

Diagnóstico de Câncer de Mama em Mamografias baseado na Análise da Diversidade de Padrões Circulares

Geovane Menezes Ramos Neto¹, Marcos Vinícius Benevides¹,
Ricardo Costa da Silva Marques¹, Simara Vieira da Rocha¹, Geraldo Braz Júnior¹

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

{geovane, markus, rikardo, simara, geraldo}@nca.ufma.br

Abstract. *This article aims to provide a better technique for analysis of texture in mammography images, to improve the diagnostic rate of breast cancer. The diagnosis made by a specialist depends on factors such as experience, quality of mammography and the patient's own characteristics. The study seeks to use computational methods in order to provide a second opinion to the expert, using image representation techniques in the form of quantized circular patterns through co-occurrence matrices. After the representation, the texture is calculated through the diversity of analysis and classified using Support Vector Machines.*

Resumo. *Este artigo tem como objetivo prover uma melhor técnica para análise de textura em imagens de mamografia, visando uma melhor taxa de diagnóstico dos casos de câncer de mama. O diagnóstico feito por especialistas depende de fatores como a experiência do profissional, a qualidade da mamografia e de características do próprio paciente. O estudo procura utilizar métodos computacionais a fim de oferecer uma segunda opinião ao especialista, utilizando-se de técnicas de representação da imagem na forma de padrões circulares quantizados através de matrizes de co-ocorrência. Após a representação, a textura é calculada através da análise de diversidade e classificada usando Máquinas de Vetores de Suporte.*

1. Introdução

O câncer de mama é o segundo tipo de câncer mais frequente no mundo, sendo mais comum entre as mulheres, respondendo por 22% dos casos novos a cada ano e sua ocorrência tem crescido 3,1% ao ano [ACS 2015]. Desde a última década tem-se observado um grande interesse na utilização de técnicas de análise e processamento de imagens para a detecção e diagnóstico auxiliado por computador (CADE e CADx respectivamente) em mamografias digitais, cuja finalidade é não só aumentar a precisão do diagnóstico, mas também servir de segunda opinião, auxiliando os especialistas na interpretação dos exames mamográficos [Kinoshita et al. 2004].

O diagnóstico de câncer de mama através de métodos computacionais consiste em classificar regiões extraídas de mamografias como sendo benignas ou malignas. O esforço consiste em auxiliar em um diagnóstico mais preciso e evitar um maior número de biópsias com resultados negativos.

Um importante fator utilizado para a diferenciação entre massas malignas e benignas é a textura. A textura representa a organização dos pixels em relação a sua localização

espacial na imagem e carrega a informação do padrão que pode ser capturado na forma de descrições estatísticas, estruturais ou baseada em modelos.

Este trabalho busca descrever uma metodologia em construção para extração de textura de regiões de massa usando *Circular Local Quinary Pattern* (CLQP) como representação básica para realização de Análise de Diversidade. O objetivo é usar a diversidade como extrator de textura, sendo que esta é calculada sobre um conjunto de novas representações da imagem com a ideia de melhorar o padrão descrito. Finalmente, essas informações são utilizadas para fazer o diagnóstico de câncer de mama baseado nas regiões de massa extraídas.

2. Metodologia

A metodologia proposta é representada na Figura 1, onde é apresentada a metodologia que foi desenvolvida. Inicialmente temos uma ROI¹. Na etapa de pré-processamento, o realce logarítmico [Gonzalez and Woods 2008] foi aplicado com fator 255, pois é o maior valor possível que um pixel pode assumir; e na quantização [Gonzalez and Woods 2008] utilizamos 2^n , $n = 3, \dots, 8$ fatores.

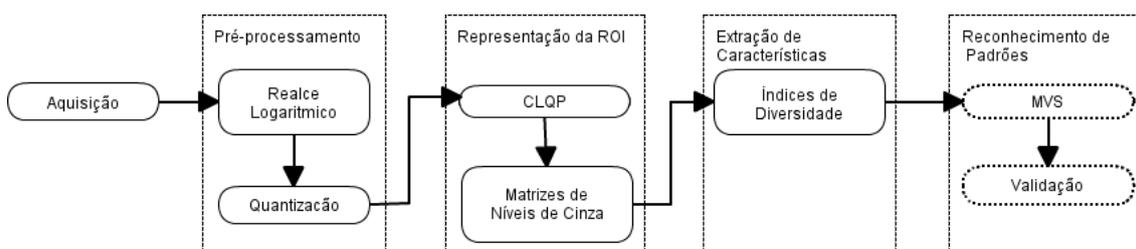


Figura 1. Fluxo da Metodologia Proposta

Na etapa Representação da ROI, o CLQP é utilizado para extrair padrões circulares da imagem. Esta é uma técnica baseada no Local Binary Pattern (LBP) [Ojala et al. 2002]. A técnica CLQP foi escolhida para esse trabalho pois a utilização de padrões quinários superam os padrões binários e em média é preferível a todas as variantes do LBP testadas em [Nanni et al. 2010]. Foram utilizados raio 1 (um) e número de vizinhos 8 (oito) além dos *thresholds* 1 e 25 (um e vinte e cinco) como parâmetros no CLQP. Os resultados gerados pelo CLQP são exemplificados na Figura 2.

O CLQP gera novas imagens que são submetidas a análise de co-ocorrência com a intenção de quantificar a importância de determinados padrões encontrados pelo CLQP e realçar a extração de textura. As Matrizes de Níveis de Cinza são bastante utilizadas para extrair características de textura. Este trabalho utiliza as seguintes abordagens de matrizes *Gray-level Co-occurrence* (GLCM) [Haralick et al. 1973], *Gray-level Run Length* (GLRLM) [Galloway 1975], *Gray-level Gap Length* (GLGLM) [Xinli et al. 1995] e *Neighboring Gray-level Dependence Matrix* (NGLDM) [Sun and Wee 1983].

De posse da importância relativa de cada padrão gerado pelo CLQP, representado através de matrizes, a textura é calculada baseada em índices de diversidade. Para [dos Santos 2009], uma medida de diversidade é um parâmetro extremamente reducionista que objetiva expressar toda a complexidade estrutural de uma comunidade ecológica

¹Region of Interest

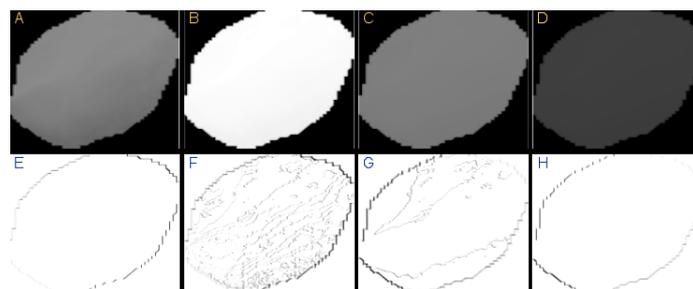


Figura 2. Imagem (A): ROI original, Imagem (B): Realce Logarítmico, Imagem (C)(D): Quantização 128 e 64 bits respectivamente, Imagem (E)(F)(G)(H): Resultados da aplicação do CLQP, em negativo para uma melhor diferenciação visual

através de um único número. O que torna extremamente simples qualquer tipo de comparação, já que uma situação se resume a um único número. Os índices de Diversidade utilizados foram os índices de McIntosh, Shannon, Simpson, Gleason, Menhinick.

Finalmente, para realizar a validação da metodologia, usamos Máquinas de Vetores de Suporte (MVS) [Chang and Lin 2011].

3. Resultados

Para os testes foi utilizado a base DDSM [Heath et al. 2000]. Ao todo foram utilizadas 131 ROI benignas e 91 malignas. As ROIS foram extraídas usando o arquivo de marcação que a própria base fornece. Os testes foram realizados usando MVS, através da ferramenta Weka. O MVS foi configurado com núcleo radial e teve o parâmetro de custo ajustado para 8,0 e γ para 2^{-8} . Os parâmetros foram ajustados usando o GridSearch do Weka. A classificação foi feita com validação cruzada com 10 *folds*, utilizando inicialmente todas as medidas geradas por todas as configurações da metodologia.

Os resultados obtidos são de 91,86% de Acurácia, 90,11% de Sensibilidade e 92,37% de Especificidade. Os resultados adquiridos se mostram promissores, pois mesmo sem uma escolha otimizada dos parâmetros do CLQP, a metodologia obteve bons resultados. Os resultados indicam uma boa discriminação entre os padrões malignos em relação aos padrões benignos, indicando que este estudo propõe uma metodologia promissora e que só tendem a aumentar as chances de um diagnóstico bem sucedido.

4. Conclusão e Próximos Passos

Este trabalho apresentou uma metodologia para descrição do padrão de massas extraídas de mamografias usando como representação a co-ocorrência de padrões espaciais capturados através do CLQP e Matrizes de Co-Ocorrência. Os resultados obtidos com a versão atual da metodologia são promissores, indicando que estes podem ser aplicados futuramente em um ferramenta CAD.

Esta metodologia ainda está em andamento, e como passos futuros pretende-se inicialmente analisar o desempenho das características separadamente para então ter como base as melhorias encontradas com o CLQP. Pretende-se ainda combinar características de geometria e textura para realizar esta tarefa, além de regular adaptativamente o CLQP, estudar outros índices para descrição de textura e forma e ainda analisar os resultados individualizando as características calculadas e com bases maiores de imagens.

Referências

- ACS (2015). Learn about breast cancer. Available at <http://www.cancer.org>, last accessed February 4, 2012.
- Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.
- dos Santos, V. K. (2009). Uma generalização da distribuição do índice de diversidade generalizado por good com aplicações em ciências agrárias. Master's thesis, Universidade Federal Rural de Pernambuco, Recife.
- Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Computer graphics and image processing*, 4(2):172–179.
- Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6):610–621.
- Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, P. (2000). The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography*, pages 212–218. Citeseer.
- Kinoshita, S., Pereira, R., Honda, M., Rodrigues, J., and Azevedo-Marques, P. (2004). An automatic method for detection of the nipple and pectoral muscle in digitized mammograms. In *Congresso Latino-Americano de Engenharia Biomédica (CLAEB'2004)*.
- Nanni, L., Lumini, A., and Brahnam, S. (2010). Local binary patterns variants as texture descriptors for medical image analysis. *Artificial intelligence in medicine*, 49(2):117–125.
- Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.
- Sun, C. and Wee, W. G. (1983). Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 23(3):341–352.
- Xinli, W., Albregtsen, F., and Foyn, B. (1995). Texture analysis using gray level gap length matrix. In *Theory and Applications of Image Analysis II: Selected Papers from the 9th Scandinavian Conference on Image Analysis*, volume 232, page 65. World Scientific.