Biology and Medicine*, vol. 39, no. 12, pp. 1063-1072, 2009.
- [18] J. Liu, et al., Improved local binary patterns for classification of masses using mammography in *Systems, Man and Cybernetics (SMC), IEEE International Conference*, pp. 2692-2695, 2011.