

Detecção de lesões em imagens termográficas da mama utilizando Índice de Similaridade de Jaccard e Artificial Crawlers

Caio Nogueira Silva Belfort¹, Aristófanés Corrêa Silva¹, Anselmo Cardoso Paiva¹

¹Núcleo de Computação Avançada
Laboratório de Processamento e Análise de Imagens
Universidade Federal do Maranhão (UFMA)
São Luís – MA – Brazil

{caiobelfort90, aricsilva, anselmo.c.paiva}@gmail.com

Abstract. *This paper proposes a methodology to detection of lesions in breast thermography images, helping specialists in early diagnosis helping specialists in early diagnosis of mamma cancer. The techniques of image processing, artificial life models and computational intelligence are used. The proposed methodology presented 78% of accuracy, 50% of sensitivity and 86% of specificity.*

Resumo. *Este trabalho propõe uma metodologia para detecção de lesões em uma imagem termográfica das mamas, auxiliando médicos especialistas no diagnóstico precoce de patologias como o câncer. São utilizadas as técnicas de processamento de imagens, modelos de vida artificial e inteligência computacional. O processo apresentou 78% de acurácia, 50% de sensibilidade e 84% de especificidade.*

1. Introdução

O câncer de mama é o tipo mais comum entre as mulheres, respondendo por 22% dos novos casos a cada ano [INCA 2014]. Os exames de imagem são largamente utilizados na medicina e tem demonstrado grandes resultados na detecção de tumores em fase precoce. Atualmente a mamografia é o exame de imagem mais utilizado para detecção de lesões mamárias, porém é contra indicado para mulheres como mamas densas, como é o caso das mulheres jovens. A termografia surge como um exame alternativo que utiliza os raios infravermelhos emitidos pelo corpo humano para detectar anomalias nas mamas [Borchardt *et al.* 2013]. Este trabalho propõe uma metodologia de processamento e análise de uma termografia mamária, para destacar regiões que podem caracterizar lesões, inclusive o câncer.

2. Metodologia

As imagens utilizadas neste trabalho são provenientes do banco de imagens do projeto PROENG [PROENG 2012] e foram adquiridas do Hospital Universitário da Universidade Federal do Pernambuco (UFPE). Foram utilizadas 34 termografias, sendo 15 de pacientes que possuem algum tipo de lesão (maligna ou benigna) e 19 de pacientes saudáveis. A metodologia proposta segue o fluxograma apresentado pela Figura 1.



Figura 1. Fluxograma da metodologia.

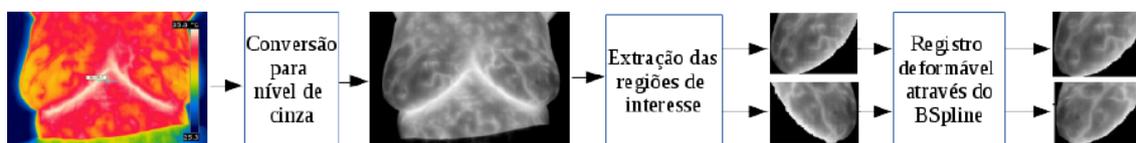


Figura 2. Etapa de pré-processamento.

2.1. Pré-processamento

A etapa de pré-processamento consiste em efetuar as operações necessárias para a aplicação da metodologia. Inicialmente é necessário a conversão da imagem original, que geralmente está em pseudocor, para uma imagem em níveis de cinza. Após a conversão, é feita a extração das regiões de interesses (mamas direita e esquerda) através da segmentação manual, realizada pelo especialista. Em seguida, efetua-se o registro não-rígido nas imagens utilizando *BSplines* [Klein *et al.* 2007], pois, é fundamental que ambas as mamas possuam formas semelhantes nas etapas futuras. Todo o procedimento pode ser visto na Figura 2.

2.2. Detecção de dissimilaridades

O aparecimento de uma lesão geralmente causa assimetria na distribuição de temperaturas entre as mamas. O objetivo desta etapa é detectar regiões onde ocorram uma maior dissimilaridade entre elas. Primeiramente aplica-se a redução da quantidade de níveis de cinza, num processo chamado quantização. Testes realizados mostraram que uma redução de 256 para 16 níveis é uma boa quantização, pois agrupa bem os pixels de valores próximos sem perder muitas características da imagem. Em seguida, percorre-se ambas as imagens resultantes da etapa de pré-processamento através de janelas 3x3, extraíndo o índice de Jaccard [Real e Vargas 1996] que indica o nível de semelhança entre elas. Pixels que fazem parte de uma região com $j < 0,85$, são marcadas como dissimilares, onde j é o índice extraído e 0,85 um limiar que foi escolhido empiricamente. Ao final do processo, uma máscara que mostra as regiões mais dissimilares entre as mamas é gerada, onde as imagens iniciais desta etapa são mapeadas. A Figura 3 exemplifica o processo.



Figura 3. Procedimento de detecção de dissimilaridades.

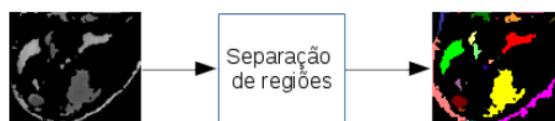


Figura 4. Etapa de separação de regiões.

2.3. Separação de regiões

O objetivo desta etapa é separar regiões de uma imagem que não possuem conexão espacial, criando novas imagens que serão analisadas separadamente. Para tal fim, foi utilizado uma variação do algoritmo de crescimento de região [Adams e Bischof 1994] onde a única condição necessária para o agrupamento de dois pixels é que eles sejam vizinhos. A Figura 4 ilustra o resultado do processo, onde as regiões de cores diferentes indicam uma nova imagem a ser analisada.

2.4. Extração de características

Nesta etapa são extraídas as características de uma imagem que serão utilizadas posteriormente na classificação. Utiliza-se o modelo dos *Artificial Crawlers* descrito em [Gonçalves *et al.* 2014] para gerar um conjunto de valores denominado vetor de características. Inicialmente duas populações de *crawlers* são inicializadas em todos os pixels que possuem valor diferente de 0 (fundo da imagem). A primeira população se move para os picos de temperatura da imagem, enquanto a segunda move-se para os vales (pixels próximos de 0, excluindo-se o 0), nesta, a absorção de energia é feita de forma inversa e é dada por:

$$t = (255 - I_{i,j} * t_x)$$

onde, t é a quantidade de energia a ser absorvida, $I_{i,j}$ é o valor do pixel na posição (i, j) e t_x é a taxa de absorção do modelo.

A lei da selva é a mesma utilizada em [Gonçalves *et al.* 2014], com a diferença que um *crawler* só compete com aqueles da mesma população que ele.

A partir de um certo instante, ambas as populações atingirão o equilíbrio, que é o estágio em que a população deixa de se alterar, com todos os *crawlers* permanecendo na mesma posição do ciclo anterior. Após o equilíbrio ser atingido, as características são extraídas como descrito em [Gonçalves *et al.* 2014]. Por fim, os parâmetros utilizados foram:

$$[E_{init} = 5, E_{max} = 35, t_x = 0, 01]$$

onde E_{init} é a energia inicial, E_{max} a energia máxima e t_x a taxa de absorção do modelo. Os raios utilizados no cálculo do volume de influência foram:

$$[\sqrt{5}, \sqrt{10}, \sqrt{15}, \sqrt{20}, \sqrt{25}, \sqrt{30}]$$

2.5. Classificação

Nesta etapa as regiões resultantes da Etapa 2.3 são classificadas como lesão ou não, de acordo com as características extraídas na Etapa 2.4. Para isto, utilizou-se a *Support Vector Machine* [Steinwart e Christmann 2008].

2.6. Resultados Parciais

Na Etapa 2.2, 26% das lesões não foram detectadas pela máscara de dissimilaridade, ocasionando a perda da lesão para as etapas seguintes. Já na Etapa 2.5 obteve-se os resultados mostrados na Tabela 1.

Acurácia	Sensibilidade	Especificidade
78%	50%	84%

Tabela 1. Resultados.

A acurácia indica a porcentagem de acerto, a sensibilidade indica o acerto de regiões que são algum tipo de lesão, e a especificidade o acerto das regiões consideradas saudáveis.

3. Conclusão

A metodologia ainda está em fase de amadurecimento, em algumas etapas (2.2 e 2.4) não foram realizados testes exaustivos e podem precisar de um maior refinamento, como na etapa de detecção de dissimilaridade onde utiliza-se um limiar encontrado empiricamente, o que pode estar ocasionando os erros mostrados. Os resultados obtidos indicam que a metodologia tem potencial e que, acrescentando novas técnicas no processo, como a extração de novas características e a automatização na escolha do limiar para a máscara de dissimilaridade, sejam encontrados resultados mais satisfatórios para o objetivo proposto.

Referências

- Adams, R. and Bischof, L. (1994). Seeded region growing. volume 16, pages 641–647. IEEE.
- Borchardt, T. B., Conci, A., Lima, R. C., Resmini, R., and Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, 93(10):2785–2803.
- Gonçalves, W. N., Machado, B. B., and Bruno, O. M. (2014). Texture descriptor combining fractal dimension and artificial crawlers. *Physica A: Statistical Mechanics and its Applications*, 395:358–370.
- INCA (2014). Tipos de cancer: Mama. http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/cancer_mama+. Acesso em: 16/10/2014.
- Klein, S., Staring, M., and Pluim, J. P. (2007). Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *Image Processing, IEEE Transactions on*, 16(12):2879–2890.
- PROENG (2012). Image processing and image analyses applied to mastology. <http://visual.ic.uff.br/en/proeng/>. Acesso em: 19/06/2012.
- Real, R. and Vargas, J. M. (1996). The probabilistic basis of jaccard's index of similarity. *Systematic biology*, pages 380–385.
- Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer.