

ANÁLISE DA INFLUÊNCIA DA SEGMENTAÇÃO SOBRE O PROBLEMA DA CLASSIFICAÇÃO DE ANOMALIAS EM IMAGENS TERMOGRÁFICAS DE MAMA

K.F.F.C. Queiroz*, M.C. Araujo*, R.C.F. de Lima*

*Departamento de Eng. Mecânica/UFPE, Recife, Brasil

e-mail: kamila942@hotmail.com

Resumo: Atualmente a termografia vem sendo proposta como uma ferramenta complementar para investigar o câncer de mama. A segmentação das imagens termográficas digitais de mama visa definir a região de interesse, diferenciando-a do ambiente externo e das demais regiões na imagem. O objetivo do presente trabalho é avaliar a eficácia de um classificador estatístico usando imagens térmicas de mama segmentadas a partir de um método semiautomático de segmentação das mamas baseado em *spline* linear. Para identificação de anormalidades na mama foram utilizadas cinquenta imagens de pacientes do Hospital das Clínicas da Universidade Federal de Pernambuco. Três técnicas foram utilizadas para classificação dos termogramas: classificação por discriminante linear, classificação por distância mínima e janela de Parzen. A eficácia da classificação foi medida através da sensibilidade, da especificidade e da taxa de acerto.

Palavras-chave: Mama, Termografia, Segmentação, *Spline* Linear, Classificação.

Abstract: Nowadays the thermography is being proposed as a complementary exam for the diagnosis of breast cancer. The segmentation of the breast infrared images intends to define regions of interest (ROIs). Those regions were used for the development of the present work. The goal of the segmentation is to separate the breast from the background and from others regions of the human body. Thus the aim of this work is the use of semiautomatic segmentation method of the breasts that was developed from linear splines. The referred segmentation was used to classify breast abnormalities using of new method of feature extraction based on symbolic variables. 50 images from patients of Hospital das Clínicas (HC) of the Federal University of Pernambuco (UFPE) were used during the tests. Three statistical techniques are used to classify the thermograms: linear discriminant classifier, minimum distance classifier and Parzen window. The influence of the segmentation on the method of feature extraction referred above was measured using the sensitivity, the specificity and the accuracy.

Keywords: Breast, Thermography, Segmentation, Linear Spline, Classification.

Introdução

A inspeção termográfica é uma técnica não-destrutiva que utiliza a radiação infravermelha emitida pelos corpos para gerar informações específicas sobre a distribuição da temperatura na superfície de um objeto. O seu uso na área médica vem se difundindo pelo fato de não emitir radiação ionizante, diferente da mamografia. Neste âmbito, é destacada sua aplicação na detecção de tumores mamários.

A computação gráfica oferece métodos que possibilitam a manipulação adequada de imagens digitais, no processo de segmentação e extração de estruturas de interesse no diagnóstico médico. Além disso, tais técnicas proporcionam uma análise objetiva e quantitativa das imagens, que são importantes para reduzir as falhas causadas normalmente por limitações da técnica de diagnóstico utilizada, por limitações do sistema visual humano, ou pelo fato de as imagens serem, na grande maioria, qualitativas e subjetivas.

Desta maneira, a segmentação da imagem da mama, tanto de modo automático quanto semiautomático, pode ser utilizada no suporte ao diagnóstico do câncer mamário. A primeira, porém, pode extrair áreas específicas das mamas de maneira errônea devido à assimetria natural entre elas e às variações anatômicas entre pessoas [1].

A proposta deste trabalho é avaliar a influência do processo de segmentação na detecção de anormalidades de mama a partir de diferentes classificadores baseados numa extração de características sugerida em [2]. Para isto, é proposta uma técnica de segmentação semiautomática por *splines* lineares. Os resultados obtidos foram comparados com aqueles gerados através da aplicação de uma segmentação por elementos elipsoidais, realizada em [2].

Materiais e métodos

As imagens termográficas tratadas neste trabalho foram obtidas com uma câmera de infravermelho Flir S45, adquirida pelo Departamento de Engenharia Mecânica (DEMEC) da Universidade Federal de Pernambuco (UFPE) através do Edital FINEP 2007 Laboratório de Multiusuário. As imagens foram obtidas de pacientes do Ambulatório de Mastologia do Hospital das Clínicas (HC) da UFPE e a amostra escolhida foi

composta por mulheres com diagnóstico concluído a partir de exames clínicos, ultrassonografia, mamografia e biópsia. Cada paciente concordou em participar da pesquisa por meio da assinatura do Termo de Consentimento Livre e Esclarecido (TCLE), documento associado ao cadastro do projeto no Ministério de Saúde, sob o Registro CEP/CCS/UFPE N° 279/05 e aprovado pelo Comitê de Ética da Universidade Federal de Pernambuco. A análise das imagens foi realizada com o auxílio do software *FLIR QuickReport*® e processadas no *Matlab*®.

Segmentação das mamas – A imagem térmica digital é representada por uma imagem indexada que possui dois componentes: matriz de temperatura e um mapa de cor. Assim, no termograma, cada pixel representa o valor de temperatura definido pela matriz de temperatura. Estes valores permanecem constantes, independentes da mudança do mapa de cor ou da escala de temperatura utilizada, viabilizando o tratamento das imagens a partir da temperatura ao invés da intensidade dos pixels.

Pelo fato de ser um processo simples, rápido e eficaz que permite ao usuário determinar a região de interesse, o método escolhido para a segmentação das mamas foi a aplicação da interpolação utilizando *spline* linear, gerada manualmente através de um programa desenvolvido no Matlab sobre a matriz temperatura citada anteriormente.

A interpolação utilizando *splines* sugere que o contorno de um objeto seja dividido em um conjunto de segmentos que podem ser representados adequadamente por polinômios de graus pequenos [3]. No caso da *spline* linear, a função interpolante $S_1(x)$ nos nós x_0, x_1, \dots, x_n , pode ser escrita em cada subintervalo $[x_{i-1}, x_i]$, $i=1, 2, \dots, n$, como:

$$S_1(x) = f(x_{i-1}) \frac{x_i - x}{x_i - x_{i-1}} + f(x_i) \frac{x - x_{i-1}}{x_i - x_{i-1}} \quad (1)$$

Nesta etapa utilizou-se a definição da função *spline* linear (Equação 1) para a formação da curva adequada à mama. A curva é definida exclusivamente por um conjunto de pontos de controle selecionados pelo usuário através do *mouse* e, entre estes pontos, a função é aplicada para ajuste da curva.

Os passos usados para a segmentação das imagens foram: 1) A matriz de temperaturas é extraída do termograma usando-se o software *FLIR QuickReport*®, e processada no *Matlab*® como uma imagem (Figura 2a); 2) O usuário seleciona a mama e, em seguida, os pontos de controle referentes ao contorno da região de interesse (Figura 2b-d); 3) Com os pontos de controle selecionados pelo usuário, a curva de interpolação por *spline* linear é gerada, definindo uma máscara que representa a região de interesse em cada uma das mamas das pacientes (Figura 2e); 4) A partir da máscara gerada, é obtida a matriz correspondente à região selecionada (Figura 2f). 5) Ao fim, a segmentação das mamas fornecerá duas matrizes de temperatura, MD e ME para a mama direita e mama esquerda, respectivamente.

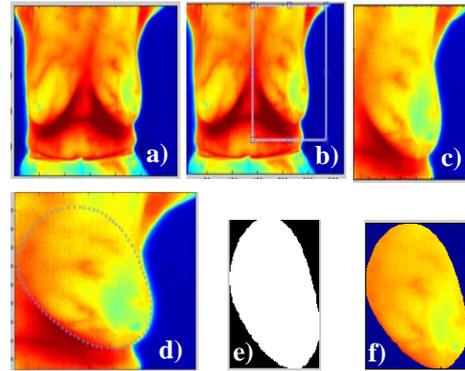


Figura 2: Etapas do algoritmo de segmentação.

Extração de características – A partir das matrizes MD e ME, geradas pelo processo de segmentação das mamas, foram obtidos os vetores de características para o processo de classificação. Sobre cada matriz de temperatura, se extraiu a informação referente ao aumento de temperatura local e da área total da mama. A etapa de extração de características segue o modelo sugerido em [2].

Classificação – Os dados foram apresentados a diferentes classificadores existentes na literatura: classificador discriminante linear; classificador de distância mínima baseado na distância de Mahalanobis, na distância Euclidiana e na distância City-Block; e um classificador do tipo janela de Parzen utilizando função de núcleo uniforme e outro utilizando função de núcleo gaussiano [2]. Os dados foram avaliados quanto aos índices de sensibilidade e especificidade para cada classe. Embora estas medidas estejam relacionadas com um teste de classificação binária, é possível generalizar este conceito para um problema com mais de duas classes.

A sensibilidade (Equação 2) representa a proporção de indivíduos com o problema em questão que são classificados corretamente pelo teste, ou seja, indica o quão bom é o teste em identificar tal problema. Para a i -ésima classe, ela pode ser calculada como:

$$\text{Sensibilidade}_i = \frac{VP_i}{VP_i + FN_i} \quad (2)$$

onde VP_i são os resultados verdadeiros positivos para a classe i e FN_i são os falsos negativos referentes à classe i .

Já a especificidade (Equação 3) é a proporção de pacientes sem o problema que são identificados corretamente pelo teste, isto é, indica o quão bom é um teste em identificar o indivíduo sem o problema em questão. Para a i -ésima classe tem-se:

$$\text{Especificidade}_i = \frac{VN_i}{VN_i + FP_i} \quad (3)$$

onde VN_i são os resultados verdadeiros negativos avaliados para a classe i e FP_i são os falsos positivos avaliados em função da classe i .

O experimento realizado utilizou uma amostragem balanceada de 50 termogramas, sendo estes 14 casos de pacientes com tumores malignos, 19 casos de tumores benignos e 17 casos de cistos. O classificador foi

validado em uma avaliação onde uma paciente é usada como teste e as outras 49 como treino (*leave one out*). Esta avaliação é repetida para as 50 pacientes, de forma que 50 diferentes grupos são usados como treino e 50 diferentes elementos de teste. A separação do grupo "não-maligno" em benigno e cisto foi uma solicitação da equipe médica participante do grupo de pesquisa.

Resultados

Entre todos os 50 pacientes, cinco destes não tiveram o processo de segmentação concluído devido a falhas apresentadas durante o processamento (Figura 3). Então, a classificação foi realizada com dois conjuntos de dados distintos:

1. 50 termogramas, sendo 45 segmentados por *spline* linear e 05 segmentados por elipses;

2. 45 termogramas segmentados por *spline* linear com uma amostra de 11 casos de tumores malignos, 18 casos de tumores benignos e 16 casos de cistos.

Após a extração do vetor das características, os resultados foram comparados com aqueles obtidos através da segmentação por elipse [2].

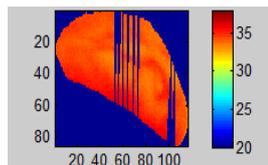


Figura 3: Erro apresentado no processo de segmentação.

Nas Tabelas 1, 2 e 3 são apresentados os dados de sensibilidade (S) e especificidade (E) encontrados para o grupo diagnosticado com tumor benigno, com tumor maligno e com cistos, respectivamente. Nestas tabelas, SSL (Segmentação por Spline Linear) representa a extração de características por *spline* linear utilizando o conjunto de dados 2; SSE (Segmentação por Spline e Elipse) representa a extração de características através da *spline*, utilizando o conjunto de dados 1 e SEL (Segmentação por Elipse) representa a extração de características por elipses, utilizando um conjunto de dados original fornecido por [2].

A Tabela 4 mostra, a partir das taxas de acerto, o desempenho de cada classificador para cada conjunto de amostra utilizado.

Tabela 1: Resultados da classificação para tumores benignos utilizando diferentes técnicas.

Técnica utilizada	SSL		SSE		SEL	
	S(%)	E(%)	S(%)	E(%)	S(%)	E(%)
Disc. Linear	83,3	91,7	84,2	96,5	84,2	96,5
Distância de Mahalanobis	77,8	96,1	78,9	96,7	78,9	96,6
Distância Euclidiana	72,2	96,1	84,2	96,7	84,2	96,6
Distância City-Block	72,2	83,3	47,4	74,2	52,6	74,2
J. Parzen Gaussiana	100,0	34,6	100,0	43,3	100,0	50,0

J. Parzen Uniforme	100,0	72,0	100,0	82,8	100,0	82,7
--------------------	-------	------	-------	------	-------	------

Tabela 2: Resultados da classificação para tumores malignos utilizando diferentes técnicas.

Técnica utilizada	SSL		SSE		SEL	
	S(%)	E(%)	S(%)	E(%)	S(%)	E(%)
Disc. Linear	54,5	86,5	78,6	92,1	78,5	92,1
Distância de Mahalanobis	81,8	80,0	85,7	86,5	85,7	86,5
Distância Euclidiana	81,8	77,1	85,7	86,5	85,7	86,5
Distância City-Block	36,4	73,0	78,6	83,3	78,6	83,3
J. Parzen Gaussiana	18,2	100,0	35,7	100,0	42,8	100,0
J. Parzen Uniforme	36,4	100,0	64,3	100,0	64,2	100,0

Tabela 3: Resultados da classificação para cistos utilizando diferentes técnicas.

Técnica utilizada	SSL		SSE		SEL	
	S(%)	E(%)	S(%)	E(%)	S(%)	E(%)
Disc. Linear	81,2	86,2	88,2	87,9	88,2	87,8
Distância de Mahalanobis	75,0	93,1	82,3	90,1	82,3	90,9
Distância Euclidiana	75,0	93,1	82,3	93,9	82,3	93,9
Distância City-Block	62,5	86,2	58,8	81,8	58,8	84,5
J. Parzen Gaussiana	46,7	100,0	50,0	100,0	56,2	100,0
J. Parzen Uniforme	93,3	96,5	93,7	97,0	93,7	96,9

Tabela 4: Taxas de acerto dos classificadores.

Técnica utilizada	SSL	SSE	SEL
	Taxa de acerto (%)	Taxa de acerto (%)	Taxa de acerto (%)
Disc. Linear	75	84	84
Distância de Mahalanobis	77	82	82
Distância Euclidiana	76	84	84
Distância City-Block	54	60	62
J. Parzen Gaussiana	62	66	65
J. Parzen Uniforme	82	88	88

Discussão

De acordo com as Tabelas 1 a 4, o método SSL apresentou resultados de sensibilidade e especificidade relativamente piores quando comparados com as demais segmentações avaliadas. Como esperado, isto se deve ao fato de a amostra ser 10% menor e, por conta disso, ter apresentado um pequeno desbalanceamento.

Para os métodos SSE e SEL, os resultados obtidos na classificação baseada em distância e para o discriminante linear mostraram que os métodos SSE e SEL apresentam valores próximos de sensibilidade e especificidade, e que apesar das condições da

amostragem, a SSL apresentou medidas aceitáveis de classificação. Dentre os classificadores avaliados, a distância de Mahalanobis apresentou os melhores índices para classe de tumor maligno, com uma diferença de 3,9 pontos percentuais para a sensibilidade e 6,5 pontos percentuais para a especificidade entre a SSL e SSE. Já a distância Euclidiana apresentou melhores resultados destes índices para as classes de tumor benigno e cisto, com uma diferença entre a SSL e SSE de 12 e 7,3 pontos percentuais de sensibilidade, e 0,6 e 0,8 pontos percentuais de especificidade para tumor benigno e cisto, respectivamente.

O classificador Janela de Parzen (Gaussiana e Uniforme) não apresentou uma boa taxa de classificação para o problema proposto. Apesar de ter atingido 100% de sensibilidade para a classe benigno, não obteve bom desempenho para a classe maligno. A classe maligno é de fato a mais preocupante visto que a sensibilidade para esta classe é um parâmetro de grande importância na detecção de câncer mamário. Para esta classe, os classificadores que se destacaram foram os classificadores baseados nas distâncias de Mahalanobis e na distância Euclidiana.

Como esperado, os resultados da Tabela 4 mostram que SSE e SEL, pelo uso dos mesmos termogramas, apresentaram resultados similares ou até mesmo iguais. Dentre os classificadores avaliados, aqueles que apresentaram melhor desempenho na classificação de anormalidades da mama a partir de imagens termográficas foram: discriminante linear e distância Euclidiana, ambas com 84% de acerto. Com relação à sensibilidade à classe maligno, pode-se destacar o classificador baseado na distância de Mahalanobis.

Na SSL, que analisou uma amostra menor, a técnica que causou menor variação em relação às amostras SSE e SEL foi a distância de Mahalanobis, com uma diferença de 5 pontos percentuais.

Conclusão

Este trabalho avaliou diferentes técnicas de segmentação de imagens térmicas de mama quanto às suas influências na classificação de anormalidades mamárias. As imagens segmentadas pela técnica de *spline* linear serviram de base para o processo de extração de características de imagens térmicas proposto em [2] e usadas para identificar anormalidades mamárias. Foram analisados os termogramas de 50 pacientes com diagnósticos concluídos que foram determinados por exame clínico, ultrassonografia, mamografia e exame de biópsia. A matriz de temperatura de cada imagem, que se encontra armazenada em um banco de dados (<http://150.161.49.17/termo>), foi obtida e, após processamento no Matlab, segmentada a fim de extrair a área de interesse (ROI) de cada mama do termograma.

Os resultados das taxas de acerto das técnicas utilizadas e das sensibilidades à classe de tumores malignos mostraram-se satisfatórios, apesar do mau desempenho da Janela de Parzen. A classificação obtida

no presente trabalho foi comparada com aquela obtida usando-se a segmentação por elipses e o desempenho se mostrou de acordo com o esperado. Dessa forma, pode-se sugerir que o tipo de segmentação aplicado, seja ela mais precisa ou mais simples, não impactou de forma significativa na eficiência desses classificadores para o problema proposto. Assim, a utilização da termografia associada a um classificador de disfunções mamárias sensível aos interesses do profissional de saúde pode ser utilizada como uma boa ferramenta na prévia detecção das anomalias mamárias antes dos métodos tradicionais (mamografia e biópsia).

Agradecimentos

Os autores agradecem a CAPES e ao CNPq pelo apoio financeiro para a pesquisa.

Referências

- [1] Araujo MC., Queiroz KFFC., Pininga MMC., Lima RCF., Santos WP. Uso de regiões elipsoidais como ferramenta de segmentação em termogramas de mama, *XXIII Congresso Brasileiro de Engenharia Biomédica*, Porto de Galinhas, Pernambuco, 2012.
- [2] Araujo, MC.; Lima, RCF.; de Souza, RMCR; Interval symbolic feature extraction for thermography breast cancer detection, *Expert Systems with Applications*, V.41, n.15, 2014.
- [3] Ruggiero MAG., Lopes VLR. *Cálculo Numérico Aspectos Teóricos e Computacionais*, Segunda Edição, PEARSON Makron.