

01

Segmentação Automática de Mamas em Imagens Infravermelhas Utilizando Limiarização com Refinamento Adaptativo em Bases Multivariadas

Matheus F. O. Baffa ¹
Deivison J. M. Cheloni ²
Lucas G. Lattari ³
Maurício A. N. Coelho ⁴

Resumo

O câncer de mama é o tipo mais comum de câncer no mundo. Anualmente são detectados milhares de casos de câncer, e destes, 25% são de mama. Sabendo que o diagnóstico precoce é crítico para o prognóstico do paciente, novas tecnologias à base de análise de imagens são desenvolvidas para guiar um diagnóstico eficaz e menos invasivo. Nesse artigo, é desenvolvido um novo método de segmentação de imagens das mamas em imagens termográficas utilizando limiarização com refinamento adaptativo. Esse método se mostrou eficaz com aproximadamente 96% de acurácia e 98% de sensibilidade na base de dados de imagens infravermelhas da Universidade Federal de Pernambuco e 93% de acurácia e 98% de sensibilidade na base DMR-UFF (Database for Mastology Research) mantidas pela Universidade Federal Fluminense. Além disso, a abordagem proposta é simples de ser implementada computacionalmente, é eficiente e apropriada para aplicações em tempo real e dispositivos móveis.

Abstract

Breast cancer is the most common type of cancer in the world. New cases of cancer are detected every year, and 25% of them are breast cancer. It is known that early diagnosis is critical for the prognosis of patients, and image based techniques has been developed to guide an effective, minimally invasive and cheaper way to diagnose breast cancer. In this work, we propose an unsupervised segmentation method based on adaptive thresholding. The experimental results show that our method is competitive with other works on both solution quality and runtime performance, achieving 96% of accuracy and 98% of sensibility on Federal Pernambuco University's infrared database and 93% of accuracy and 98% of sensibility on Database for Mastology Research (DMR) from Federal Fluminense University (UFF). Furthermore, it is simpler to implement, computationally efficient and suitable for real time and mobile applications.

Keywords: computer-aided diagnosis, adaptative thresholding, breast infrared imaging, image processing, image segmentation, computer vision.

¹IF Sudeste MG - Campus Rio Pomba, Rio Pomba, MG - Brasil, E-mail: mfreitas826@gmail.com

²IF Sudeste MG - Campus Rio Pomba, Rio Pomba, MG - Brasil, E-mail: deivison.junior@gmail.com

³IF Sudeste MG - Campus Rio Pomba, Rio Pomba, MG - Brasil, E-mail: lucas.lattari@ifsudestemg.edu.br

⁴IF Sudeste MG - Campus Rio Pomba, Rio Pomba, MG - Brasil, E-mail: mauricio.coelho@ifsudestemg.edu.br

1 Introdução

Para o ano de 2015, o Instituto Nacional do Câncer (INCA) previu um total de 500 mil novos casos de câncer somente no Brasil. Desse total, estima-se que 25% são de câncer de mama. Essas doenças afetam homens e mulheres das mais diversas idades, sendo mais frequente em mulheres acima dos 50 anos. Segundo o INCA (INCA, 2015), estima-se que, em 2013, o câncer de mama foi responsável pela morte de aproximadamente 14 mil mulheres.

A detecção precoce do câncer é muito significativa para aumentar as chances de cura do paciente. Em média, os pacientes diagnosticados no início do desenvolvimento do tumor podem ter até 95% de chance de cura (HCB, 2015). O INCA divulga que não há apenas uma causa para a doença, mas diversos fatores que desencadeiam o câncer, como fatores genéticos, hormonais e ambientais.

Atualmente, existem diversas formas de se diagnosticar o câncer de mama. Um dos exemplos mais conhecidos é o autoexame, em que a própria pessoa verifica se possui alguma anomalia. Outras maneiras são através de exames de imagens como a mamografia, o ultrassom e a ressonância magnética, em que através da verificação dessas imagens, o médico irá identificar a existência de tumores. Destas, as imagens infravermelhas aplicadas a saúde têm se mostrado promissoras, sendo utilizadas por vários pesquisadores para se detectar anomalias nas mamas (ARORA et al., 2008; MOTTA, 2010). Segundo Gore e Xu (GORE; XU, 2014), o corpo humano possui alta simetria em relação a sua distribuição térmica, que se mantém estável mesmo após longos períodos de observação. Qualquer alteração na simetria térmica pode indicar a existência de anomalias.

Para se diagnosticar câncer em imagens infravermelhas, torna-se necessário identificar a região de interesse (ROI) que compõe as mamas, removendo o fundo da cena que não é necessário para a análise. Somente a ROI é avaliada posteriormente para a identificação de

doenças.

O trabalho de (MOTTA, 2010) realiza uma segmentação automática das mamas em imagens infravermelhas. Seu algoritmo, representado na Figura 1, realiza a segmentação utilizando morfologia matemática, detecção de contornos usando filtros de Canny, detecção de formas através da transformada de Hough e limiarização adaptativa. A acurácia média da avaliação quantitativa foi de aproximadamente 96% em relação ao *ground truth*.

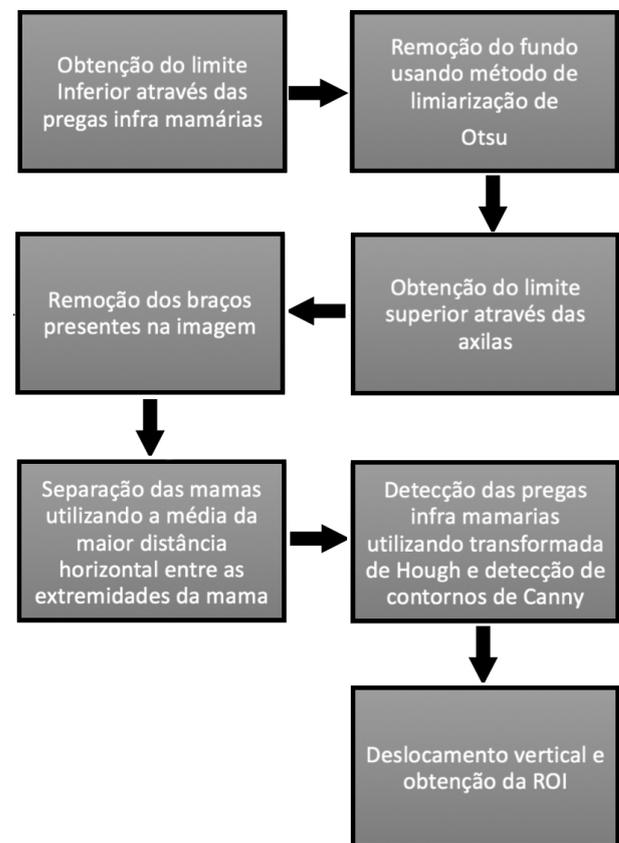


Figura 1: Fluxograma do método de segmentação proposto em (MOTTA, 2010).

O trabalho de (MARQUES, 2012) também apresenta uma abordagem diferente à segmentação automática das mamas. Seu método, descrito na Figura 2, inclui detecção de contornos, limiarização, crescimento de regiões, ajuste de curvas por mínimo quadrado e *B-Splines*.

O método de (MARQUES, 2012) apresenta uma acurácia de 97%. As imagens da ROI (*Region of Interest*) segmentadas são comparadas com um *ground truth* gerado a partir de um

aplicativo usando tablets para a segmentação manual das imagens.



Figura 2: Fluxograma do método de segmentação proposto em (MARQUES, 2012).

Nesse trabalho, é apresentado um novo método para a segmentação da região das mamas em imagens infravermelhas, usando o conceito de binarização com refinamento adaptativo. Este método foi aplicado em duas diferentes bases de imagens mastológicas infravermelhas, mantidas pela Universidade Federal de Pernambuco (UFPE) (PROENG, 2008) e pela Universidade Federal Fluminense (UFF) (SILVA et al., 2014). Os resultados obtidos nesse método se mostram muito promissores, obtendo taxas médias de 96% de acurácia e 98% de sensibilidade na base de imagens infravermelhas da UFPE e 93% de acurácia e 98% de sensibilidade na base da UFF.

Este procedimento de segmentação possui como premissa essencial a análise da simetria térmica da imagem, identificando a região que representa as pregas inframamárias. Essa área é de grande importância para a localização das mamas, pois localiza-se imediatamente abaixo

dos seios.

Essa descrição computacional mostra-se simples e intuitiva de implementar e portar em diversas arquiteturas. Isso decorre do fato dessa implementação usar somente algoritmos muito conhecidos de processamento e análise de imagens, como binarização, filtro laplaciano, morfologia matemática etc. Além disso, o tempo de execução médio e a pouca demanda por memória física permite aplicar o método proposto em outros dispositivos, como plataformas móveis. Assim, por exemplo, seria possível viabilizar a construção de um sistema CAD via *smartphone*.

Em relação ao artigo publicado anteriormente (BAFFA; CHELONI; LATTARI, 2016), o presente trabalho apresenta as seguintes contribuições:

- Duas novas etapas são introduzidas ao método original: uma etapa de remoção de membros superiores e a separação das mamas esquerda e direita. Essas etapas tem por intuito remover segmentos do corpo que não serão relevantes para posterior análise de anomalias das mamas. Elas são elucidadas na seção 2.1.
- Novos experimentos foram realizados para a base DMR-IR UFF (SILVA et al., 2014), que consiste em um total de 229 imagens infravermelhas frontais definidas para um protocolo estático, desconsiderando outros ângulos e protocolos. Dessas imagens, 80 foram selecionadas randomicamente para serem segmentadas. Esses são demonstrados junto a seção 3.3.
- A descrição de uma proposta de aprendizado de máquina para a seleção de um conjunto de parâmetros (apresentados na seção 2.1) que sejam ótimos locais. Essa abordagem é apresentada na seção 3.2.
- Melhor detalhamento e correções do método proposto foram realizados no decorrer do texto.

2 Método Proposto

As imagens termográficas das mamas são introduzidas às etapas do método proposto que é apresentado na Figura 3.

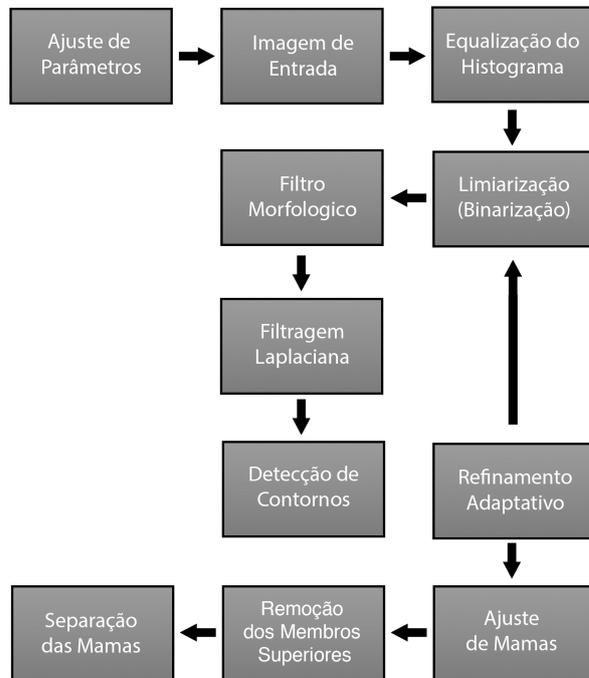


Figura 3: Fluxograma do método proposto neste trabalho.

2.1 Limiarização por Refinamento Adaptativo

Segundo (MARQUES, 2012), as pregas inframamárias são indicadores dos limites inferiores das mamas. Além disso, esta região possui maior temperatura devido a sobreposição entre a parte inferior da mama e a parede torácica. Esta característica facilita a identificação computacional da prega inframamária e das mamas.

Dessa forma, utilizando uma limiarização com limiar L elevado, torna-se possível realizar a segmentação das pregas inframamárias. Entretanto, experimentalmente, foi observado que o valor de L ideal para a correta segmentação das pregas inframamárias é variável, dependendo do conteúdo da imagem infravermelha I . Para o cálculo desse valor ideal, propõe-se uma abordagem de refinamento adaptativo,

que tem por objetivo calcular o melhor valor de L para qualquer I considerado.

No início do método, I é limiarizada com limiar elevado ($L = 254$), gerando a imagem I_{lim} . Após essa etapa, a imagem infravermelha I_{lim} é filtrada usando um operador morfológico de abertura. O objetivo da aplicação desse filtro é eliminar artefatos de I_{lim} que dificultam a identificação das pregas inframamárias.

Na próxima etapa, é realizada a filtragem laplaciana de I_{lim} . Isso é feito para a obtenção dos contornos do corpo como um todo e das pregas, permitindo uma análise geométrica que ofereça subsídios para a identificação da região mamária.

A partir dos contornos obtidos, a área A_i de cada componente conexo $i \in I_{lim}$ é calculada e comparada a um conjunto de limiares A_{l1} e A_{l2} fornecidos a priori. Estes representam áreas com dimensões ideais para a representação das pregas inframamárias.

Essa avaliação é realizada da seguinte maneira:

- É procurado um segmento $i_1 \in I_{lim}$, tal que sua área $A_{i1} > A_{l1}$.
- São procurados dois segmentos $i_1, i_2 \in I$, tal que suas áreas A_{i1} e A_{i2} sejam, respectivamente, $A_{i1} > A_{l2}$ e $A_{i2} > A_{l2}$.

Se qualquer uma das duas condições listadas anteriormente forem verdadeiras, então as pregas inframamárias foram identificadas e a etapa de refinamento adaptativo é encerrada. Caso contrário, a etapa de refinamento adaptativo é reiniciada a partir da limiarização. No entanto, o valor de L é decrementado em uma unidade ($L = 253$). A ideia é repetir iterativamente esse procedimento até que a região das pregas inframamárias seja identificada.

Logo, a redução iterativa de L permite maximizar a área pertencente às pregas inframamárias, minimizando o efeito causado pela introdução de artefatos indesejados.

Dois limiares associados às áreas são introduzidos devido a uma avaliação empírica observada durante a criação do método. Em uma série de imagens de teste consideradas,

observou-se que a região de alta temperatura associada às pregas inframamárias é composta por uma única componente conexa. No entanto, para um conjunto de imagens, notou-se que a região das pregas inframamárias é composta por duas componentes conexas de menor tamanho. Dessa forma, com a introdução das duas condições mencionadas para os limiares A_{I1} e A_{I2} , foram obtidas às regiões de prega inframamária para um número elevado de imagens advindas das bases de dados utilizadas na literatura.

Uma vez obtida a região das pregas inframamárias, a etapa a seguir, denominada ajuste de curvas, tem por intuito representar de maneira fidedigna às curvas que compõem as mamas. Essa representação é feita a partir da modelagem realizada por um polinômio cúbico, que interpola todos os pontos da região inferior das pregas obtidas na etapa anterior.

Após calculada a etapa de ajuste de mamas, inicia-se o estágio da remoção dos membros superiores da imagem. Realizando uma varredura para cada linha horizontal de I binarizada, nota-se na região corporal no mínimo a ocorrência de duas regiões de alta frequência por linha. Neste trabalho, define-se uma região de alta frequência ra como uma linha horizontal com um pixel $p(x, y) = \text{branco}$ e $p(x + 1, y) = \text{preto}$, ou $p(x, y) = \text{preto}$ e $p(x + 1, y) = \text{branco}$.

No segmento inferior da imagem, é natural ocorrer apenas duas regiões ra . No entanto, em áreas próximas aos braços e a cabeça, essas regiões tendem a ser superiores a duas por linha, pois os braços e a cabeça introduzem mais pixels brancos com elementos adjacentes de cor escura. Dessa forma, sempre que uma linha horizontal possuir mais do que $2ra$, essa linha é inteiramente colorida de preto e não será mais considerada durante a análise. Convém mencionar que essa proposta é muito similar a que foi apresentada por (MARQUES, 2012).

Empiricamente, observou-se que uma grande parcela das regiões contendo os braços e pescoços foram removidas. No entanto, como as bases de dados ainda consideram

esses elementos em seu interior, não foi possível verificar quantitativamente esse resultado. Ainda assim, é importante realizar esse procedimento para realizar a separação definitiva das mamas da esquerda e da direita.

Ademais, após a remoção dos membros superiores e do pescoço, na última etapa (denominada separação das mamas), é realizado um recorte das mamas em mama esquerda e direita através da análise do histograma vertical da imagem. Para cada coluna c da imagem, calcula-se seu histograma vertical $hist(c)$, que contabiliza o total de pixels claros de I após binarização.

Finalmente, é realizada uma varredura para cada coluna c da imagem considerada. Verifica-se sempre que $hist(c) > 0.3 \times h_I$, em que h_I é a altura de I . A partir daí, são assinaladas as colunas c_{min} e c_{max} que são, respectivamente, a menor e a maior coordenada de coluna que atende a condição supracitada.

Finalmente, é obtida a coordenada média que determina o eixo que realiza a separação mamária (Equação 1). Um exemplo do resultado final obtido pela proposta referida é visível na Figura 4.

$$c_{med} = \frac{c_{min} + c_{max}}{2} \quad (1)$$

3 Desenvolvimento

3.1 Descrição da Metodologia

A metodologia apresentada nesse trabalho foi desenvolvida utilizando o Python 2.7.8 e a biblioteca OpenCV 3.1. Para a geração de resultados e análise de tempos computacionais, foi utilizado um notebook Asus N46V com processador Intel I7, 8 GB de RAM e sistema operacional Windows 10.

Para a análise de resultados foram empregadas duas bases de dados públicas conhecidas (UFPE e DMR-UFF), totalizando 363 imagens térmicas, sendo 282 provenientes da base da UFPE e 80 da base DMR-UFF, segmentadas separadamente (SILVA et al., 2014). As imagens possuem resolução espacial de 320 x

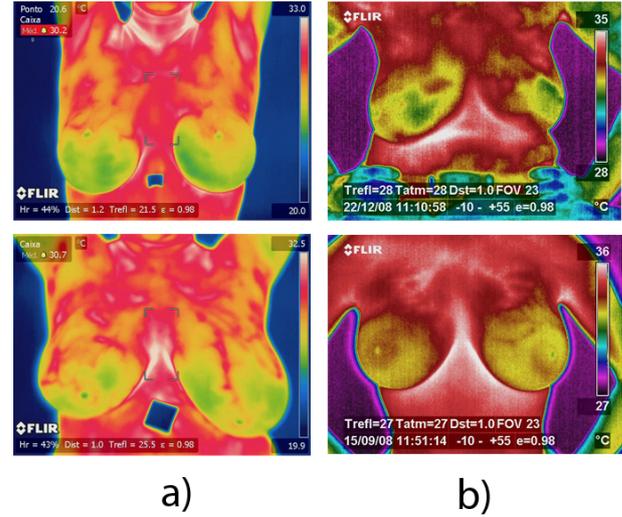
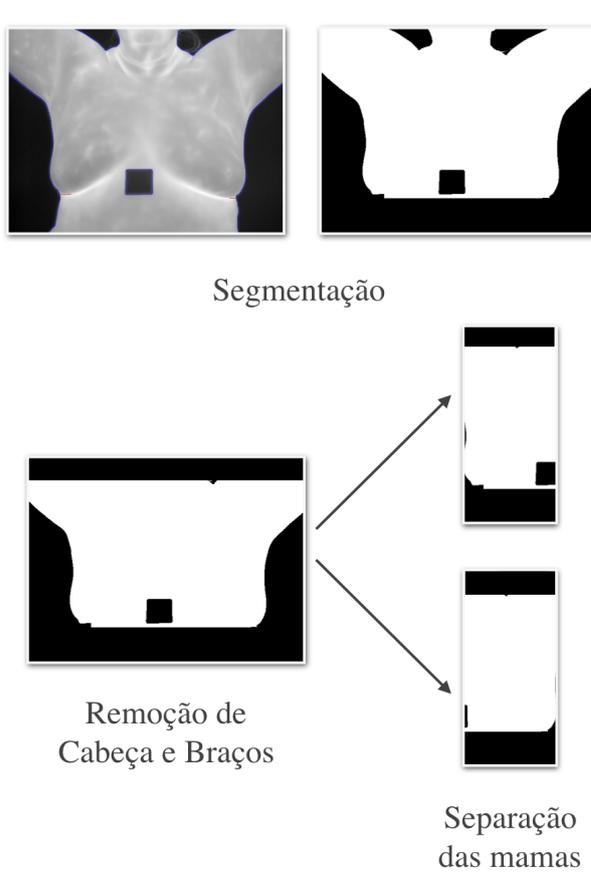


Figura 5: Algumas imagens advindas das bases de dados. Em a) uma representação da base de dados DMR-UFF apresentada em (SILVA et al., 2014) e em b) uma representação das imagens provenientes da base da UFPE disponível em (PROENG, 2008).

Figura 4: Resultado obtido na base DMR-UFF para a imagem T0023.1.1.S.2012-10-31.00.

240 pixels, além de 256 níveis de resolução tonal. Maiores informações sobre a captura das imagens utilizadas podem ser vistas no mesmo trabalho. Alguns exemplos de imagens da base podem ser vistos na Figura 5.

Para cada imagem de teste disponível na base de dados, é gerada sua versão *ground truth*. Essas imagens são construídas a partir de um método de política de votos apresentado em Li et al. (LI et al., 2011), que utiliza diferentes imagens *ground truths* obtidas manualmente. Essas imagens são comparadas com as obtidas pelo método proposto nesse artigo.

Para avaliar a eficiência desse método, diferentes métricas estatísticas serão consideradas: acurácia (ACC), sensibilidade (SEN), especificidade (ESP), preditividade positiva (PDP) e negativa (PDN). Tais medidas são definidas pelas equações:

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (2)$$

$$SEN = \frac{VP}{VP + FN} \quad (3)$$

$$ESP = \frac{VN}{VN + FP} \quad (4)$$

$$PDP = \frac{VP}{VP + FP} \quad (5)$$

$$PDN = \frac{VN}{VN + FN}, \quad (6)$$

em que VP representa as amostras de verdadeiros positivos (pixel p identificado como parte da região da mama, tanto na imagem teste quanto em seu *ground truth* na mesma posição (x, y)), VN representa os verdadeiros negativos considerados (pixel p identificado como parte da região do fundo, tanto na imagem teste quanto em seu *ground truth*), FP denota os falsos positivos avaliados (pixel identificado incorretamente como parte da ROI na imagem teste) e, finalmente, FN assinala as amostras de falsos negativos obtidas

(pixel da ROI erroneamente reconhecido como parte do fundo).

Para maximizar os resultados obtidos, foram introduzidos três parâmetros avaliados experimentalmente e empiricamente. A utilização desses parâmetros é necessária pois foi constatado, durante a análise do método, que as pregas inframamárias identificadas podem variar significativamente em dimensões, dependendo das mamas consideradas. Esses parâmetros estão diretamente relacionados com o procedimento de limiarização com refinamento adaptativo.

O primeiro parâmetro verifica se existe uma única componente conexa que compõe a região das pregas inframamárias, sendo denominado la_1 . Esse valor determina um limiar mínimo de área medida em pixels que a prega inframamária ideal deve ter, antes da etapa de ajustes das mamas.

Analogamente, o segundo parâmetro analisa se existem duas componentes conexas que definem as pregas inframamárias, sendo esse chamado la_2 . Essa informação determina uma área mínima a ser medida em pixels para as duas componentes distintas ideais, uma para cada mama.

Finalmente, o último parâmetro se baseia no conceito de que a prega inframamária se localiza nos quadrantes inferiores da imagem térmica. Ele é denominado la_r . Dessa maneira, após a etapa de limiarização, todos os pixels $I(x, y)$, tal que

$$I(x, y) = \begin{cases} 0, & \text{se } x, y \geq 0 \text{ e } x, y \leq la_r, \\ 1, & \text{se } x, y > la_r, \end{cases} \quad (7)$$

assumindo que $I(x, y) = 0$ define um pixel de cor preta na posição x, y , e $I(x, y) = 1$ torna o pixel daquela posição de cor branca.

No presente trabalho foram realizados diversos experimentos avaliando empiricamente cada parâmetro apresentado. Para os parâmetros apresentados, os seguintes intervalos foram considerados: $la_1 = \{1000, 1100, 1200, 1300, 1400, \dots, 2500, 2600\}$, $la_2 = \{500, 600, 700, 800, 900, \dots, 1500, 1600\}$, $la_r = \{80, 90, 100, 110, 120, \dots, 140, 150\}$.

Isso foi feito para que fosse possível determinar qual o conjunto de parâmetros que maximiza as métricas estatísticas apresentadas.

No entanto, é importante dizer que não é necessário realizar o teste exaustivo de verificar a combinação de todos os valores possíveis dos parâmetros la_1 , la_2 e la_r . A seleção dos melhores parâmetros foi feita de acordo com a estratégia de aprendizado de máquina discutida na seção 3.2. Essa técnica culminou no melhor conjunto de parâmetros verificado nesse trabalho e que está apresentado na Tabela 1.

Tabela 1: Conjunto de parâmetros que maximiza o resultado das métricas estatísticas consideradas.

Parâmetro	Base UFPE	Base UFF
la_1	2200	2200
la_2	1500	1700
la_r	130	300

3.2 Seleção de Parâmetros Ótimos Locais

A aplicabilidade do método de segmentação proposto para diferentes bases de dados está condicionada a seleção e a adaptação dos valores dos parâmetros la_1 , la_2 e la_r . Tendo em vista que cada banco de imagens termográficas possui diferentes protocolos de aquisição de imagem, para cada base de dados é necessário selecionar um conjunto de parâmetros que melhor se adeque às imagens dos bancos de dados analisados. Conforme discutido anteriormente, estes são: a área de cada prega inframamária encontrada (la_1); o somatório de todas as áreas das pregas encontradas (la_2); e o quadrante em que as pregas inframamárias estão localizadas na imagem (la_r).

Para realizar tal tarefa, utiliza-se diferentes algoritmos de classificação para gerar um valor aproximado de acurácia para um determinado conjunto de parâmetros.

O treinamento foi realizado com base nos valores de acurácia obtidos experimentalmente em 250 conjuntos de parâmetros aplicados às bases de dados de imagens infravermelhas.

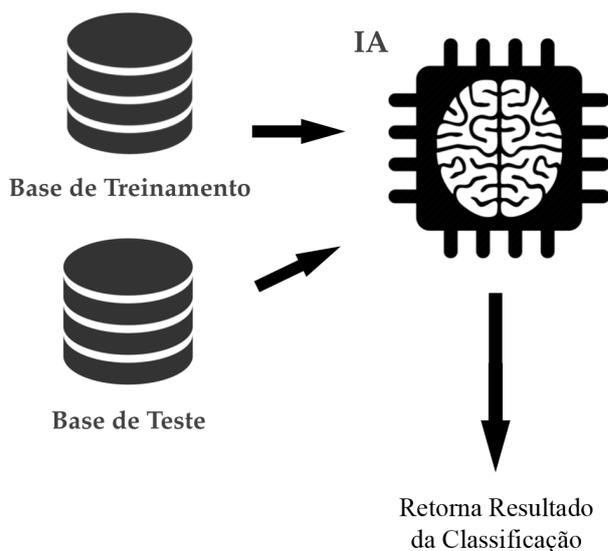


Figura 6: Representação básica do funcionamento do algoritmo de seleção de parâmetros ótimos locais mediante classificação.

Como base de dados de teste, variou-se os parâmetros a fim de se construir 133.000 conjuntos de parâmetros. Desse modo, ao aplicar-se um algoritmo de aprendizado, uma estimativa da acurácia correspondente ao conjunto de parâmetros testado seria obtida.

Para cada algoritmo de aprendizado, foram avaliados o coeficiente de correlação entre os parâmetros e a taxa de erro do algoritmo. Os resultados foram representados na Tabela 2.

Tabela 2: Resultados obtidos no treinamento dos algoritmos de classificação utilizando o software Weka 3.8.

Classificador	Taxa de Correlação	Taxa de Erro
Decision Table	0.9753	0.0033
M5P	0.9759	0.0049
Linear Regression	0.7642	0.0160
M5 Rules	0.9773	0.0043
SMOreg	0.7367	0.0156
Random Forest	0.9989	0.0007

Posteriormente, os parâmetros que geraram as melhores estimativas de acurácia foram testados experimentalmente, no intuito de verificar se tais valores realmente correspondiam à acurácia predita. O principal selecionado foi o

algoritmo Random Forest, devido ao alto coeficiente de correlação e a baixa taxa média de erros. Através desta metodologia foi possível prever o melhor conjunto de parâmetros para cada base de dados utilizada neste trabalho.

3.3 Resultados e Discussão

A partir do conjunto de parâmetros da Tabela 1 e do método proposto, o presente trabalho será comparado com outros trabalhos importantes apresentados na literatura.

Na Tabela 3, são apresentados os resultados médios de métodos de segmentação automática das mamas, considerando as taxas estatísticas apresentadas em todas as imagens disponibilizadas na base de dados. Observando os resultados obtidos, é possível concluir que o método apresentado obteve taxas médias altas para todas as medidas estatísticas analisadas. Alguns resultados dos trabalhos referidos são apresentados nas Figuras 7 e 8.

Tabela 3: Resultados normalizados obtidos pelos trabalhos apresentados na literatura e o método proposto.

	Marques UFPE	Motta UFPE	Método Proposto UFPE	Método Proposto UFF
ACC	0.97	0.96	0.96	0.93
SEN	0.97	0.88	0.98	0.98
ESP	0.97	0.99	0.95	0.86
PDP	0.97	0.96	0.95	0.91
PDN	0.98	0.96	0.98	0.95

Dos três trabalhos considerados, o método apresentado possui as maiores taxas médias de sensibilidade. Isso significa que, dos trabalhos avaliados, a solução proposta possui maior capacidade de identificação de amostras positivas (pertencentes a região de interesse). Essa diferença é superior ao trabalho de (MOTTA, 2010) em uma margem de 10%.

A diferença de acurácia entre (MARQUES, 2012) e o método proposto é de aproximadamente 0,5%, devido a introdução de amostras de falsos positivos na ROI. No entanto, nosso

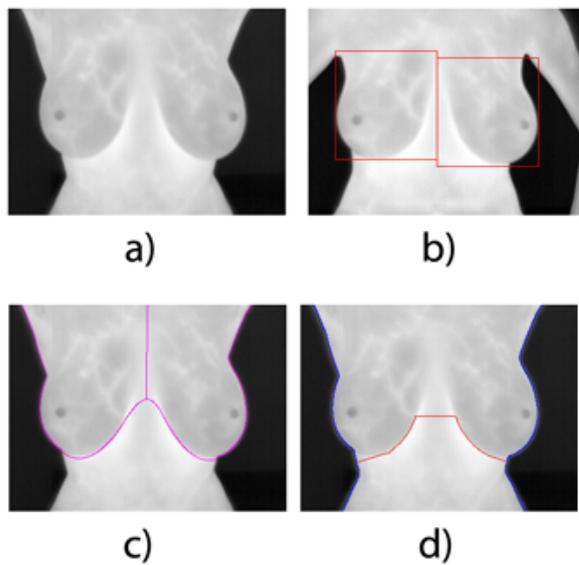


Figura 7: Resultados obtidos para a imagem IR_0100, proveniente da base da UFPE: a) Imagem original de entrada; b) Resultado obtido por (MOTTA, 2010); c) Resultado obtido por (MARQUES, 2012); d) Resultado adquirido pelo método proposto nesse trabalho.

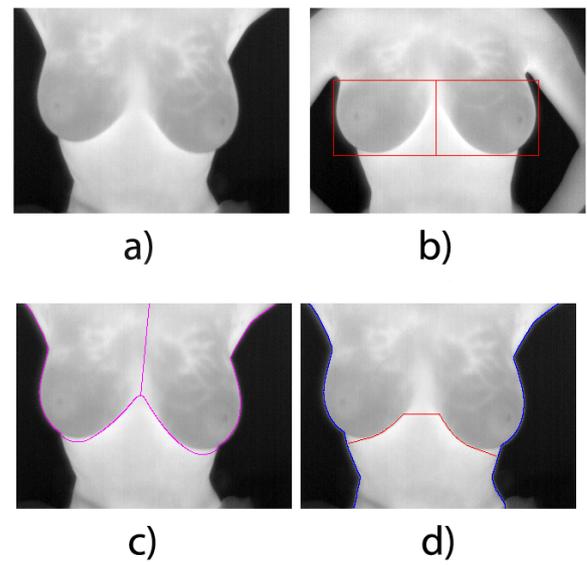


Figura 8: Resultados obtidos para a imagem IR_5986: a) Imagem original de entrada; b) Resultado obtido por (MOTTA, 2010); c) Resultado obtido por (MARQUES, 2012); d) Resultado adquirido pelo método proposto nesse trabalho.

método é bem mais simples de ser implementado, possuindo menos etapas que os trabalhos anteriores. Essa simplicidade sem acarretar em discrepâncias nas métricas estatísticas torna o método muito competitivo e vantajoso para utilização na literatura e na sua utilização em diversas arquiteturas computacionais.

Tabela 4: Comparativo de tempo de execução entre os métodos comparados e o trabalho proposto.

Trabalhos Considerados	Tempo de Execução (s)
(MOTTA, 2010)	36,63
(MARQUES, 2012)	1,55
Método Proposto	1,94

O método de (MARQUES, 2012) possui tempos computacionais de execução muito similares ao trabalho proposto. Isso é visível na Tabela 4, em que os tempos médios em segundos de todos os métodos são apresentados. Se comparado a (MOTTA, 2010), a diferença torna-se muito mais significativa, sendo apro-

ximadamente 18 vezes mais lento do que o método apresentado nesse trabalho.

Para algumas imagens da base de dados, o método não apresenta resultados de maior qualidade. Isso é visível, por exemplo, quando as pacientes possuem pouco volume nas mamas. Isso torna a identificação da região das pregas inframamárias mais complexa. No entanto, como é visível na Figura 9, o impacto desse problema é menos significativo do que em alguns resultados de (MARQUES, 2012) devido a abordagem de refinamento adaptativo, que iterativamente busca uma região de prega inframamária com maior contraste de temperatura, de acordo com a imagem de teste considerada.

Outros problemas comuns são: a identificação inapropriada de uma dobra abdominal como parte da região das pregas inframamárias (Figura 10) e a assimetria entre as mamas (Figura 11). No primeiro caso, a identificação de uma dobra do abdômen como parte das pregas inframamárias introduz uma grande quantidade de falsos positivos, que poderiam ser corrigidos se o ponto de interseção entre

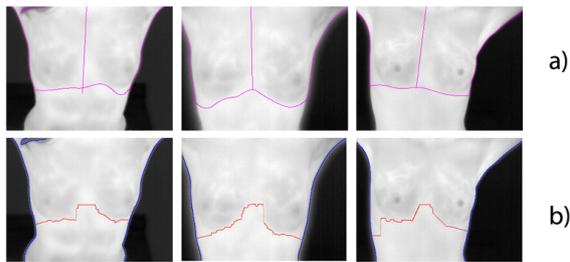


Figura 9: Imagens em que as pacientes possuem mamas de menor volume (respectivamente: IR_5451, IR_7446 e IR_5506): a) Resultados obtidos em (MARQUES, 2012); b) Resultados obtidos pelo método proposto nesse trabalho.

os seios fosse melhor assinalado. Com relação à assimetria, o resultado é superior a alguns exemplos apresentados em (MARQUES, 2012), devido a premissa de se assumir que as pregas inframamárias localizam-se nos quadrantes inferiores da imagem e pelo fato de se extrair uma região de prega inframamária de maior densidade de pixels. No entanto, erros ainda ocorrem com certa frequência. É possível corrigi-los adicionando restrições que impeçam que a distância entre as regiões inferiores da mama seja tão contrastante.

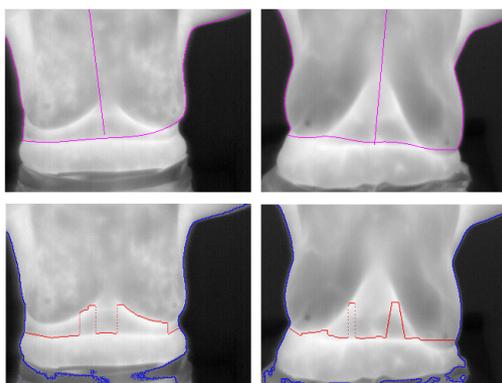


Figura 10: Imagens em que as pacientes possuem dobras abdominais passíveis de confusão com as pregas inframamárias (respectivamente: IR_0889 e IR_1330): a) Resultados obtidos em (MARQUES, 2012); b) Resultados obtidos pelo método proposto nesse trabalho.

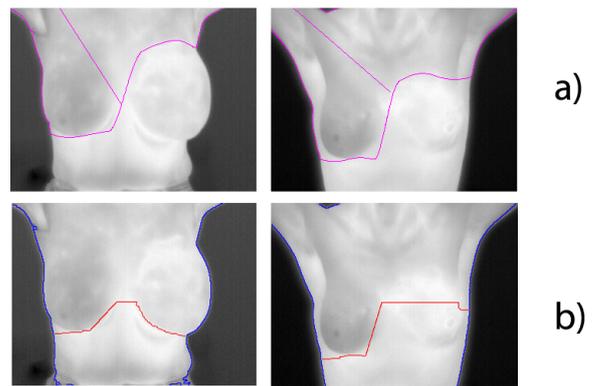


Figura 11: Imagens em que as pacientes possuem mamas com significativa assimetria (respectivamente: IR_8304 e IR_3786): a) Resultados obtidos em (MARQUES, 2012); b) Resultados obtidos pelo método proposto nesse trabalho.

4 Considerações Finais

A detecção precoce do câncer de mama é essencial para melhorar os prognósticos de cura da paciente. Sistemas CAD baseados na simetria da distribuição térmica das mamas se mostram eficazes, baratos e pouco invasivos, podendo se tornar relevantes no diagnóstico de doenças da mama. No entanto, para que isso seja feito, torna-se necessário extrair a região de interesse que será avaliada por sistemas de diagnóstica de doenças.

A partir dos resultados apresentados, é visível que o método apresentado torna-se bastante competitivo para uso em situações reais, devido a acurácia do resultado final, bem como o tempo de execução médio que é relativamente baixo. Além disso, a metodologia é fácil de ser desenvolvida computacionalmente, pois possui muito menos etapas que os outros trabalhos da literatura e requer poucos recursos computacionais (tanto de hardware quanto de software) para o seu funcionamento.

Para trabalhos futuros, espera-se:

- Adaptar o método para imagens com câmeras em ângulos distintos;
- Utilizar técnicas para extração de atributos e características da mama para diag-

nóstico de doenças. Esses atributos podem ser geométricos ou de textura;

- Desenvolver sistemas CAD para diagnóstico de doenças diversas da mama a partir do algoritmo de segmentação proposto.

Referências

- ARORA, N. M. et al. Effectiveness of a noninvasive digital infrared thermal imaging systems in the detection of breast cancer. *American Journal of Surgery*, v. 196, n. 4, p. 523–526, 2008.
- BAFFA, M. F. O.; CHELONI, D. J. M.; LATTARI, L. G. Segmentação automática de imagens térmicas da mama utilizando limiarização com refinamento adaptativo. *Anais do XXXVI Congresso da Sociedade Brasileira de Computação*, p. 2499–2508, 2016.
- GORE, J. P.; XU, L. X. *Thermal Imaging for Biological and Medical Diagnostics*. [S.l.]: CRC Press, 2014. 517–530 p.
- HCB. *Informação: Saiba quais são os tipos de câncer mais comuns no Brasil*. 2015. [Http://www.hcancerbarretos.com.br/82-institucional/noticias-institucional/1300-informacao-saiba-quais-sao-os-tipos-de-cancer-mais-comuns-no-brasil](http://www.hcancerbarretos.com.br/82-institucional/noticias-institucional/1300-informacao-saiba-quais-sao-os-tipos-de-cancer-mais-comuns-no-brasil). Acesso em: 29 de outubro de 2015.
- INCA. *Tipos de Câncer*. 2015. <<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama>>. Acesso em: 07 de novembro de 2015.
- LI, X. et al. Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to skin lesion segmentation. In: *ISBI*. IEEE, 2011. p. 1438–1441. ISBN 978-1-4244-4128-0. Disponível em: <<http://dblp.uni-trier.de/db/conf/isbi/isbi2011.html#LiAFR11>>.
- MARQUES, R. S. *Segmentação Automática das mamas em imagens térmicas*. Dissertação (Mestrado) — Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil, 2012.
- MOTTA, L. S. *Obtenção automática da região de interesse em termogramas frontais da mama para o auxílio à detecção precoce de doenças*. Dissertação (Mestrado) — Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil, 2010.
- PROENG. *Processamento e Análise de Imagens Aplicadas à Mastologia*. 2008. <<http://visual.ic.uff.br/proeng/>>. Acesso em: 20 de setembro de 2016.
- SILVA, L. F. et al. A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, v. 4, n. 1, p. 92–100, 2014.