

UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

**ANTÔNIO MARCOS VIEIRA SALES**

**DETECÇÃO DE REGIÕES COM LESÃO EM MAMOGRAFIAS USANDO  
COEFICIENTE DE CORRELAÇÃO CRUZADA, DISTÂNCIA EUCLIDIANA E  
ÍNDICE DE GETIS-ORD**

São Luís  
2012

**ANTÔNIO MARCOS VIEIRA SALES**

**DETECÇÃO DE REGIÕES COM LESÃO EM MAMOGRAFIAS USANDO  
COEFICIENTE DE CORRELAÇÃO CRUZADA, DISTÂNCIA EUCLIDIANA E  
ÍNDICE DE GETIS-ORD**

Monografia apresentada ao Curso de Ciência da  
Computação da Universidade Federal do Maranhão,  
como parte dos requisitos necessários para obtenção  
do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Aristófanés Corrêa Silva.

São Luís  
2012

Sales, Antônio Marcos Vieira.

Detecção de regiões com lesão em mamografias usando coeficiente de correlação cruzada, distância euclidiana e índice de Getis-Ord / Antônio Marcos Vieira Sales. – 2012.

53 f.

Impresso por computador (Fotocópia).

Orientador: Aristófanês Corrêa Silva.

Monografia (graduação) – Universidade Federal do Maranhão, Curso de Ciência da Computação, 2012.

1. Processamento de imagem médica. 2. Mamografia – Processamento de imagem. 3. Índice de Getis-Ord. 4. Máquina de vetores de suporte. I. Título.

CDU 004.932:61

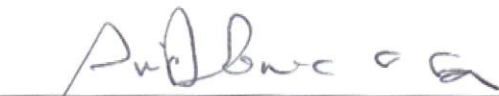
ANTÔNIO MARCOS VIEIRA SALES

**DETECÇÃO DE REGIÕES COM LESÃO EM MAMOGRAFIAS USANDO  
COEFICIENTE DE CORRELAÇÃO CRUZADA, DISTÂNCIA EUCLIDIANA E  
ÍNDICE DE GETIS-ORD**

Monografia apresentada ao curso de Ciência da  
Computação da Universidade Federal do Maranhão,  
como parte dos requisitos necessários para obtenção  
do grau de Bacharel em Ciência da Computação.

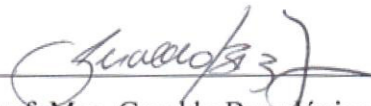
Aprovada em 11/06/2012

BANCA EXAMINADORA



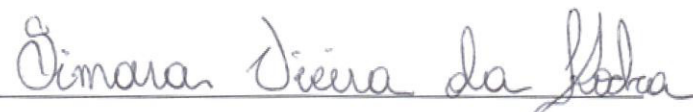
---

Prof. Dr. Aristófanes Corrêa Silva (Orientador)  
Doutor em Informática  
Universidade Federal do Maranhão



---

Prof. Msc. Geraldo Braz Júnior (Membro da Banca)  
Mestre em Engenharia de Eletricidade  
Universidade Federal do Maranhão



---

Prof. Msc. Simara Vieira da Rocha (Membro da Banca)  
Mestra em Engenharia de Eletricidade  
Universidade Federal do Maranhão

## AGRADECIMENTOS

A Deus, pois sem ele absolutamente nada seria possível.

Aos meus pais, por sempre me apoiarem na busca dos meus objetivos.

Ao Prof. Dr. Aristófanês Corrêa Silva, pela orientação fornecida.

Aos professores da banca examinadora.

Enfim, a todos que contribuíram de alguma forma para realização deste trabalho.

## RESUMO

O câncer de mama é o tipo de câncer que tem início nas células das mamas, mas pode se espalhar para outras partes do organismo. A principal forma de prevenção e diagnóstico precoce da doença é através de exames de mamografia, um tipo de radiografia das mamas. Este trabalho tem como objetivo principal apresentar uma metodologia de auxílio à detecção de lesões em mamografias partindo da determinação de regiões suspeitas por nível de simetria. Primeiramente as mamografias passam por processos de redução de ruído e segmentação, para isolar a mama na imagem. Após isso, são feitos o registro rígido e o registro deformável, para deixar a mama direita mais próxima o possível do formato da mama esquerda. O nível de simetria é determinado através da similaridade entre regiões correspondentes na mama esquerda e na direita, com a utilização do coeficiente de correlação cruzada e da distância euclidiana. As medianas encontradas em cada par de mamografias, para essas duas medidas, são usadas como limiares de similaridade do par. Uma vez determinadas as regiões mais assimétricas dos casos de teste, são extraídas características de cada uma delas, com índice de Getis-Ord na sua forma geral. Então são montados vetores de características, compostos de valores de índice de Getis-Ord e da média dos níveis de cinza de cada região. Também são montados vetores a partir de outras mamografias selecionadas para formar uma base de treinamento para uma Máquina de Vetores de Suporte. Depois de feito este treinamento, as regiões suspeitas (assimétricas) dos casos de teste são classificadas em lesão e não lesão. A metodologia, de modo geral, apresentou 80,11% de sensibilidade, 84,41% de especificidade e 84,38% de acurácia.

Palavras-chave: Câncer de mama. Mamografia. Coeficiente de correlação cruzada. Distância euclidiana. Índice de Getis-Ord. Máquina de vetores de suporte.

## ABSTRACT

Breast cancer is cancer that begins in cells of the breast, but can spread to other parts of the body. The main form of prevention and early diagnosis is through mammograms, a type of x-ray of the breasts. This work has as main objective to present a methodology to aid the detection of lesions in mammograms based on the determination of suspicious regions by level of symmetry. First the mammograms undergo processes of noise reduction and segmentation, to isolate the breast in the image. After that, rigid and deformable registrations are made, to leave the right breast the closest possible to the shape of the left breast. The degree of symmetry is determined by the similarity between corresponding regions in the left breast and the right breast, using the cross-correlation coefficient and the Euclidean distance. The median values found in each pair of mammograms for these two measures are used as thresholds of similarity of the pair. After determination of the asymmetric regions of the test cases, features are extracted for each one of them, with Getis-Ord index in its general form. Then feature vectors are assembled, composed of values of Getis-Ord index and the average gray level of each region. Also vectors are assembled from other mammograms selected to form a training database for a Support Vector Machine. After doing this training, the suspicious regions (asymmetric) of the test cases are classified as non-lesion and lesion. The methodology, in general, showed 80.11% sensitivity, 84.41% specificity and 84.38% accuracy.

Keywords: Breast cancer. Mammography. Cross-correlation coefficient. Euclidean distance. Getis-Ord index. Support vector machine.

## LISTA DE ILUSTRAÇÕES

	p.
Figura 1 – Anatomia da mama	17
Figura 2 – Projeções CC e MLO de mamografia	18
Figura 3 – Diversos tipos de lesão	19
Figura 4 – Conjuntos linearmente separáveis e não linearmente separáveis	25
Figura 5 – Representação gráfica da metodologia utilizando imagens do DDSM	28
Figura 6 – Mamografia antes da redução de ruído	29
Figura 7 – Mamografia após a redução de ruído	29
Figura 8 – Mamografia antes e após segmentação	30
Figura 9 – Mamografia antes e após redimensionamento	31
Figura 10 – Registro rígido da mama direita com base na mama esquerda	32
Figura 11 – Registro rígido e registro deformável da mama direita com base na mama esquerda	32
Figura 12 – Divisão das mamografias em regiões de 16x16 <i>pixels</i>	33
Figura 13 – Regiões correspondentes no par de mamografias	34
Figura 14 – Correspondência entre <i>pixels</i> de regiões hipotéticas de 4x4 <i>pixels</i>	34
Figura 15 – Determinação de regiões suspeitas por assimetria usando CCC e DE	35
Figura 16 – Distância 1 para cálculo de índice de Getis-Ord	36
Figura 17 – Caso A1639 do banco DDSM	37
Figura 18 – Caso A1639 do banco DDSM após pré-processamento e registro	38
Figura 19 – Regiões classificadas como suspeitas no caso A1639 do banco DDSM	38



Figura 20 – Caso A1134 do banco DDSM	39
Figura 21 – Caso A1134 do banco DDSM após pré-processamento e registro	39
Figura 22 – Regiões classificadas como suspeitas no caso A1134 do banco DDSM	40
Figura 23 – Caso A0074 do banco DDSM	40
Figura 24 – Caso A0074 do banco DDSM após pré-processamento e registro	41
Figura 25 – Regiões classificadas com suspeitas no caso A0074 do banco DDSM	41
Figura 26 – Caso A1627 do banco DDSM	42
Figura 27 – Caso A1627 do banco DDSM após pré-processamento e registro	42
Figura 28 – Regiões classificadas como suspeitas no caso A1627 do banco DDSM	43
Figura 29 – Classificação final do caso A1639	44
Figura 30 – Classificação final do caso A1134	44
Figura 31 – Classificação final do caso A0074	45
Figura 32 – Classificação final do caso A1627	46

## LISTA DE SIGLAS

ACS – *American Cancer Society*

CAD – *Computer-Aided Detection*

CADx – *Computer-Aided Diagnosis*

CC – *Crânio-Caudal*

CCC – *Coeficiente de Correlação Cruzada*

DDSM – *Digital Database for Screening Mammography*

DE – *Distância Euclidiana*

INCA – *Instituto Nacional do Câncer*

MLO – *Médio-Lateral Oblíqua*

MVS – *Máquina de Vetores de Suporte*

NCI – *National Cancer Institute*

RBF – *Radial Basis Function*

## SUMÁRIO

	p.
<b>1 INTRODUÇÃO</b>	<b>12</b>
<b>1.1 Motivação e Objetivos</b>	<b>12</b>
<b>1.2 Trabalhos Relacionados</b>	<b>13</b>
<b>1.3 Organização do Trabalho</b>	<b>15</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>16</b>
<b>2.1 Câncer de Mama</b>	<b>16</b>
<b>2.2 Mamografia</b>	<b>17</b>
<b>2.3 Processamento de Imagem</b>	<b>19</b>
2.3.1 Redução de Ruído	19
2.3.2 Segmentação	20
2.3.2.1 Morfologia Matemática	20
2.3.2.2 Watershed	21
2.3.3 Registro de Imagem	21
2.3.3.1 Registro Rígido	22
2.3.3.2 Registro Deformável com Algoritmo “Demons”	22
<b>2.4 Coeficiente de Correlação Cruzada</b>	<b>22</b>
<b>2.5 Distância Euclidiana</b>	<b>23</b>
<b>2.6 Índice de Getis-Ord</b>	<b>24</b>
<b>2.7 Máquina de Vetores de Suporte</b>	<b>24</b>
<b>3 METODOLOGIA</b>	<b>27</b>

<b>3.1 Aquisição das Mamografias</b>	<b>28</b>
<b>3.2 Pré-processamento e Redimensionamento das Mamografias</b>	<b>29</b>
3.2.1 Redução de Ruído	29
3.2.2 Segmentação das Mamas	30
3.2.3 Redimensionamento das Mamografias	30
<b>3.3 Registro das Mamografias</b>	<b>31</b>
3.3.1 Registro Rígido	31
3.3.2 Registro Deformável	32
<b>3.4 Divisão das Mamografias em Regiões de 16x16 pixels</b>	<b>33</b>
<b>3.5 Determinação de Regiões Suspeitas com CCC e DE.</b>	<b>33</b>
<b>3.6 Extração de Características com Índice de Getis-Ord</b>	<b>35</b>
<b>3.7 Classificação Final com MVS</b>	<b>36</b>
<b>4 RESULTADOS</b>	<b>37</b>
<b>4.1 Pares de Regiões Suspeitas</b>	<b>37</b>
4.1.1 Pares de Regiões Suspeitas: Teste 1	37
4.1.2 Pares de Regiões Suspeitas: Teste 2	39
4.1.3 Pares de Regiões Suspeitas: Teste 3	40
4.1.4 Pares de Regiões Suspeitas: Teste 4	41
<b>4.2 Detecção de Regiões com Lesão</b>	<b>43</b>
4.2.1 Detecção de Regiões com Lesão: Teste 1	44
4.2.2 Detecção de Regiões com Lesão: Teste 2	44
4.2.3 Detecção de Regiões com Lesão: Teste 3	45
4.2.4 Detecção de Regiões com Lesão: Teste 4	45

<b>4.3 Análise dos Resultados</b>	<b>46</b>
4.3.1 Análise da Determinação de Regiões Suspeitas	46
4.3.2 Análise da Classificação em Lesão e Não Lesão	47
4.3.3 Análise Geral da Metodologia	48
<b>5 CONCLUSÃO</b>	<b>50</b>
<b>REFERÊNCIAS</b>	<b>52</b>

## 1 INTRODUÇÃO

Câncer é o nome dado ao crescimento desordenado de células que ataca tecidos e órgãos, podendo se tornar muito agressivo e causando o surgimento de tumores malignos. O câncer de mama é o surgimento de tumores malignos que se iniciam nas células da mama. De acordo com a *American Cancer Society* (ACS), este é o segundo tipo de câncer a causar mais mortes em mulheres (ACS, 2012a). Também é o tipo de câncer que mais acomete mulheres em todo o mundo, respondendo por, aproximadamente, 22% dos casos novos a cada ano, de acordo com o Instituto Nacional do Câncer (INCA). Segundo o mesmo instituto a estimativa para o ano de 2012 é de 52.680 novos casos no Brasil, onde as taxas de mortalidade por essa doença continuam elevadas, muito provavelmente porque a doença ainda é diagnosticada em estágios avançados (INCA, 2011).

Infelizmente o câncer de mama não possui causa específica determinada, mas, segundo o *National Cancer Institute* (NCI), existem fatores considerados de risco, por exemplo: histórico da doença na família, submissão a tratamentos radioterápicos, obesidade, consumo de álcool, nível de atividade física, etc. É possível diminuir o risco de se desenvolver a doença praticando atividade física frequentemente e mantendo um peso saudável (NCI, 2012).

O fato de a chance de cura ser bem maior quando o câncer é detectado precocemente e as estimativas de que grande parte das lesões deixa de ser detectada no exame feito pelo especialista, devido à dificuldade de identificar padrões na imagem, são um grande incentivo à pesquisa e ao desenvolvimento de métodos e sistemas que visam auxiliar o especialista na detecção e diagnóstico da doença (sistemas CAD - *Computer-Aided Detection* e CADx - *Computer-Aided Diagnosis*, respectivamente), indicando áreas suspeitas, assim como anormalidades mascaradas (ROCHA et al., 2011).

### 1.1 Motivação e Objetivos

A mamografia ou radiografia da mama é o exame mais usado na prevenção e diagnóstico do câncer de mama e permite a descoberta de lesões em fase inicial, imperceptíveis ao exame do toque, o que aumenta bastante as chances de cura. É difícil determinar a efetividade da mamografia, que depende de fatores relacionados às características da própria mama, da lesão, dos recursos disponíveis e da interpretação do especialista (INCA, 2012). A análise de imagens mamográficas pode ser melhorada com o

auxílio de processamento digital de imagens e visão computacional. A principal contribuição das pesquisas na área de detecção e diagnósticos baseados em processamento de imagens vêm da dificuldade humana em localizar e classificar áreas lesionadas.

Estudos anteriores veem analisando o nível de assimetria entre mamografias esquerda e direita da mesma paciente e o risco de desenvolver câncer. Um trabalho realizado em (SCUTT; LANCASTER; MANNING, 2006) detectou uma assimetria de volume maior em pacientes que chegaram a desenvolver câncer até o fim da pesquisa do que em pacientes que permaneceram normais. Uma metodologia de detecção de massas baseada na análise de simetria foi proposta em (ERICEIRA, 2011). Na dissertação, o autor utilizou registro bilateral das mamografias da mesma paciente e a função conhecida como variograma cruzado para extração de características de pares de regiões correspondentes das imagens.

Este trabalho tem como objetivo geral apresentar uma metodologia para detecção de regiões lesionadas em mamografias, partindo da determinação de pares de regiões suspeitos correspondentes na mama esquerda e na direita, com posterior classificação individual de tais regiões. Para isso, pares de mamografias são processados de forma que se possa medir a similaridade entre regiões correspondentes do par. Assim, regiões suspeitas são determinadas pelo nível de assimetria encontrado entre regiões da mama esquerda e da mama direita da mesma paciente através do índice de similaridade conhecido como coeficiente de correlação cruzada (*cross-correlation*) e da distância euclidiana. Após esta determinação, o índice de Getis-Ord, em sua forma geral, é utilizado em cada região suspeita isoladamente para extração de características e posterior classificação, com uma Máquina de Vetores de Suporte, em “lesão” e “não lesão”.

## **1.2 Trabalhos Relacionados**

Nesta seção estão relacionados alguns trabalhos com ênfase em análise de simetria, registro, e detecção e classificação de anormalidades em imagens de mamografia. Estes trabalhos serviram como base de referência para o desenvolvimento desta monografia.

No estudo apresentado em (SCUTT; LANCASTER; MANNING, 2006) foram comparados exames de mamografias de dois grupos de mulheres. O primeiro grupo consistia de 252 mulheres assintomáticas, mas que chegaram a desenvolver câncer até o final do estudo. Já o segundo era constituído de 252 mulheres saudáveis que continuaram assim até o final do estudo. Foi calculado o volume das mamas a partir de mamografias do tipo crânio-

caudal e observada uma relação entre a assimetria de volume e o risco de desenvolver câncer. O grupo que chegou a desenvolver a doença apresentava uma assimetria de volume média de 63,17 ml enquanto que o grupo que permaneceu saudável apresentou uma assimetria média de 52,99 ml.

Técnicas de registro com imagens médicas tem sido alvo de vários estudos e aplicadas em várias metodologias propostas. Em (LEE, 1997), o autor propôs uma metodologia de registro de mamografias para fazer um mapeamento entre duas imagens através do cálculo do fluxo óptico, ou da mudança de intensidade aparente, entre a imagem fonte e a imagem destino. A eficiência da metodologia proposta foi melhorada através de uma abordagem multirresolução, pela qual as imagens podiam ser comparadas em mais de uma escala, e cujo resultado foi considerado consistente com o esperado.

Em (RODRIGUES, 2010), o autor realizou uma comparação entre diversas métricas de similaridade no contexto de corregristo não rígido de imagens médicas em relação a diferentes parâmetros, como dimensão da região a ser comparada entre as imagens, dimensão da região de busca por similaridade, numero de tons de cinza e um parâmetro entrópico. Dentre as medidas estudadas, o coeficiente de correlação cruzada foi considerado um dos mais robustos.

A utilização de medidas conhecidas com descritores espaciais para extração de características em imagens médicas apresenta constante crescimento, especialmente quando utilizadas em conjunto com técnicas de aprendizado de máquinas. Um estudo feito por (BRAZ JÚNIOR, 2008) utilizou várias dessas medidas (Índice de Moran, Coeficiente de Geary, Índice de Getis-Ord e função K de Ripley), onde as características extraídas eram submetidas para treinamento e classificação com Máquina de Vetores de Suporte.

Um estudo preliminar feito por (ERICEIRA; SILVA; PAIVA, 2010) aplicava o descritor espacial variograma cruzado (um descritor espacial bivariado) para detecção de regiões suspeitas em pares de mamografias registrados (mama direita registrada conforme mama esquerda correspondente). Posteriormente em (ERICEIRA, 2011), o autor propôs um método para detecção de massas em mamografias registradas, através da busca de regiões suspeitas, caracterizando pares de regiões correspondentes na mama esquerda e na direita pelos seus valores de variograma cruzado. Em outra etapa, as regiões suspeitas eram, então, submetidas à classificação final pelos seus valores de variograma de forma isolada (a direita



separada da esquerda). Em ambas as etapas a metodologia utilizava Máquina de Vetores de Suporte para classificação das regiões.

Os trabalhos relacionados acima mostram um crescente desenvolvimento na pesquisa por métodos de extração de características, análise de simetria e classificação (segundo algum aspecto) de regiões em imagens médicas com utilização de Máquina de Vetores de Suporte.

### **1.3 Organização do Trabalho**

O trabalho está organizado em mais quatro capítulos. No Capítulo 2, é apresentada a fundamentação teórica necessária à compreensão da metodologia proposta. As secções deste capítulo falam de câncer de mama, mamografia, medidas estatísticas utilizadas, técnicas de processamento de imagem e Máquina de Vetores de Suporte. No Capítulo 3, a metodologia é explicada, exibindo a aplicação das técnicas, citadas no Capítulo 2, sobre as mamografias. O Capítulo 4 mostra os resultados obtidos com a aplicação da metodologia em casos de teste selecionados e o Capítulo 5 exhibe a conclusão obtida com os resultados do trabalho realizado e sugere melhorias que podem ser feitas em trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção são apresentados assuntos necessários à compreensão da metodologia deste trabalho, eles abordam: câncer de mama, mamografia, técnicas de processamento de imagem utilizadas, medidas utilizadas e classificação por Máquina de Vetores de Suporte.

### 2.1 Câncer de Mama

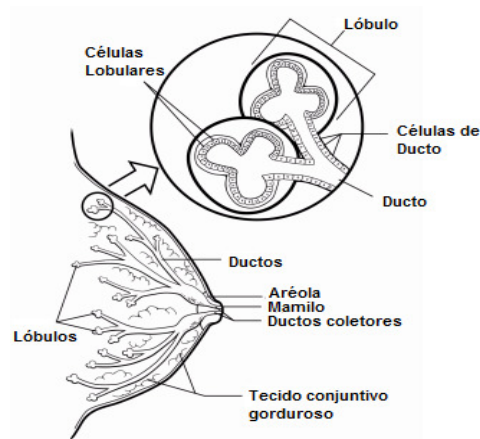
Câncer de mama é o desenvolvimento de tumores malignos que se inicia nas células da mama, ocorrendo, quase que na totalidade dos casos, em mulheres, mas pode haver casos em homens. Esses tumores podem crescer e invadir tecidos circundantes e se espalhar para outras partes do organismo. Quando isso ocorre, diz-se que houve uma metástase. Se as células cancerígenas entrarem nos vasos linfáticos (parte importante do sistema imunológico) e chegarem aos linfonodos (glândulas ligadas pelos vasos linfáticos) há uma maior probabilidade de caírem na corrente sanguínea e se espalharem para outras áreas do organismo (ACS, 2012b).

Dentre os sintomas que podem indicar câncer de mama, podem ser citados: alterações na pele da mama, como abaulamentos ou retrações; secreção no mamilo; nódulos palpáveis no seio, com ou sem dor; nódulos palpáveis na axila; etc. A doença é relativamente rara antes dos 35 anos de idade, porém, acima desta faixa etária, a incidência sobe progressivamente. Esta incidência tem aumentado tanto nos países desenvolvidos como nos países em desenvolvimento (INCA, 2011).

A mama tem sua anatomia composta principalmente de lóbulos, ductos e estroma. Os lóbulos são as glândulas produtoras de leite, os ductos são tubos que transportam o leite dos lóbulos para o mamilo. O estroma, por sua vez, é composto de tecido adiposo e tecido conjuntivo que envolve os ductos e os lóbulos, vasos sanguíneos e vasos linfáticos. A Figura 1 ilustra a anatomia da mama. A maioria dos casos de câncer de mama se inicia nas células que revestem os ductos, enquanto alguns casos se iniciam nas células que revestem os lóbulos ou em outros tecidos (ACS, 2012b).

O câncer ainda pode ser classificado como *insitu*, quando permanece confinado aos ductos, e invasivo, quando invade o tecido gorduroso que envolve os ductos e, possivelmente, se espalha para outros órgãos.

Figura 1 – Anatomia da mama



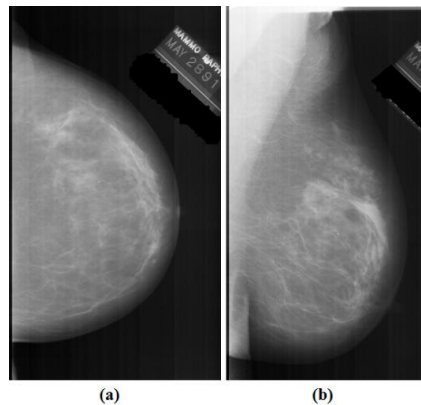
Adaptado de (ACS, 2012b).

## 2.2 Mamografia

A mamografia (radiografia da mama) permite a detecção precoce do câncer, pois é capaz de mostrar lesões em fase inicial, muito pequenas (medindo milímetros). Deve ser realizada a cada dois anos por mulheres entre 50 e 69 anos, ou segundo recomendação médica. É realizada em um aparelho de raios-X específico, chamado mamógrafo. Neste aparelho, a mama é comprimida de forma a fornecer melhores imagens e melhor capacidade de diagnóstico (INCA, 2012). Para mulheres pertencentes aos grupos com maior risco de desenvolver câncer (por exemplo, com história na família de casos de câncer de mama, principalmente em parentes de primeiro grau) são recomendados o exame clínico das mamas e a mamografia a partir dos 35 anos de idade.

As imagens de raios-X da mama podem ser obtidas com diferentes projeções, de diferentes ângulos, onde as mais comuns são: crânio-caudal (CC) e médio-lateral oblíqua (MLO). A Figura 2 exemplifica os dois tipos de projeções com imagens retiradas do projeto *The Digital Database for Screening Mammography* (DDSM) (HEATH et al., 2001).

Figura 2 – Projeções CC e MLO de mamografia



(a) Projeção CC. (b) Projeção MLO. Adaptado do banco DDSM (HEATH et al., 2001).

É difícil distinguir a sensibilidade da mamografia como estratégia isolada de rastreamento, pois estudos sobre a efetividade do método sempre utilizam o exame clínico como exame adicional. Os resultados dependem de fatores tais como: tamanho e localização da lesão, densidade do tecido mamário (mulheres mais jovens apresentam mamas mais densas), qualidade dos recursos técnicos e habilidade de interpretação do radiologista. Cerca de 30% da mortalidade em mulheres acima dos 50 anos é reduzida com o uso da mamografia (INCA, 2012).

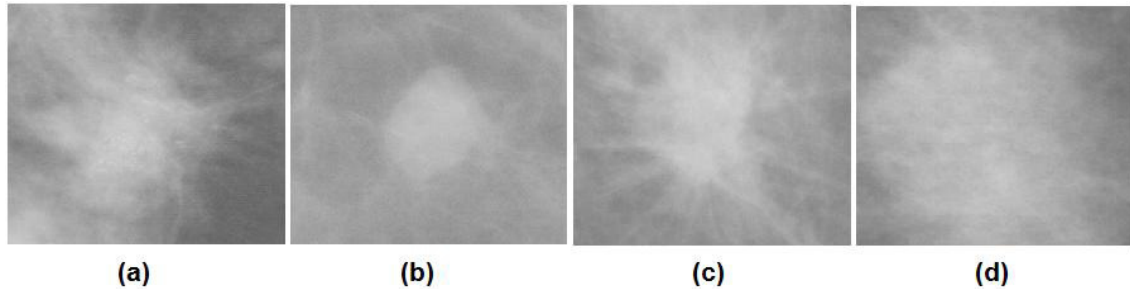
As lesões mamárias que o especialista comumente procura em mamografias podem ser principalmente calcificações ou massas, com diferentes formatos e distribuições. Além disso, sua natureza pode ser maligna ou benigna.

Calcificações são depósitos minerais dentro do tecido mamário. Eles se parecem com pequenas manchas brancas em uma mamografia. As calcificações vistas no exame podem ou não estar associadas com o câncer da mama. Há principalmente dois tipos de calcificações: macrocalcificações (grossos depósitos de cálcio que são provavelmente associados com alterações nas mamas causadas pelo envelhecimento das artérias do peito, lesões antigas, ou inflamação) e microcalcificações (partículas minúsculas de cálcio na mama, podem aparecer sozinhas ou em grupos, são mais preocupantes, mas nem sempre significam que o câncer está presente).

Massas são áreas (aglomerado de células) anormais que podem ser muitas coisas, incluindo cistos e tumores não cancerosos. Vários fatores como o tamanho, forma, e as

margens (bordas) da massa são levados em consideração para determinar se a mesma é cancerígena. Há, ainda, lesões referentes a distorções de arquitetura. Algumas massas podem ser estudadas somente com a mamografia, porém outras precisam de uma biópsia para determinar a natureza da lesão (ACS, 2012c). A Figura 3 exemplifica alguns tipos de lesão pertencentes a imagens do banco DDSM.

Figura 3 – Diversos tipos de lesão



(a) Calcificação. (b) Massa bem definida. (c) Massa espiculada. (c) Massa mal definida. Adaptado do banco DDSM (HEATH et al., 2001).

## 2.3 Processamento de Imagem

O Processamento de Imagens vem na realidade do Processamento de Sinais. O sinal é um suporte físico que carrega no seu interior uma informação. A informação pode estar associada a um fenômeno físico ou a um nível cognitivo. “Processar uma imagem consiste em transformá-la sucessivamente com o objetivo de extrair mais facilmente a INFORMAÇÃO nela presente” (ALBUQUERQUE, Marcio; ALBUQUERQUE, Marcelo, 2000, p. 1).

Em se tratando de processamento de imagem, muitos procedimentos se destacam como de grande importância ou de uso comum em determinadas atividades. Podemos citar a aplicação de filtros de imagem (por exemplo, para redução de ruído), a segmentação (que tem a função de isolar um objeto de interesse na imagem) e o registro (que visa tornar a imagem alvo mais próxima o possível da imagem fonte).

### 2.3.1 Redução de Ruído

Muitas imagens nem sempre estão adequadas ao processamento que se deseja realizar por apresentarem falhas conhecidas como ruídos, que geralmente podem ter surgido

na própria obtenção da imagem (por exemplo, na digitalização de uma mamografia) ou posteriormente por causas externas.

Uma das técnicas mais aplicadas para redução de ruído e que se mostra muito efetiva na remoção de ruídos de imagens de mamografia é o filtro da mediana (LEE, 1997). A mediana é uma das medidas estatísticas que representam uma tendência central para um conjunto de dados. O filtro da mediana é um filtro de vizinhança onde o valor de um *pixel* da imagem é substituído pela mediana dos valores dos pixels de sua vizinhança que pode ser uma região circular de raio  $r$  com centro no *pixel* em questão.

### 2.3.2 Segmentação

O processo de segmentação consiste em separar na imagem o objeto sobre o qual se quer trabalhar do restante da imagem, removendo aqueles que poderiam alterar os resultados de processamentos futuros. Em outras palavras, consiste em isolar o objeto de interesse na imagem através da divisão da imagem em regiões constituintes, divisão tal que deve parar quando o objeto de interesse tiver sido isolado (GONZALEZ, 2002).

Imagens de mamografia costumam apresentar elementos chamados de “artefatos”, que ficam no fundo da imagem e contêm informações sobre o exame como rótulos, dados sobre filme utilizado, etc. Estes objetos devem ser removidos, pois podem prejudicar o processo realizado por algoritmos de registro. Assim, a segmentação da mama torna o processo de registro mais rápido e preciso por remover áreas que não devem ser analisadas pelo algoritmo.

#### 2.3.2.1 Morfologia Matemática

A morfologia matemática é o estudo quantificado da forma em um conjunto de pontos que visa revelar a forma dos objetos formados pelos pontos. Em imagens esses pontos são os *pixels*. Os operadores morfológicos utilizam-se de elementos estruturantes. Estes elementos são constituídos por um conjunto de pixels, cuja forma e tamanho dependem do objeto estudado na imagem. Há vários operadores morfológicos, porém todos eles são formados pela interação de duas operações básicas: erosão e dilatação (ALVARENGA et al., 2003).

A erosão de uma imagem  $A$  para determinado *pixel*  $x$  é definida pelo valor mínimo em uma janela definida pelo elemento estruturante  $B$  (Equação 1), que é usado para encolher o objeto encontrado na imagem.

$$(A \ominus B)(x) = \min_{b \in B} f(x + b) \quad (1)$$

onde  $b$  é um pixel pertencente ao elemento estruturante  $B$ .

A dilatação, por sua vez, utiliza o elemento estruturante para expandir as formas do objeto que se encontra na imagem  $A$ , e é representada pela Equação 2:

$$(A \oplus B)(x) = \max_{b \in B} f(x + b) \quad (2)$$

### 2.3.2.2 Watershed

Na segmentação por *Watershed*, a imagem é comparada a um relevo topográfico, onde os níveis de cinza mais baixos representam vales e os níveis de cinza mais altos, montanhas. A técnica simula uma inundação desses vales (que pode ser imaginada como chuva ou através de buracos nos pontos mais baixos). A cada iteração do algoritmo, ocorre um incremento nessa inundação. À medida que as bacias vão sendo inundadas, águas de diferentes bacias tendem a se encontrar no pico da montanha entre elas. Quando isso acontece, é formada uma barreira. Estas barreiras correspondem às linhas divisórias encontradas pelo algoritmo.

O algoritmo é descrito com mais detalhes em (GONZALEZ, 2002) e faz uso de morfologia matemática para construção de barreiras que podem ser criadas com elementos estruturantes de dilatação de 3x3 pixels como descrito em (ERICEIRA, 2011). Neste trabalho, aplica-se o algoritmo da mesma forma.

### 2.3.3 Registro de Imagem

O registro de imagens é o processo que corresponde ao alinhamento espacial entre duas ou mais imagens. Este alinhamento é feito através da correspondência ponto a ponto entre as imagens. Vários fatores podem contribuir para a existência de diferenças espaciais em mamografias, dentre os quais podemos citar: a estrutura de tecidos moles deformável, móvel, anisotrópica e não homogênea da mama, o posicionamento da mama, a compressão aplicada durante o exame, etc. Normalmente se trabalha no processamento de imagens mamográficas com registro rígido e registro deformável, que são apresentados nas subseções a seguir.

### 2.3.3.1 Registro Rígido

O registro rígido baseia-se em operações de translação, rotação e escala, com objetivo de reduzir as diferenças globais, como posicionamento e tamanho, entre as imagens, aplicando uma transformação linear na imagem toda. A imagem fonte e a imagem alvo podem ser vistas, respectivamente, como duas funções  $f=(x,y)$  e  $f' = (x',y')$ , onde a relação entre as intensidades dos *pixels* das imagens é dada pelas Equações 3 e 4.

$$x' = S(x \cos\theta + y \sen\theta) + h \quad (3)$$

$$y' = S(-x \sen\theta + y \cos\theta) + k \quad (4)$$

onde  $S$ ,  $\theta$  e  $(h,k)$  são as diferenças de escala, rotação e translação entre as imagens (GOSHTASBY, 2005).

### 2.3.3.2 Registro Deformável com Algoritmo “*Demons*”

O algoritmo “*Demons*” é bastante implementado para a realização do registro deformável de imagens (com uma imagem fonte e uma imagem alvo) (THIRION, 1995). “A ideia principal é a de que uma ‘rede’ regular de forças deforma a imagem empurrando os contornos na direção normal. A direção e magnitude do deslocamento são derivadas da equação final do fluxo óptico” (ERICEIRA, 2011, p. 37):

$$D(X) \cdot \nabla f(X) = -(m(X) - f(X)) \quad (5)$$

sendo  $f(X)$  a imagem fixa,  $m(X)$  a imagem alvo a ser registrada, e  $D(X)$  o campo de fluxo óptico (ou campo de deformação) entre as imagens

O campo de fluxo óptico é estimado através da aplicação de novas restrições à Equação 5, como explanado em (ERICEIRA, 2011) sendo usado em uma transformação não linear para transformar a imagem alvo em relação à imagem fonte.

## 2.4 Coeficiente de Correlação Cruzada

O Coeficiente de Correlação Cruzada (que neste trabalho será chamado de CCC) possui aplicações em estudo de sinais, reconhecimento de padrões e processamento de imagem. Por exemplo, em (RODRIGUES, 2010) foi utilizado como métrica de similaridade para corregristo não rígido. É uma medida para estimar o grau de correlação entre duas séries ou funções. Em processamento de imagem, estas funções nada mais são do que as próprias



imagens (ou regiões das imagens). O CCC é indicado para ser usado em imagens que tenham sido capturadas pelo mesmo sensor (MITCHELL, 2010), por exemplo, um par de mamografias da mesma paciente.

Sejam  $A$  e  $B$  duas imagens (ou duas regiões correspondentes das imagens) compostas pela mesma quantidade de *pixels*, onde  $a_k$  é o pixel de índice  $k$  na imagem  $A$  e  $b_k$  é o pixel de índice  $k$  na imagem  $B$ , o CCC pode ser usado como uma medida de similaridade entre as imagens e é descrito pela seguinte equação (MITCHELL, 2010):

$$\rho = \frac{\sum_k a_k b_k}{\sqrt{\sum_k a_k^2 \sum_k b_k^2}} \quad (6)$$

A medida gerada pela Equação 6 do CCC (representado por  $\rho$  na equação) varia de 0 a 1, onde valores mais próximos de 1 significam imagens mais similares. Neste trabalho o CCC é usado para medir a similaridade entre regiões correspondentes de mamografias, ou seja, entre regiões na mama esquerda e suas correspondentes na mama direita da mesma paciente.

## 2.5 Distância Euclidiana

A Distância Euclidiana (chamada neste trabalho pela expressão DE) é uma medida utilizada para definir a distância entre dois pontos em um espaço  $n$ -dimensional. Esta medida é generalizada pela distância de Minkowski. Sejam dois pontos genéricos de  $n$  dimensões  $O(o_1, o_2, \dots, o_n)$  e  $P(p_1, p_2, \dots, p_n)$ , a distância de Minkowski entre os dois é expressa pela Equação 7:

$$DM_{OP} = \sqrt[k]{|o_1 - p_1|^k + |o_2 - p_2|^k + \dots + |o_n - p_n|^k} \quad (7)$$

onde  $k \geq 1$ .

A DE pode ser vista como um caso especial da distância de Minkowski, onde  $k = 2$ . Considerando os pontos como sendo imagens e as coordenadas como sendo os valores possíveis para os pixels das imagens, pode-se, analogamente, medir a DE entre essas imagens através dos níveis de intensidade de seus pixels. Assim, sendo  $A(a_1, a_2, \dots, a_n)$  e  $B(b_1, b_2, \dots, b_n)$  duas imagens de mesma dimensão (ou duas regiões correspondentes de duas imagens), compostas por  $n$  pixels cada e dispostas linearmente, a DE entre  $A$  e  $B$  pode ser escrita, analogamente, como:

$$DE_{AB} = \sqrt{((a_1 - b_1)^2) + ((a_2 - b_2)^2) + \dots + ((a_n - b_n)^2)} \quad (8)$$

Quanto maior for a divergência entre os valores dos *pixels* da imagem *A* e da imagem *B*, maior será o valor gerado pela Equação 8.

## 2.6 Índice de Getis-Ord

O índice de Getis-Ord é uma medida de associação espacial que se apresenta em duas formas, local e geral. A sua forma local é bastante utilizada, por exemplo em (BRAZ JÚNIOR, 2008), para extração de características de imagens mamográficas e em (WULDER, 1998) para agrupamento de vegetação a partir de imagens de satélite. Esta forma mede a associação a partir da concentração dos pontos que se encontram dentro de uma distância  $d$  do ponto original. A sua forma geral considera todos os pares de pontos possíveis a uma distância  $d$  entre si (GETIS;ORD, 1992). Esta última forma pode ser usada como índice de associação entre os pontos de uma região e servir como um descritor geral da mesma, por isso é utilizada neste trabalho para tal fim e se apresenta como segue:

$$G(d) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}, \quad j \neq i \quad (9)$$

Na Equação 9,  $G(d)$  equivale ao quociente entre o somatório de todos os pares de *pixels* (o produto entre esses *pixels*), que estejam dentro da (ou seja, até a) distancia  $d$  entre si, pelo somatório de todos os pares de *pixels* possíveis, com o *pixel*  $i$  diferente do *pixel*  $j$ . A matriz de vizinhança  $w_{ij}$  recebe o valor 1 quando o par de pixels  $(x_i, x_j)$  está a uma distância menor ou igual a  $d$  um do outro.

O conceito de distância, neste caso, se torna relativo e pode variar de acordo com a forma que a medida for utilizada. Neste trabalho a distância  $d$  é definida como os deslocamentos na vertical e na horizontal, determinando uma janela quadrada ao redor de um ponto, neste caso um *pixel*. Por exemplo,  $d$  igual a 1 determina uma janela de 3x3 *pixels* ao redor do *pixel* em questão, assim como  $d$  igual a 2 determina uma janela de 5x5 *pixels*.

## 2.7 Máquina de Vetores de Suporte

Uma Máquina de Vetores de Suporte (MVS) é uma forma de aprendizado de máquina que busca classificar um conjunto de dados (conjunto de teste) a partir da criação de um modelo baseado nas características obtidas de outro conjunto (conjunto de treinamento). Estando esses dados dispostos em um espaço dimensional e pertencendo a classes de dados

distintas, tenta-se determinar um hiperplano (uma função) que separe ao máximo estas classes uma da outra.

Seja o conjunto de dados de treinamento formado por pares  $(x_i; y_i)$ , sendo  $x_i$  o vetor de entrada (esse vetor representa um ponto no espaço de entrada) e  $y_i$  a classificação correta do vetor, onde  $x_i$  pertence a  $\mathbb{R}^n$  e  $y_i$  pertence a  $\{-1, 1\}$ . Cada ponto  $x$  no espaço de entrada é mapeado para um ponto  $z = \phi(x)$  em um espaço de características, onde os dados podem ser separados linearmente por um hiperplano.

Como apresentado em (ERICEIRA, 2011), a MVS exige a solução para o seguinte problema de minimização:

$$\min_{w, \xi} \frac{1}{2} w^2 + C \sum_{i=1}^n \xi_i \quad (10)$$

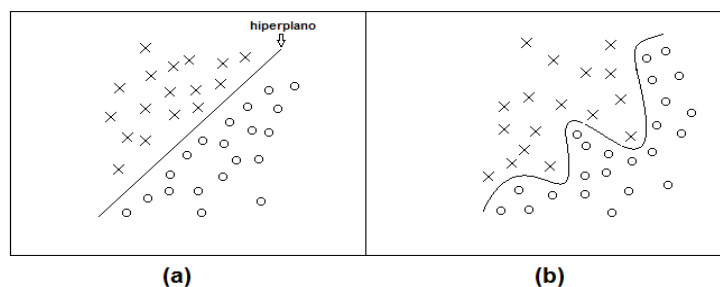
sujeito a:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (11)$$

$C > 0$  é um parâmetro que correspondendo à penalidade do erro e os  $\xi_i$ 's, às variáveis de folga que penalizam os erros de treinamento.

Um conjunto de dados pode ser linearmente separável ou não e é dito linearmente separável quando seus dados podem ser separados por um hiperplano, como a Figura 4 tenta exemplificar de forma simplificada. O que torna possível o trabalho com conjuntos não linearmente separáveis é o uso de funções chamadas *kernel*, capazes de realizar o mapeamento dos dados para um espaço de dimensão mais elevada.

Figura 4 – Conjuntos linearmente separáveis e não linearmente separáveis



(a) Conjuntos linearmente separáveis. (b) Conjuntos não linearmente separáveis. Fonte: O autor.

Neste trabalho, a função *kernel* utilizada foi a *Radial Basis Function* (RBF) da Equação 12, que geralmente apresenta bons resultados, como analisado em (KEERTHI; LIN, 2003).

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0 \quad (12)$$

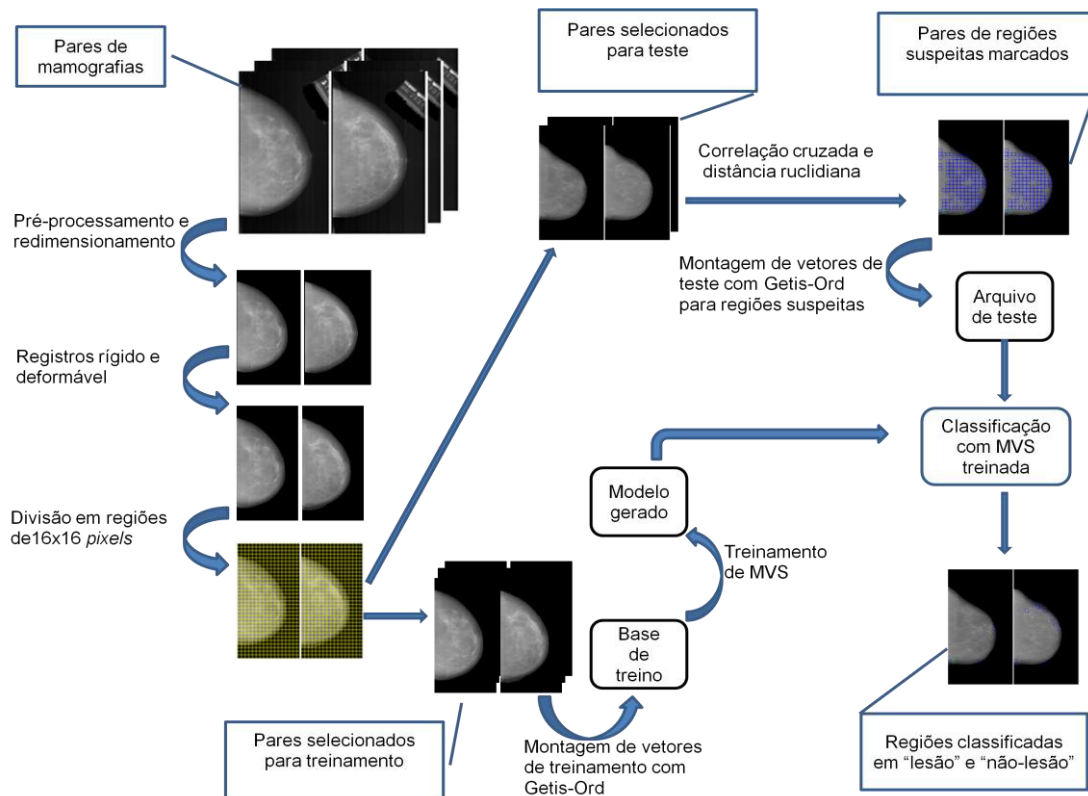
### 3. METODOLOGIA

Esta seção descreve a metodologia desenvolvida pelo trabalho. Primeiramente é feita a aquisição de pares de imagens de mamografias (mama esquerda e mama direita) de várias pacientes. Em seguida é aplicado um pré-processamento, composto de uma fase de redução de ruídos e outra de segmentação para isolar o objeto de interesse, no caso a mama. A próxima etapa é um redimensionamento das imagens para o tamanho de 512 *pixels* de largura e uma altura proporcional à original, para tornar as etapas posteriores mais viáveis computacionalmente e evitar distorções desnecessárias na imagem. Depois de feito o redimensionamento, as imagens passam pelos processos de registro rígido e registro deformável, que têm a função de aproximar ao máximo o formato da mama direita do formato da mama esquerda. Então as imagens são divididas em regiões de 16x16 *pixels*.

Assim, para cada par de regiões correspondentes nas mamas esquerda e direita dos casos selecionados para teste, são calculados os valores de *CCC* e de *DE* entre os níveis de cinza dos *pixels* do par de regiões. Após isso são obtidos os valores medianos de *CCC* e *DE*. Então cada par de regiões é, de acordo como o nível de simetria, previamente classificado como suspeito ou não suspeito baseado nas medianas encontradas. Entende-se aqui simetria como sendo o nível de similaridade indicado pelo *CCC* e pela *DE*.

Uma base de treinamento é montada para a *MVS* a partir dos valores de índice de *Getis-Ord* de cada região dos casos separados para treino. As regiões dos casos de teste previamente determinadas, aos pares, como as mais assimétricas (suspeitas) são testadas, desta vez individualmente (unilateralmente), e classificadas como lesão ou não lesão. A Figura 5 demonstra a metodologia proposta neste trabalho utilizando imagens do *DDSM* e as subseções seguintes explicam cada etapa da mesma.

Figura 5 – Representação gráfica da metodologia utilizando imagens do DDSM



Fonte: O autor.

### 3.1 Aquisição das Mamografias

Foram utilizadas imagens do banco DDSM que contém 2620 casos adquiridos a partir de hospitais e instituições (HEATH et al., 2001). É uma base de imagens de mamografias disponibilizada na Internet para propósitos de pesquisa e bastante utilizada pela comunidade científica. A base é dividida em vários volumes, cada volume possui vários casos e cada caso possui seu diagnóstico para imagens em projeções crânio-caudal (CC) e médio-lateral oblíqua (MLO), além de detalhes referentes ao caso (como idade da paciente e nível de densidade das mamas que varia de 1 a 4) e a eventuais lesões presentes na mama. Imagens que possuem lesão contêm uma indicação da área lesionada. Foram selecionados no total 499 pares de imagens do tipo CC, por terem apresentado melhor resultado quanto às técnicas de processamento de imagem a serem citadas, provavelmente por causa da área muscular, com altos níveis de cinza, existente no tipo MLO. Todas as imagens de mamografias utilizadas na metodologia deste trabalho pertencem ao DDSM.

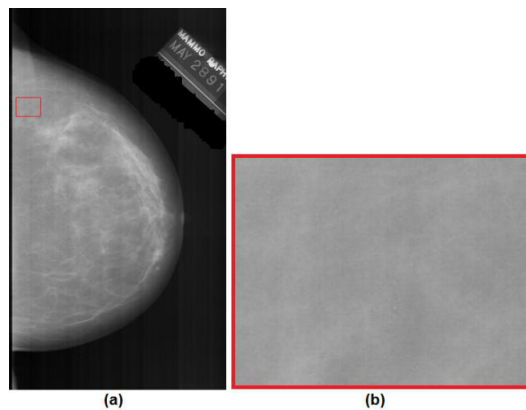
### 3.2 Pré-Processamento e Redimensionamento das Mamografias

Nesta etapa as imagens passaram por processos de redução de ruído e segmentação e depois foram redimensionadas. A manipulação das imagens foi feita com uso da biblioteca OpenCV (BRADSKI; KAEHLER, 2008). Os processos são descritos a seguir.

#### 3.2.1 Redução de Ruído

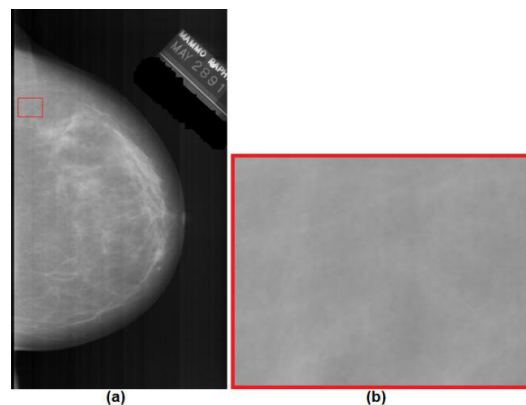
A redução de ruído foi feita com a aplicação de um filtro de mediana de tamanho 5x5 pixels em toda a imagem. Este processo é realizado devido à presença de eventuais ruídos na imagem adquiridos, possivelmente, durante a digitalização. As Figuras 6 e 7 mostram, respectivamente, uma mamografia antes e depois do processo.

Figura 6 – Mamografia antes da redução de ruído



(a) Mamografia.(b) Área em destaque ampliada.

Figura 7 – Mamografia após a redução de ruído

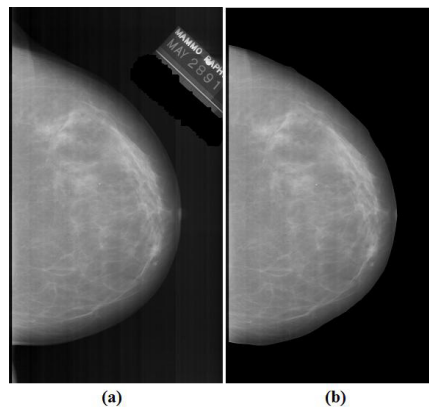


(a) Mamografia.(b) Área em destaque ampliada.

### 3.2.2 Segmentação das Mamas

A etapa de segmentação das mamas, utilizou uma implementação da técnica de *Watershed* com aplicação de operador morfológico de dilatação para estimar a borda da mama. Esta etapa visa isolar o objeto de interesse, no caso a mama, e remover artefatos não pertencentes à mama em si presentes na imagem, além de aumentar a precisão e rapidez do registro. A Figura 8 mostra o resultado da etapa de segmentação.

Figura 8 – Mamografia antes e após segmentação



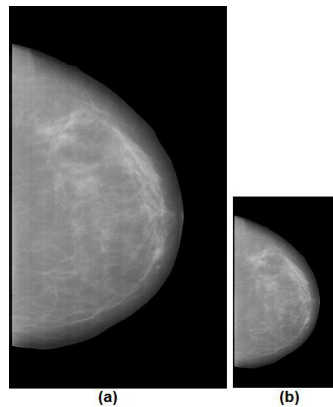
(a) Mamografia antes da segmentação. (b) Mamografia após segmentação.

### 3.2.3 Redimensionamento das Mamografias

Etapa em que todas as imagens foram redimensionadas para o tamanho de 512 *pixels* de largura por uma altura proporcional à original (às imagens originais possuem em média dimensões da ordem de 4000 *pixels* de largura por 6000 *pixels* de altura aproximadamente), para tornar os cálculos posteriores mais viáveis computacionalmente. A perda de informação causada pela redução das imagens não demonstrou ter grande impacto na metodologia, comparado ao tempo ganho com o processamento das imagens reduzidas. A Figura 9 mostra a imagem antes e após o redimensionamento.



Figura 9 – Mamografia antes e após redimensionamento



(a) Mamografia antes do redimensionamento. (b) Mamografia após o redimensionamento. Escala alterada para visualização neste trabalho.

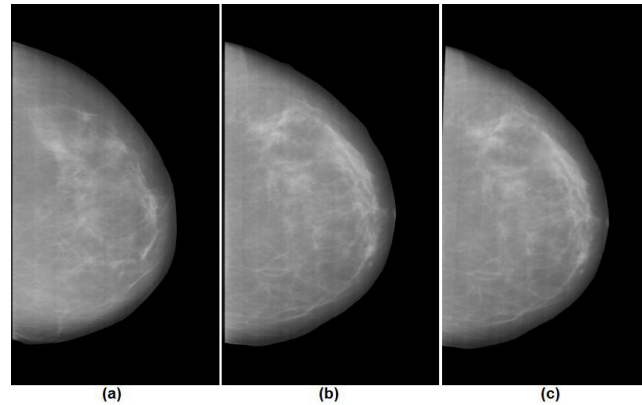
### 3.3 Registro das Mamografias

O registro das mamografias é dividido em registro rígido e registro deformável, ambos tentam transformar a mama direita tendo a esquerda como base. O primeiro busca diminuir diferenças como escala e posicionamento entre as mamas e o segundo, partindo da saída do primeiro, aplica uma serie de deformações na mama direita para aproximá-la ao máximo do formato da esquerda. Para aplicação das técnicas de registro, foi utilizado o conjunto de ferramentas ITK (*The Insight Segmentation and Registration Toolkit*) (ITK, 2012).

#### 3.3.1 Registro Rígido

Foram utilizadas no registro rígido transformações de escala, rotação e translação com intuito de diminuir as diferenças globais entre mama esquerda e mama direita, como posição das mamas, rotação e tamanho, que podem ter sido criadas na própria realização do exame. As transformações são aplicadas em toda a imagem. A Figura 10 demonstra o resultado obtido com o registro rígido.

Figura 10 – Registro rígido da mama direita com base na mama esquerda

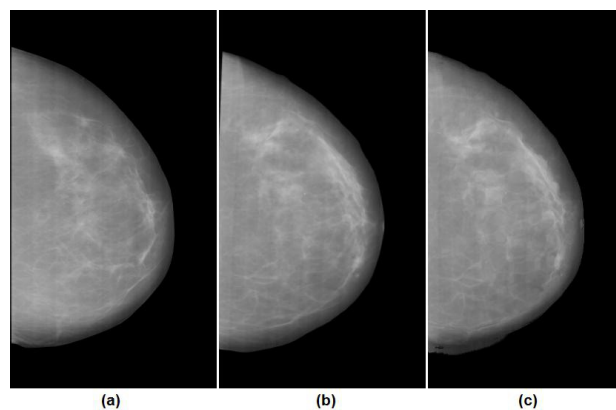


(a) Mama esquerda.(b) Mama direita. (c) Mama direita após registro rígido.

### 3.3.2 Registro Deformável

O registro deformável visa melhorar a saída do registro rígido, visto que as transformações lineares aplicadas pelo registro rígido modificam somente alguns aspectos (como posicionamento e escala) da mamografia, e foi feito aplicando-se o algoritmo conhecido como “Demons”. A Figura 11 exibe um par de imagens após os registros rígido e deformável.

Figura 11 – Registro rígido e registro deformável da mama direita com base na mama esquerda

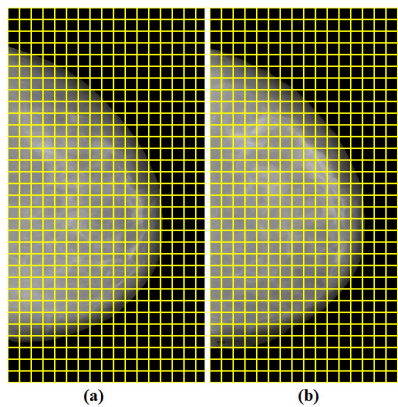


(a) Mama esquerda.(b) Mama direita após registro rígido.(c) Mama direita após registro deformável.

### 3.4 Divisão das Mamografias em Regiões de 16x16 *pixels*

Todas as imagens foram divididas em regiões de tamanho 16x16 *pixels*, como pode ser visto na Figura 12, de forma que se pudesse fazer a correspondência entre regiões nas mamas esquerda e direita da mesma paciente (somente pares de regiões internas às mamas). Este tamanho foi escolhido por que com regiões maiores (por exemplo, de 32x32 *pixels*) foi constatado empiricamente que muitas lesões ficavam localizadas em fronteiras de regiões o que dificultava (dependendo do tamanho da lesão) a caracterização da região como área lesionada. Já tamanhos menores reduziriam muito a informação (características) de vizinhança presentes na região.

Figura 12 – Divisão das mamografias em regiões de 16x16 *pixels*

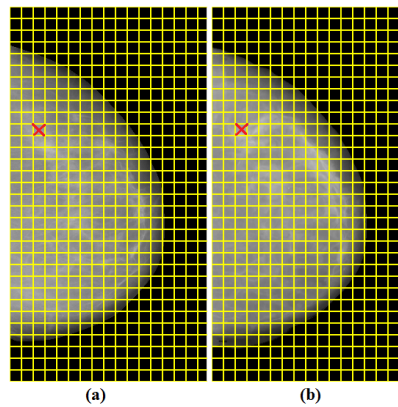


(a) Mama esquerda.(b) Mama direita.

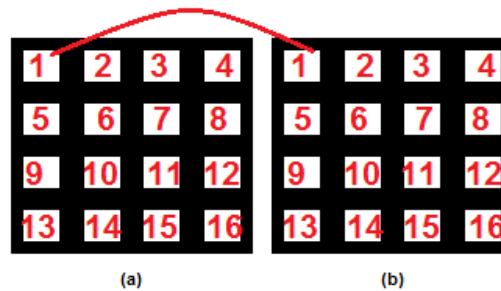
### 3.5 Determinação de Regiões Suspeitas com CCC e DE

Deseja-se submeter à classificação final somente as regiões com alto grau de assimetria. Para isso, esta etapa visa classificar previamente os pares de regiões de cada par de mamografias em suspeitos (assimétricos) e não suspeitos (simétricos). Para cada par de regiões (Figura 13) é calculado um valor de CCC e um valor de DE. As variáveis  $a$  e  $b$  das Equações 6 e 8, mostradas anteriormente, são os níveis de cinza dos *pixels* correspondentes na região esquerda e na direita, como exemplifica a Figura 14.

Figura 13 – Regiões correspondentes no par de mamografias



(a) Mama esquerda. (b) Mama direita. Regiões correspondentes marcadas com um X vermelho.

Figura 14 – Correspondência entre *pixels* de regiões hipotéticas de 4x4 *pixels*

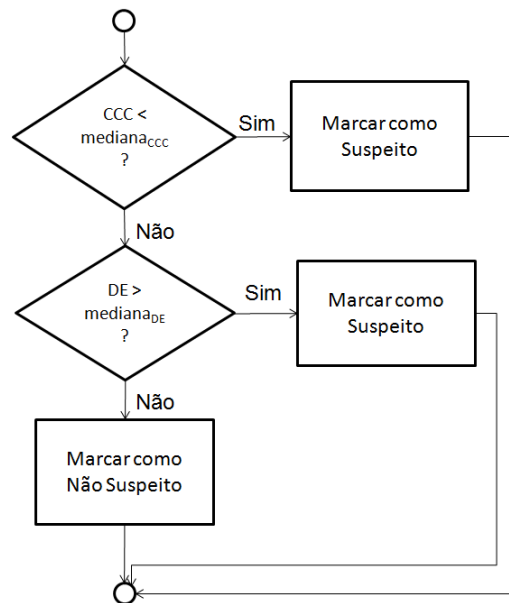
(a) Região esquerda. (b) Região direita. Fonte: O autor.

Uma vez determinados os valores de CCC e de DE para cada par de regiões, usam-se os valores medianos de cada medida como limiares de similaridade. A utilização da mediana como limiar foi definida empiricamente e pelo fato de ela ser menos influenciada por valores extremos do que, por exemplo, a média.

Quanto maior o grau de assimetria, menor o valor de CCC e maior a DE. Assim, os pares de regiões com valores abaixo da mediana de CCC foram classificados como suspeitos. Porém, regiões visivelmente assimétricas, mas com uma distribuição homogênea de níveis de cinza, tendem a possuir altos valores de CCC. Neste caso usa-se a mediana da DE como contraprova, sendo que os pares com DE maior que a mediana da mesma medida para o par de mamas são classificados como suspeitos. O restante dos pares de regiões são

classificados como não suspeitos (simétricos). A Figura 15 demonstra o processo descrito acima.

Figura 15 – Determinação de regiões suspeitas por assimetria usando CCC e DE

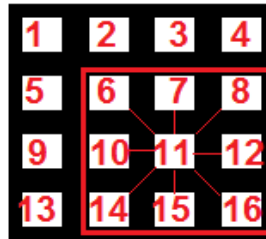


Fonte: O autor.

### 3.6 Extração de Características com Índice de Getis-Ord

Como foi mostrado na Seção 2.6 o índice de Getis-Ord, na sua forma geral, mede o nível de associação entre os pontos, neste caso os *pixels*, de uma região. Para o cálculo do índice de Getis-Ord de uma região são considerados todos os pares de *pixels* possíveis dentro da mesma região para distâncias de 1 a 14 *pixels*. Considera-se distância aqui como uma janela ao redor do pixel que cresce na vertical e na horizontal como mostrado na Figura 16, uma região hipotética de 4x4 *pixels*. Nesta figura, por exemplo, é exibida a janela com distância igual a 1 ao redor do *pixel* 11. Considerando a Equação 9, se na iteração do cálculo o valor de  $x_i$  estiver no nível de cinza do pixel 11, os pares  $(x_i, x_j)$  incluídos na equação serão os níveis de cinza dos *pixels* (11,6), (11,7), (11,8), (11,10), (11,12), (11,14), (11,15) e (11,16).

Figura 16 – Distância 1 para cálculo de índice de Geti-Ord



Pares possíveis com *pixel* 11 para distância 1 em destaque. Fonte: O autor.

A distância 15, para regiões de 16x16 *pixels*, que englobaria todos os *pixels* da região (de acordo com o a forma de distância e matriz de vizinhança utilizadas), não foi considerada, visto que ela sempre gera um índice de valor 1 para qualquer caso, portanto não serve como medida de distinção. No lugar da ultima distância foi incluída a média dos níveis de cinza da região. Esta medida foi utilizada por que regiões homogêneas tendem a possuir índices de Getis-Ord parecidos, mesmo que possuam níveis de cinza bem diferentes. Assim á media foi escolhida como uma característica a mais para ajudar a distinguir regiões homogêneas mas com uma diferença muito grande entre os seus níveis de cinza. Os valores de Getis-Ord variam de 0 a 1 e a média foi colocada na mesma escala.

Então para cada região isolada (não mais aos pares), montou-se um vetor com 15 características, sendo as 14 primeiras os índices de Getis-Ord para distâncias de 1 a 14 e a última a média dos níveis de cinza da região. Com os vetores de características das regiões obtidas a partir dos casos selecionados para treino, foi formada uma base de treinamento para uma MVS. E a para os casos selecionados para teste foram gerados vetores de características somente para as regiões consideradas suspeitas.

### 3.7 Classificação Final com MVS

Para realização da classificação final das regiões que foram previamente consideradas suspeitas, foi gerado um modelo a partir da base de treinamento para ser usado na classificação das regiões suspeitas em lesão e não lesão. Nesta etapa os vetores de características das regiões suspeitas são submetidos à MVS treinada, lembrando que a classificação é feita agora unilateralmente, ou seja, para cada região, não mais para o par de regiões. Foi utilizado o *software* LIBSVM (CHANG;LIN, 2011) para esta etapa.

## 4. RESULTADOS

Neste capítulo são apresentados e discutidos os resultados de testes realizados para validação da metodologia proposta no Capítulo 3. Como proposto pela metodologia, a classificação de regiões é dividida em duas fases. Primeiramente os pares de regiões são determinados como suspeitos e não suspeitos. Posteriormente, as regiões dos pares suspeitos são submetidas individualmente à segunda fase para classificação em lesão e não lesão.

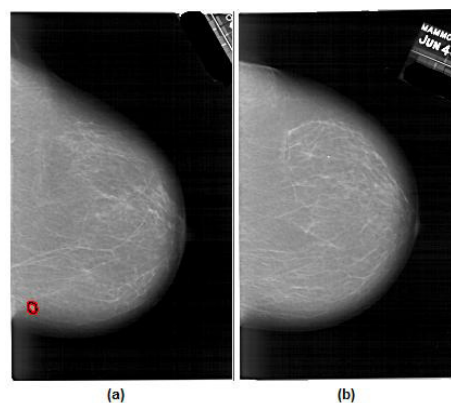
### 4.1 Pares de Regiões Suspeitas

Foram selecionados 60 pares de mamografias para servirem de casos de testes, sendo que 30 deles apresentavam algum tipo de lesão e os outros 30 casos com diagnóstico normal, todos na projeção crânio-caudal. Todas as imagens passaram pelas etapas de pré-processamento, registro e divisão em regiões. Nesta seção são discutidos os resultados de 4 pares de imagens, sendo dois com lesões detectadas pela metodologia, um normal e um com lesão não detectada pela metodologia.

#### 4.1.1 Pares de Regiões Suspeitas: Teste 1

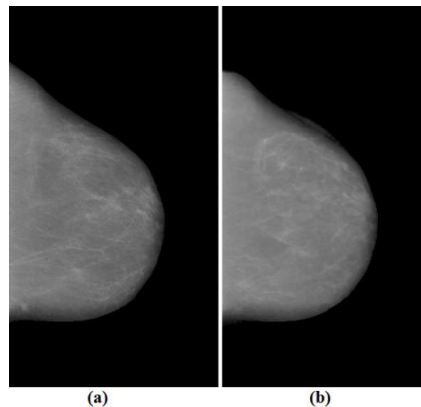
Um dos pares selecionados para teste é exibido na Figura 17 com a área lesionada destacada em vermelho na mama esquerda. É o par A1639 do banco DDSM, pertence a uma paciente de 73 anos de idade com densidade de nível 2. A lesão exibida é uma massa de formato irregular e margens mal definidas. O par também é exibido na Figura 18 após as etapas de pré-processamento e registro.

Figura 17 – Caso A1639 do banco DDSM



(a) Mama esquerda com área da lesão destacada em vermelho. (b) Mama direita.

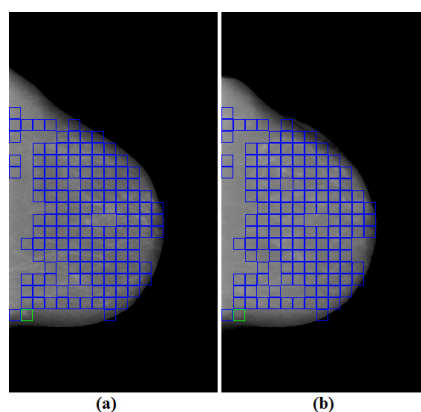
Figura 18 – Caso A1639 do banco DDSM após pré-processamento e registro



(a) Mama esquerda.(b) Mama direita.

O par de imagens foi dividido em regiões de 16x16 pixels, o que gerou 195 pares de regiões internas às mamas, com 194 pares normais e 1 com lesão. Após o cálculo do coeficiente de correlação cruzada e da distância euclidiana para cada par de regiões, os pares mais assimétricos foram classificados como suspeitos, conforme o explicado na seção 3.5. A Figura 19 mostra os pares de regiões suspeitos para este par de imagens. Neste caso 140 pares de regiões apresentaram um nível de assimetria mais elevado e foram considerados suspeitos, incluindo o par de regiões com a lesão marcada no DDSM.

Figura 19 – Regiões classificadas como suspeitas no caso A1639 do banco DDSM



(a) Mama esquerda. (b) Mama direita. Destacam-se os pares de regiões considerados suspeitos em azul e verde.

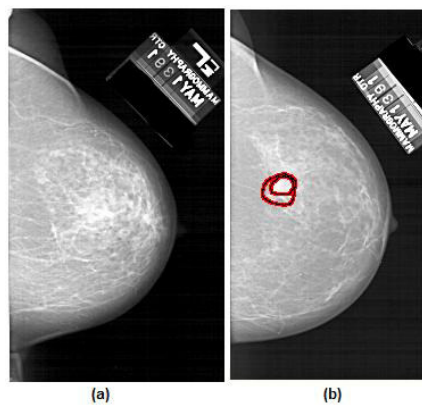
O par de regiões verde corresponde à lesão.



#### 4.1.2 Pares de Regiões Suspeitas: Teste 2

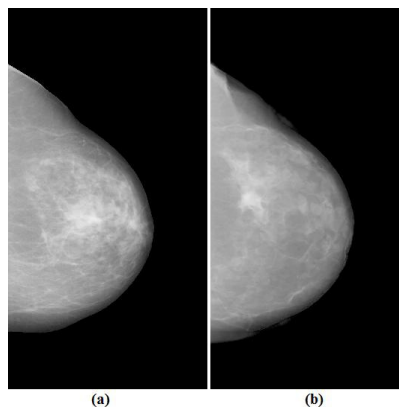
Outro par de imagens usado como teste foi o correspondente ao caso A1134 do banco DDSM. Este é de uma paciente de 69 anos, com densidade de nível 2. A lesão possui formato irregular e margens espiculadas. A Figura 20 mostra o par de mamografias com a lesão na mama direita destacada em vermelho. Já a Figura 21 mostra o par após pré-processamento e registro.

Figura 20 – Caso A1134 do banco DDSM



(a) Mama esquerda. (b) Mama direita com área lesionada destacada em vermelho.

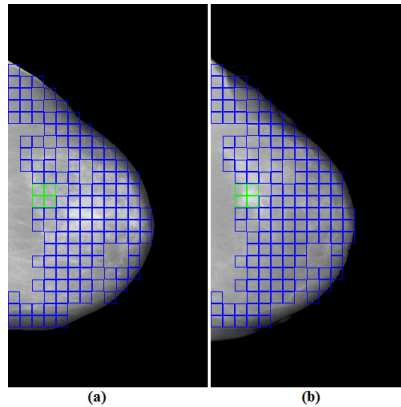
Figura 21 – Caso A1134 do banco DDSM após pré-processamento e registro



(a) Mama esquerda. (b) Mama direita.

Este par de imagens gerou 187 pares de regiões internas à mama com 183 pares normais e 4 pares com lesão. Foram classificados como suspeitos 136 pares de regiões, incluindo os 4 pares com a lesão. Estes podem ser vistos na Figura 22.

Figura 22 – Regiões classificadas como suspeitas no caso A1134 do banco DDSM

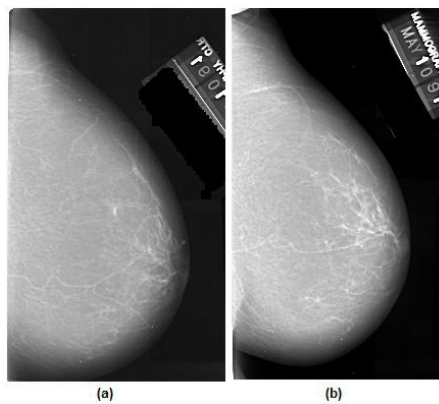


(a) Mama esquerda. (b) Mama direita. Destacam-se os pares de regiões considerados suspeitos em azul e verde. Os pares de regiões em verde correspondem à lesão.

#### 4.1.3 Pares de Regiões Suspeitas: Teste 3

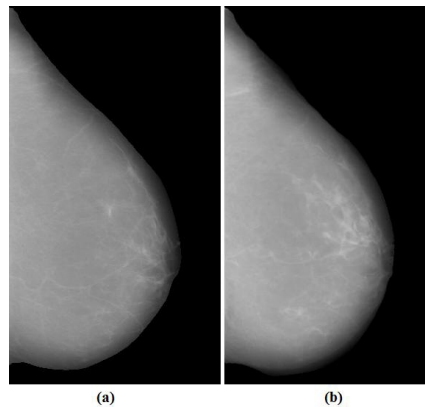
O próximo caso de teste possui diagnóstico normal no banco DDSM. Trata-se do caso A0074, paciente com idade de 53 anos e densidade das mamas de nível 1. A Figura 23 exibe o par de mamografias e a Figura 24, o mesmo após o pré-processamento e registro.

Figura 23 – Caso A0074 do banco DDSM



(a) Mama esquerda. (b) Mama direita.

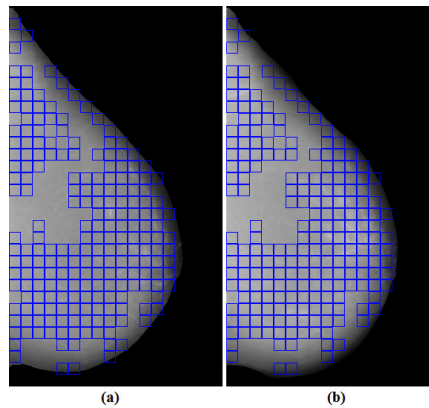
Figura 24 – Caso A0074 do banco DDSM após pré-processamento e registro



(a) Mama esquerda. (b) Mama direita.

Na Figura 25 são exibidos os 211 pares de regiões considerados assimétricos (suspeitos) de um total de 287 pares internos às mamas do caso A0074, lembrando que essa classificação é feita baseada nos valores medianos de correlação cruzada e distância euclidiana do próprio par de mamografias. Neste caso não há lesão.

Figura 25 – Regiões classificadas como suspeitas no caso A0074 do banco DDSM



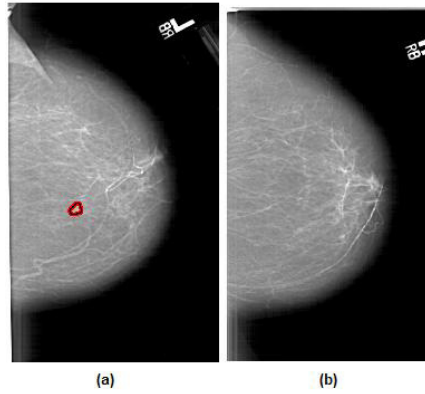
(a) Mama esquerda. (b) Mama direita. Destacam-se os pares de regiões considerados suspeitos em azul.

#### 4.1.4 Pares de Regiões Suspeitas: Teste 4

O quarto teste exibido corresponde ao caso A1627 do banco DDSM e possui uma lesão com formato irregular e margens mal definidas exibida na Figura 26, paciente com 72

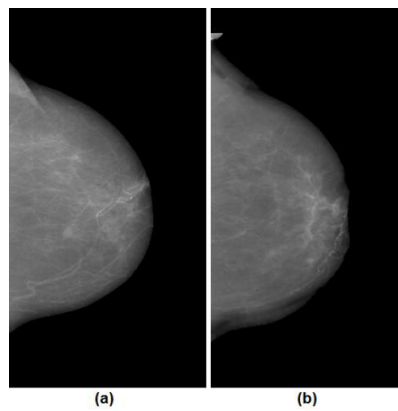
anos e densidade das mamas de nível 1 A Figura 27 exibe o par de mamografias após pré-processamento e registro.

Figura 26 – Caso A1627 do banco DDSM



(a) Mama esquerda com área lesionada destacada em vermelho. (b) Mama direita.

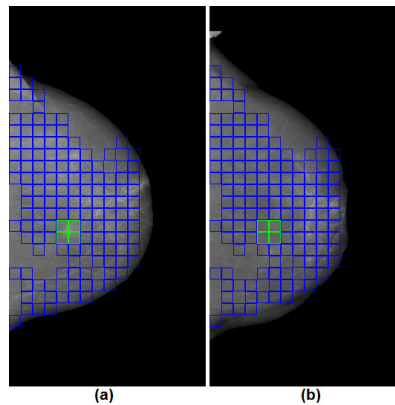
Figura 27 – Caso A1627 do banco DDSM após pré-processamento e registro



(a)Mama esquerda. (b)Mama direita.

Foram considerados suspeitos 141 pares de regiões para este caso, de um total de 183 pares internos. A área marcada como lesão no DDSM está incluída nos pares de regiões suspeitos, destacados na Figura 28.

Figura 28 – Regiões classificadas como suspeitas no caso A1627 do banco DDSM



(a)Mama esquerda. (b)Mama direita. Destacam-se os pares de regiões considerados suspeitos em azul e verde.  
Os pares de regiões em verde correspondem à lesão.

#### 4.2 Detecção de Regiões com Lesão

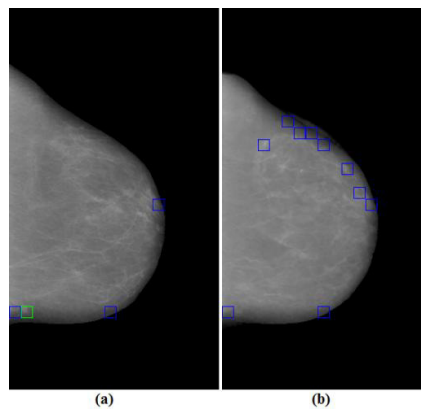
Para esta etapa dos testes, foi necessário montar uma base de treinamento para a MVS. Esta base contém características de regiões individuais, não de pares. Para isso foram selecionados 439 pares de mamografias, também em projeção crânio-caudal. Destes, 309 pares eram de casos que apresentavam lesão em uma das mamas, dos quais os lados com lesão foram usados para gerar um total de 1808 regiões lesionadas. Para manter uma proporção de 1:2, dos outros 130 pares (260 imagens) normais restantes, foram geradas 3616 regiões normais. A partir de cada uma dessas regiões (lesionadas e normais) foi extraído um vetor com 15 características, como mostrado no Capítulo 3, gerando um total de 5424 vetores de características.

Os vetores extraídos foram usados para montar a base de treinamento da MVS. Com esta base foram estimados os parâmetros ( $C=512,0$  e  $\gamma=2,0$ ) necessários para criar o modelo de treinamento, como mostrado no Capítulo 2. As mesmas imagens usadas na primeira fase da classificação, ou seja, na detecção de regiões suspeitas, foram submetidas à MVS treinada. Porém desta vez, somente as regiões consideradas previamente como suspeitas foram submetidas, e as imagens foram testadas isoladamente, ou seja, a esquerda separada da direita.

#### 4.2.1 Detecção de Regiões com Lesão: Teste 1

As regiões suspeitas do caso A1639 do banco DDSM foram submetidas à MVS treinada para classificação em “lesão” e “não lesão”. O resultado pode ser visto na Figura 29. As regiões destacadas em azul e verde foram classificadas como lesão, sendo que a região em verde corresponde à lesão marcada no DDSM.

Figura 29 – Classificação final do caso A1639

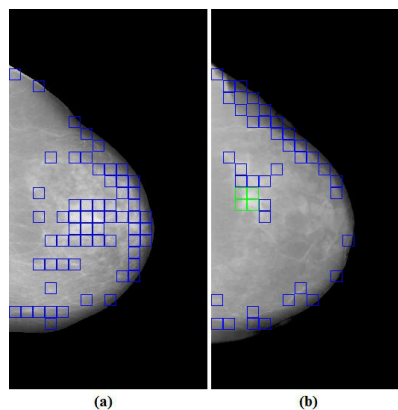


(a)Mama esquerda.(b) Mama direita. Regiões classificadas como “lesão” em azul e verde, sendo a região em verde a correspondente à lesão marcada no DDSM.

#### 4.2.2 Detecção de Regiões com Lesão: Teste 2

O resultado da submissão à MVS treinada das regiões suspeitas do caso A1134 pode ser visto na Figura 30.

Figura 30 – Classificação final do caso A1134

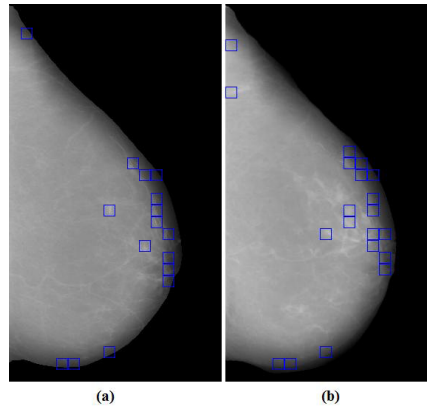


(a)Mama esquerda. (b)Mama direita. Regiões classificadas como “lesão” em azul e verde, sendo as regiões em verde as correspondentes à lesão marcada no DDSM.

#### 4.2.3 Detecção de Regiões com Lesão: Teste 3

Na Figura 31 é exibida a classificação final do caso A0074 do banco DDSM, que possui diagnóstico normal. Como este caso não possui lesão em nenhuma das mamas, todas as regiões destacadas foram classificadas como “lesão” pela MVS erroneamente.

Figura 31 – Classificação final do caso A0074

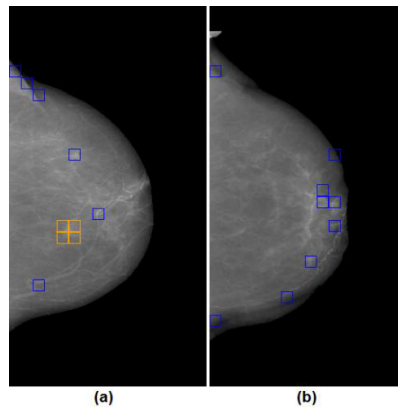


(a)Mama esquerda.(b)Mama direita. Regiões classificadas como “lesão” em azul. Não é indicada lesão no banco DDSM.

#### 4.2.4 Detecção de Regiões com Lesão: Teste 4

A Figura 32 exibe o resultado da classificação em lesão e não lesão para o caso A1627. Neste caso, a lesão não foi detectada (regiões destacadas em laranja). Na mesma figura, pode-se observar que a lesão esta localizada na fronteira entre regiões (mais especificamente, entre 4 regiões). Como estas regiões possuem, cada uma, pequena parte da lesão, ficaram pouco caracterizadas como área lesionada, provocando uma classificação errada por parte da MVS.

Figura 32 – Classificação final do caso A1627



(a)Mama esquerda. (b)Mama direita. Regiões classificadas como “lesão” em azul. Regiões com lesão no DDSM, porém não detectadas, em laranja.

### 4.3 Análises dos Resultados

Como a classificação de regiões possui duas etapas e estas possuem objetivos específicos, são feitas três análises com relação aos resultados: uma análise da primeira etapa, cujo objetivo é encontrar pares de regiões suspeitos baseados no nível de assimetria calculado a partir do CCC e da DE; uma análise da segunda etapa, que visa encontrar regiões com lesão a partir das regiões consideradas suspeitas anteriormente; e finalmente uma análise da metodologia de forma geral, que busca detectar regiões com lesão a partir de todas as regiões internas da mama submetidas desde o início até o fim do processo.

#### 4.3.1 Análise da Determinação de Regiões Suspeitas

Esta etapa trabalhou com pares de regiões, medindo seu nível de simetria, portanto a análise é feita da mesma forma. Foram testados, no total, 11865 pares de regiões, dentre os quais 11684 eram normais e 181 possuíam algum tipo de lesão em um dos lados (baseado no diagnóstico do DDSM). Foram considerados suspeitos 8040 pares de regiões. Destes, 177 correspondiam a lesões no DDSM e 7863 correspondiam a regiões normais. Foram considerados não suspeitos 3825 pares de regiões. Destes, 3821 correspondiam a regiões normais no DDSM e 4 correspondiam a lesões. Estes resultados geraram sensibilidade de 97,79%, especificidade de 32,70% e acurácia de 33,69%, como mostrado na Tabela 1. Os baixos valores de especificidade e acurácia devem-se ao alto número de falso-positivos.



Tabela 1 – Resultados da determinação de pares de regiões suspeitas

INTERNOS	NORMAIS	LESÃO	FP	FN	VP	VN	SE(%)	ES(%)	AC(%)
11865	11684	181	7863	4	177	3821	97,79	32,70	33,69

INTERNOS: pares de regiões internas às mamas.

NORMAIS: pares de regiões normais, segundo DDSM.

LESÃO: pares de regiões com lesão, segundo DDSM.

FP: falso-positivos (pares determinados como suspeitos, porém sem lesão no DDSM).

FN: falso-negativos (pares determinados como não suspeitos, porém com lesão no DDSM).

VP: verdadeiro-positivos (pares determinados como suspeitos, contendo lesão no DDSM).

VN: verdadeiro-negativos (pares determinados como não suspeitos, sem lesão no DDSM).

SE: sensibilidade (capacidade de identificar pares de regiões com lesão).

ES: especificidade (capacidade de identificar pares de regiões normais).

AC: acurácia (taxa de acerto).

Vale ainda destacar que 100% dos pares de mamografias que possuíam lesão (segundo o DDSM) tiveram a lesão (ou pelo menos parte dela) incluída nos pares de regiões consideradas suspeitas.

#### 4.3.2 Análise da Classificação em Lesão e Não Lesão

A segunda etapa da classificação trabalhou com regiões individuais, a partir das regiões consideradas suspeitas na etapa anterior, testando regiões esquerdas e direitas separadamente. Foram submetidas à MVS treinada 16080 regiões suspeitas (a partir de 8040 pares de regiões considerados suspeitos), dentre as quais 15903 eram normais e 177 possuíam lesão, de acordo com o DDSM. A MVS treinada classificou como lesão 3814 regiões. Destas, 145 correspondiam a algum tipo de lesão e 3669 correspondiam a áreas normais no DDSM. Foram classificadas como não lesão, pela MVS treinada, 12266 regiões. Destas, 12234 correspondiam a áreas normais e 32 correspondiam a algum tipo de lesão no DDSM. Os resultados geraram 81,92% de sensibilidade, 76,92% de especificidade e 76,98% de acurácia, conforme Tabela 2.

Embora a sensibilidade tenha sido menor em relação à detecção de regiões suspeitas, houve um aumento na especificidade e na acurácia, devido à redução de falso-positivos.

Tabela 2 – Resultados da classificação em lesão e não lesão

SUSPEITAS	NORMAIS	LESÃO	FP	FN	VP	VN	SE(%)	ES(%)	AC(%)
16080	15903	177	3669	32	145	12234	81,92	76,92	76,98

SUSPEITAS: regiões suspeitas submetidas.

NORMAIS: regiões normais, segundo DDSM.

LESÃO: regiões com lesão, segundo DDSM.

FP: falso-positivos (regiões classificadas erroneamente como lesão).

FN: falso-negativos (regiões classificadas erroneamente como não lesão).

VP: verdadeiro-positivos (regiões classificadas corretamente como lesão).

VN: verdadeiro-negativos (regiões classificadas corretamente como não lesão).

SE: sensibilidade (capacidade de identificar regiões com lesão, dentre as suspeitas).

ES: especificidade (capacidade de identificar regiões normais, dentre as suspeitas).

AC: acurácia (taxa de acerto).

Destaca-se que, das 30 mamografias com lesão (segundo o DDSM), 29 (ou seja, 96,66%) tiveram a lesão detectada total ou parcialmente. Em apenas 1 (uma) mamografia a lesão não foi, nem mesmo parcialmente, detectada.

#### 4.3.3 Análise Geral da Metodologia

Sabendo que o total de regiões de falso-positivos foi de 3669 e que foram testadas 120 mamografias (60 pares), pode-se considerar que a metodologia apresentou uma média de 30,575 regiões de falso-positivos por mamografia. Porém, como foi visto, após a classificação final com a MVS, 96,66% das mamografias com lesão tiveram suas lesões detectadas, mesmo que parcialmente em alguns casos.

Visto que as duas fases de classificação (em pares suspeitos e não suspeitos e em regiões de lesão e não lesão) foram analisadas separadamente, faz-se agora uma análise geral, considerando as regiões submetidas à metodologia do principio ao fim.

A metodologia visa classificar, de modo geral, regiões em lesão e não lesão a partir da busca prévia de assimetrias. Assim, pode-se considerar que foi submetido à metodologia um total de 23730 regiões internas de mamografias (total de regiões individuais internas às mamas dos 60 pares de mamografias), onde 23549 possuíam diagnóstico normal e 181 possuíam algum tipo de lesão (de acordo com o DDSM). A metodologia classificou como lesão 3814 regiões. Destas, 145 correspondiam a algum tipo de lesão e 3669 correspondiam a áreas normais no DDSM. Foram classificadas como não lesão 19916 regiões. Destas, 19880 regiões correspondam a áreas normais e 36 correspondiam a áreas lesionadas no DDSM. A

metodologia demonstrou desta forma 80,11% de sensibilidade, 84,41% de especificidade e 84,38% de acurácia. A Tabela 3 contém os resultados da metodologia como um todo.

Tabela 3 – Resultados gerais da metodologia

INTERNAS	NORMAIS	LESÃO	FP	FN	VP	VN	SE(%)	ES(%)	AC(%)
23730	23549	181	3669	36	145	19880	80,11	84,41	84,38

INTERNOS: regiões internas às mamas.

NORMAIS: regiões normais, segundo DDSM.

LESÃO: regiões com lesão, segundo DDSM.

FP: falso-positivos (regiões classificadas erroneamente como lesão).

FN: falso-negativos (regiões classificadas erroneamente como não lesão).

VP: verdadeiro-positivos (regiões classificadas corretamente como lesão).

VN: verdadeiro-negativos (regiões classificadas corretamente como não lesão).

SE: sensibilidade (capacidade de identificar regiões com lesão).

ES: especificidade (capacidade de identificar regiões normais).

AC: acurácia (taxa de acerto).

## 5. CONCLUSÃO

Este trabalho apresentou o desenvolvimento de uma metodologia de detecção de regiões lesionadas em mamografias, utilizando Índice de Getis-Ord na sua forma geral, partindo da determinação inicial de pares de regiões suspeitos através do uso conjunto das medidas CCC e DE entre as regiões de cada par. O CCC foi usado para indicar regiões assimétricas (consideradas suspeitas) determinando o nível de correlação entre região esquerda e região direita correspondente. Como o CCC gera valores muito parecidos para pares de regiões homogêneas, porém, com níveis de cinza bem diferentes entre si, foi usada a DE como contraprova de assimetria. As regiões consideradas suspeitas, foram submetidas individualmente (a direita separada da esquerda) a uma MVS treinada a partir de características extraídas com o Índice de Getis-Ord de outras regiões selecionadas para treino.

A análise dos resultados obtidos com a validação da metodologia, foi feita de três maneiras: a primeira examinando a efetividade da determinação de regiões suspeitas; a segunda examinando a classificação das regiões suspeitas em lesão e não lesão e a terceira, por sua vez, examinando a metodologia como um todo. A etapa de determinação de regiões suspeitas apresentou 97,79% de sensibilidade, 32,70% de especificidade e 33,69% de acurácia. Os baixos valores de especificidade e acurácia devem-se ao grande número de falso-positivos, levando-se em conta que muitos pares de regiões normais foram considerados suspeitos. Esta etapa teve o resultado considerado bom, por visar fornecer à etapa seguinte regiões que possuíssem maior nível de assimetria, onde provavelmente estariam inclusas as regiões com lesão. A etapa de classificação com MVS em lesão e não lesão apresentou 81,92% de sensibilidade, 76,92% de especificidade e 76,98% de acurácia. Embora a sensibilidade desta etapa não tenha sido tão alta quanto à da anterior, houve uma considerável redução de falso-positivos que proporcionou um aumento de especificidade e acurácia, contribuindo para o resultado final. Por fim, foi percebido que a metodologia apresentou, de modo geral, 80,11% de sensibilidade, 84,41% de especificidade e 84,38% de acurácia.

Com o intuito de melhorar os resultados, trabalhos futuros podem considerar o uso de outras medidas estatísticas em conjunto com as utilizadas ou separadamente, tanto na determinação de regiões suspeitas quanto na detecção de lesões com um classificador, assim como a distinção entre áreas de massa e não massa, ou quanto à determinação da natureza de lesões (benignas ou malignas). Também podem ser feitos testes com técnicas de segmentação

e registro que propiciem bons resultados com imagens do tipo MLO. Além disso, podem ser consideradas outras estratégias de divisão em regiões.

O trabalho apresentado mostrou a viabilidade da detecção de regiões lesionadas em mamografias a partir da busca de regiões suspeitas por nível de similaridade entre regiões correspondentes, podendo sua metodologia ser expandida ou adaptada, desde que com o uso de medidas e técnicas adequadas, a outras situações e outros tipos de imagem.

## REFERÊNCIAS

- ACS – American Cancer Society. **What are the key statistics about breast cancer?**, 2012a. Disponível em: <<http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-key-statistics>>. Último acesso em: 31 maio 2012.
- ACS – American Cancer Society. **What is breast cancer?**, 2012b. Disponível em: <<http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-what-is-breast-cancer/>>. Último acesso em: 31 maio 2012.
- ALBUQUERQUE, Marcio; ALBUQUERQUE, Marcelo. **Processamento de Imagens: Métodos e Análises**. Coordenação de Atividades Técnicas, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, 2000.p. 1.
- ALVARENGA, A. V.; INFANTOSI, A. F. C.; AZEVEDO, C. M.; PEREIRA, W. C. A. Aplicação de operadores morfológicos na segmentação e determinação do contorno de tumores de mama em imagens por ultra-som. **Revista Brasileira de Engenharia Biomédica**, v. 19, n. 2, p. 91-101, agosto 2003. ISSN 1517-3151.
- BRADSKI, G; KAEHLER, A. **Learning OpenCV: Computer Vision with the OpenCV Library**, O'Reilly Press, October, 2008.
- BRAZ JUNIOR, G. **Classificação de Regiões de Mamografias em Massa e Não Massa usando Estatística Espacial e Máquina de Vetores de Suporte**. Dissertação de Mestrado, Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2008.
- CHANG, C. –C; LIN, C. –J. LIBSVM : a library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- ERICEIRA, D. R. **Detecção de Regiões Suspeitas e Classificação de Massas em Mamografias Digitais Utilizando Descrição Espacial com Função Variograma**. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2011.
- ERICEIRA, D. R.; SILVA, A. C.; PAIVA, A. C. **Detecção de Regiões Suspeitas em Mamografias Digitais Utilizando Descrição Espacial com Variograma Cruzado**. In: XII BRAZILIAN CONGRESS OF HEALTH INFORMATICS, CBIS, outubro de 2010.
- GETIS, A.; ORD, J. K. The Analysis of Spatial Association by Use of Distance Statistics. **Geographical Analysis**, Vol. 24, No. 3, Pages 189–206, July 1992.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Prentice Hall. Ed. 2, 2002.
- GOSHTASBY, A. A. **2-D and 3-D image registration for medical, remote sensing, and industrial applications**. Wiley-Interscience publication, 2005.
- HEATH, M.; BOWYER, K.; KOPANS, D.; MOORE, R.; KEGELMEYER, W. P. The Digital Database for Screening Mammography. In: **Proceedings of the Fifth International**

**Workshop on Digital Mammography**, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, 2001. ISBN 1-930524-00-5.

INCA – Instituto Nacional Do Câncer. **Detecção precoce do câncer de mama**, 2012. Disponível em: <[http://www.inca.gov.br/conteudo\\_view.asp?id=1932](http://www.inca.gov.br/conteudo_view.asp?id=1932)>. Último acesso em: 31 maio 2012.

INCA – Instituto Nacional do Câncer. **Estimativa 2012: Incidência de Câncer no Brasil**, 2011 Disponível em: <<http://www1.inca.gov.br/estimativa/2012/>>. Último acesso em: 31 maio 2012.

ITK – **The Insight Segmentation and Registration Toolkit**. Disponível em: <[www.itk.org](http://www.itk.org)>. Último acesso em: 31 maio 2012.

KEERTHI, S. S.; LIN C. –J. Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. **Neural Computation**, Vol.15, No. 7, Pages 1667-1689, July 2003.

LEE, K. A. **A Mammographic Registration Method Based On Optical Flow and Multiresolution Computing**. Thesis, School of Engineering of the Air Force Institute of Technology, United States, 1997.

MITCHELL, H. B. **Image Fusion: Theories, Techniques and Applications**. Springer. Ed. 1, 2010.

NCI – National Cancer Institute. **Breast Cancer: Prevention, Genetics, Causes**, 2012. Disponível em: <<http://www.cancer.gov/cancertopics/prevention-genetics-causes/breast>>. Último acesso em: 31 maio 2012.

ROCHA, S. V.; BRAZ JÚNIOR, G.; SILVA, A. C.; PAIVA, A. C. Detecção e Diagnóstico de Massas em Mamografia: revisão bibliográfica. **Cadernos de Pesquisa**, v. 18, n. especial, dez 2011. ISSN 2178-2229. Disponível em : <<http://www.periodicoeletronicos.ufma.br/index.php/cadernosdepesquisa/article/view/735>>. Último acesso em: 31 maio 2012.

RODRIGUES, E. P. **Avaliação de métricas para o corregristo não rígido de imagens médicas**. Tese de Doutorado, Pós-Graduação em Física Aplicada à Medicina e Biologia, Universidade de São Paulo, Ribeirão Preto, 2010.

SCUTT, D.; LANCASTER, G. A.; MANNING, J. T. Breast asymmetry and predisposition to breast cancer. **Breast Cancer Research**, 8:R14, licensee BioMed Central Ltd., 2006.

THIRION, J.P. **Fast Non-Rigid Matching Of 3D Medical Image**. Technical report, Research Report RR-2547, Epidure Project, INRIA Sophia, May 1995.

WULDER, M. A. **Spatial Dependence Clusters in the Estimation of Forest Structural Parameters**. Thesis. University of Waterloo, Ontario, Canada, 1998.