

UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

**DANIEL RODRIGUES ERICEIRA**

**DETECÇÃO DE REGIÕES SUSPEITAS E CLASSIFICAÇÃO DE MASSAS  
EM MAMOGRAFIAS DIGITAIS UTILIZANDO DESCRIÇÃO ESPACIAL  
COM FUNÇÃO VARIOGRAMA**

São Luís  
2011

**DANIEL RODRIGUES ERICEIRA**

**DETECÇÃO DE REGIÕES SUSPEITAS E CLASSIFICAÇÃO DE MASSAS  
EM MAMOGRAFIAS DIGITAIS UTILIZANDO DESCRIÇÃO ESPACIAL  
COM FUNÇÃO VARIOGRAMA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Eletricidade na área de concentração Ciência da Computação.

Orientador: Prof. Dr. Aristófanês Corrêa Silva

Co-orientador: Prof. Dr. Anselmo Cardoso de Paiva

Ericeira, Daniel Rodrigues.

Detecção de regiões suspeitas e classificação de massas em mamografias digitais utilizando descrição espacial com função variograma / Daniel Rodrigues Ericeira. – São Luís, 2011.

80 f.

Impresso por computador (fotocópia).

Orientador: Aristófanês Corrêa Silva.

Co-orientador: Anselmo Cardoso de Paiva.

Dissertação (Mestrado) – Universidade Federal do Maranhão, Programa de Pós-Graduação em Engenharia de Eletricidade, 2011.

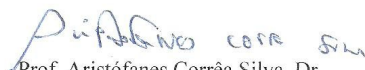
1. Rede neural - Mamografia. 2. Variograma. 3. Máquina de vetores de suporte. I. Título.

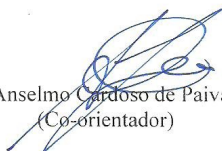
CDU 004.032.26:618.19-073

**DETECÇÃO DE REGIÕES SUSPEITAS E CLASSIFICAÇÃO  
DE MASSAS EM MAMOGRAFIAS DIGITAIS UTILIZANDO  
DESCRIÇÃO ESPACIAL COM FUNÇÃO VARIOGRAMA**


**Daniel Rodrigues Ericeira**

Dissertação aprovada em 15 de março de 2011.

  
Prof. Aristófares Corrêa Silva, Dr.  
(Orientador)

  
Prof. Anselmo Cardoso de Paiva, Dr.  
(Co-orientador)

  
Prof. Aura Conci, Dra.  
(Membro da Banca Examinadora)

  
Prof. Rita de Cássia Fernandes de Lima, Dra.  
(Membro da Banca Examinadora)

## AGRADECIMENTOS

Agradeço a meus familiares, especialmente à minha mãe. A eles dedico esta conquista. Agradeço aos professores do curso, especialmente aos meus orientadores Aristófanés e Anselmo pelos ensinamentos e pelo apoio.

## RESUMO

A mamografia é um exame de mama, utilizado de forma preventiva ao câncer de mama e também como método diagnóstico. Este exame, que consiste em uma radiografia das mamas, permite a detecção do câncer. O objetivo deste trabalho é utilizar técnicas de processamento de imagens e visão computacional para auxiliar especialistas na detecção de regiões suspeitas e detecção de massas mamárias em mamografias digitais. A primeira etapa da metodologia consiste em pré-processar as imagens de forma a torná-las mais apropriadas ao registro, através de redução de ruído, segmentação e re-dimensionamento. A etapa seguinte apresenta o registro bilateral de pares de mamas esquerda e direita. Para corrigir as diferenças de posicionamento e compressão ocorridas no momento do exame, o método de registro rígido foi aplicado (seguido do método de registro deformável com fluxo óptico) para cada par de imagens. Pares de regiões correspondentes foram relacionados e suas variações foram medidas através do descritor espacial variograma cruzado. Na etapa seguinte, foi criado um modelo para treinamento de uma Máquina de Vetores de Suporte (MVS) utilizando como características os valores de variograma cruzado de cada par de janelas de 180 casos. Esta MVS foi testada em 100 novos casos. Os pares que continham lesões foram classificados como “regiões suspeitas”; as demais, como “regiões não-suspeitas”. Destas regiões suspeitas, foram extraídas características de variograma como descritores de textura de tecido. As regiões que continham massas foram classificadas como “regiões de massa” e as demais como “regiões de não-massa”. Análise linear discriminante *stepwise* foi aplicada para selecionar as características mais significativas para treinamento de uma segunda MVS. Foram realizados testes com 30 novos casos para a classificação final pela MVS treinada em “massa” e “não-massa”. O melhor resultado apresentou na classificação final: 96% de acurácia, 100% de sensibilidade e 95,34% de especificidade. O pior caso apresentou: 70% de acurácia, 100% de sensibilidade e 67,56% de especificidade. Em média, os 30 casos apresentaram: 90% de acurácia, 100% de sensibilidade e 85% de especificidade.

Palavras-chave: Câncer de Mama. Mamografia. Variograma. Variograma Cruzado. Máquina de Vetores de Suporte.

## ABSTRACT

Mammography is the exam of the breast, used as breast cancer prevention and also as a diagnostic method. This exam, which consists in an X-Ray of the breast, allows cancer detection. The purpose of this work is to use image processing techniques and computer vision to help specialists in detecting suspect regions and masses in digital mammographies. The first stage of the methodology consists in pre-processing the images to make them more suitable to registration, through noise reduction, image segmentation and re-scale. The next stage presents bilateral left and right breast image pairs registration. In order to correct position and compression differences that occur during the exams, rigid registration (followed by optic flow deformable registration) was applied in each image pair. Corresponding pairs of regions were related and their mutual variations were measured through cross-variogram spatial description. On the next stage, a training model for a Support Vector Machine (SVM) was created using as characteristics the cross-variogram values of each pair of regions of 180 cases. This SVM was tested for 100 new cases. The region pairs that contained lesions were classified as “suspect regions”, and the other regions as “non-suspect regions”. From the suspect regions, variogram characteristics were extracted as tissue texture descriptors. The regions that contained masses were classified as “mass regions” and the other regions as “non-mass regions”. Stepwise linear discriminant analysis was applied to select the most significant characteristics to train the second SVM. Tests with 30 new cases were performed for the trained SVM final classification in “mass” or “non-mass”. The best case presented on the final classification: 96% accuracy, 100% sensitivity and 95,34% specificity. The worst case presented: 70% accuracy, 100% sensitivity and 67,56% specificity. On average, the 30 cases presented: 90% accuracy, 100% sensitivity and 85% specificity.

Keywords: Breast Cancer. Mammography. Variogram. Cross Variogram. Support Vector Machine.

## LISTA DE FIGURAS

Figura 1:	Visão lateral da anatomia da mama.	22
Figura 2:	Câncer confinado aos dutos ( <i>in-situ</i> ) e câncer que invade o tecido gorduroso (invasivo).	22
Figura 3:	Projeções de exames de mamografia.	23
Figura 4:	Calcificações encontradas em mamografias.	25
Figura 5:	Massas encontradas em mamografias.	25
Figura 6:	Distorção de arquitetura em mamografias.	26
Figura 7:	Resultado do processo de segmentação.	30
Figura 8:	h-scatterplots para 4 distâncias de separação na direção vertical entre pares de valores de uma variável V.	41
Figura 9:	h-scatterplots para 4 distâncias de separação na direção vertical entre pares de valores de duas variáveis: V e U.	43
Figura 10:	Distância entre as médias projetadas entre duas classes.	46
Figura 11:	Projeção das duas classes e de suas médias na análise linear discriminante.	47
Figura 12:	Metodologia desenvolvida pelo trabalho.	49
Figura 13:	Processo de redução de ruído aplicado a uma imagem.	52
Figura 14:	Processo de redução de ruído aplicado a uma imagem.	53
Figura 15:	Processo de segmentação aplicado a um par de imagens.	54
Figura 16:	Resultado do processo de registro rígido.	55
Figura 17:	Imagens de diferença entre as mamas resultantes do registro rígido.	55
Figura 18:	Resultado do processo de registro deformável.	56
Figura 19:	Imagens de diferença entre as mamas resultantes do registro deformável.	56
Figura 20:	Par de imagens com suas 256 regiões consideradas para cálculo de variograma cruzado	58
Figura 21:	Região 5x5 ilustrativa para cálculo de variograma na direção vertical.	58
Figura 22:	Região 5x5 ilustrativa para cálculo de variograma na direção horizontal.	59
Figura 23:	Região 5x5 ilustrativa para cálculo de variograma na direção diagonal 45°.	59
Figura 24:	Região 5x5 ilustrativa para cálculo de variograma na direção diagonal 135°.	60
Figura 25:	Mama esquerda e direita do par B3499 do DDSM, escolhidas para o Teste 1.	63
Figura 26:	Regiões destacadas no par escolhido para o Teste 1.	64
Figura 27:	Gráficos de variograma cruzado nas 8 direções.	64
Figura 28:	Regiões suspeitas resultantes do Teste 1.	65
Figura 29:	Mama esquerda e direita do par A1581 do DDSM, escolhido para o Teste 2.	66
Figura 30:	Processo de segmentação aplicado ao par de imagens do Teste 2.	66



Figura 31:	Processo de registro aplicado ao par de imagens do Teste 2.	67
Figura 32:	Regiões destacadas no par escolhido para o Teste 2.	67
Figura 33:	Gráficos de variograma cruzado nas 8 direções.	68
Figura 34:	Regiões suspeitas resultantes do Teste 2.	68
Figura 35:	Mama esquerda e direita do par A0323 do DDSM, escolhido para o Teste 3.	69
Figura 36:	Regiões suspeitas resultantes do Teste 3.	70
Figura 37:	Regiões suspeitas do Teste 1 escolhidas para exibição de função variograma.	72
Figura 38:	Gráfico de variograma nas 8 direções.	73
Figura 39:	Classificação final das regiões suspeitas para o Teste 1.	73
Figura 40:	Classificação final das regiões suspeitas para o Teste 2.	74
Figura 41:	Classificação final das regiões suspeitas para o Teste 3 (imagem da mama esquerda).	75
Figura 42:	Classificação final das regiões suspeitas para o Teste 3 (imagem da mama direita).	76

## SUMÁRIO

1.	<i>INTRODUÇÃO</i>	11
1.1 -	<b>Motivação e Objetivos</b>	11
1.2 -	<b>Trabalhos Relacionados</b>	14
1.3 -	<b>Organização do Trabalho</b>	20
2.	<i>FUNDAMENTAÇÃO TEÓRICA</i>	21
2.1 -	<b>Câncer de Mama</b>	21
2.2 -	<b>Mamografia</b>	23
2.3 -	<b>Tipos de Câncer de Mama: Massas e Calcificações</b>	24
2.4 -	<b>Detecção Auxiliada Por Computador</b>	26
2.5 -	<b>Processamento de Imagens</b>	27
2.5.1 -	Redução de Ruído	28
2.5.2 -	Segmentação	29
2.5.2.1 -	Processamento de Imagens Com Morfologia Matemática	30
2.5.2.2 -	Watershed	31
2.5.3 -	Registro de Imagens	33
2.5.3.1 -	Registro Bilateral	33
2.5.3.2 -	Registro Rígido	34
2.5.4 -	Fluxo Óptico	35
2.5.4.1 -	Algoritmo “ <i>Demons</i> ” de Thirion	37
2.5.5 -	Funções de Transformação Não-Lineares	38
2.5.5.1 -	Transformação Não-Linear Através de Fluxo Óptico	39
2.6 -	<b>Descrição Espacial Com Variograma e Variograma Cruzado</b>	39
2.6.1 -	Função Variograma	40
2.6.2 -	Função Variograma Cruzado	42
2.7 -	<b>Máquina de Vetores de Suporte (MVS)</b>	43
2.8 -	<b>Análise Linear Discriminante Stepwise (Stepwise LDA)</b>	44
2.9 -	<b>Medidas de Avaliação da Metodologia</b>	47
3.	<i>METODOLOGIA PROPOSTA</i>	49
3.1 -	<b>Seleção das Imagens</b>	51
3.2 -	<b>Pré-Processamento e Re-dimensionamento das Imagens</b>	51
3.2.1 -	Redução de Ruído	52
3.2.2 -	Segmentação das Imagens de Mamografia	53
3.2.3 -	Re-Dimensionamento das Imagens de Mamografia	54
3.3 -	<b>Registro das Imagens</b>	54
3.3.1 -	Registro Rígido	55
3.3.2 -	Registro Deformável	56
3.4 -	<b>Divisão das Imagens e Cálculo de Variograma Cruzado</b>	57
3.5 -	<b>Extração de Características Unilaterais: Variograma</b>	60
4.	<i>RESULTADOS</i>	62
4.1 -	<b>Regiões Suspeitas</b>	62
4.1.1 -	Regiões Suspeitas: Teste 1	63
4.1.2 -	Regiões Suspeitas: Teste 2	65
4.1.3 -	Regiões Suspeitas: Teste 3	69
4.2 -	<b>Classificação Entre Massa e Não-Massa</b>	70
4.2.1 -	Classificação Final: Teste 1	71
4.2.2 -	Classificação Final: Teste 2	74
4.2.3 -	Classificação Final: Teste 3	75

5.	<i>CONCLUSÃO</i>	.....	77
	REFERÊNCIAS BIBLIOGRÁFICAS	.....	79

## *1. INTRODUÇÃO*

Câncer de mama é o crescimento irregular e descontrolado de células que se originam no tecido mamário. Um grupo destas células pode formar uma massa de tecido extra (tumor). De acordo com a American Cancer Society (ACS), o câncer de mama é o tipo de câncer mais comum entre as mulheres, sendo no geral o segundo tipo de câncer a causar mais mortes (ficando atrás apenas do câncer de pulmão). Segundo a ACS, em países ocidentais, os casos de câncer de mama aumentaram em aproximadamente 30% nos últimos 25 anos. Este aumento pode ser explicado pelo melhoramento dos sistemas de detecção, que são capazes de detectar o câncer em estados iniciais. Ainda de acordo com a ACS, as taxas de morte por câncer de mama têm caído constantemente desde 1990. Este fato também é decorrente de melhores tratamentos e sistemas de detecção (ACS, 2010).

Segundo o Instituto Nacional do Câncer (INCA), no Brasil, o câncer de mama é o que mais causa mortes entre as mulheres. De acordo com a Estimativa de Incidência de Câncer no Brasil para 2010, realizada pelo INCA em 2009, esperavam-se 253.030 novos casos de câncer para o sexo feminino, sendo que o câncer de mama seria o segundo mais incidente, com 49.000 novos casos (INCA, 2010).

### **1.1 – Motivação e Objetivos**

A mamografia é um exame de mama, mais utilizado de forma preventiva para detectar câncer de mama e também como método diagnóstico, quando já há a suspeita da

existência de uma anomalia. Este exame, que consiste em uma radiografia das mamas, permite a detecção precoce do câncer, ao mostrar lesões em fase inicial (muito pequenas, da ordem de milímetros).

Estudos sobre a efetividade da mamografia sempre utilizam o exame clínico como exame adicional, o que torna difícil distinguir a sensibilidade do método como estratégia isolada de rastreamento. A sensibilidade varia de 46% a 88% e depende 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. A especificidade varia entre 82% e 99% e é igualmente dependente da qualidade do exame.

Resultados de ensaios clínicos aleatórios (por amostragem) que comparam a mortalidade em mulheres convidadas para rastreamento mamográfico com mulheres não submetidas a nenhuma intervenção são favoráveis ao uso da mamografia como método de detecção precoce capaz de reduzir a mortalidade por câncer de mama. As conclusões de estudos de meta-análise demonstram que o principal benefício do uso da mamografia é a diminuição de cerca de 30% da mortalidade em mulheres acima dos 50 anos, depois de sete a nove anos de implementação de ações organizadas de rastreamento (INCA, 2010).

Apesar do fato de um exame de mamografia poder detectar pequenas formações de câncer até dois anos antes de serem palpáveis em exames físicos, estima-se que a maior parte das lesões passa sem ser detectada na avaliação feita pelos especialistas. No Brasil, as taxas de mortalidade por câncer de mama continuam elevadas, muito provavelmente porque a doença ainda é diagnosticada em estágios avançados.

A evolução lenta e sutil do câncer pode ser mais fácil e previamente identificada com ajuda de técnicas de visão computacional associadas a processamento de imagens, fato que iria melhorar a porcentagem de eficiência dos exames preventivos.

Mamografias da mama esquerda e direita da mesma paciente tendem a apresentar alto grau de simetria (SCUTT *et al.*, 2006). Fazendo-se o alinhamento espacial (registro) e a comparação entre estas duas imagens de mamografia, seria possível avaliar o grau de simetria e determiná-lo como normal ou não-normal. Segundo o estudo realizado em (SCUTT *et al.*, 2006), o grupo considerado “normal” (ou seja, que não chegou a desenvolver câncer) apresentava assimetria de volume com média de 52,99 ml, enquanto o grupo que desenvolveu câncer apresentava média de 63,17 ml. Na análise de assimetria flutuante (AF), que denota pequenos desvios da simetria perfeita de qualquer caráter de organismo com simetria bilateral (PARSONS, 1990), o grupo “normal” apresentou AF de mama média de 2,5%, enquanto o grupo do câncer apresentou média de 2,7%. Portanto, valores que desviem do grau “normal” de simetria podem ser indicadores de possíveis anomalias. Nas regiões nas quais o alinhamento apresentar as maiores disparidades, pode-se considerar que a presença de um tumor é mais provável.

Este trabalho apresenta uma metodologia para detecção de massas através da determinação de regiões suspeitas entre regiões de pares de mamografias esquerda e direita, e posterior classificação das mesmas. As regiões suspeitas são detectadas por variações estruturais entre regiões correspondentes, definidas por um descritor espacial de conjunto de dados conhecido como variograma cruzado. Após a determinação de regiões suspeitas em um par de imagens, o variograma é utilizado em cada região suspeita isoladamente, para classificação em massa e não-massa.

## 1.2 – Trabalhos Relacionados

Nesta seção estão relacionados alguns trabalhos dedicados ao registro rígido, não-rígido e bilateral de imagens de mamografia, e detecção e classificação de anormalidades baseando-se em diferenças estruturais existentes em regiões correspondentes do par de imagens. Estes trabalhos serviram como referência e apoio para desenvolvimento desta dissertação de mestrado.

Vários estudos foram desenvolvidos com objetivo de buscar a detecção de tumores através do registro bilateral de mamas e análise de assimetria. Em (LAU & BISCHOF, 1990), foi apresentado um método para detecção de tumores através da busca de assimetrias estruturais intensas entre pares de mamografias esquerda e direita. Primeiramente as imagens eram alinhadas através de registro rígido, seguido de ajuste por pontos de controle. Em seguida, a assimetria era avaliada através de medidas de brilho, direcionalidade e aspereza. O método foi testado em 10 pares de mamografias que continham 13 áreas suspeitas, avaliadas por especialistas. Destas, 12 foram detectadas pelo método.

Em (YIN *et al.*, 1993) foi feito o registro bilateral através de pontos de controle (mamilos e borda). Um método de subtração não-linear foi aplicado ao par, buscando a identificação inicial de possíveis massas em locais de maior assimetria. 154 pares de mamografias foram testados e a técnica de análise bilateral obteve melhores resultados de detecção em relação a um esquema de detecção que trabalhava em uma só imagem.

Em (STAMATAKIS *et al.*, 1996), dois métodos de comparação entre mama esquerda e direita foram propostos. O primeiro método definia um limiar de intensidade de diferenciação que determinava as áreas correspondentes do par. Os resultados de testes com 50 pares de imagens apresentaram precisão de 86,8% e 4,9 falsos positivos por imagem. O

segundo método envolvia a geração de 8 pares de imagens, a partir de cada par, a serem comparados bilateralmente. Este processo de comparação determinava áreas suspeitas no par original. Um conjunto de 10 características era calculado para cada área e análise discriminante foi utilizada para determinar 5 características finais do classificador. Este método, testado com os mesmos 50 pares do primeiro método, apresentou melhor desempenho, com precisão de 89,2% e 4,3 falsos positivos por imagem.

Em (LEE, 1997), o autor propôs um método para registro e avaliação de similaridade em imagens de mamografias de uma mesma paciente ao longo do tempo. Este método utilizava fluxo óptico, tanto para realizar o registro deformável quanto para avaliar as diferenças entre as imagens registradas. O método também apresentava uma abordagem voltada para multi-resolução, na qual o campo de fluxo óptico era calculado para a imagem em várias escalas. Os campos de fluxo óptico produzidos nos testes corresponderam ao esperado, e a técnica de multi-resolução se mostrou superior e com execução mais rápida em relação ao método comum.

A relação entre assimetria das mamas e a ocorrência de câncer tem sido alvo de análise em vários estudos. Em um primeiro estudo (SCUTT *et al.*, 1997) foi realizada uma observação inicial sobre a ligação entre assimetria e o risco de câncer. Após a comparação entre 250 casos de pacientes com câncer e 250 casos de pacientes saudáveis, ocupando a mesma faixa etária, chegaram à conclusão de que o grupo com câncer apresentava maior assimetria (média de 87,39 ml) em relação ao grupo saudável (média de 59,27 ml). O objetivo do estudo era obter primeiras evidências que relacionassem a assimetria entre as mamas como fator indicativo de risco de câncer. O estudo subsequente (SCUTT *et al.*, 2006) confirmou esta relação, através da avaliação de 252 casos de mulheres normais que não desenvolveram câncer e 252 casos de mulheres normais que depois desenvolveram câncer. Foi observado o grupo considerado “normal” (ou seja, que não chegou a desenvolver câncer) apresentava



assimetria de volume com média de 52,99 ml, enquanto o grupo que desenvolveu câncer apresentava média de 63,17 ml. O estudo também chegou à conclusão de que, para cada 100 ml de assimetria de volume, a chance de desenvolver câncer aumenta em 50%.

Também buscando massas malignas baseando-se em subtração bilateral para identificar assimetrias, em (MENDEZ *et al.*, 1998) foi desenvolvido um método para alinhar mamogramas direitos e esquerdos e depois subtraí-los, delimitando as regiões assimétricas com um algoritmo de crescimento de região. 70 pares de mamografias foram testados, com taxa de 71% de verdadeiros positivos e 0,67 falsos negativos por imagem.

Em (SALLAM & BOWYER, 1999), foi proposto um método para analisar diferenças entre pares correspondentes de mamogramas. A imagem de diferença gerada pelo par registrado utilizava diferenças de intensidades multiplicadas por pesos, para destaque das regiões com diferenças mais intensas (regiões suspeitas). Para avaliar a imagem de diferença, foram utilizados 124 pares bilaterais contendo 77 anormalidades de diferentes tipos previamente diagnosticadas: se ao menos 40% da região da imagem de diferença estivesse dentro do círculo determinado como anormalidade pelo radiologista, isto era considerado um verdadeiro positivo; caso contrário, era considerado falso positivo. Resultados mostraram que entre 16 e 20% das anormalidades não eram detectadas, dependendo do tipo de imagem.

Em (GEORGSSON, 2003), foram propostos dois métodos para analisar diferenças bilaterais entre mamogramas. Um método era baseado na diferença absoluta de intensidades, e o outro era baseado em diferenças estatísticas entre propriedades de vizinhos correspondentes. Em teste realizado em uma base de dados de 100 pares de imagens com 59 anormalidades de 4 tipos diferentes, o método estatístico apresentou taxa de falsos positivos menor.

A utilização de análise de funções geoestatísticas para sugerir sobre malignidade ou benignidade de nódulos pulmonares em imagens de tomografia computadorizada foi

realizada em (CORRÊA, 2004). Na tese, o autor propôs a utilização de funções geoestatísticas clássicas (semivariograma, semimadograma, covariograma e correlograma) para análise de textura com objetivo de diagnosticar o nódulo. Em comparação a outros métodos de análise dos nódulos, as funções geoestatísticas obtiveram ótimos resultados: a classificação com Análise Linear Discriminante de Fisher (ALDF) mostrou porcentagem de acerto de 93,1% para nódulos benignos e 100% para malignos; a precisão foi de 94,4%. A classificação com Rede Neural Perceptron de Múltiplas Camadas (MLP) mostrou porcentagem de acerto de 96,5% para nódulos benignos e 100% para malignos; a precisão foi de 97,2%.

Uma aplicação de função semivariograma para classificar tecidos de mamografia foi desenvolvida em (CORRÊA *et al.*, 2006). Regiões de interesse quadradas que envolviam lesões foram extraídas de um determinado grupo de imagens e delas foram extraídas 80 medidas de variograma. Um método passo-a-passo foi usado para determinar quais medidas eram mais significativas na determinação de malignidade ou benignidade. Um conjunto com as três medidas escolhidas (direção 0° com distância 19 e 20 e direção 90° com distância 10) foi dado como entrada para uma rede neural perceptron multi-camadas que iria fazer a classificação. Testes foram realizados com 117 amostras de tecidos (67 benignas e 50 malignas). Os resultados obtidos mostraram que o semivariograma é um bom caracterizador de texturas, com melhor caso com sensibilidade de 92,8%, especificidade de 83,3% e precisão acima de 88%.

Outro trabalho que combinava registro bilateral e detecção de anormalidades foi apresentado em (HACHAMA *et al.*, 2006). A metodologia se baseava em definir os pixels das imagens em duas classes: tecido normal e anormalidades. As constantes utilizadas para registrar as imagens eram baseadas na descrição estatística dos níveis de cinza de ambas as classes, com abordagem Bayesiana. Um mapa de probabilidade de anormalidade era definido para cada pixel das imagens, baseando-se também na distribuição estatística. A integração

estatística entre processo de registro de detecção de anormalidades apresentou desempenho superior a outros métodos, com taxa de detecção de 0,947 para 10000 pixels falsos positivos.

Em (VAN ENGELAND, 2006), o autor apresentou um método para determinar regiões de interesse suspeitas correspondentes entre diferentes visões de exames de mamografia. As regiões suspeitas em uma das imagens eram determinadas pela extração de duas medidas morfológicas que eram dadas como entrada a uma rede neural *backpropagation* de 3 camadas. A rede classificava os pixels e determinava uma imagem com medidas de probabilidade de massa. Uma segunda rede neural *backpropagation* recebia como entrada para treinamento medidas de contraste das regiões suspeitas. Esta rede treinada classificava as regiões suspeitas de um par de mamogramas em falso positivo e verdadeiro positivo, determinando suas probabilidades de malignidade. O classificador “*two-view*”, que iria utilizar a informação das duas visões diferentes, utilizava análise linear discriminante *stepwise* para definir quais as melhores características a serem utilizadas. Estas eram utilizadas para treinar uma terceira rede neural, que dava como saída a determinação de probabilidades de malignidade. O classificador “*two-view*” apresentou desempenho melhor do que o classificador que utilizava características apenas de uma das imagens (testado em uma base de quase 1000 casos, a sensibilidade aumentava de 52% a 61% para taxas de 0,1 falsos positivos por imagem).

Em (YI-TA WU *et al.*, 2007) foi desenvolvido um método para redução de falsos positivos em sistemas de detecção de massas baseados em registro bilateral de mamas. Para cada massa em potencial detectada eram definidas regiões de interesse em ambas as mamas das quais eram extraídas características de textura e morfologia. A região que continha a massa em si gerava características unilaterais, e a região que continha a massa combinada com sua região correspondente no outro par de mamogramas definia as características bilaterais (características de assimetria). Análise linear discriminante *stepwise* definia as

melhores características unilaterais, e outra análise linear discriminante *stepwise* definia as melhores características bilaterais. Uma terceira análise linear discriminante *stepwise* combinava a saída dos dois primeiros e foi utilizado para classificar massas verdadeiras e falsos positivos. Aplicado em 341 casos, o sistema apresentou redução de falsos positivos em aproximadamente 40% em relação a um sistema que utilizava apenas o classificador unilateral.

Em (ERICEIRA *et al.*, 2010) foi desenvolvido um primeiro estudo sobre aplicação de variograma cruzado na detecção de regiões suspeitas em pares registrados de mamografias. A classificação realizada em 10 casos por uma MVS treinada com valores de variograma cruzado não apresentou casos de falsos negativos. Um número relativamente elevado de falsos positivos, porém, foi apresentado no trabalho devido a alguns aspectos melhorados em estudos futuros, como aumento da amostra de treino e acréscimo de mais uma direção de variograma.

O estudo dos trabalhos anteriores mostrou que o desenvolvimento de técnicas de registro bilateral e comparação estrutural entre as mamas é motivado por bons resultados no auxílio a diagnósticos. A contribuição deste trabalho está na utilização medidas de variância, que se mostraram bons descritores de textura, relacionadas ao par de mamas. Para isso, será utilizada uma descrição espacial bivariada (variograma cruzado) que determinará a continuidade espacial de uma região em uma das mamas em relação à sua região correspondente no outro par. Este procedimento será utilizado para a determinação das regiões suspeitas. Para a classificação final das mesmas, a descrição espacial com variograma será aplicada a cada imagem do par isoladamente, com intenção de detectar massas. O objetivo, portanto, é testar a aplicabilidade e a qualidade do variograma como descritor de padrões de tecidos mamários e como classificador de lesões.

### 1.3 – Organização do Trabalho

O trabalho descrito está organizado em mais quatro capítulos. O Capítulo 2 aborda a fundamentação teórica necessária ao desenvolvimento da metodologia proposta. Apresenta conceitos sobre câncer de mama, as técnicas de processamento de imagens digitais utilizadas, e teoria sobre a função variograma, máquinas de vetores de suporte e análise linear discriminante *stepwise* (Stepwise LDA).

O Capítulo 3 descreve a metodologia proposta, apresentada em oito etapas: o pré-processamento das imagens, registro, divisão e cálculo de variograma cruzado, montagem de vetores e treino da primeira Máquina de Vetores de Suporte (MVS), extração de características de variograma das regiões suspeitas, análise linear discriminante *stepwise*, montagem de vetores e treino da segunda MVS e a classificação final das regiões suspeitas em massas ou não-massas.

O Capítulo 4 apresenta e discute os resultados obtidos com a metodologia proposta.

O Capítulo 5 apresenta a conclusão sobre o trabalho, mostrando a eficiência dos métodos utilizados, a relevância dos resultados obtidos e perspectivas de evolução para trabalhos futuros.

## 2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção é apresentada a fundamentação teórica necessária à compreensão da metodologia do trabalho. São apresentados os seguintes temas: câncer de mama e suas características, exames de mamografia, sistemas de detecção auxiliada por computação, técnicas de processamento de imagens utilizadas, descrição espacial através da função geoestatística de variograma e classificação realizada por Máquinas de Vetores de Suporte (MVS) e por Análise Linear Discriminante *Stepwise* (*Stepwise LDA – Linear Discriminant Analysis*).

### 2.1 – Câncer de Mama

O câncer de mama é uma doença caracterizada pela disfunção ocorrida nas células que compõem o tecido mamário, que crescem ou se multiplicam descontroladamente causando o aparecimento de estruturas (massas) de caráter maligno ou benigno. Em determinadas ocasiões e por razões ainda desconhecidas, estas células que se reproduzem com uma velocidade maior causam o aparecimento de massas celulares denominadas neoplasias; tecido de formação nova, mas de origem patológica. Estas massas são designadas tumores.

A anatomia da mama é composta principalmente por: lóbulos (glândulas produtoras de leite), dutos de leite (vasos que conectam os lóbulos ao mamilo), tecido gorduroso (que envolve lóbulos e canais), vasos sanguíneos e linfáticos (Figura 1). A maior

parte dos casos de câncer tem sua origem nas células dos dutos, enquanto alguns têm sua origem nos lóbulos (Cancer Help, UK – 2010).

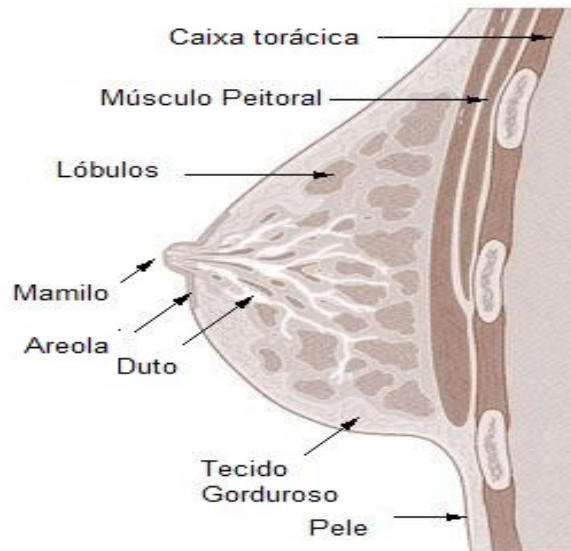


Figura 1: Visão lateral da anatomia da mama. Fonte: (CANCERHELP UK , 2010).

Quando o câncer permanece confinado aos dutos, refere-se a ele como câncer “*in-situ*”. Quando o câncer invadiu o tecido gorduroso que envolve os dutos e, possivelmente, espalhou-se para demais órgãos, refere-se a ele como “invasivo” (Figura 2).

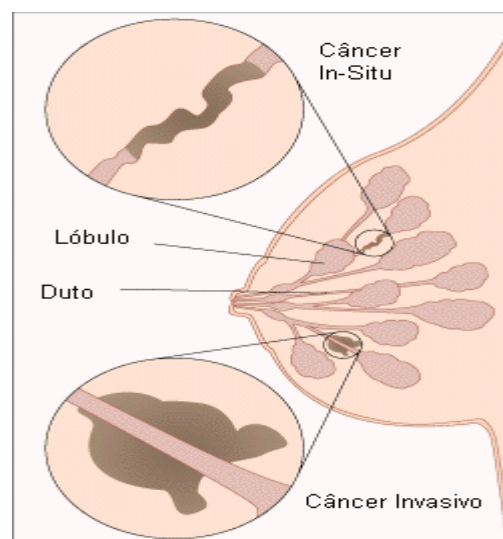


Figura 2: Câncer confinado aos dutos (*in-situ*) e câncer que invade o tecido gorduroso (invasivo). Fonte: (CANCERHELP UK , 2010).

## 2.2 - Mamografia

Para detectar os tumores que causam câncer de mama é necessário realizar uma radiografia das mamas, conhecida como mamografia. A mamografia é atualmente a melhor técnica de detecção prévia de lesões não palpáveis na mama. Para que o tratamento seja eficiente, o câncer deve ser detectado ainda em fase inicial. Mulheres acima dos quarenta anos apresentam maior probabilidade de serem afetadas pela doença, e, portanto, devem ter maiores preocupações com os devidos exames preventivos (ACS, 2010).

É recomendável que mulheres façam exame de mamografia de ambas as mamas regularmente. As imagens de raio-X das mamas podem ser obtidas de diferentes ângulos. As projeções mais comuns são: crânio-caudal (CC) e médio-lateral oblíqua (MLO) (Figura 3).

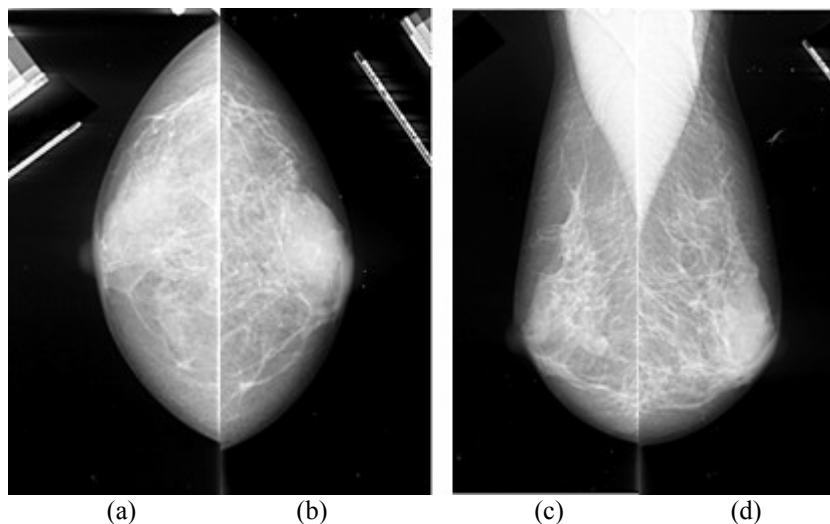


Figura 3: Projeções de exames de mamografia. (a) Crânio-caudal esquerda. (b) Crânio-caudal direita. (c) Médio-lateral oblíqua esquerda. (d) Médio-lateral oblíqua direita.

Estudos sobre a efetividade da mamografia sempre utilizam o exame clínico como exame adicional, o que torna difícil distinguir a sensibilidade do método como estratégia



isolada de rastreamento. A sensibilidade varia de 46% a 88% e depende de fatores tais como: tamanho e localização da lesão, densidade do tecido mamário, qualidade dos recursos técnicos e habilidade de interpretação do radiologista. A especificidade varia entre 82% e 99% e é igualmente dependente da qualidade do exame (INCA, 2010). Resultados de ensaios clínicos aleatórios (por amostragem) que comparam a mortalidade em mulheres convidadas para rastreamento mamográfico com não submetidas a nenhuma intervenção são favoráveis ao uso da mamografia como método de detecção precoce capaz de reduzir a mortalidade por câncer de mama. As conclusões de estudos de meta-análise demonstram que o principal benefício do uso da mamografia é a diminuição de cerca de 30% da mortalidade em mulheres acima dos 50 anos, depois de sete a nove anos de implementação de ações organizadas de rastreamento (INCA, 2010).

### **2.3 – Tipos de Câncer de Mama: Massas e Calcificações**

A manifestação do câncer de mama nas imagens de mamografias se dá nas formas de massas e calcificações.

Uma massa é um aglomerado de células que se unem de forma mais densa em relação ao tecido que a envolve. Este aglomerado pode ser causado por câncer de mama, assim como também por condições benignas. Algumas características das massas são determinantes para determinar suas probabilidades de malignidade: tamanho, forma e disposição de suas margens.

Calcificações são depósitos de cálcio que aparecem como pontos brancos no mamograma. São de dois tipos: microcalcificações e macrocalcificações. As

microcalcificações são depósitos pequenos e indicam, dependendo de sua forma, uma possível presença cancerígena. As macrocalcificações são grandes depósitos de cálcio e normalmente estão associadas com condições benignas, causadas, por exemplo, por inflamações ou envelhecimento das artérias.

Os tipos mais comuns de anormalidades presentes em imagens de mamografia são: calcificações (benignas e malignas), massas circulares e bem-definidas, massas espiculadas, massas mal-definidas e distorção de arquitetura (HEATH, 2001). A Figura 4 exibe anormalidades do tipo “calcificação”. A Figura 5 exibe anormalidades do tipo “massa”. Um exemplo de distorção de arquitetura é exibido na Figura 6. As imagens são provenientes do banco DDSM (*Digital Database For Screening Mammography*) (HEATH, 2001).

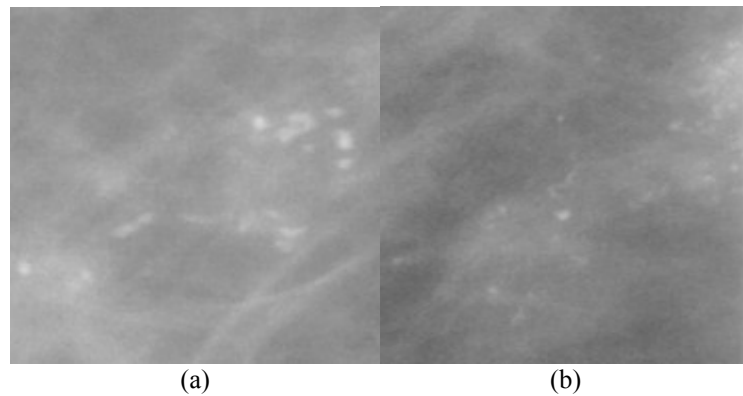


Figura 4: Calcificações encontradas em mamografias. (a) Benignas. (b) Malignas.

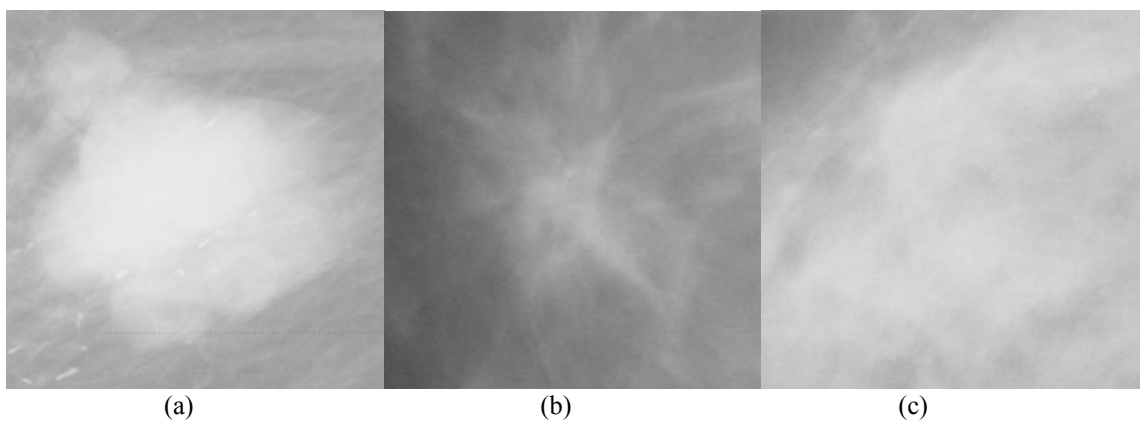


Figura 5: Massas encontradas em mamografias. (a) Bem-definida. (b) Espiculada. (c) Mal-definida.

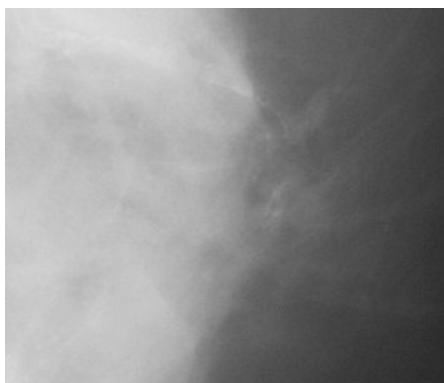


Figura 6: Distorção de arquitetura em mamografias.

## 2.4 – Detecção Auxiliada por Computador

Casos de câncer podem passar despercebidos em diagnósticos devido a erros de interpretação por parte de um especialista. Em muitos casos, lesões só são detectadas em estágios evoluídos. No Brasil, as taxas de mortalidade por câncer de mama continuam elevadas, muito provavelmente porque a doença ainda é diagnosticada em estádios avançados (INCA, 2010).

A evolução lenta e sutil do câncer pode ser mais fácil e previamente identificada com ajuda de técnicas computacionais de visão associadas a processamento de imagens, fato que iria melhorar a porcentagem de eficiência dos exames preventivos. Para melhorar a detecção, sistemas de detecção auxiliado por computação (sistemas CAD) são constantemente desenvolvidos e aprimorados. Em mamografias, sistemas CAD utilizam técnicas de reconhecimento de padrões para identificar características de lesões mamárias. A intenção dos sistemas CAD não é a de substituir a avaliação do especialista, mas sim de auxiliá-la na determinação de áreas suspeitas e na interpretação de lesões detectadas. O principal desafio no desenvolvimento de um sistema deste tipo é diminuir o número de erros de detecção.

Vários estudos dedicaram-se à observação dos benefícios da utilização de sistemas CAD. Em (FREER & ULISSEY, 2001), por exemplo, foi feita uma comparação entre diagnósticos feitos em 12.860 mamogramas com e sem utilização de sistemas CAD. Os números indicaram um aumento de 19,5% do número de casos de câncer detectados em fases iniciais quando foram utilizados os sistemas CAD. Em (BURHENNE *et al.*, 2000), na análise feita por especialistas em 427 casos de mamografias que sabidamente continham lesões, 27% foram consideradas suspeitas. Destes casos suspeitos, o sistema CAD aplicado diagnosticou corretamente 77%. Os resultados mostram que a utilização de sistemas CAD poderia ter reduzido a taxa de falsos negativos em 77%.

## **2.5 – Processamento de Imagens**

Processamento de imagens é a tarefa de processar sinais na qual o sinal de entrada é uma imagem, que pode ser considerada um sinal bidimensional (uma função  $f(x,y)$ , onde  $x$  e  $y$  são as coordenadas no plano e a amplitude de  $f$  em um ponto  $(x,y)$  é chamada de intensidade ou nível de cinza). A saída deste processamento pode ser também uma imagem ou um conjunto de características relacionadas a ela. Um paradigma para classificar os tipos de processamento pode ser considerar os processamentos como sendo de baixo, médio e alto nível (GONZALEZ, 2002).

Em processamentos de baixo nível, tanto a entrada quanto a saída são imagens. Algumas dessas tarefas incluem: redução de ruído, melhoramento de contraste e aguçamento. Em processamentos de nível médio, a entrada é uma imagem, mas a saída normalmente é dada por características extraídas desta imagem: bordas, contornos, objetos reconhecidos.

Algumas dessas tarefas incluem: segmentação, descrição e classificação (reconhecimento). Processamentos de alto nível envolvem tarefas de visão computacional e interpretação dos dados extraídos (GONZALEZ, 2002).

### 2.5.1 – Redução de Ruído

Imagens de mamografias normalmente apresentam sinais indesejados, que podem interferir na informação desejada. Estes sinais constituem o ruído, inerente às imagens de mamografia devido ao próprio processo de aquisição e digitalização dos filmes de radiografia. Na etapa de pré-processamento deste trabalho, a redução de ruído foi aplicada às imagens, objetivando a remoção destas informações indesejadas.

O processo de suavização de imagens através de filtros tem por objetivo a redução de ruído. Como o ruído contribui para frequências espaciais mais altas na imagem, um filtro de suavização deve ser capaz de reduzir as magnitudes destas frequências (GOSHTASBY, 2005). Filtro da mediana é muito utilizado e se mostrou eficiente na remoção de ruídos de imagens de mamografia (LEE, 1997).

Considerando a imagem antes da aplicação do filtro como a função  $f(x,y)$  e a imagem após a aplicação do filtro como sendo a função  $\bar{f}(x,y)$ :

$$\bar{f}(x,y) = \text{MEDIAN}(f, x, y, r) \quad (1)$$

onde  $r$  é o raio do filtro em *pixels* e MEDIAN é uma função que retorna a intensidade mediana em uma janela circular de raio  $r$  centrada em  $(x,y)$ .

### 2.5.2 – Segmentação

O processo de segmentação subdivide uma imagem em seus objetos e regiões constituintes. O nível a que esta subdivisão é levada depende do problema a ser resolvido, isto é, a segmentação deve parar quando os objetos de interesse tiverem sido isolados (GONZALEZ, 2002).

A precisão da tarefa de segmentação pode determinar o eventual sucesso ou falha de um sistema de processamento e análise de imagens. Imagens de exames de mamografia costumam apresentar uma série de elementos (chamados de “artefatos”), que ficam na região de fundo e contêm informações sobre o exame (rótulos, dados sobre filme utilizado, etc.). Estes artefatos podem interferir nos algoritmos de registro de imagens, e, portanto, devem ser preferencialmente removidos. Esta remoção foi feita na fase de pré-processamento deste trabalho.

A remoção de artefatos e a segmentação da região da mama podem ser realizadas em processamento de mamografias com a utilização da técnica de *Watershed* e com a aplicação de morfologia matemática para estimativa da borda. A segmentação da região da mama tem por objetivo diminuir a região percorrida pelos algoritmos aplicados, tornando as operações a serem realizadas nas imagens das mamografias mais rápidas e precisas. A Figura 7 mostra o resultado de um processo de segmentação. A etapa de segmentação faz parte do pré-processamento deste trabalho, anterior à etapa de registro, realizada com aplicação de dilatação e *Watershed*.

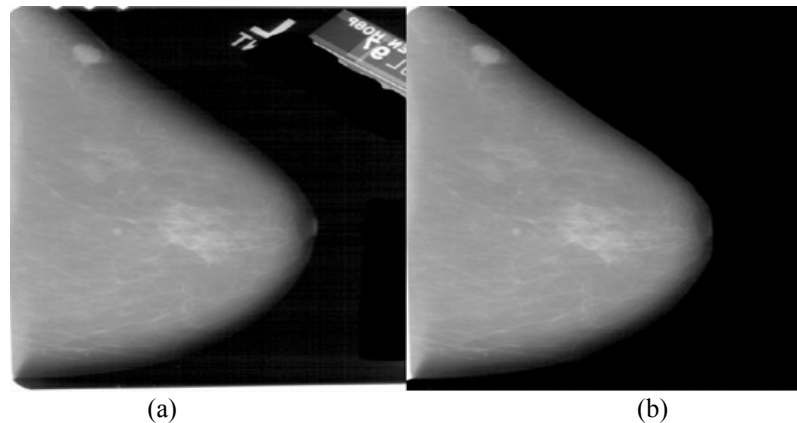


Figura 7: Resultado do processo de segmentação. (a) Imagem original, com artefatos. (b) Imagem da mama segmentada. Os artefatos são removidos, ficando apenas a informação de interesse referente à mama.

#### 2.5.2.1 – Processamento de Imagens com Morfologia Matemática

Morfologia matemática é uma ferramenta utilizada para extrair componentes de imagens que são úteis na representação e descrição de formas de regiões, como fronteiras, esqueletos e fecho convexo. Técnicas morfológicas de pré e pós-processamento envolvem: filtragem, dilatação, erosão, *thinning*, *pruning*, etc. (GONZALEZ, 2002).

A noção de uma estrutura geométrica não é puramente objetiva. A morfologia matemática quantifica esta “intuição” através do conceito de elementos estruturantes. Estes elementos interagem com o objeto em estudo, modificando sua forma e reduzindo-o a uma espécie de caricatura, mais expressiva do fenômeno inicial (SERRA, 1982). Ou seja, o elemento estruturante é um conjunto bem definido que é utilizado para transformar um conjunto desconhecido (imagem) e assim dele extrair informações de geometria e topologia. A linguagem da morfologia matemática é a teoria de conjuntos (GONZALEZ, 2002).

A dilatação utiliza um elemento estruturante para expandir as formas contidas em uma imagem de entrada. Sendo  $E$  um espaço euclidiano,  $A$  uma imagem binária em  $E$  e  $B$  o

elemento estruturante, a dilatação de  $A$  por  $B$  é dada pela Equação 2.  $A$  e  $B$  são conjuntos em  $Z^2$  e  $B^s$  é o simétrico de  $B$ .

$$A \oplus B = \{z \in E \mid (B^s)_z \cap A \neq \emptyset\} \quad (2)$$

#### 2.5.2.2 – Watershed

A ideia de segmentação *Watershed* baseia-se em uma interpretação topográfica de uma imagem. A altura das “montanhas” é proporcional ao valor de nível de cinza de um ponto da imagem. Assim, uma imagem seria formada por três tipos de pontos: pontos pertencentes a um mínimo local, pontos nos quais uma gota d’água, caso neles fossem colocada, cairia em um único ponto mínimo e pontos nos quais a gota d’água poderia cair em mais de um mínimo local. Para um dado mínimo local, o conjunto de pontos nos quais as gotas d’água certamente cairiam nesse mínimo é chamado *watershed* deste mínimo. Supondo que em cada ponto de mínimo local seja feito um buraco e que toda a topografia seja inundada com água vindo de baixo e subindo pelos buracos em padrão uniforme, quando as águas provenientes de lugares pontos estiverem prestes a se encontrar, uma barragem é construída para evitar que isto aconteça. Estas barragens correspondem às linhas divisórias e formam as fronteiras extraídas pelo algoritmo (GONZALEZ, 2002).

O funcionamento do algoritmo se dá da seguinte forma: a topografia (imagem) é inundada em incrementos inteiros que vão de  $n = \min + 1$  até  $n = \max + 1$  ( $n$  é a iteração atual do processo de inundação,  $\min$  é o valor mínimo de intensidade de nível de cinza da imagem e  $\max$  é o valor máximo de intensidade da imagem). A cada iteração  $n$ , o algoritmo precisa saber o número de pontos que estão “inundados”. Seja  $C_n(M_i)$  o conjunto de



coordenadas de pontos associados com a *watershed* do conjunto de mínimos  $M_i$  de coordenadas que estão inundadas na iteração  $n$ ,  $C_n(M_i)$  pode ser visto como uma imagem binária dada por:  $C_n(M_i) = 1$  em  $(x,y)$  se  $(x,y) \in C(M_i)$  E  $(x,y) \in T[n]$ ;  $C_n(M_i) = 0$ , nos demais casos.  $T[n]$  representa o conjunto de coordenadas de pontos na imagem que estão abaixo do plano  $f(x,y) = n$ . Sendo  $C[n]$  a união de porções de *watershed* inundadas na iteração  $n$ :

$$C[n] = \bigcup_{i=1}^R C_n(M_i) \quad (3)$$

Então  $C[\text{Max}+1]$  é a união de todas as *watersheds*:

$$C[\text{max}+1] = \bigcup_{i=1}^R C(M_i) \quad (4)$$

O algoritmo de busca de linhas de *watersheds* é inicializado com  $C[\text{min}+1] = T[\text{min}+1]$ . O procedimento para obtenção de  $C[n]$  recursivamente a partir de  $C[n-1]$  se dá da seguinte forma. Sendo  $Q$  o conjunto de componentes conectados em  $T[n]$ , para cada componente conectado  $q$  pertencente a  $Q[n]$  existem três possibilidades: (a)  $q \cap C[n-1]$  está vazio; (b)  $q \cap C[n-1]$  contém um componente conectado de  $C[n-1]$  e (c)  $q \cap C[n-1]$  contém mais de um componente conectado de  $C[n-1]$ .

A possibilidade (a) ocorre quando um novo mínimo é encontrado. Neste caso, o componente  $q$  é incorporado a  $C[n-1]$  para formar  $C[n]$ . A possibilidade (b) ocorre quando  $q$  está dentro da *watershed* de algum mínimo local. Neste caso, o componente  $q$  também é incorporado a  $C[n-1]$  para formar  $C[n]$ . Na possibilidade (c), uma próxima iteração do algoritmo faria o nível de água de diferentes *watersheds* se unirem. Neste caso, a barreira deve ser construída dentro de  $q$ . Uma barreira de 1 pixel de espessura pode ser construída através da dilatação de  $q \cap C[n-1]$  com elemento estruturante 3x3 de 1's, e restringindo a

dilatação a  $q$ . Os valores de  $n$ ,  $min$  e  $max$  utilizados pelo algoritmo podem ser obtidos pelo histograma da imagem.

### 2.5.3 – Registro de Imagens

Registro de imagens é o processo de alinhar espacialmente duas ou mais imagens, através da determinação de uma correspondência ponto-a-ponto entre elas. O objetivo deste processo é encontrar uma função de transformação ótima que irá alinhar uma imagem a outra. Registro de imagens é um componente fundamental de sistemas de análise de imagens médicas e industriais. Este processo permite: fusão de informação, determinação de mudanças de uma seqüência de imagens, identificação de objetos e suas localizações em outra imagem, etc. (GOSHTASBY, 2005). É preciso ter uma metodologia para otimizar a medida de similaridade entre as imagens. A similaridade é determinada com a utilização de certas características das imagens.

#### 2.5.3.1 – Registro Bilateral

O processo de registro busca minimizar as diferenças espaciais entre duas imagens. Entre os fatores que contribuem para diferenças espaciais entre mamografias, tem-se aqueles relativos ao processo de aquisição, tem-se o posicionamento da mama, a compressão

aplicada, etc. Em relação ao comportamento da mama, tem-se a estrutura de tecidos moles deformável, móvel, não-homogênea e anisotrópica.

O registro bilateral envolve a aplicação das técnicas de registro para análise comparativa bilateral de mamas. Esta análise consiste na comparação entre imagens de mamografia da mama esquerda e direita da mesma paciente, representando a mesma visão e obtidas durante a mesma sessão de exame.

Mamografias da mama esquerda e direita tendem a apresentar um alto grau de simetria. Estudos mostram que valores que desviem deste grau “normal” de simetria podem indicar presença de anormalidades. Um estudo realizado por Scutt et al. (2006), por exemplo, analisou 252 pacientes com exames normais e que mais tarde iriam desenvolver câncer, e 252 pacientes com exames normais que não chegaram a desenvolver câncer durante a conclusão do estudo. O grupo que desenvolveu câncer apresentava assimetria maior do que o grupo que não desenvolveu câncer. Assimetria entre mama esquerda e direita foi considerada então um dos fatores principais na previsão do câncer.

Neste trabalho, para realizar o registro bilateral de pares de mamografia, foi utilizada a seguinte metodologia: primeiramente é feito o registro rígido, e, em seguida, o registro não-rígido através de fluxo óptico.

#### 2.5.3.2 – Registro Rígido

O registro rígido é composto por operações de translação, rotação e escala. Este tipo de registro é realizado utilizando métodos baseados na relação linear que existe entre as intensidades dos pixels de imagem para outra. O objetivo do registro rígido é diminuir as

diferenças globais que existem entre as duas imagens, através de uma transformação linear aplicada em toda a imagem.

Considerando a imagem fonte como sendo a função  $f(x,y)$  e a imagem alvo como sendo a função  $\bar{f} = (x',y')$ , a relação linear entre as intensidades de uma imagem para outra pode ser dada por:

$$x' = S(x \cos \theta + y \operatorname{sen} \theta) + h \quad (5)$$

$$y' = S(-x \operatorname{sen} \theta + y \cos \theta) + k \quad (6)$$

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

O comportamento deformável e a estrutura não-homogênea da mama tornam a aplicação unicamente de registro rígido insuficiente e com resultados insatisfatórios.

#### 2.5.4 – Fluxo Óptico

O campo de velocidade que descreve o movimento de padrões de intensidade em uma seqüência de imagens é dado pela análise do fluxo óptico. Existem vários métodos para a estimativa de fluxo óptico. Os métodos diferenciais baseiam-se em aproximações locais do sinal da imagem em Série de Taylor, utilizando as derivadas parciais calculadas deste sinal nas coordenadas espaciais e temporais.

Dadas duas imagens em tempos  $t$  e  $t + \delta t$ , os métodos de fluxo óptico tentam calcular as mudanças de intensidade ocorridas para cada posição  $x,y$  da imagem. Considerando uma intensidade  $I(x,y,t)$  de um local de coordenadas  $x,y$  no tempo  $t$ , ela sofrerá

uma mudança de  $\delta x$ ,  $\delta y$  e  $\delta t$  de um quadro para outro em uma comparação entre duas imagens. Os métodos de estimativa de fluxo óptico fazem a restrição de que as mudanças de intensidade entre os quadros são constantes, o que nos traz a seguinte equação de restrição:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (7)$$

Considerando o movimento da intensidade como sendo pequeno, a restrição da imagem em  $I(x, y, t)$  pode ser desenvolvida em série de Taylor para chegar a:

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t + h.o.t. \quad (8)$$

Sendo “h.o.t.” termos de alta ordem, pequenos o suficiente para serem ignorados, observando-se as duas equações anteriores podemos chegar facilmente ao próximo conjunto de igualdades:

$$\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t = 0 \quad (9)$$

$$\frac{\partial I}{\partial x} \frac{\delta x}{\delta t} + \frac{\partial I}{\partial y} \frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} = 0 \quad (10)$$

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} = 0 \quad (11)$$

A idéia básica é estimar o vetor de movimento de um pixel utilizando as derivadas de intensidade no tempo e no espaço. Na Equação 11, os valores  $V_x$  e  $V_y$  são os componentes do vetor de velocidade de  $I(x, y, t)$ . Pode-se escrever então a equação final da forma:

$$\nabla I \times \vec{v} + I_t = 0 \quad (12)$$

onde  $\nabla I = \left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) = (I_x, I_y)$  e  $\vec{v} = \left( \frac{\delta x}{\delta t}, \frac{\delta y}{\delta t} \right) = (V_x, V_y)$  e  $I_t = \frac{\partial I}{\partial t}$ .

A equação final é insuficiente para determinar as duas componentes do vetor de velocidade. Este problema é conhecido como “*aperture problem*” dos algoritmos de fluxo

óptico. Para que se possa encontrar as duas componentes, deve-se desenvolver mais algum conjunto de equações, fazendo-se necessário, para tal, aplicar novas restrições. As técnicas de estimativa de fluxo óptico introduzem novas condições para aplicar estas restrições.

#### 2.5.4.1 – Algoritmo “*Demons*” de Thirion

Para realizar o registro de deformação entre duas imagens (imagem fixa e imagem alvo), um algoritmo muito implementado é o “*demons*” de J.P. Thirion. Esta “deformação” indica mudanças que ocorreram entre as imagens, que são utilizadas para desenhar o campo de fluxo óptico (THIRION, 1995). A idéia 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:

$$D(X) \cdot \nabla f(X) = -(m(X) - f(X)) \quad (13)$$

onde  $f(X)$  é a imagem fixa,  $m(X)$  é a imagem alvo a ser registrada, e  $D(X)$  é o campo de fluxo óptico (ou campo de deformação) entre ambas.

Devido ao “*aperture problem*”, a Equação 13 é insuficiente para especificar  $D(X)$  e é normalmente determinada utilizando algum tipo de regularização. Para registro, são utilizadas projeções dos vetores do campo na direção do gradiente de intensidade:

$$D(X) = -\frac{(m(X) - f(X))\nabla f(X)}{\|\nabla f\|^2} \quad (14)$$

A Equação 14 se torna instável para pequenos valores do gradiente da imagem, resultando em grandes valores de deslocamento. Para resolver este problema, o algoritmo re-normaliza a equação:

$$D(X) = -\frac{(m(X) - f(X))\nabla f(X)}{\|\nabla f\|^2 + (m(X) - f(X))^2 / K} \quad (15)$$

onde  $K$  é o fator de normalização, dado pela média dos quadrados dos espaçamentos entre os pixels. A inclusão de  $K$  torna a computação invariável à escala de pixels das imagens.

Iniciando com um campo de deformação  $D^0(x)$ , o algoritmo atualiza iterativamente o campo de fluxo de forma que o campo na  $N$ -ésima iteração é dado por:

$$D^N(X) = D^{N-1}(X) - \frac{(m(X + D^{N-1}(X)) - f(X))\nabla f(X)}{\|\nabla f\|^2 + (m(X + D^{N-1}(X)) - f(X))^2} \quad (16)$$

Existem muitas soluções para o problema de reconstrução do campo de deformação. Este problema envolve a combinação das imagens fixa e alvo. Por exemplo, é possível que todos os pixels de um valor particular na imagem fixa sejam mapeados para um único pixel de mesmo valor na imagem alvo. Ou seja, estes problemas podem causar a reconstrução de campos enganosos quando aplicados a problemas reais. Uma solução aplicada pelo algoritmo é forçar um comportamento elástico, suavizando o campo de deformação com um filtro Gaussiano entre as iterações.

### 2.5.5 – Funções de Transformação Não-Lineares

Funções de transformação utilizam coordenadas de pontos correspondentes para fazer a relação entre eles. Dadas as coordenadas de  $N$  pontos de uma imagem fonte  $(x,y)$  e uma imagem alvo  $(X,Y)$ :

$$\{(x_i, y_i), (X_i, Y_i) : i = 1, \dots, N\} \quad (17)$$

Deve-se achar uma função de transformação  $f(x,y)$  com componentes  $f_x$  e  $f_y$  que satisfaçam:

$$\begin{aligned} X_i &\approx f_x(x_i, y_i) \\ Y_i &\approx f_y(x_i, y_i) \end{aligned} \quad (18)$$

O tipo de função selecionada depende de: tipo de diferença geométrica entre as imagens, precisão da correspondência entre os pontos e densidade e organização dos pontos. As imagens a serem registradas muitas vezes apresentam diferenças geométricas não-lineares que requerem uma composição de transformações locais não-rígidas (GOSHTASBY, 2005).

#### 2.5.5.1 – Transformação Não-Linear Através de Fluxo Óptico

A estimativa do campo de fluxo óptico pode ser utilizada para transformar a imagem alvo em relação à imagem fonte, caracterizando a seguinte transformação não-linear (LEE, 1997):

$$f'(x, y) = f(i - \Delta u, j - \Delta v) \quad (19)$$

onde:  $f'$  é a imagem alvo,  $f$  é a imagem fonte e  $\Delta u$  e  $\Delta v$  são os componentes do campo vetorial de fluxo óptico estimado.

## 2.6 – Descrição Espacial com Variograma e Variograma Cruzado



Neste trabalho, a descrição espacial com variograma é utilizada para determinar a variação de intensidade de *pixels* em uma determinada região de uma imagem de mamografia. Um determinado padrão de variação em uma região pode ser empregado como descritor de textura de tecido. Já a descrição espacial com variograma cruzado é utilizada para relacionar a continuidade espacial de uma região de uma das mamas de um par com a região correspondente no outro par. As variáveis consideradas são os valores de intensidade de *pixel* das imagens. Uma região que varia muito de intensidade com relação à sua região correspondente gera valores de variograma cruzado com maiores magnitudes em comparação com regiões menos descontínuas.

### 2.6.1 – Função Variograma

A continuidade espacial de uma região pode ser descrita por *h-scatterplots*, que são diagramas bivariados que mostram todos os possíveis valores de dados cujas localizações são separadas por uma distância em uma determinada direção (**h**). Pontos muito concentrados indicam uma região contínua.

A forma da nuvem de pontos de um *h-scatterplot* determinam quão contínuos são os valores de dados em uma certa distância e uma direção particular. Se os valores de dados separados por **h** forem similares, eles ficam próximos da linha de 45° que cruza a origem. Quanto menos similares forem os valores, eles irão se afastar desta linha. A Figura 8 exibe os *h-scatterplots* para 4 distâncias de separação entre os valores de uma variável *V*. A localização no espaço (x,y) da variável é dada por um vetor **t**, e a distância entre as variáveis é dada por um vetor **h**. Assim, por exemplo,  $h=(0,1)$  significa que a distância é na direção

vertical, com 1 dado de separação. No *h-scatterplot* da Figura 8, cada valor da variável  $V$  é relacionado com o valor da mesma variável, separado por  $h$ . Percebe-se que, com o aumento das distâncias de separação, a similaridade entre pares de valores diminui e os pontos tornam-se mais espalhados em relação ao eixo diagonal.

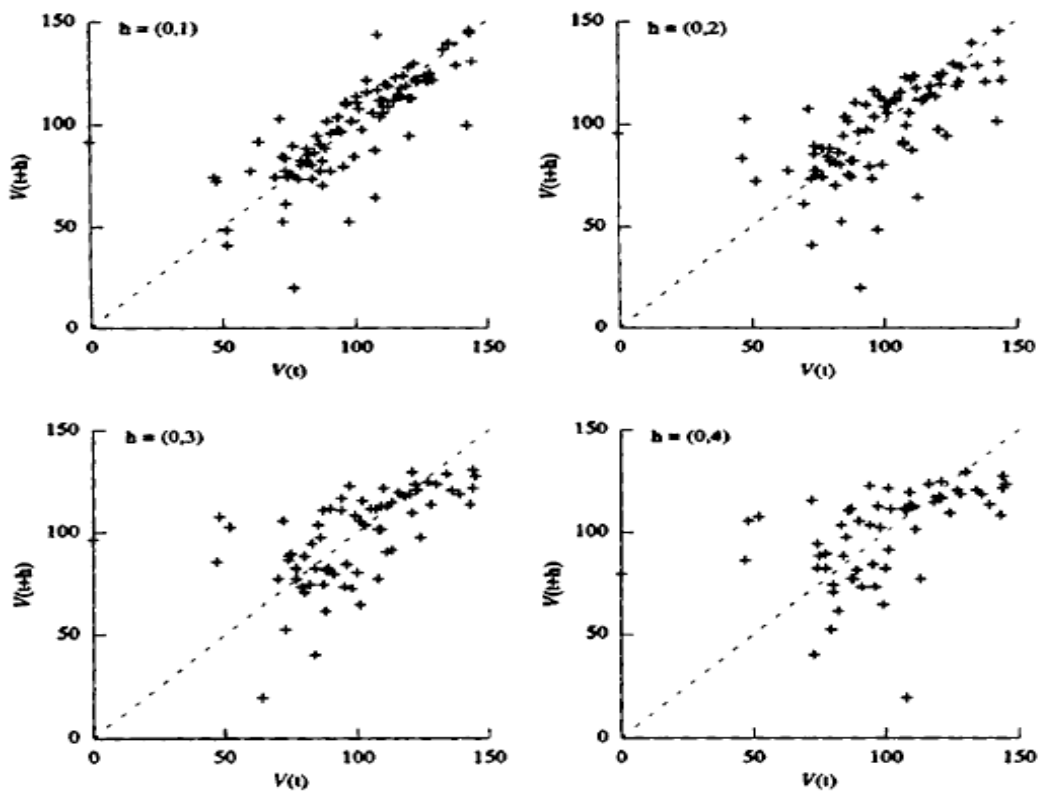


Figura 8: *h-scatterplots* para 4 distâncias de separação na direção vertical entre pares de valores de uma variável  $V$ . O aumento das separações reflete em valores mais dispersos em relação ao eixo diagonal, devido à diminuição de similaridades entre os pares de valores. Fonte: (ISAAKS & SRIVASTAVA, 1989).

Uma forma de medir esta concentração de pontos é o momento de inércia, que relaciona diferenças entre pares de dados. A função variograma descreve essas diferenças entre pares de dados da forma dada pela Equação 20.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)_{h_{ij}=h}} (v_i - v_j)^2 \quad (20)$$

onde  $h$  são os vetores que indicam distância e direção que separam os dados,  $N(h)$  é o número de pares de dados cujas localizações são separadas por  $h$ , e  $v$  são os valores da variável considerada (ISAAKS & SRIVASTAVA, 1989).

### 2.6.2 – Função Variograma Cruzado

Para se comparar a continuidade espacial relativa a duas regiões, utiliza-se o *h-scatterplot* cruzado, diagrama no qual duas variáveis diferentes (no caso, as intensidades da mama esquerda e direita) são relacionadas. Pontos muito concentrados no *h-scatterplot* cruzado indicam maior similaridade entre o comportamento das duas variáveis e, portanto, indicam menor variação entre as duas regiões correspondentes.

Considerando-se duas variáveis (U e V), a Figura 9 exibe os *h-scatterplots* cruzados entre estas variáveis considerando 4 valores de separação na direção vertical. Assim como na Figura 8, com a distância entre as variáveis, a relação de similaridade entre seus comportamentos diminui, e a nuvem de pontos torna-se mais dispersa em relação ao eixo diagonal.

Uma forma de medir a concentração dos pontos é o momento de inércia, que relaciona a diferença entre as duas variáveis de acordo com a equação seguinte:

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{(i,j)_{h_{ij}=h}} (u_i - u_j)(v_i - v_j) \quad (21)$$

onde  $h$  são os vetores que indicam distância e direção que separam os dados,  $N(h)$  é o número de pares de dados cujas localizações são separadas por  $h$ , e  $u$  e  $v$  são os valores das duas variáveis (ISAAKS & SRIVASTAVA, 1989).

Quanto menos concentrados estiverem os pontos do *h-scatterplot* cruzado (regiões menos variantes), maiores os valores dos momentos de inércia entre as duas variáveis. Um variograma cruzado é um gráfico que mostra os valores calculados pela Equação 21 em função da distância e direção que separa as variáveis.

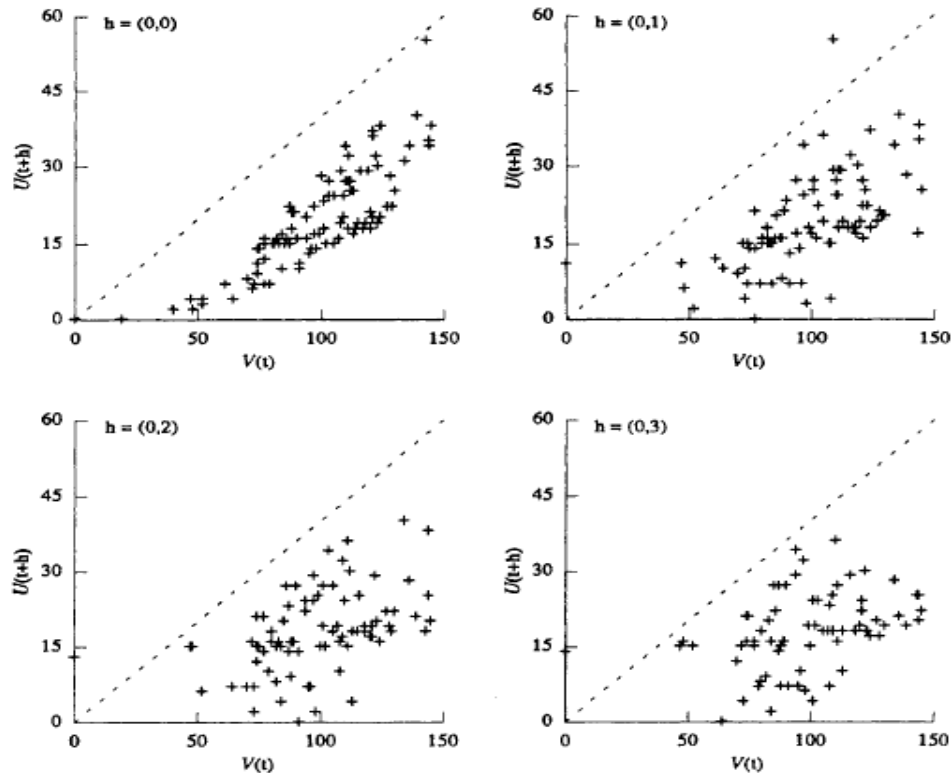


Figura 9: *h-scatterplots* para 4 distâncias de separação na direção vertical entre pares de valores de duas variáveis: V e U. O aumento das separações reflete em valores mais dispersos em relação ao eixo diagonal, devido à diminuição de similaridades entre os pares de valores. Fonte: (ISAAKS & SRIVASTAVA, 1989).

## 2.7 – Máquinas de Vetores de Suporte (MVS)

Máquina de vetores de suporte (MVS) é uma técnica de classificação que produz um modelo (baseado em um conjunto de dados de treinamento) cujo objetivo é prever a classificação de um conjunto de testes baseado em seus atributos. Uma tarefa de classificação

normalmente envolve a separação de dados em: conjunto de treino e conjunto de testes. Cada elemento do conjunto de treino contém uma classificação e vários atributos (características observadas). O treinamento da MVS produz um modelo capaz de prever a qual classe pertencem os elementos que formam o conjunto de testes.

Dado um conjunto de rótulos de classe  $(x_i, y_i)$ ,  $i = 1, \dots, l$ , onde  $x_i \in R^n$  e  $y \in \{1, -1\}^l$ , a MVS exige solução para o seguinte problema de otimização:

$$\min_{w, \xi} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right\} \quad (22)$$

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

Os vetores de treinamento  $x_i$  são mapeados em um espaço dimensional mais elevado pela função  $\phi$ . A MVS encontra um hiperplano linear de separação com fronteira localizada neste espaço.  $C > 0$  é o parâmetro de penalidade do termo de erro (CORTES & VAPNIK, 1995).

Os hiperplanos no espaço mais elevado são definidos como um conjunto de pontos cujos produtos cruzados com um vetor neste espaço são constantes. Os produtos cruzados neste espaço são definidos em termos de uma função de kernel  $K(x,y)$  que pode ser selecionada para um determinado problema. A função de kernel é definida como:

$$K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j) \quad (23)$$

O kernel utilizado pelo método deste trabalho foi o RBF (Radial Basis Function), que é definido pela Equação 24 (CHIH-CHUNG C e CHIH-JEN L, 2001):

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (24)$$

## 2.8 – Análise Linear Discriminante Stepwise (Stepwise LDA)

Um classificador linear é aquele que trabalha com uma combinação linear dos vetores de características de um objeto. Análise Linear Discriminante Stepwise (Stepwise LDA) é um método utilizado como classificador linear (ou como redutor de dimensionalidade) para realizar tarefas de reconhecimentos de padrões, através do encontro de uma combinação linear de características que determinam classes de objetos.

O objetivo da Stepwise LDA neste trabalho é realizar a redução de dimensionalidade das variáveis preservando o máximo possível de informação discriminatória da classe.

Considerando-se um conjunto de amostras de dimensão  $D$  e duas classes apenas ( $N_1$  pertencendo à classe  $\omega_1$  e  $N_2$  pertencendo à classe  $\omega_2$ ). Deseja-se obter um escalar  $y$  através da projeção das amostras  $x$  em uma linha:

$$y = w^T x \quad (25)$$

De todas as linhas possíveis, a mais conveniente seria aquela que maximizasse a separabilidade dos escalares. Para que se possa encontrar um bom vetor de projeção, é necessário definir uma medida de separação entre as projeções. Os valores médios de  $x$  e  $y$  são:

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x \quad (26)$$

$$\bar{\mu}_i = \frac{1}{N_i} \sum_{y \in \omega_i} y = \frac{1}{N_i} \sum_{x \in \omega_i} w^T x = w^T \mu_i \quad (27)$$

A distância entre as médias projetadas não seria uma boa medida para separar as classes por não levar em consideração os respectivos desvios padrões. Na Figura 10, observa-

se que o eixo vertical rende melhor separabilidade entre as classes 1 e 2, enquanto o eixo horizontal apresenta maior distância entre as médias.

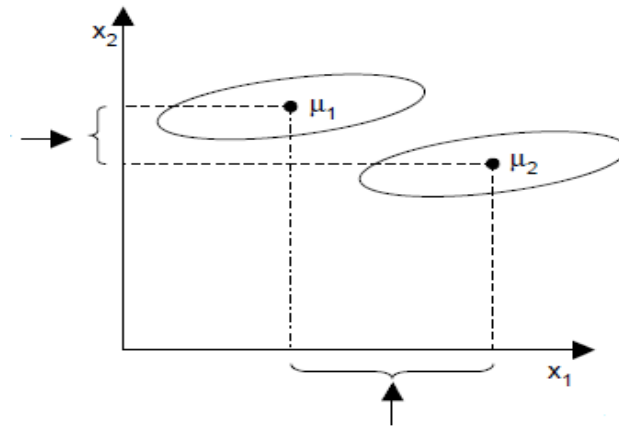


Figura 10: Distância entre as médias projetadas entre duas classes.

A solução proposta pela análise de Fisher é maximizar a função que representa a diferença entre as médias, normalizada por uma medida de dispersão interna das classes. A dispersão para cada classe é definida pela Equação 28:

$$s_i^2 = \sum_{y \in \omega_i} (y - \bar{\mu}_i)^2 \quad (28)$$

O discriminante linear de Fisher é definido como a função linear  $w^T x$  que maximiza a função:

$$J(w) = \frac{|\bar{\mu}_1 - \bar{\mu}_2|^2}{s_1^2 + s_2^2} \quad (29)$$

Com isto, estará sendo procurada uma projeção onde exemplos da mesma classe são projetados muito próximos uns dos outros e, ao mesmo tempo, as médias projetadas são as mais afastadas possíveis (Figura 11).

A seleção das variáveis em si é feita através do método *stepwise*. O objetivo é a construção de um modelo passo a passo, que inclui ou remove, em cada passo, uma variável do modelo. O modelo inicial é definido pelas variáveis iniciais. A cada passo, novos modelos são gerados com a inclusão de cada variável que não está no modelo, e com a exclusão de cada variável que está no modelo. A performance resultante destes modelos é estimada (por validação cruzada) e, se o valor máximo do critério escolhido for melhor que a menor performance estimada, a variável correspondente é excluída ou incluída. O processo para se o novo melhor valor não for bom o bastante, ou se um número máximo especificado de variáveis é alcançado.

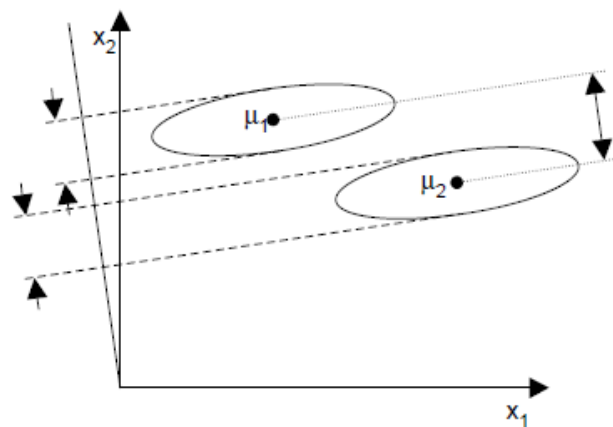


Figura 11: Projeção das duas classes e de suas médias na análise linear discriminante.

## 2.9 – Medidas de Avaliação da Metodologia

Em problemas de reconhecimento de padrões em processamento de imagens costuma-se medir o desempenho da metodologia calculando-se algumas medidas de desempenho sobre os resultados dos testes: acurácia, sensibilidade e especificidade.



Dada uma amostra com casos positivos e negativos para uma determinada detecção, os resultados dos testes de classificação dos casos analisados podem ser divididos em quatro grupos: VP (Verdadeiros Positivos): número de casos corretamente classificados como positivos; FP (Falsos Positivos): número de casos erroneamente classificados como positivos; VN (Verdadeiros Negativos): número de casos corretamente classificados como negativos; e FN (Falsos Negativos): número de casos erroneamente classificados como negativos. Estes valores são utilizados para gerar as medidas adotadas por este trabalho.

A acurácia (A) mede a percentagem total de casos corretamente classificados e é dada pela Equação 30:

$$A = \frac{VP + VN}{VP + VN + FP + FN} \quad (30)$$

A sensibilidade (S) mede o desempenho da classificação em relação aos casos positivos e é dada pela Equação 31:

$$S = \frac{VP}{VP + FN} \quad (31)$$

A especificidade (E) mede o desempenho da classificação em relação aos casos negativos e é dada pela Equação 32:

$$E = \frac{VN}{VN + FP} \quad (32)$$

### 3. METODOLOGIA PROPOSTA

Nesta seção é descrita a metodologia desenvolvida pelo trabalho. A Figura 12 apresenta um esquema ilustrativo dos procedimentos (que foram descritos na Seção 2): aquisição das imagens, pré-processamento e re-dimensionamento, registro rígido e deformável, divisão das imagens em janelas quadradas, cálculo de função variograma cruzado, montagem dos vetores de características de variograma cruzado e treinamento de MVS, obtenção de regiões suspeitas, extração de características de textura através de cálculo de função variograma das regiões suspeitas, análise linear discriminante *stepwise* para seleção de características e montagem de vetores para treino de MVS.

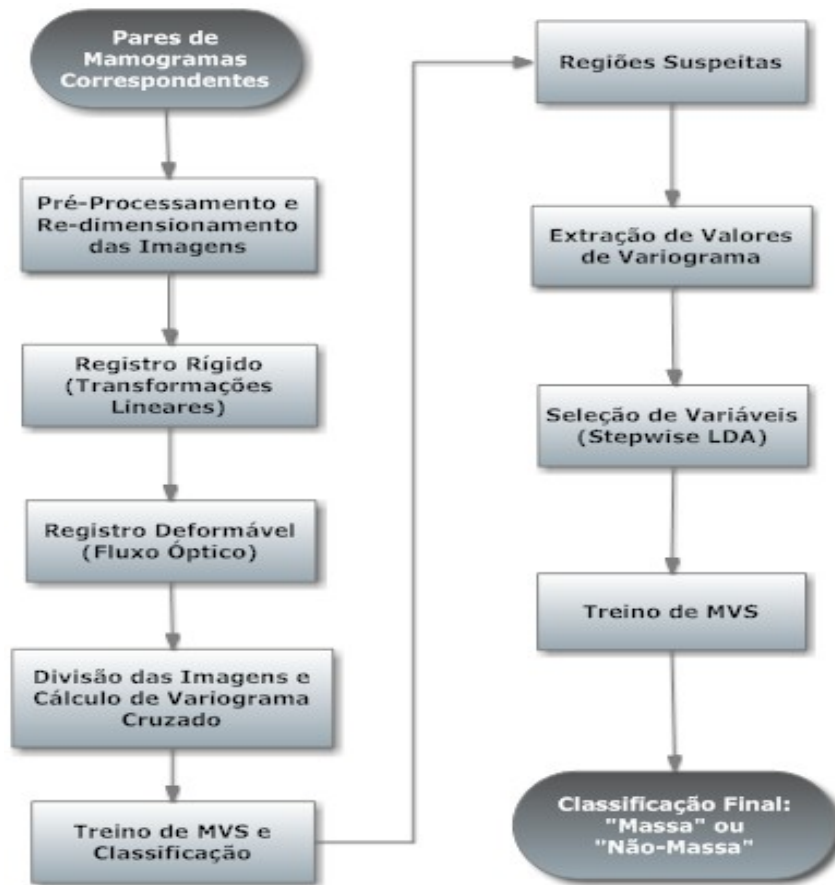


Figura 12: Metodologia desenvolvida pelo trabalho.

A aquisição das imagens utilizadas por este trabalho foi feita através do banco digital DDSM (*Digital Database for Screening Mammography*). A etapa de pré-processamento envolve um processo de redução de ruídos, feito através da aplicação de um filtro de mediana em toda a imagem, e um processo de segmentação, que tem por objetivo extrair a região da imagem correspondente à mama, excluindo a parte do fundo e os artefatos presentes. A segmentação foi implementada com a técnica de *Watershed*, com a aplicação de filtro morfológico de dilatação para estimar a borda da mama. O re-dimensionamento consiste em transformar todas as imagens para o tamanho de 512x512 pixels.

Para corrigir as diferenças de posicionamento e compressão ocorridas no momento do exame, o método de registro rígido foi aplicado (seguido do método de registro deformável) para cada par de imagens. As imagens registradas foram divididas em janelas quadradas de tamanho 32x32 pixels, e o cálculo de variograma cruzado foi realizado para

cada par de janelas correspondentes. Na etapa seguinte, foi criado um modelo para treinamento de um classificador MVS utilizando como características dos vetores os valores de variograma de cada par de janelas. Os pares que continham lesões (de acordo com diagnóstico do DDSM) foram classificados como “regiões suspeitas”. As demais, como “regiões não-suspeitas”. Novos pares de imagens foram utilizados para teste e obtenção de uma nova base de dados, formada por regiões suspeitas. Destas regiões suspeitas, foram extraídas características de variograma. As regiões que continham massas foram classificadas como “regiões de massa” e as demais como “regiões de não-massa”. Foi realizada análise linear discriminante *stepwise* para seleção das características mais significativas. Em seguida, foi criado um modelo para treinamento de uma segunda MVS utilizando vetores com estas características. Foram realizados testes com 30 novos casos para a classificação final pela MVS treinada em “massa” e “não-massa”.

### **3.1 – Seleção das Imagens**

A aquisição das imagens utilizadas por este trabalho foi feita através do banco digital DDSM, que contém 2.620 casos adquiridos do Hospital Geral de Massachusetts, Universidade Wake Forest e Universidade de Washington na Escola de Medicina de St. Louis (HEATH *et al.*, 2001).

Cada caso possui 4 imagens (mama esquerda e direita nas visões crânio-caudal e médio-lateral). O banco DDSM contém descrições de lesões de mamografias de acordo com a Faculdade Americana de Radiologia, além de informações sobre as imagens (tipo de filme e digitalizador utilizados, número de *pixels*, número de bits por *pixel*, etc.).

## 3.2 – Pré-Processamento e Re-dimensionamento das Imagens

A etapa de pré-processamento envolve: redução de ruído e segmentação das imagens. As imagens a serem processadas são re-dimensionadas para o tamanho comum de 512x512, de forma que todas as imagens tenham o mesmo tamanho e que o processo de registro seja viável em termos de tempo de processamento.

### 3.2.1 – Redução de Ruído

As imagens utilizadas passaram por um processo de redução de ruídos, presentes nas imagens devido ao processo de digitalização. Neste trabalho, o processo foi realizado através da aplicação de um filtro de mediana de tamanho 5x5 em toda a imagem.

A Figura 13 mostra uma imagem utilizada no trabalho antes do processo de redução de ruído. A região destacada aparece ampliada à direita para melhor visualização. A Figura 14 mostra a mesma imagem após o processo de redução de ruído.

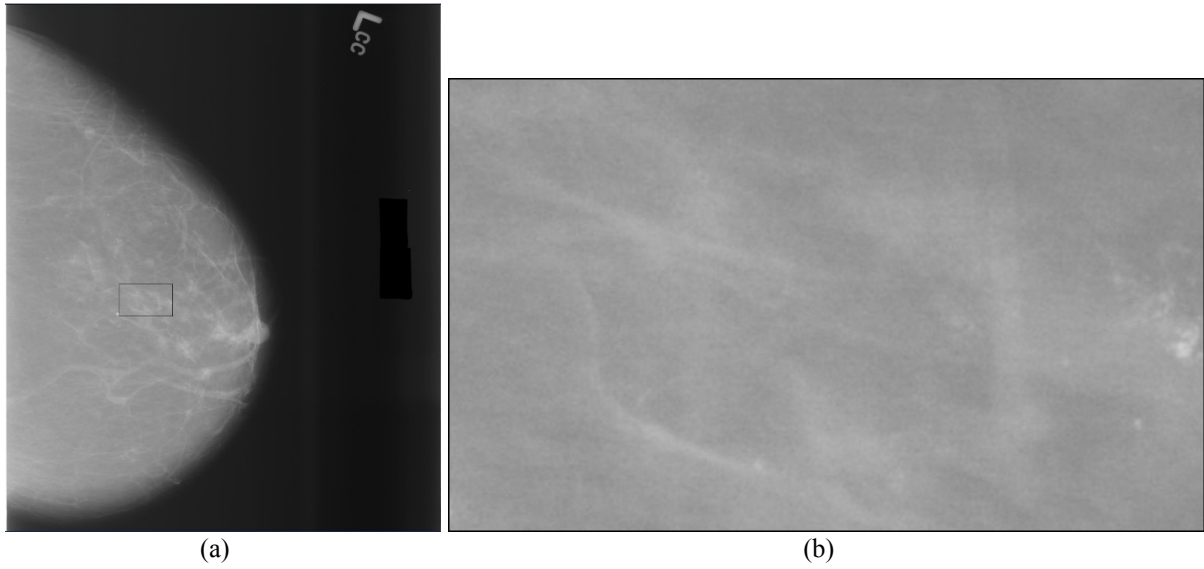


Figura 13: Processo de redução de ruído aplicado a uma imagem. (a) Imagem original com região destacada. (b) Região destacada mostrada para melhor visualização.

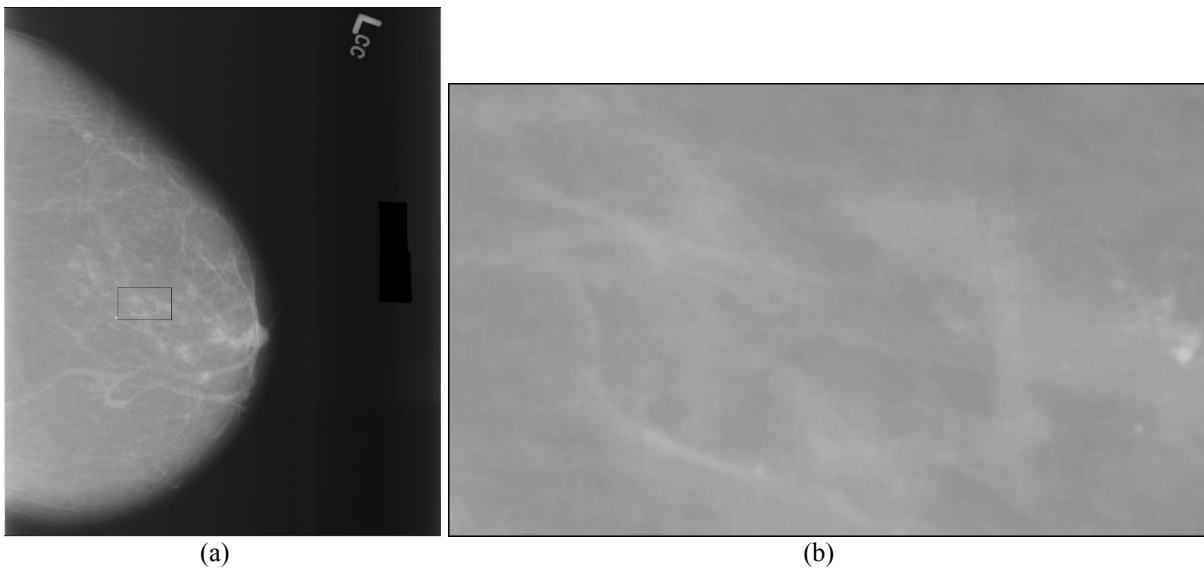


Figura 14: Processo de redução de ruído aplicado a uma imagem. (a) Imagem após o processo de redução de ruído com região destacada. (b) Região destacada mostrada para melhor visualização.

### 3.2.2 – Segmentação das Imagens de Mamografia

A segmentação das imagens foi realizada com a técnica de *Watershed*, com a aplicação de morfologia matemática (dilatação) para estimar a borda da mama. Esta etapa tem por objetivo tornar o registro mais preciso e rápido, pois evita o processamento de partes não pertencentes à mama em si. A Figura 15 mostra um exemplo do processo de segmentação aplicado a um dos pares de imagens utilizados neste trabalho.

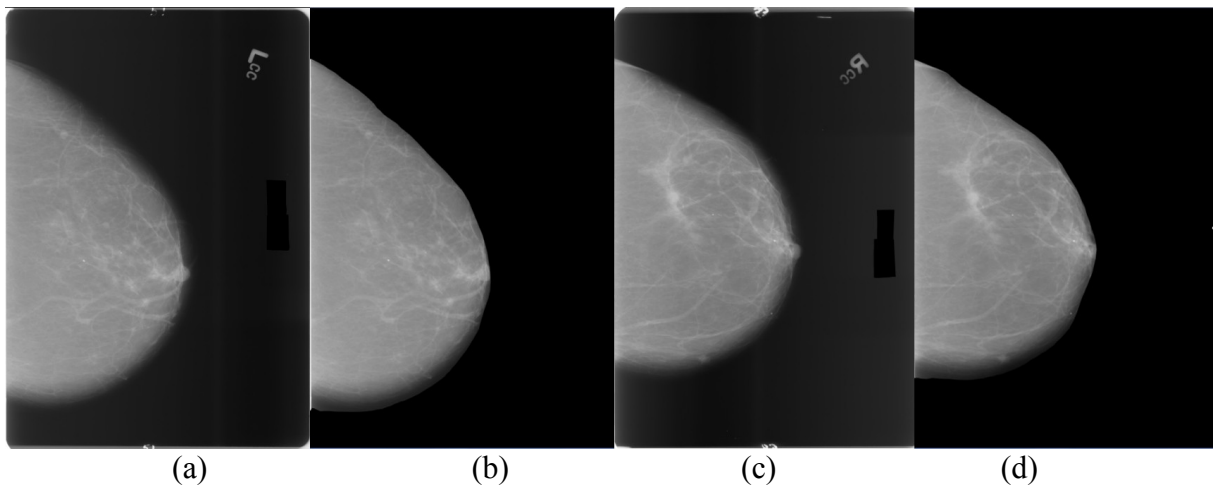


Figura 15: Processo de segmentação aplicado a um par de imagens. (a) Imagem original da mama esquerda. (b) Imagem da mama esquerda segmentada. (c) Imagem original da mama direita. (d) Imagem da mama direita segmentada.

### 3.2.3 – Re-dimensionamento das Imagens de Mamografia

Cada par de imagens de mamografia pré-processado foi re-dimensionado para o tamanho de 512x512 *pixels* (as imagens do banco DDSM possuem tamanhos originais da ordem de 5000x5000 *pixels* aproximadamente), para tornar mais viável computacionalmente (em termos de tempo) o processo de registro e cálculo de variograma.

### 3.3 – Registro das Imagens

O registro das imagens é feito em duas etapas: registro rígido e registro deformável. Neste trabalho, a imagem da mama esquerda serve como imagem referência, e a mama direita é registrada para aproximar-se de seu posicionamento.

#### 3.3.1 – Registro Rígido

O registro rígido foi realizado por operações de translação, rotação e escala (Seção 2.5.3.2). O objetivo do registro rígido é diminuir as diferenças globais que existem entre as duas imagens, através de uma transformação linear aplicada em toda a imagem. A Figura 16 mostra o resultado de um processo de registro rígido. A Figura 17 exhibe as imagens de diferença entre as duas mamas antes e depois do processo.

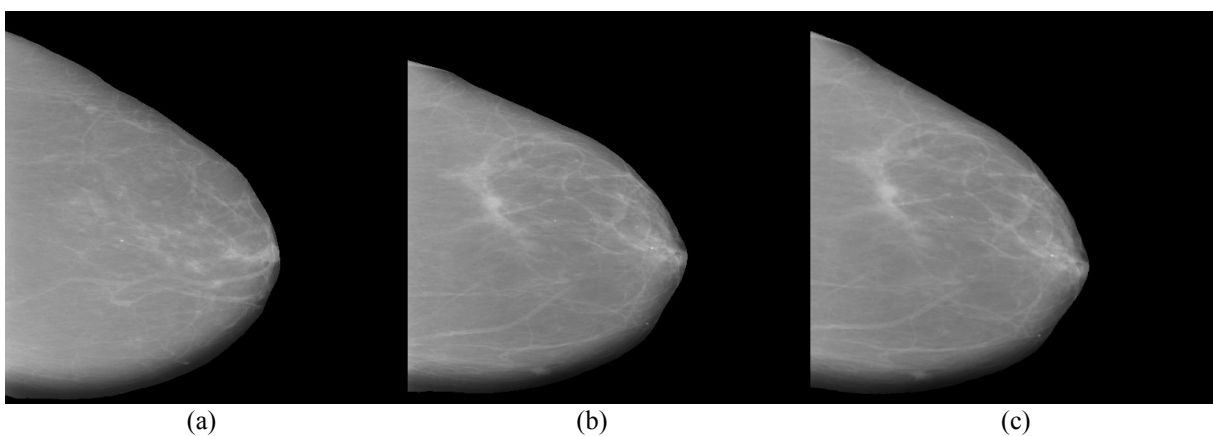


Figura 16: Resultado do processo de registro rígido. (a) Mama esquerda. (b) Mama direita. (c) Mama direita registrada com base na mama esquerda.



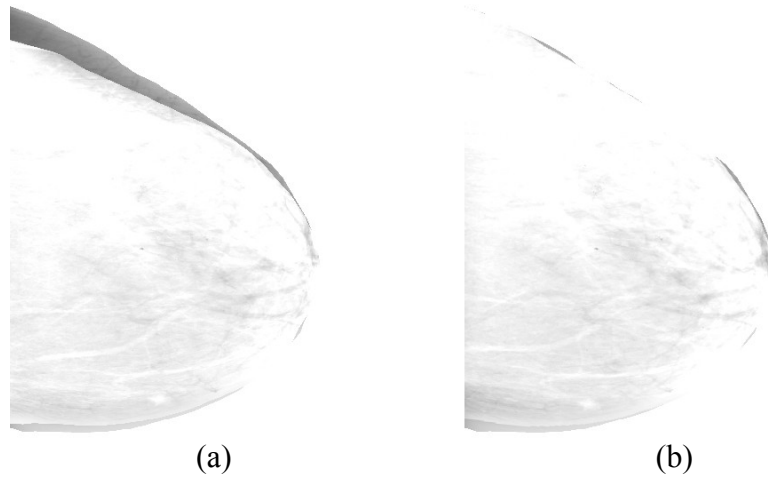


Figura 17: Imagens de diferença entre as mamas resultantes do registro rígido. (a) Diferença antes do registro. (b) Diferença após o registro.

### 3.3.2 – Registro Deformável

O registro deformável foi aplicado utilizando-se o algoritmo de fluxo óptico “*Demons*” de Thirion. A Figura 18 mostra o resultado de um processo de registro deformável, realizado após o registro rígido exibido na Figura 16.

A Figura 19 exibe as imagens de diferença entre as duas mamas antes e depois do processo, assim como as magnitudes dos vetores de fluxo óptico gerados pela deformação que gerou o registro. Os vetores com maior magnitude mostram onde as deformações sofridas pela imagem registrada foram maiores.

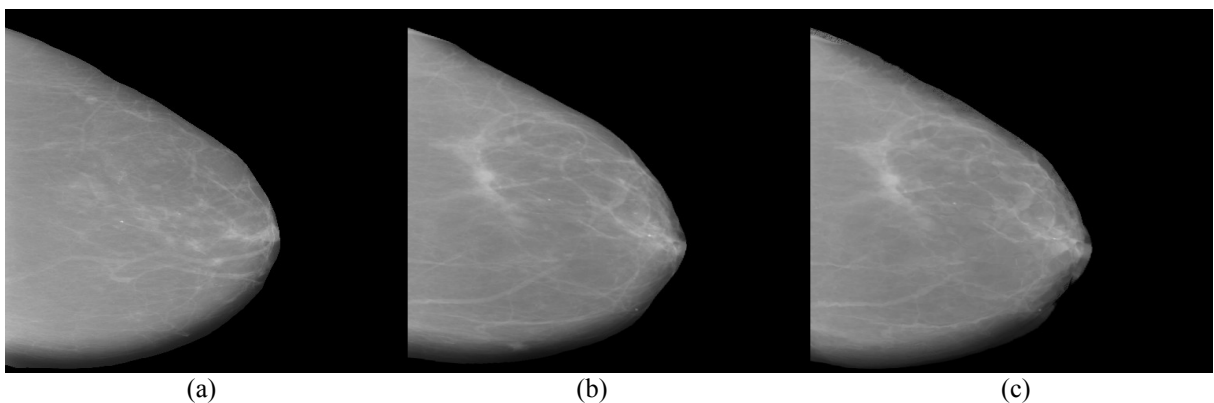


Figura 18: Resultado do processo de registro deformável. (a) Mama esquerda. (b) Mama direita resultante do registro rígido. (c) Mama direita registrada à forma da mama esquerda.

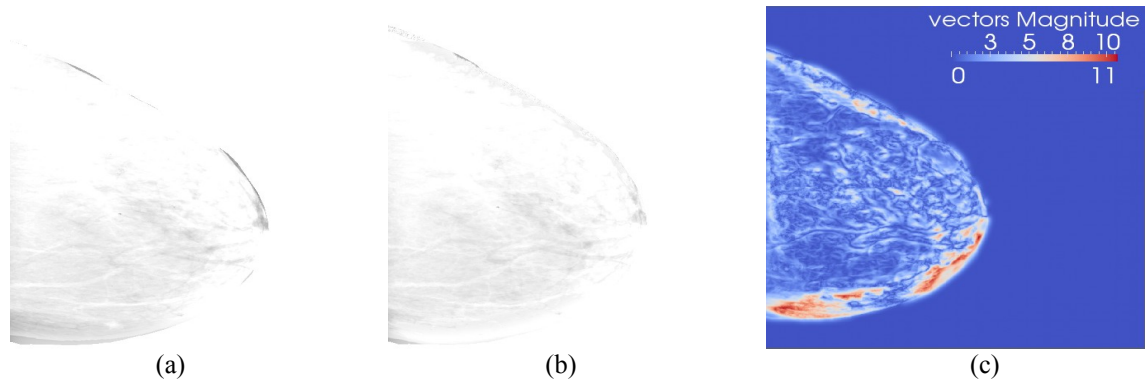


Figura 19: Imagens de diferença entre as mamas resultantes do registro deformável. (a) Diferença antes do registro. (b) Diferença após o registro. (c) Magnitudes dos vetores de fluxo óptico.

### 3.4 – Divisão das Imagens e Cálculo de Variograma Cruzado

Os pares de imagens registrados foram divididos em 256 janelas de tamanho 32x32. Este tamanho de janela apresentou melhores resultados e gerou um número mais conveniente de características de variograma em comparação a outros tamanhos testados. Para cada par de janelas foram calculados 64 valores de variograma cruzado: 16 na direção horizontal, 16 na direção vertical, 16 na direção diagonal 45° e 16 na direção diagonal 135°. A Figura 20 mostra um par de imagens com as 256 janelas consideradas para cálculo de variograma cruzado com uma seta indicando exemplo de região correspondente considerada.

A Figura 21 mostra uma região de tamanho 5x5 que serve como exemplo para ilustrar as direções utilizadas no cálculo dos variogramas. A direção vertical considera que a intensidade de um *pixel* é relacionada com a do *pixel* de posição imediatamente abaixo. O cálculo do variograma considera apenas uma variável (no caso, a intensidade de *pixel* da região mostrada), como ilustrado na Figura 21(a). Já o cálculo do variograma cruzado, (Figura 21(b)) considera duas variáveis (a intensidade de *pixel* da região esquerda e a intensidade da região direita). Portanto, enquanto o variograma relaciona as distâncias verticais (um

determinado *pixel* com o *pixel* de posição inferior) na mesma região, o variograma cruzado relaciona as distâncias verticais correspondentes nas duas regiões (um determinado *pixel* com o *pixel* inferior correspondente na outra região). Esta relação é utilizada pelas Equações 20 e 21 da Seção 2.6. A relação indicada pelas setas da Figura 21 é aplicada para todas as colunas da região. Da mesma forma, a Figura 22 exibe os variogramas horizontais, a Figura 23 exibe os variogramas de 45° e a Figura 24 exibe os variogramas de 135°.

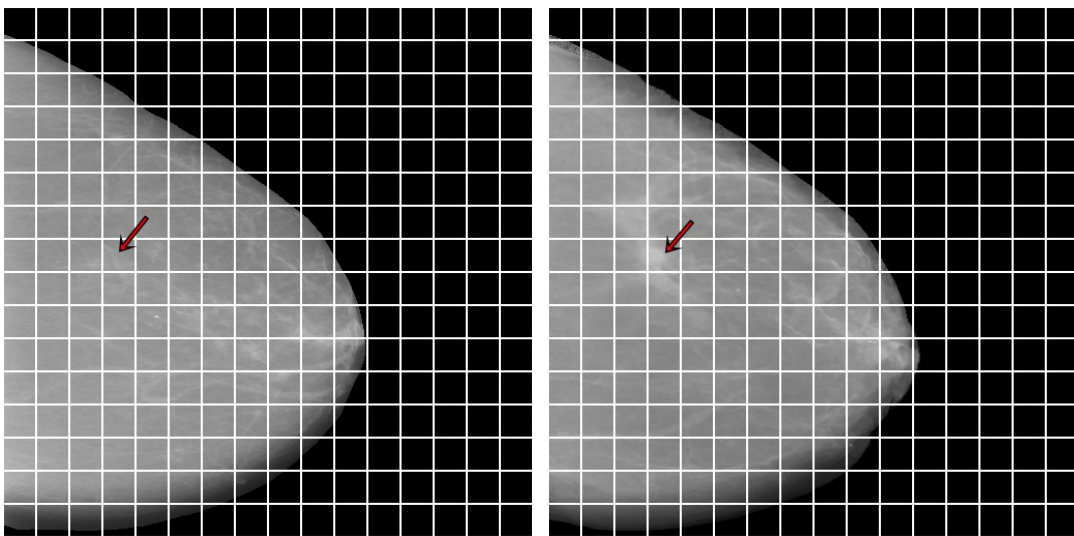


Figura 20: Par de imagens com suas 256 regiões consideradas para cálculo de variograma cruzado. A seta vermelha aponta para um exemplo de duas regiões correspondentes.

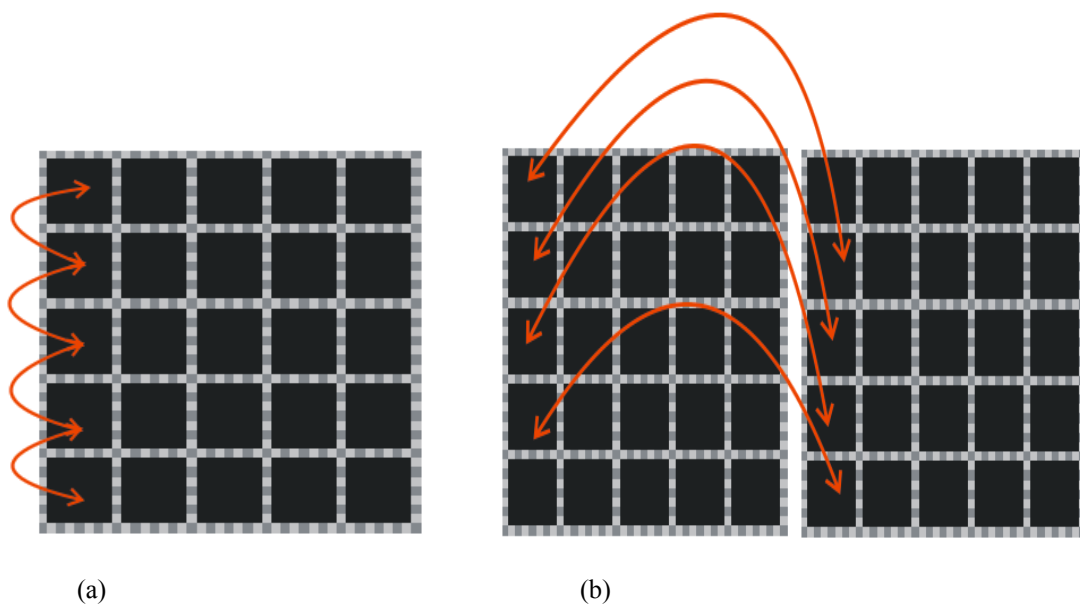


Figura 21: Região 5x5 ilustrativa para cálculo de variograma na direção vertical. (a) *Pixels* relacionados para cálculo de variograma. (b) *Pixels* relacionados em duas regiões correspondentes para cálculo de variograma cruzado.

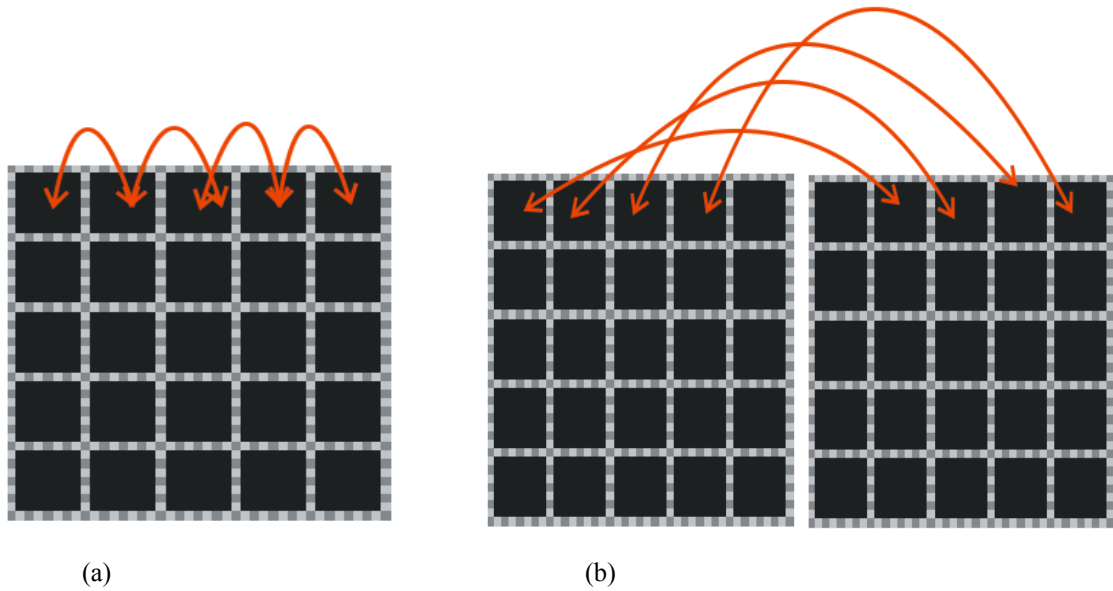


Figura 22: Região 5x5 ilustrativa para cálculo de variograma na direção horizontal. (a) *Pixels* relacionados para cálculo de variograma. (b) *Pixels* relacionados em duas regiões correspondentes para cálculo de variograma cruzado.

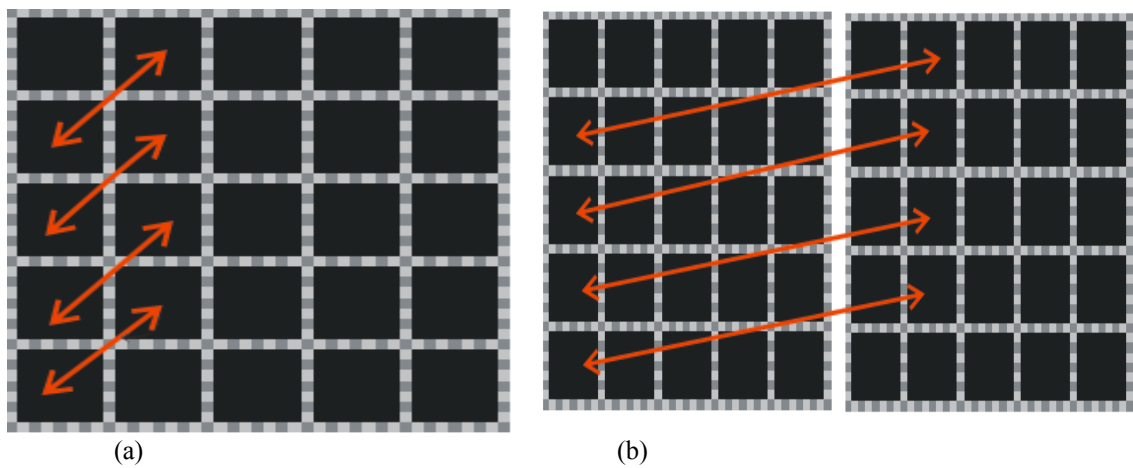


Figura 23: Região 5x5 ilustrativa para cálculo de variograma na direção diagonal 45°. (a) *Pixels* relacionados para cálculo de variograma. (b) *Pixels* relacionados em duas regiões correspondentes para cálculo de variograma cruzado.

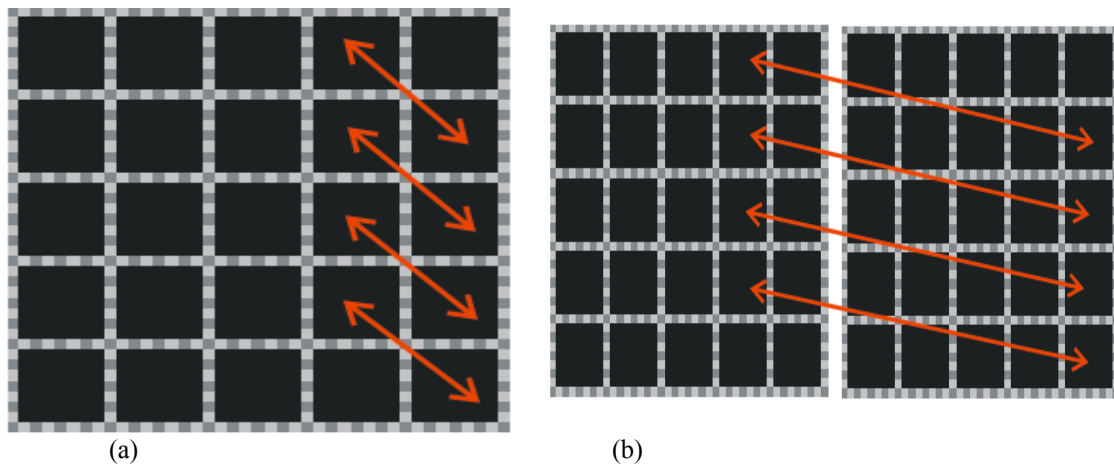


Figura 24: Região 5x5 ilustrativa para cálculo de variograma na direção diagonal 135°. (a) *Pixels* relacionados para cálculo de variograma. (b) *Pixels* relacionados em duas regiões correspondentes para cálculo de variograma cruzado.

Foi criado um modelo para treinamento de um classificador MVS utilizando como características dos vetores os valores de variograma cruzado de cada par de janelas. Os pares que continham lesões (de acordo com diagnóstico do DDSM) foram classificados como “regiões suspeitas”. As demais, como “regiões não-suspeitas”. Novos pares de imagens foram utilizados para teste e obtenção de uma nova base de dados, formada por regiões suspeitas.

### 3.5 – Extração de Características Unilaterais: Variograma

A primeira MVS treinada foi testada para obtenção de regiões consideradas suspeitas. Estes novos casos geraram um novo conjunto de treinamento de pares de regiões suspeitas. Destas regiões suspeitas, foram extraídas características de variograma. As regiões que continham massas foram classificadas como “regiões de massa” e as demais como “regiões de não-massa”. Novamente, a classificação foi feita utilizando-se o diagnóstico

original do banco DDSM. Cada vetor é formado por 64 valores de variograma: 16 na direção horizontal, 16 na direção vertical, 16 na direção diagonal 45° e 16 na direção diagonal 135°. As janelas utilizadas para cálculo de variograma pertencem sempre à imagem do par que contém a massa.

Uma redução de variáveis usando análise linear discriminante *stepwise* foi realizada para seleção das características de variograma mais significativas. Com o conjunto de valores de variograma calculados das regiões suspeitas, foi criado um modelo para treinamento de uma segunda MVS. As regiões que continham massas foram classificadas para treinamento como sendo “massa”, e as demais como “não-massa”. A MVS treinada foi utilizada para classificar novas regiões suspeitas em “massa” ou “não-massa”. Para validação do método, foram realizados testes para a classificação final pela MVS treinada.

#### 4. RESULTADOS

Este capítulo apresenta e discute os resultados de testes realizados com objetivo de avaliar a metodologia proposta no Capítulo 3.

##### 4.1 – Regiões Suspeitas

Para o treinamento da MVS utilizada para determinar as regiões suspeitas, foram pré-processados e registrados 180 pares (esquerda e direita) de imagens, a maioria contendo presença de massa (150 pares) e outros com diagnóstico normal (30 pares), todos na projeção crânio-caudal. Neste conjunto de 180 casos, cada par de imagens foi utilizado para gerar os vetores de características de variograma cruzado. Foram consideradas apenas regiões internas das mamas. Um conjunto com 2700 vetores de 64 características foi criado a partir destes casos (cada vetor é formado por 64 valores de variograma: 16 na direção horizontal, 16 na direção vertical, 16 na direção diagonal 45° e 16 na direção diagonal 135°). Destes vetores, 1800 foram classificados como normais e 900 foram classificados como regiões suspeitas (de acordo com o diagnóstico original da base DDSM). Este conjunto de 2700 vetores foi utilizado para treinar uma MVS que classifica novos casos de teste. O número total de 2700 vetores foi baseado no número de regiões suspeitas consideradas. Nos 180 pares, 900 regiões foram consideradas suspeitas. Para que fosse observada a proporção 2:1 no arquivo de treino entre regiões não-suspeitas e suspeitas, outras 1800 regiões foram escolhidas (não-suspeitas),

totalizando 2700 regiões. Nesta seção são exibidos resultados das regiões suspeitas de três testes, todos pares de imagens obtidos do banco DDSM.

#### 4.1.1 – Regiões Suspeitas: Teste 1

Um dos pares de imagens escolhido para determinação de regiões suspeitas é exibido na Figura 25. Trata-se do par B3499 do banco DDSM. A massa encontra-se indicada na imagem direita do par. Os processos de segmentação e registro realizados para este par são mostrados nas Figuras 15, 16 e 18 do Capítulo 3.

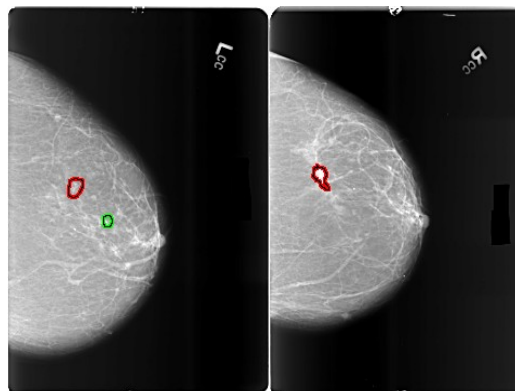


Figura 25: Mama esquerda e direita do par B3499 do DDSM, escolhidas para o Teste 1. A massa encontra-se na mama direita. As marcas na mama esquerda referem-se a calcificações.

Como exemplo, duas regiões do par foram destacadas e seus gráficos de função variograma-cruzado foram exibidos. A Figura 26 mostra uma região normal (marcada em branco) e uma região considerada suspeita (marcada em vermelho). A Figura 27 mostra os gráficos de variograma cruzado para as duas regiões marcadas na Figura 26. Os valores de variograma cruzado exibidos vêm da aplicação da Equação 21. Percebe-se que as magnitudes



dos valores de variograma cruzado da região normal (que estão sobrepostos ao eixo horizontal) possuem valores significativamente menores em relação aos valores da região suspeita. Os valores de variograma cruzado refletem, portanto, as variações existentes de uma região para outra.

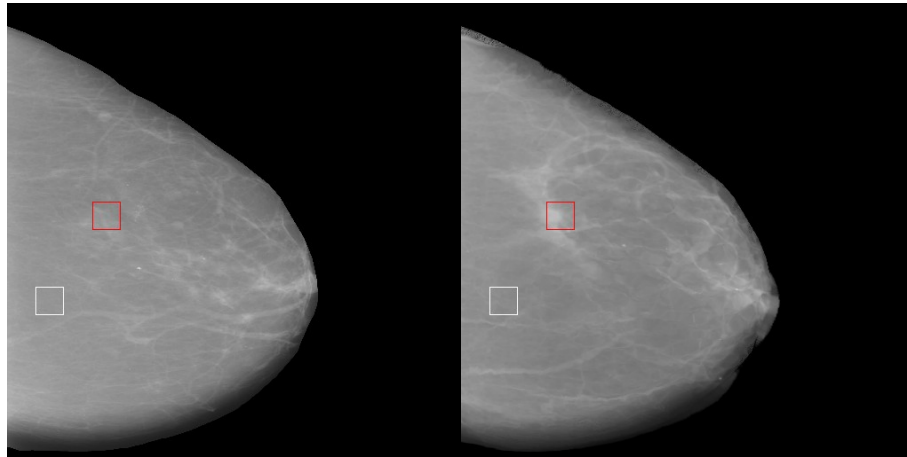


Figura 26: Regiões destacadas no par escolhido para o Teste 1. Em cor branca, a região normal. Em cor vermelha, a região considerada suspeita.

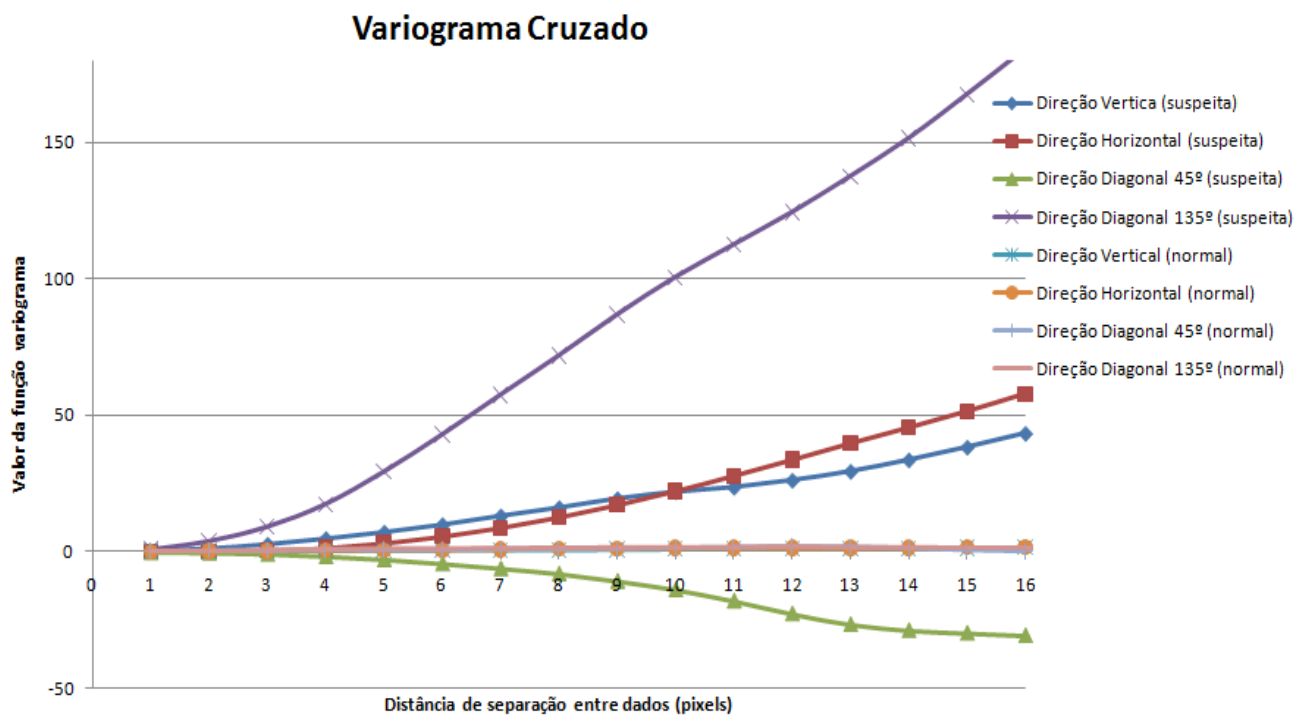


Figura 27: Gráficos de variograma cruzado nas 8 direções: 4 direções para a região vermelha da Figura 26 (região suspeita) e 4 direções para a região branca da Figura 26 (região normal).

As regiões suspeitas definidas pela MVS para o par B3499 são exibidas na Figura 28. Em cor vermelha estão as 3 regiões que correspondem à massa indicadas no diagnóstico do DDSM (Figura 25). Em cor amarela estão as demais regiões, que apresentam variações entre si, mas que não são indicadas no banco DDSM como sendo massa. Mesmo assim, a MVS treinada captou as variações existentes entre essas regiões.

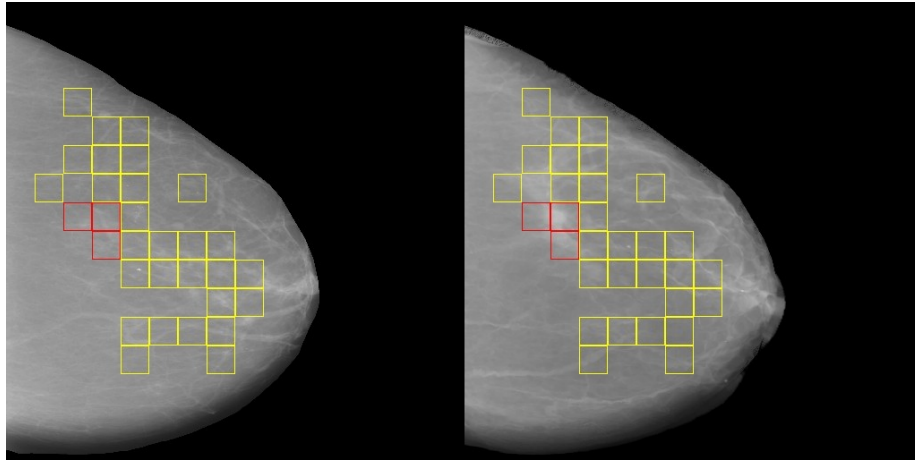


Figura 28: Regiões suspeitas resultantes do Teste 1.

#### 4.1.2 – Regiões Suspeitas: Teste 2

Outro par de imagens escolhido para determinação de regiões suspeitas é exibido na Figura 29. Trata-se do par A1581 do banco DDSM. A massa encontra-se indicada na imagem da mama esquerda.

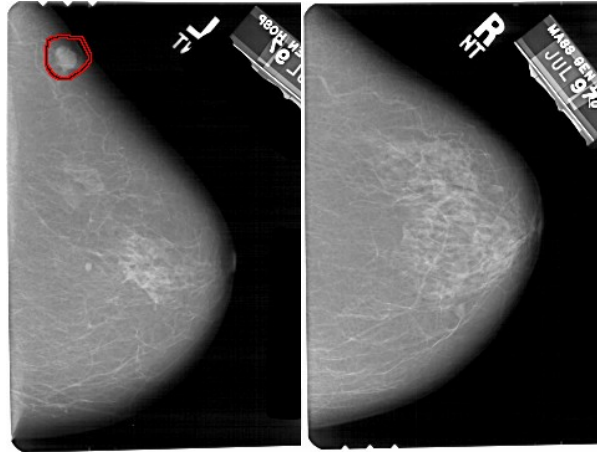


Figura 29: Mama esquerda e direita do par A1581 do DDSM, escolhido para o Teste 2.

O processo de segmentação realizado para este par é exibido na Figura 30. São mostradas as duas imagens originais seguidas de suas imagens processadas.

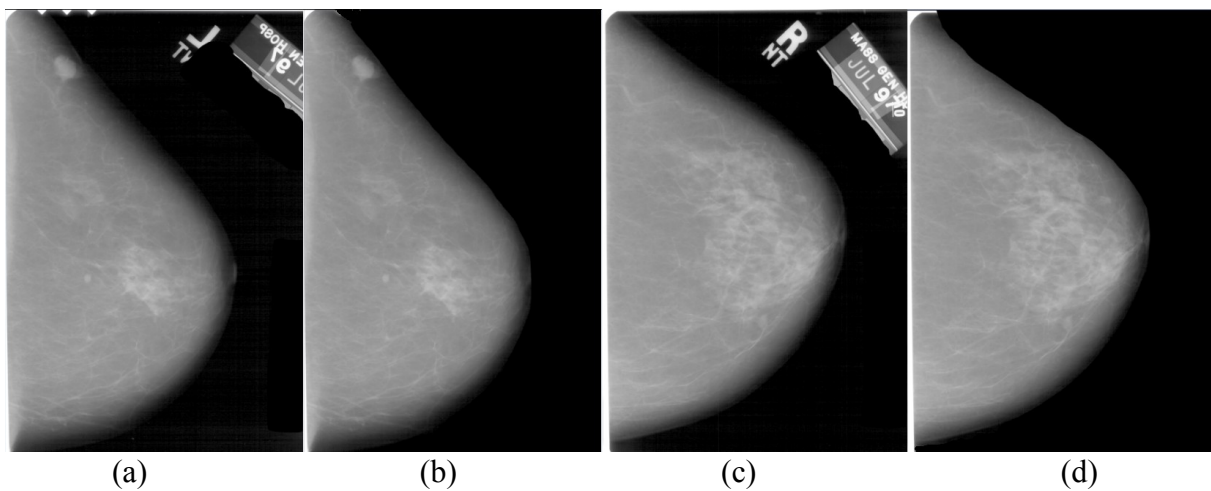


Figura 30: Processo de segmentação aplicado ao par de imagens do Teste 2. (a) Imagem original da mama esquerda. (b) Imagem da mama esquerda segmentada. (c) Imagem original da mama direita. (d) Imagem da mama direita segmentada.

O processo de registro realizado para este par é exibido na Figura 31. São mostradas as imagens da mama esquerda e da direita seguidas da imagem da mama direita registrada à forma da esquerda.

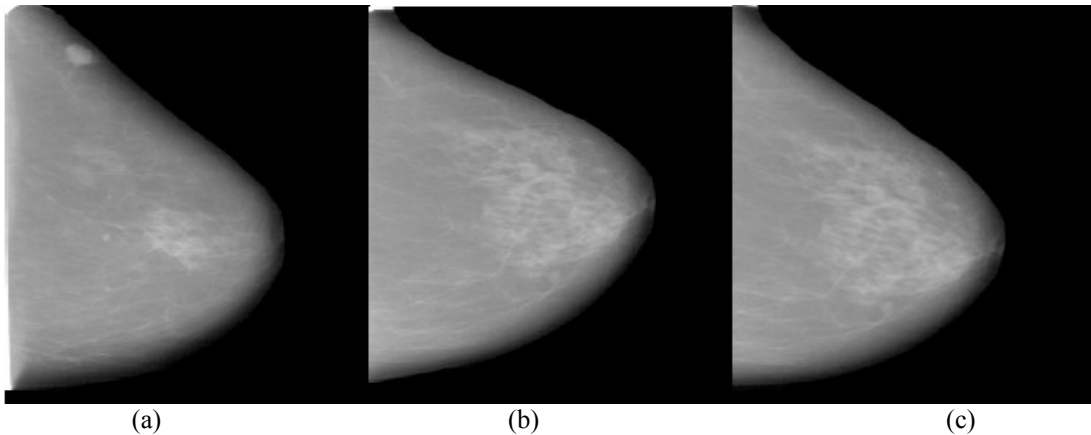


Figura 31: Processo de registro aplicado ao par de imagens do Teste 2. (a) Imagem original da mama esquerda. (b) Imagem original da mama direita. (c) Imagem da mama direita registrada à forma da esquerda.

Assim como foi feito para o Teste 1, duas regiões do par foram destacadas e seus gráficos de função variograma-cruzado foram exibidos. A Figura 32 mostra uma região normal (marcada em cor branca) e uma região considerada suspeita (marcada em cor vermelha). A Figura 33 mostra os gráficos de variograma cruzado para as duas regiões marcadas na Figura 32. Os valores de variograma exibidos no eixo vertical vêm da aplicação da Equação 21. Percebe-se que, assim como no Teste 1, as magnitudes dos valores de variograma cruzado da região normal (que estão sobrepostos ao eixo horizontal) possuem valores significativamente menores em relação aos valores da região suspeita.

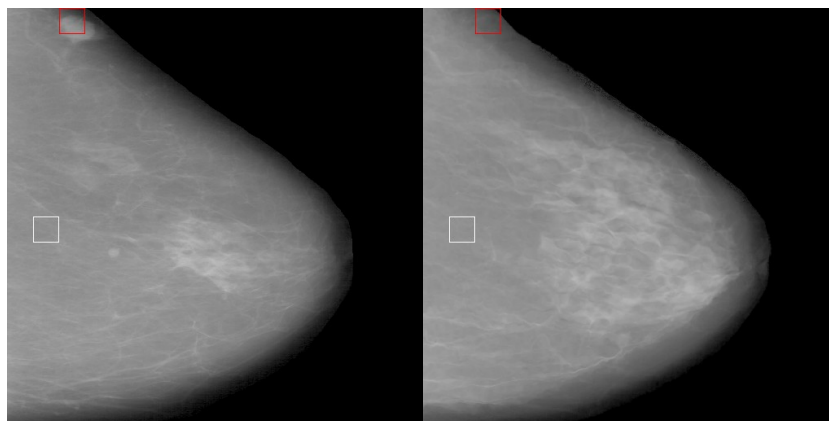


Figura 32: Regiões destacadas no par escolhido para o Teste 2. Em cor branca, a região normal. Em cor vermelha, a região considerada suspeita.

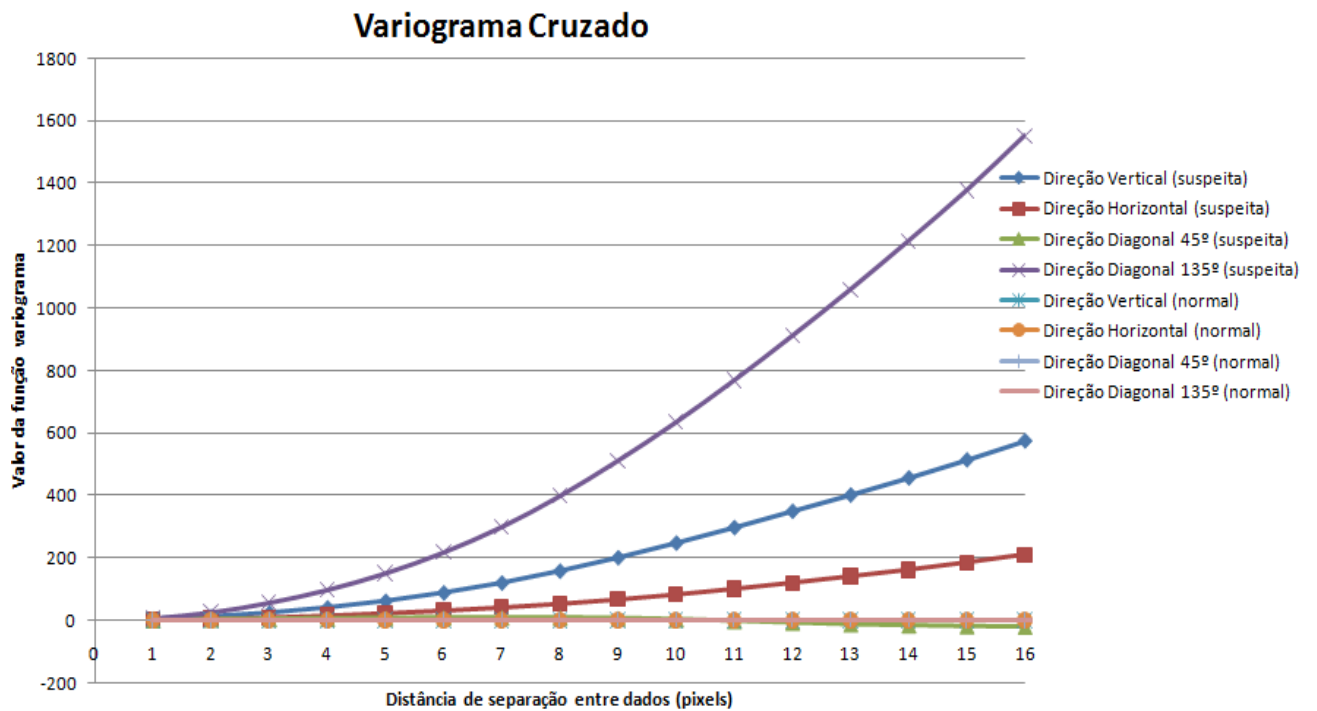


Figura 33: Gráficos de variograma cruzado nas 8 direções: 4 direções para a região vermelha da Figura 32 (região suspeita) e 4 direções para a região branca da Figura 32 (região normal).

As regiões suspeitas definidas pela MVS para o par A1581 são exibidas na Figura 34. Em cor vermelha estão as 3 regiões que correspondem à massa indicadas no diagnóstico do DDSM (Figura 29). Em cor amarela estão as demais regiões. A MVS treinada captou as variações existentes entre essas regiões, apesar de elas não serem indicadas no banco DDSM como sendo massa.

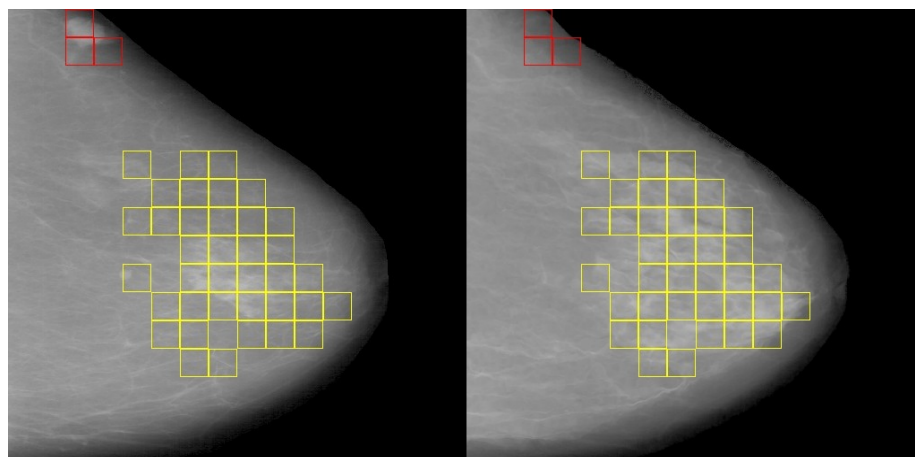


Figura 34: Regiões suspeitas resultantes do Teste 2.

#### 4.1.3 – Regiões Suspeitas: Teste 3

Outro par de imagens escolhido para determinação de regiões suspeitas é exibido na Figura 35. Trata-se do par A0323 do banco DDSM. Este par foi escolhido para um terceiro teste por não apresentar nenhum tipo de lesão.

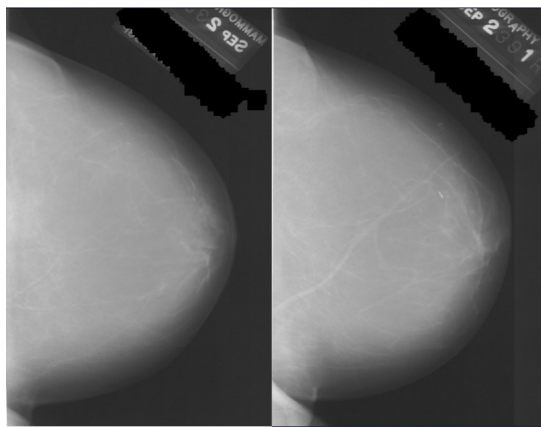


Figura 35: Mama esquerda e direita do par A0323 do DDSM, escolhido para o Teste 3.

Da mesma forma feita para os testes anteriores, o par de imagens foi pré-processado e registrado. As 12 regiões suspeitas definidas pela MVS para o par A0323 são exibidas na Figura 36. Em cor amarela estão regiões que apresentam variações entre si, mas que não são indicadas no banco DDSM como sendo massa. Mesmo assim, a MVS treinada captou as variações existentes entre essas regiões.

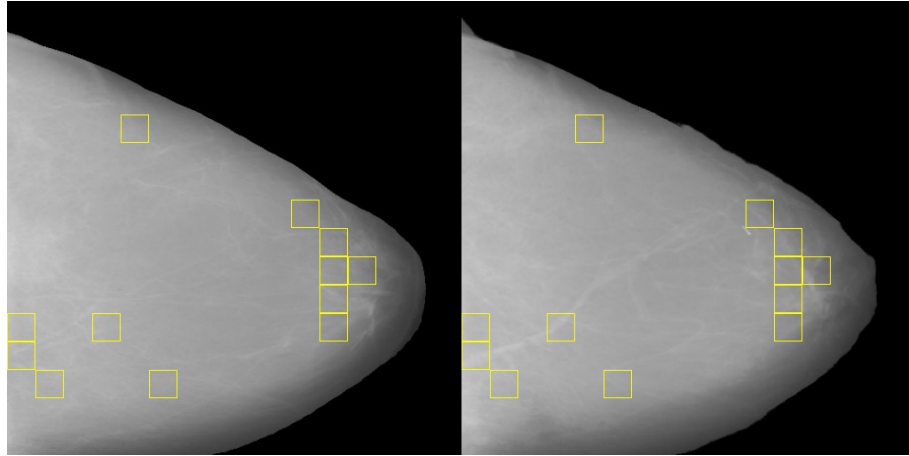


Figura 36: Regiões suspeitas resultantes do Teste 3.

#### 4.2 – Classificação Entre Massa e Não-Massa

Para a obtenção das características de variograma das regiões suspeitas e treinamento da MVS utilizada para classificar as massas, foram pré-processados e registrados outros 100 pares de imagens. A primeira MVS treinada foi testada com estes 100 novos casos, para obtenção de regiões consideradas suspeitas. Estes novos casos geraram um novo conjunto de treinamento, de 1050 regiões suspeitas. O novo conjunto de vetores foi composto com 700 vetores classificados como normais (não-massa) e 350 classificados como massa. Novamente, a classificação foi feita utilizando-se o diagnóstico original do banco DDSM. Cada vetor é formado por 64 valores de variograma: 16 na direção horizontal, 16 na direção vertical, 16 na direção diagonal  $45^\circ$  e 16 na direção diagonal  $135^\circ$ . As janelas utilizadas para cálculo de variograma pertenciam sempre à imagem do par que continha a massa.

Do conjunto de 1050 regiões suspeitas, foram obtidas 5 características de variograma de cada região (resultantes da análise linear discriminante *stepwise* sobre as 64 características originais). Estas características foram: direção vertical (distâncias de 1,2, e 10

*pixels*), direção diagonal 45° (distância 9) e direção diagonal 135° (distância 1). Um novo conjunto de treino para uma segunda MVS foi então montado com estes valores. As regiões que continham massas foram classificadas para treinamento como sendo “massa”, e as demais como “não-massa”.

Para o teste da classificação final, foram pré-processados e registrados 30 novos casos. A segunda MVS treinada foi utilizada para classificar novas regiões suspeitas em “massa” ou “não-massa”. Os 30 novos pares de imagens foram utilizados para determinação de regiões suspeitas e posterior classificação final de lesões. Foram realizados, no total, testes com 30 casos para a classificação final. Os Testes 1 e 2 foram escolhidos para serem mostrados por representarem um caso com bom resultado (Teste 1, com acurácia de 90%, sensibilidade de 100% e especificidade de 89,28%) e um caso com resultado ruim (Teste 2, com acurácia de 70%, sensibilidade de 100% e especificidade de 67,56%). O melhor resultado apresentou na classificação final: 96% de acurácia, 100% de sensibilidade e 95,34% de especificidade. O pior caso apresentou 70% de acurácia, 100% de sensibilidade e 67,56% de especificidade. Nesta seção são exibidos resultados das classificações das regiões suspeitas obtidas nos Testes 1, 2 e 3.

#### 4.2.1 – Classificação Final: Teste 1

Para a classificação final das regiões suspeitas, um dos pares de imagens utilizado foi o par B3499 do banco DDSM (regiões suspeitas do Teste 1 mostradas na Seção 4.1.1 – Figura 28). A imagem escolhida foi a imagem direita do par, por ser a imagem que contém a massa. A Figura 37 exibe duas regiões suspeitas da imagem direita do par da Seção 4.1.1



(Figura 28): uma que contém massa (janela vermelha) e uma que não contém massa (janela amarela).

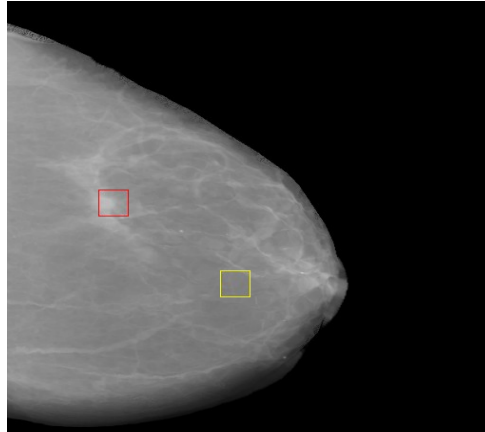


Figura 37: Regiões suspeitas do Teste 1 escolhidas para exibição de função variograma. A região amarela não contém massa. A região vermelha contém massa.

A Figura 38 exibe os gráficos de função variograma correspondentes às regiões da Figura 37. Os valores de variograma exibidos no eixo vertical vêm da aplicação da Equação 20. Pode-se observar novamente a diferença entre as magnitudes de variograma das regiões. Nos gráficos de variograma, por existirem apenas valores positivos, o eixo y pode ser colocado em escala logarítmica de base 10 para que se possa visualizar melhor o comportamento e diferenças entre os dois tipos de região.

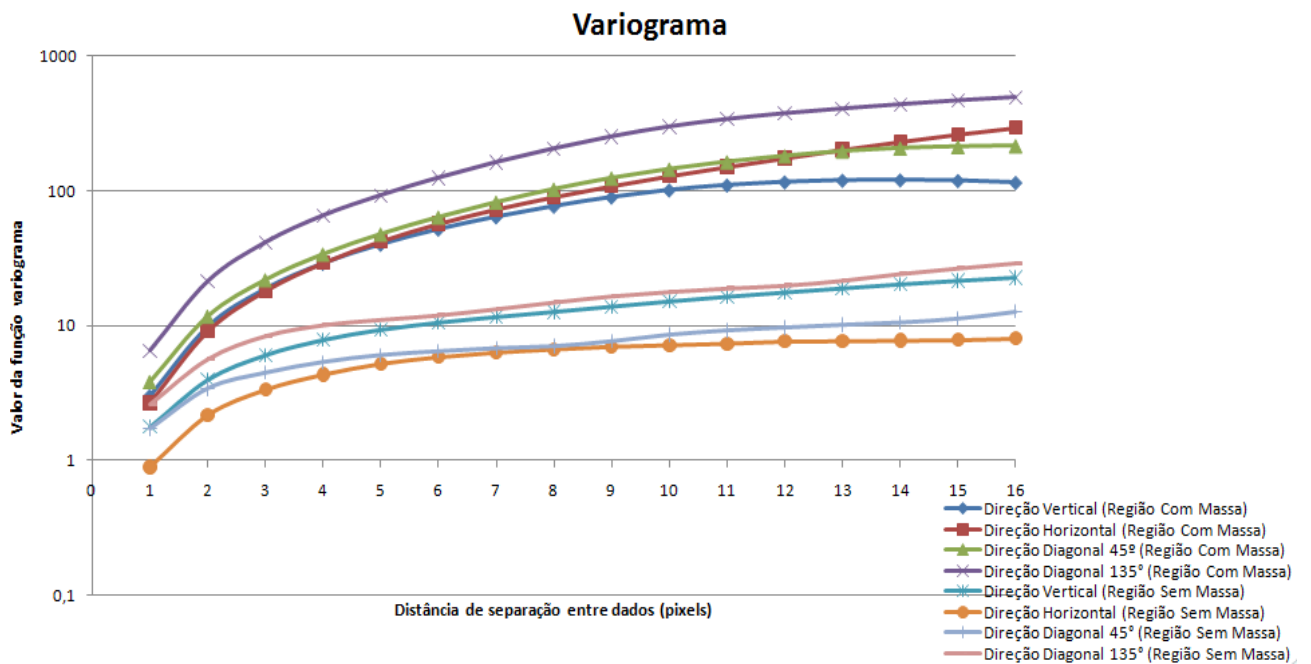


Figura 38: Gráficos de variograma nas 8 direções: 4 direções para a região vermelha da Figura 37 (região com massa) e 4 direções para a região amarela da Figura 37 (região sem massa).

Das 31 regiões suspeitas determinadas para este par (Seção 4.1.1 – Figura 28), 28 foram classificadas corretamente. As três regiões determinadas como sendo massa (de acordo com o diagnóstico do banco DDSM) foram classificadas corretamente, e estão indicadas em cor azul na Figura 39. Três outras regiões foram classificadas erradamente como sendo massa (regiões em cor amarela na Figura 39). Isto resulta em acurácia de 90%, sensibilidade de 100% e especificidade de 89,28%.

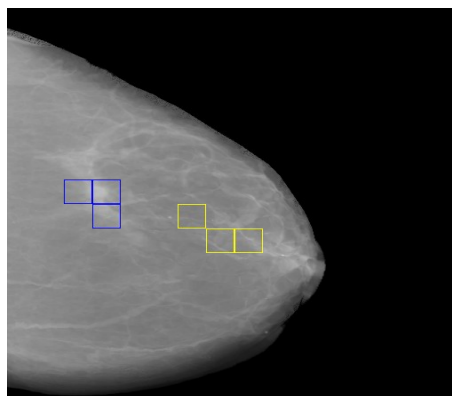


Figura 39: Classificação final das regiões suspeitas para o Teste 1. Em cor azul, as classificações de massa corretas. Em cor amarela, as classificações de massa erradas.

#### 4.2.2 – Classificação Final: Teste 2

Para o Teste 2, o par de imagens utilizado foi o par A1581 do banco DDSM. A imagem escolhida foi a imagem esquerda do par, por ser a imagem que contém a massa. Das 40 regiões suspeitas determinadas para este par (Seção 4.1.2 – Figura 34), 28 foram classificadas corretamente. As três regiões determinadas como sendo massa (de acordo com o diagnóstico do banco DDSM) foram classificadas corretamente, e estão indicadas em cor azul na Figura 40. Doze outras regiões foram classificadas erradamente como sendo massa (regiões em cor amarela na Figura 40). Isto resulta em acurácia de 70%, sensibilidade de 100% e especificidade de 67,56%.

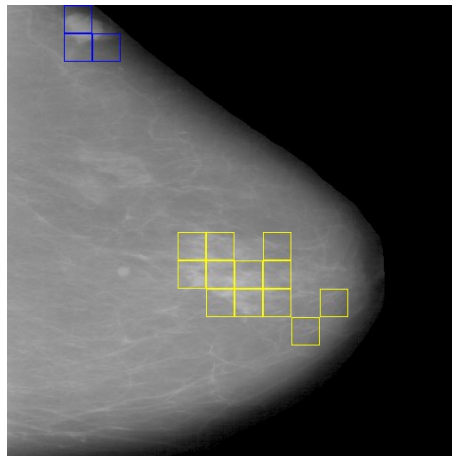


Figura 40: Classificação final das regiões suspeitas para o Teste 2. Em cor azul, as classificações de massa corretas. Em cor amarela, as classificações de massa erradas

#### 4.2.3 – Classificação Final: Teste 3

Para o Teste 3, o par de imagens utilizado foi o par A0323 do banco DDSM. As regiões suspeitas para este par foram determinadas na Seção 4.1.3 (Figura 36). Como este par não contém massas, o teste para classificação final foi feito em cada imagem (os testes anteriores foram realizados apenas na imagem que continha a massa).

Na imagem esquerda do par, das 12 regiões suspeitas determinadas para este par (Seção 4.1.3), 11 foram classificadas corretamente. Uma região foi classificada erradamente como sendo massa (região amarela na Figura 41). Isto resulta em acurácia de 91,66%.

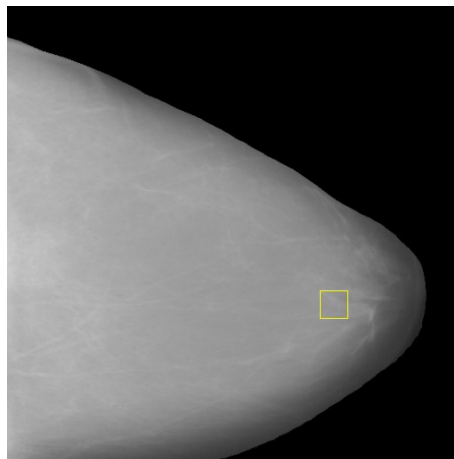


Figura 41: Classificação final das regiões suspeitas para o Teste 3 (imagem da mama esquerda). Em cor amarela, as classificações de massa erradas.

Na imagem direita do par, das 12 regiões suspeitas determinadas para este par (Seção 4.1.3 – Figura 36), 10 foram classificadas corretamente. Duas regiões foram classificadas erradamente como sendo massa (região em cor amarela na Figura 42). Isto resulta em acurácia de 83,33%.

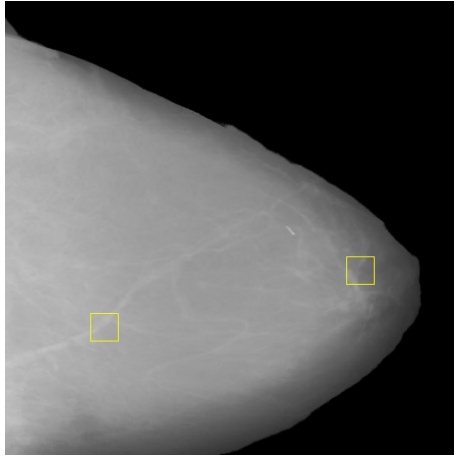


Figura 42: Classificação final das regiões suspeitas para o Teste 3 (imagem da mama direita). Em cor amarela, as classificações de massa erradas.

## 5. CONCLUSÃO

O trabalho desenvolvido apresentou o desenvolvimento de uma metodologia para detecção de regiões suspeitas em pares de imagens de mamografia esquerda e direita e posterior classificação das mesmas em massa e não-massa. As regiões suspeitas foram detectadas por variações estruturais entre regiões correspondentes, definidas por um descritor espacial de conjunto de dados conhecido como variograma cruzado. Após a determinação de regiões suspeitas em um par de imagens, o variograma foi utilizado em cada região suspeita isoladamente, para classificação em massa e não-massa. Tanto a detecção de regiões suspeitas quanto a classificação final foram realizadas por classificadores MVS treinados com valores de variograma cruzado e variograma. Os valores de treino foram obtidos do banco de imagens DDSM.

Os resultados exibidos no Capítulo 4 demonstram o bom desempenho da metodologia desenvolvida. A etapa de determinação de regiões suspeitas envolveu 100% das massas da amostra de treino, ou seja, nenhuma massa deixou de ser determinada como região suspeita. A etapa de classificação das regiões suspeitas também obteve bom desempenho para os 30 casos de teste: o melhor resultado apresentou na classificação final: 96% de acurácia, 100% de sensibilidade e 95,34% de especificidade. O pior caso apresentou 70% de acurácia, 100% de sensibilidade e 67,56% de especificidade. Os valores médios para todos os testes foram: 90% de acurácia, 100% de sensibilidade e 85% de especificidade. Destaca-se a classificação em relação a casos positivos (sensibilidade), com desempenho de 100% nos testes realizados. É importante ressaltar que estes valores consideram uma confiabilidade de 100% em diagnósticos dos exames de mamografia. Na realidade, a confiabilidade destes exames é variante, dependendo de uma série de fatores como qualidade dos recursos técnicos

na aquisição das imagens e habilidade de interpretação do médico. Algumas classificações consideradas erradas, portanto, poderiam ser regiões que não foram corretamente diagnosticadas.

Embora a metodologia tenha apresentado resultados satisfatórios, algumas outras ideias poderiam ser testadas e incluídas no processo desenvolvido, podendo possivelmente melhorar alguns aspectos e resultados. Alguns outros descritores espaciais, como função de correlação e função de covariância, poderiam ser incluídos para, em conjunto com a função variograma, possivelmente proporcionar melhor descrição de padrões de textura de tecido. Outras medidas geométricas e de textura poderiam ser combinados com os descritores espaciais para melhor caracterização de massas. A metodologia de detecção de regiões suspeitas de tamanhos fixos poderia ser substituída por uma metodologia de regiões de interesse segmentadas nas imagens. A metodologia poderia ter sua detecção ampliada para regiões de calcificação, de forma que fosse feita a distinção entre massas e calcificações. As massas detectadas poderiam ter sua classificação refinada quanto à natureza (maligna ou benigna). Testes com outros classificadores poderiam ser feitos para comparação de resultados em relação à MVS. A classificação final poderia ser aplicada às duas mamas, e não apenas à mama que contém a massa.

O presente trabalho explora a possibilidade de utilização de descritores espaciais como forma de analisar padrões de textura de tecidos, podendo a metodologia desenvolvida ter sua aplicabilidade estendida para outros tipos de lesões encontradas em imagens médicas digitais.

## REFERÊNCIAS BIBLIOGRÁFICAS

ACS - **American Cancer Society** Disponível em: <<http://www.cancer.org/>>. Último acesso em: 16 jan. 2011.

BURHENNE, Linda J. Warren; WOOD, Susan A.; D'ORSI, Carl J.; FEIG, Stephen A.; KOPANS, Daniel B.; O'SHAUGHNESSY, Kathryn F.; SICKLES, Edward A.; TABAR, Laszlo; VYBORNY, Carl J.; CASTELLINO, Ronald A.. **Potential Contribution of Computer-aided Detection to the Sensitivity of Screening Mammography**. From the 1998 RSNA scientific assembly. Received Jul 19, 1999; revision requested Aug 26; revision received Sep 21; accepted Oct 21. 2000.

**CancerHelp UK**. Disponível em: <<http://www.cancerhelp.org.uk/>>. Último acesso em: 18 jan. 2011.

CHIH-CHUNG, Chang e CHIH-JEN, Lin. **LIBSVM: a library for support vector machines**, 2001.

CORRÊA, Aristófanés. **Algoritmos para diagnóstico assistido de nódulos pulmonares solitários em imagens de tomografia computadorizada**. Tese de Doutorado, Departamento de Informática, PUC-Rio. 2004

CORRÊA, Aristófanés; SILVA, Valdeci Ribeiro; PAIVA, Anselmo Cardoso; OLIVEIRA, Alexandre Cesar Muniz. **Semivariogram Applied for Classification of Benign and Malignant Tissues**. Universidade Federal do Maranhão, Departamento de Ciência de Computação. Av. dos Portugueses, SN, Campus do Bacanga, Bacanga 65085-580, São Luís, MA, Brasil. 2006.

CORTES, Corinna e VAPNIK, Vladimir. **Support-vector network**. *Machine Learning*, 20:273{297, 1995.

ERICEIRA, Daniel; CORRÊA, Aristófanés e PAIVA, Anselmo Cardoso. **Deteção de Regiões Suspeitas em Mamografias Digitais Utilizando Descrição Espacial com Variograma Cruzado**. Universidade Federal do Maranhão, Departamento de Ciência de Computação. Av. dos Portugueses, SN, Campus do Bacanga, Bacanga 65085-580, São Luís, MA, Brasil. Publicado em "XII Brazilian Congress of Health Informatics", CBIS, Outubro de 2010.

FREER, Timothy W. e ULISSEY, Michael J.. **Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center**. From the Women's Diagnostic and Breast Health Center, 3800 W 15th St, Suite 111, Plano, TX 75075. From the 2000 RSNA scientific assembly. Received July 21, 2000; revision requested September 6; revision received March 9, 2001.

GEORGSSON, Fredrik. **Differential analysis of bilateral mammograms**. In *Int. J. Pattern Recognition and Artificial Intelligence*, Vol. 17, pp. 1207-1226, July 2003.



GONZALEZ, Rafael C. e WOODS, Richard E.. **Digital Image Processing**. Prentice Hall. Ed. 2; 2002.

GOSHTASBY, Ardeshir A.. **2-D and 3-D image registration for medical, remote sensing, and industrial applications**. Wiley-Interscience publication; 2005.

HACHAMA, Mohamed; DESOLNEUX, Agnès e RICHARD, Frédéric. **Combining Registration and Abnormality Detection in Mammography**. MAP5, University Paris 5, 45, rue des Saints P`eres, 75006 Paris, France. 2006.

HEATH, Michael; BOWYER, Kevin; KOPANS, Daniel; MOORE, Richard e KEGELMEYER, W. Phillip. **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**. Disponível em: <<http://www.inca.gov.br/>>. Último acesso em: 28 jan. 2011.

ISAAKS, Edward H. e SRIVASTAVA, R. Mohan. **An Introduction to Applied Geostatistics**. New York: Oxford University Press, Inc.;1989.

LAU, Tin-Kit e BISCHOF, Walter F.. **Automated Detection of Breast Tumors Using The Asymmetry Approach**. Department of Computing Science, University Of Alberta, Canada. 1990.

LEE, Kevin A.. **A Mammographic Registration Method Based On Optical Flow and Multiresolution Computing**. Escola de Engenharia do Instituto de Tecnologia da Força Aérea dos Estados Unidos, 1997.

MENDEZ, Arturo J.; TAHOCES, Pablo G.; LADO, Maria J.; SOUTO, Miguel e VIDAL, Juan J.. **Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms**. University of Santiago de Compostela, Spain. 1998.

PARSONS, P.A.. **Fluctuating asymmetry: an epigenetic measure of stress**. Biological Review, 65: 131–145, 1990.

SALLAM, Maha Y. e BOWYER, Kevin W.. **Fluctuating Registration and difference analysis of corresponding mammogram images**. Medical Image Analysis, volume 3, number 2, pp 103–118. 1999.

SCUTT, Diane; LANCASTER, Gillian A. e MANNING, John T.. **Breast asymmetry and predisposition to breast cancer**. Breast Cancer Research, 8:R14, licensee BioMed Central Ltd., 2006.

SCUTT, D.; MANNING, J.T.; WHITEHOUSE, G.H.; LEINSTER, S.J. e MASSEY, C.P.. **The relationship between breast asymmetry, breast size and the occurrence of breast cancer**. The British Journal of Radiology, 1997.

SERRA, Jean. **Image analysis and mathematical morphology**. London: Academic Press, 1982.

STAMATAKIS, E.A.; RICKETTS, I.W.; CAIRNS, A.Y.; WALKER, T. e PREECE, P.E.. **Detecting abnormalities on mammograms by bilateral comparison.** Dundee Univ. London, UK, 1996.

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

VAN ENGELAND, Saskia. **Detection of mass lesions in mammograms by using multiple views.** Tese de Doutorado, Universidade de Eindhoven, 2006.

WU, Yi-Ta; WEI, Jun; HADJIISKI, Lubomir M.; SAHINER, Berkman; ZHOU, Chuan; GE, Jun; SHI, Jiazheng; ZHANG, Yiheng e CHAN, Heang-Ping. **Bilateral analysis based false positive reduction for computer-aided mass detection.** Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109, 2007.

YIN, Fang-Fang; GIGER, Maryellen L.; DOI, Kunio; VYBORNÝ, Carl J. e SCHMIDT, Robert A.. **Computerized detection of masses in digital mammograms: automated alignment of breast images and its effect on bilateral-subtraction technique.** Med. Phys. v21. 445-452. 1993.