

Artigo Original

recebido: 25/04/2000 e aceito: 20/12/2000

**Detecção de agrupamentos
de microcalcificações mamárias
através de processamento
de imagens mamográficas
com detecção semi-automática
de regiões de interesse**

*Breast clustered microcalcifications
detection by mammographic image
processing with regions of interest
semi-automatic detection*

Fátima L. S. Nunes

Depto. de Física e Informática
IFSC/USP e Depto. de Eng. Elétrica, LADI (Laboratório
de Análise e Digitalização de Imagens) – EESC/USP
Av. Trabalhador São-carlense, 400
13566-590 - São Carlos - São Paulo
e-mail: fatima@fc.unesp.br.

Homero Schiabel

Professor Associado, Departamento de Eng. Elétrica,
LADI (Lab. Análise e Digitalização de Imagens)
EESC/USP, Av. Trabalhador São-carlense, 400
13566-590 - São Carlos - São Paulo
e-mail: homero@sel.eesc.sc.usp.br.

Resumo

A necessidade de detecção precoce do câncer de mama tem levado centros de pesquisas ao desenvolvimento de esquemas de diagnóstico auxiliado por computador, pois, se diagnosticado na fase inicial do desenvolvimento, esse tipo de câncer apresenta grandes chances de cura. Este trabalho apresenta um esquema computacional para diagnosticar um dos indícios da possível existência de um tumor: os agrupamentos (“clusters”) de microcalcificações mamárias. Foi desenvolvido um método de detecção semi-automática das regiões de interesse em mamogramas digitalizados. O sistema utiliza a técnica de transformação área-ponto e possibilita a identificação dos “clusters” através de dois métodos de processamento: mascaramento e crescimento de região, além de indicar a localização dos “clusters” detectados. Os testes indicaram um acerto de 94% na identificação das aglomerações em um particular conjunto de mamogramas reais.

Palavras-chave: Diagnóstico auxiliado por computador, Mamografia, Processamento de imagens.

Abstract

The need of early detection of breast cancer has taken many research groups to developing computer-aided diagnostic schemes, since the chances of cure are greater if the disease is detected earlier. This work describes a computer scheme intended to diagnose one of the signal of a possible tumor: the clusters of breast microcalcifications. A method for regions of interest semi-automatic detection was developed for digitized mammograms. The procedure uses the area-point transformation technique and it provides the clusters detection by means of two processing methods: masking and region growing. In addition, it states the location where the detected cluster is. Tests have shown a rate of 94% of correct results in identifying clusters in a particular set of actual mammograms.

Keywords: Computer-aided diagnosis, Image processing, Mammography.

Extended Abstract

Introduction

Breast cancer detection in early stages can increase the cure possibilities. Currently, mammography is still not only the most used technique to detect breast cancer but also the one which is considered the best to provide this detection by imaging diagnostic. However, viewing structures and details of clinical interest is often a hard task to the physician since film and equipment limitations can compromise the image quality. Many computer systems and schemes have been developed since the 80's, but only in 1998, FDA has approved the first commercial clinical system developed by R2 Technology, Inc (Los Altos, CA) based on the Computer-aided Diagnosis (CAD) scheme largely tested by researchers from the University of Chicago. CAD schemes indeed make use of several processing image techniques in order to enhance the quality of digitized mammograms and to provide the detection, quantification and classification of significant signals in the image, helping the radiologist diagnosis.

Breast microcalcifications are among the main structures of interest since they can provide information about a lesion malignity. An important feature regarding these structures is when they are grouped in a cluster, which is an increase in the probability of malignity. Thus, the current work describes a computer scheme, developed with purpose of detecting clustered microcalcifications in digitized conventional mammographic images. The detection is performed according 4 steps: regions of interest (ROI) selection, ROI segmentation, area-point transformation and clusters identification. Two sets of mammographic images were tested and the scheme performance was evaluated based on medical reports associated to each case in the databases.

Material and Methods

For one of the images sets used, 16 original mammograms were digitized by a UMAX UC1260-Pro scanner, with 600 dpi and 8 bits (42 μ m and 256 gray levels). These mammograms were obtained from Hospital das Clínicas at Ribeirão Preto, SP, Brazil. The second images set was composed by other 16 digital images, with the same resolution characteristics, from the National Expert and Training Centre for Breast Cancer Screening at the University of Nijmegen, the Netherlands, obtained by Internet. For all the images, cases with clusters were accurately reported.

The first procedure was ROIs determination to be processed in each image in order to reduce the processing time. One particular region was selected if its gray level average was larger than a given general average percentage of the digital mammogram. This parameter is variable according to the images set characteristics. The original image then is reduced to $\frac{1}{4}$ of its original size; all the further procedures however are applied to the ROIs digitized with the original resolution. Thus, a thresholding filter is applied to enhance the pixels of interest, the enhanced image is subtracted from the original, and a global thresholding is applied so that only a small percentage of pixels remains white (about 2%). After this, ROIs are delimited and stored in a vector. The segmentation is performed according to 4

steps: (1) image subtraction (to enhance the signal-noise ratio); (2) global thresholding (to eliminate the image background); (3) mathematical morphology (to eliminate all signals with less than 3 pixels); (4) local thresholding (to eliminate the rest of false positive signals from the image). Finally, a procedure based on the area-point transformation previously presented by Nishikawa et al. was implemented in order to convert each structure present at the segmented image in only one pixel. The purpose is to facilitate the signals counting in order to recognize possible clusters.

The final step is the cluster detection. This was provided by two different methods: masking – where masks are sequentially positioned along the image and the quantity of signals is verified for each positioning – and region growing – where the signals located in a maximum distance from each other selected are checked and analyzed. All the images resultant from each step above are successively recorded so that, at the end, the detected cluster is stressed by a rectangle in the final image and its position and quantity of microcalcifications are shown.

Results

Testing the first images set (from HCRP), and comparing with the medical reports information, the results have shown a 94% of true-positive rate and 6% of false-positive rates – all the images had one cluster. This is correspondent to get correct detection of the cluster existence in 15 of the 16 cases. For this test, 64 ROIs were processed and evaluated. The performance obtained for testing the second images set (from the Netherlands) was worse, results being registered 72% of true-positive rate with 18% of true-negative rate (90% of correct data) against 10% of false-positive rate. 50 ROIs were processed in this second evaluation.

Conclusions

The processing performance is quite dependent on the database used for the tests, which can be denoted by comparing the resultant different efficacy percentages. Although the small number of mammograms evaluated, the large quantity of selected ROIs processed shows that the correct results level is consistent. During the tests of clusters detection, the region growing technique revealed a larger tendency to divide one cluster in two or more, depending on the signal positioning compared to the masking technique. As a consequence, in some ROIs, this method has detected two clusters against one – correct data – detected by the masking technique.

Great variations in each image set usually leads to the need of using specific parameters instead of generic ones. This increases the complexity of the development of CAD schemes and the partial processing procedures. With this feature in mind, this research has emphasis in allowing changes in key-parameters (as standard deviation and gray levels average, for example) in order to allow the processing to be adapted to the images set characteristics.

The average efficacy of 92% among the images set processed by the current scheme can be confirmed with a series of more tests with other images sets. Anyway, the results presented allow us to conclude that this procedure can be a useful tool to computerized schemes to aid the early detection of breast cancer

Introdução

A detecção do câncer de mama no estágio inicial de desenvolvimento aumenta as chances de cura da doença. A mamografia, ainda hoje, é a técnica mais utilizada para a detecção, pois registra no filme radiográfico as estruturas que podem indicar a presença de tumores malignos. Estudos indicam que a realização de mamografias periódicas em mulheres com idade superior a 40 anos, por auxiliar a detecção precoce do câncer, pode diminuir em até 40% a taxa de mortalidade devida a este tipo de câncer (Dengler, Behrens & Desaga, 1993). No entanto, a visualização das estruturas de interesse nem sempre é tarefa fácil para o médico, visto que as limitações do filme e do aparelho mamográfico podem comprometer a qualidade da imagem radiológica. Além disso, a intensidade da luz emitida pelos negatoscópios pode prejudicar a visualização da imagem, dificultando a identificação dos sinais procurados.

Apesar de vários sistemas computacionais estarem sendo desenvolvidos para a detecção precoce do câncer de mama, até 1988 nenhum deles ainda apresentava resultados suficientemente satisfatórios para aplicação clínica (Chan *et al.*, 1988a). Em 1998, O FDA (*Food and Drugs Administration*) aprovou para uso clínico comercial um sistema CAD desenvolvido pela empresa R2 Technology, Inc. (Los Altos CA) em conjunto com pesquisadores da Universidade de Chicago. Técnicas de processamento de imagem são desenvolvidas para aumentar a qualidade das imagens dos mamogramas, realçar aspectos e torná-los mais visíveis aos radiologistas, a fim de que possam identificar sinais significantes e quantificá-los (Rangayyan, 1996). Entre as estruturas que podem indicar a presença de tumores estão as microcalcificações – compostos de cálcio com tamanho inferior a 0,7mm de diâmetro e formatos variados (Egan, McSweeney & Sewell, 1980). A indicação de tumores malignos ou benignos a partir de microcalcificações pode ser baseada em certos fatores, tais como tamanho, forma e distribuição dessas estruturas na mama (Chan *et al.*, 1987; Chan *et al.*, 1988a; Chan *et al.*, 1988b; Davies & Dance, 1990). Quando as microcalcificações encontram-se agrupadas, formando os “clusters”, aumenta a probabilidade de indicação de tumores malignos. Observa-se que quando ocorre a presença de mais de 10 microcalcificações na mesma região, a probabilidade de se tratar de um carcinoma é maior que 60% (Le Gal, Chavanne & Pellier, 1984).

A detecção de microcalcificações é há muito tempo objeto de estudos de pesquisadores. W. G. Wee *et al.*

(1975) publicaram resultados de trabalhos cujo objetivo era avaliar calcificações mamográficas utilizando programas computacionais. Chan *et al.* (1987, 1988a, 1988b) desenvolveram filtros para a detecção de microcalcificações, verificando aspectos como tamanho, contraste e relação sinal-ruído. Nishikawa *et al.* (1993) desenvolveram técnicas para identificar agrupamentos de microcalcificações, através da extração de atributos como textura, distribuição espacial, tamanho e contraste. Posteriormente esses mesmos autores desenvolveram a transformação área-ponto recursiva e novos métodos para identificação das microcalcificações agrupadas. D. H. Davies e D. R. Dance (1990) propuseram algoritmos computacionais baseados em “thresholding” para a detecção de agrupamentos de microcalcificações. W. B. Richardson *et al.* (1995) propuseram a utilização de análise “wavelets” para extração de aspectos da mamografia.

O trabalho aqui descrito consistiu no desenvolvimento de um sistema computacional para a detecção de “clusters”, utilizando mamogramas digitalizados como fontes de dados. Foi adotada a definição médica que indica a existência de um “cluster” quando são verificadas 3 ou mais microcalcificações em uma área de 1 cm² (Sickles, 1986; Jin, 1993). No sistema desenvolvido a detecção dos agrupamentos é realizada em 4 fases: escolha das regiões de interesse, segmentação, transformação área-ponto e identificação das aglomerações. Foram processados dois conjuntos de mamogramas reais e armazenados os resultados e imagens após a execução de cada fase.

Determinação das Regiões de Interesse

O esquema computacional desenvolvido para a detecção de “clusters” está representado na Figura 1. O sistema foi implementado utilizando a linguagem Delphi 3.0, para execução sob plataforma Windows. O usuário seleciona a imagem desejada e executa os passos necessários para obter a identificação de “clusters”, tendo a possibilidade de alterar os parâmetros de acordo com as características particulares de cada imagem. As etapas percorridas são descritas a seguir.

Devido à alta resolução utilizada na digitalização das imagens (600 dpi – 42 microns) e, em consequência disso, à grande quantidade de pontos utilizados no processamento, é inviável – por limitações de equipamento – aplicar o sistema desenvolvido para o processamento de mamogramas inteiros. Foi desenvolvido, então, um procedimento para identificação

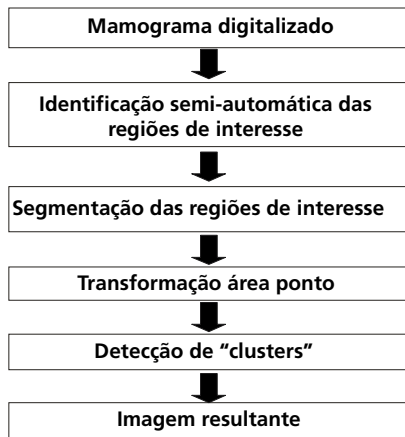


Figura 1. Esquema do sistema computacional desenvolvido para a detecção de "clusters". **Figure 1.** Developed computer scheme for clusters detection.

semi-automática das regiões de interesse a fim de que somente essas áreas fossem processadas. A constituição das microcalcificações (fundamentalmente cálcio) faz com que absorvam maior número de fótons de raios-X, quando comparadas com os demais tecidos que compõem a mama. Esse fato causa uma diferença de densidade ótica no filme mamográfico, fazendo com que as regiões do filme contendo microcalcificações fiquem mais claras após o processo de revelação. A consequência dessa propriedade é uma diferenciação em termos de nível de cinza na imagem digitalizada, pois as microcalcificações correspondem a níveis mais próximos do branco. Desse fato decorre que, em condições ideais, as regiões com microcalcificações devem apresentar maior média de níveis de cinza em relação ao restante da imagem.

Porém, geralmente as imagens não apresentam condições ideais; muitas vezes contêm variação na distribuição de intensidade ou incluem objetos que não fazem parte da estrutura da mama. Dessa forma, não basta selecionar apenas os trechos que apresentam média de níveis de cinza maior que a média geral da imagem, pois as deformações presentes influenciam no valor da média geral de cinza, tornando-o maior e impedindo que as regiões de interesse sejam identificadas. Por esse motivo, na implementação da técnica de identificação dessas regiões, atentou-se para oferecer a possibilidade de escolher qual a porcentagem da média geral de cinza que deve ser usada como parâmetro de comparação para a identificação das regiões de interesse.

No esquema desenvolvido, uma porção do mamograma digitalizado é considerada de interesse quando sua média de níveis de cinza é maior que uma

determinada porcentagem da média geral do mamograma. Através da informação de alguns parâmetros é possível definir algumas características das regiões de interesse, conforme observado no fluxograma da Figura 2. Além do tamanho da região, pode-se escolher qual porcentagem da média de níveis de cinza da imagem será utilizada para a definição de uma região como sendo de interesse. Deixar este parâmetro variável torna o programa flexível e fácil de contornar as distorções e artefatos comumente presentes nas imagens. Na Figura 3 é ilustrada uma tela do programa referente a essa etapa.

São executados, desse modo, os seguintes procedimentos:

- **redução da imagem:** a imagem é reduzida a $\frac{1}{4}$ do seu tamanho original a fim de diminuir o tempo de processamento da identificação das regiões de interesse. Salienta-se que a redução aqui efetuada é temporária, sendo considerada somente para a identificação das regiões de interesse. O procedimento de redução utiliza recursos próprios de componentes da linguagem sob a qual foi implementado o sistema computacional (Delphi). Considerando que o objetivo é comparar a média de níveis de cinza de uma porção da imagem com

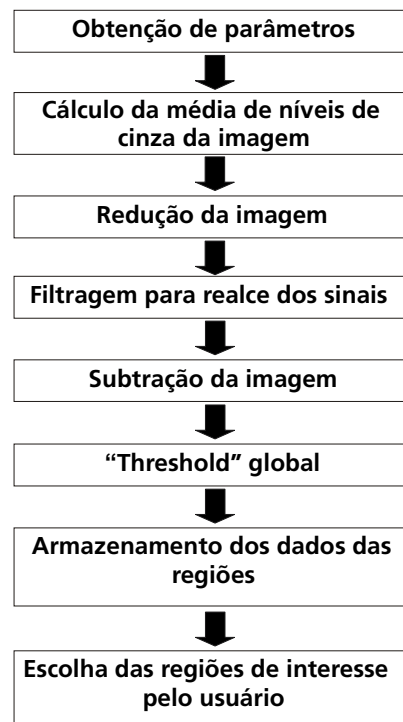


Figura 2. Fluxograma dos procedimentos realizados para identificação das regiões de interesse. **Figure 2.** Algorithm of the procedures for regions of interest identification.

a média da imagem como um todo, não há influência na resolução do processamento porque os procedimentos posteriores são executados sobre as regiões de interesse digitalizadas com a resolução inicial;

- **filtragem para realce de sinais:** um filtro que realiza um limiar local para realce dos pixels de interesse é aplicado à imagem selecionada;
- **subtração:** a imagem realçada é subtraída digitalmente da imagem original;
- **Thresholding global:** um procedimento de limiarização é aplicado à imagem, de modo a permanecer branca somente uma porcentagem de pixels pré-estabelecida. Esta porcentagem é inicialmente fixada em 2%, porém, seu valor pode ser configurado pelo usuário, através da tela inicial para a definição de parâmetros (Figura 3);
- **armazenamento dos dados das regiões:** em um vetor são armazenadas somente as regiões que possam despertar interesse, de acordo com parâmetros fornecidos previamente (“tamanho da região” e “porcentagem da média geral de nível de cinza considerada”);
- **gravação das regiões:** faz a demarcação da região de interesse na tela e espera o usuário confirmar seu desejo de armazenar a região indicada. A imagem é percorrida inteiramente e todas as regiões que possuem as características selecionadas são indicadas como sendo de interesse, a fim de que o usuário confirme se realmente deseja armazená-la. A Figura 4 mostra a tela na qual o usuário informa se deseja considerar ou descartar a região identificada como sendo de interesse. As regiões de interesse têm um tamanho fixo, de acordo com os parâmetros apresentados na Figura 3. Na Figura 5 é apresentado um mamograma origi-

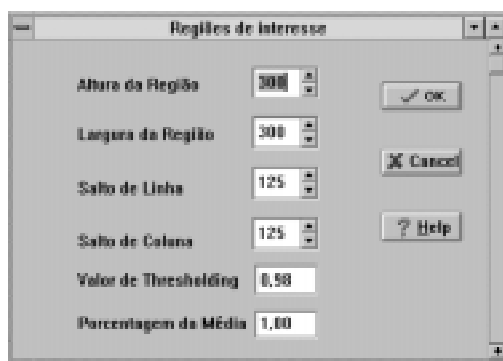


Figura 3. Tela com os parâmetros solicitados para identificação das regiões de interesse. **Figure 3.** Screen showing the parameters required for regions of interest identification.

nal digitalizado e 3 regiões de interesse extraídas dele através do esquema descrito.

Segmentação

Na segunda fase do processamento é realizada a segmentação, que consiste em destacar as estruturas de interesse (microcalcificações) e eliminar os ruídos e artefatos em cada região de interesse. O algoritmo utilizado foi proposto por Nishikawa *et al.* (1993), e sua estrutura pode ser observada na Figura 6. Os passos percorridos neste procedimentos são os seguintes:

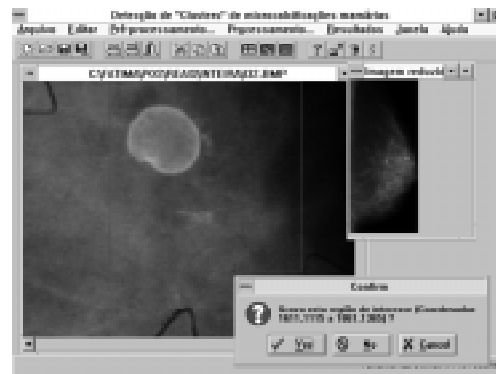


Figura 4. Tela na qual o usuário informa se deseja considerar ou descartar a região identificada como sendo de interesse. **Figure 4.** Screen for selecting or not a given region as of interest.

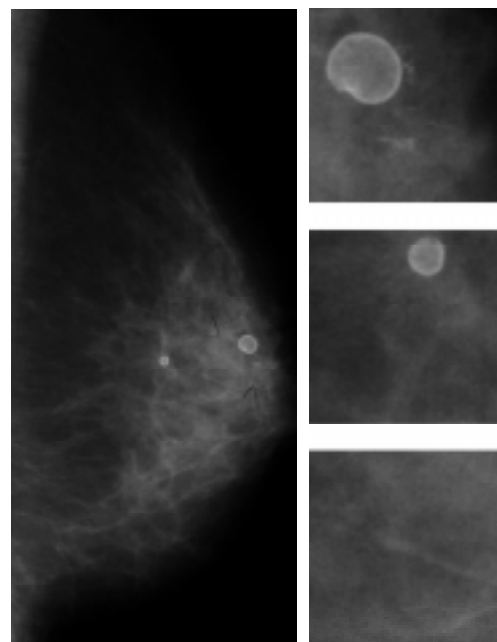


Figura 5. Mamograma original digitalizado e 3 regiões de interesse extraídas dele através da identificação semi-automática de regiões de interesse. **Figure 5.** Digitized original mammogram and 3 regions extracted from it by the semi-automated determination of regions of interest.

1. subtração da imagem: é um pré-processamento com o objetivo de incrementar a relação sinal-ruído; consiste em subtrair uma imagem com os sinais realçados de uma imagem com os sinais suavizados;
2. “thresholding” global: consiste em aplicar a técnica de limiarização, baseando-se em parâmetros pré-determinados pelo usuário; esse processamento deixa na imagem somente uma porcentagem dos pixels com os maiores níveis de cinza; o objetivo é eliminar o fundo da imagem;
3. morfologia matemática – operação de erosão: tem o objetivo de eliminar da imagem todos os sinais com menos de três pixels na sua composição, visto que, provavelmente, relacionam-se a ruídos;
4. “thresholding” local: consiste em eliminar os sinais falsos-positivos restantes da imagem, através da verificação do contraste de cada estrutura em relação à sua vizinhança.

Para que essas etapas sejam realizadas, o valor de limiar global (porcentagem dos pixels que deverão permanecer brancos durante o procedimento de limiarização global) e o desvio-padrão a serem utilizados devem ser inseridos como dados de entrada antes do processamento. O valor do desvio-padrão será somado à média de nível de cinza da imagem para determinar se um pixel ficará claro ou escuro na imagem binária resultante. Por exemplo: para valor igual a 3 - se o nível de cinza de um pixel for menor que a soma da média + 3, então este pixel será transformado em preto; caso contrário, será transformado em branco. Na Figura 7 observa-se uma região de interesse original e a mesma região após o processo de segmentação.

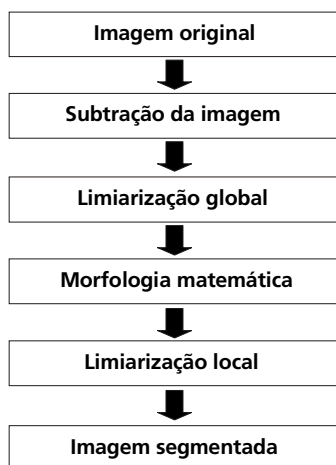


Figura 6. Fluxograma com o algoritmo utilizado para segmentação das regiões de interesse. **Figure 6.** Algorithm used for regions of interest segmentation.

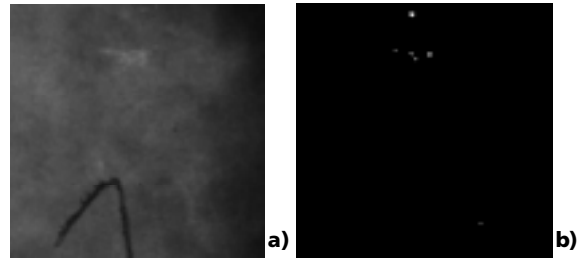


Figura 7. (a) Região de interesse original; (b) mesma região após o processo de segmentação. **Figure 7.** (a) Original region of interest; (b) same region after segmentation procedure.

Transformação Área-Ponto

Com base na idéia da transformação área-ponto, apresentada por *Nishikawa et al.* (1993), foi desenvolvido um procedimento com o objetivo de transformar cada estrutura de interesse presente na imagem segmentada em um único pixel. A transformação é realizada da seguinte maneira: um máscara de 3 x 3 pixels é passada sobre a imagem começando do canto superior esquerdo. Para cada posição, é contabilizado o número de pixels com nível de cinza superior ao do fundo da imagem. Havendo dois ou mais pixels com essa característica, todos os pixels sob a máscara são ajustados para um valor igual ao do fundo, menos o pixel do canto inferior direito do núcleo, que é ajustado para o valor máximo. O núcleo é então movido 2 pixels para a direita e o processo é repetido. O processo salta de 2 em 2 linhas até atingir o fim da imagem. Um exemplo do andamento da técnica pode ser observado na Figura 8, onde os retângulos escuros representam os pixels constituintes de uma microcalcificação. A cada passo, os pixels têm seu valor alterado (na figura, transformam-se em retângulos hachurados) e, quando o processo termina, a microcalcificação transforma-se em um único pixel. A finalidade desse processo é facilitar a contagem das estruturas para a identificação dos pos-

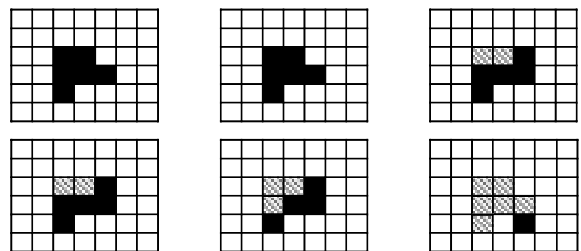


Figura 8. Exemplo esquemático da “transformação recursiva área-ponto”. **Figure 8.** Example of the recursive area-point transformation.

síveis “clusters” da imagem. Testes com variação no tamanho das máscaras utilizadas e nos saltos de linhas e colunas necessários levaram-nos a confirmar o bom funcionamento desta rotina, tornando-a um passo decisivo para a identificação dos “clusters”.

Um exemplo da técnica de transformação área-ponto pode ser observado na Figura 9: em (a) vê-se uma imagem binarizada, resultante do processo de segmentação e, em (b), está a imagem resultante após a aplicação da transformação área-ponto, na qual cada estrutura é transformada em um único ponto. Após a submissão da região de interesse ao procedimento descrito neste item, a imagem encontra-se na situação ideal para a contagem dos sinais a fim de identificar os possíveis “clusters” existentes, o que é realizado na fase seguinte.

Detecção de Agrupamentos de Microcalcificações

A detecção de microcalcificações agrupadas (“clusters”) é o processo que efetivamente identifica os agrupamentos que porventura existam na imagem, indicando também a sua localização e a quantidade de microcalcificações neles contidas. Dois métodos foram implementados para essa finalidade: um utilizando máscaras (aqui denominado “mascaramento”) e outro utilizando a técnica de crescimento de região. Em ambos, o usuário pode alterar os parâmetros que indicam a quantidade mínima de microcalcificações (o padrão utilizado são 3 microcalcificações em uma área de 1 cm²) necessárias para um agrupamento ser detectado e o tamanho da área considerada para a detecção.

O método de mascaramento consiste em posicionar “janelas” seqüencialmente ao longo da imagem e verificar em cada posicionamento a quantidade de

sinais existentes. Se forem contados mais sinais do que uma quantidade mínima pré-estabelecida, um “cluster” é identificado. Com os parâmetros fornecidos pelo usuário, o método de “mascaramento” posiciona a janela no início da imagem (linha 0, coluna 0) e, para cada posicionamento da máscara, verifica se há uma quantidade igual ou maior que a quantidade mínima de sinais pré-estabelecida. Caso esta condição seja satisfeita, incrementa o contador de “clusters”, armazena a posição de cada sinal e apaga os sinais localizados sob a máscara para que não sejam contados repetidamente. Ocorre, então, o deslocamento da máscara (horizontal e verticalmente) e o processo se repete até que toda a imagem tenha sido percorrida.

Após a execução desse procedimento, são aplicados alguns métodos para corrigir possíveis falhas ocorridas durante o processo: (a) agrupamento de sinais isolados – a fim de agrupar sinais que ocasionalmente tenham ficado isolados durante a aplicação do algoritmo de “mascaramento”; para isso, é considerada a menor distância entre o sinal isolado e os sinais presentes em todos os “clusters”; se o valor resultante for menor ou igual ao tamanho da área considerada, o sinal é incluído no “cluster”; (b) agrupamento de “clusters” próximos – a fim de minimizar o efeito de divisão de um “cluster” em dois ou mais conjuntos durante a detecção; os “clusters” são testados dois a dois, sendo calculadas as distâncias entre todos os seus sinais e armazenada a menor delas; se esse valor for menor ou igual à distância mínima estabelecida nos parâmetros, os agrupamentos envolvidos são transformados em uma única aglomeração.

O método de crescimento de região, por outro lado, consiste em verificar quais sinais estão localizados a uma distância máxima de um outro sinal selecionado. As estruturas que atendem a esse critério de distância são alocadas em um mesmo “cluster”. Para tanto, o procedimento armazena a posição de todos os sinais encontrados e, para cada sinal que ainda não foi considerado como parte de um “cluster”, realiza o seguinte procedimento: incrementa o contador de “clusters” e inclui no “cluster” atual todos os sinais com distância menor que uma distância máxima pré-estabelecida, eliminando-os simultaneamente da imagem. Após esta fase, para cada “cluster” identificado, é verificado se a quantidade de sinais alocados nele é menor que uma quantidade mínima pré-estabelecida; se isto acontecer, o “cluster” é eliminado e o contador de “clusters” é diminuído de uma unidade.

A fim de tornar mais visível a identificação dos “clusters” e armazenar os dados resultantes da

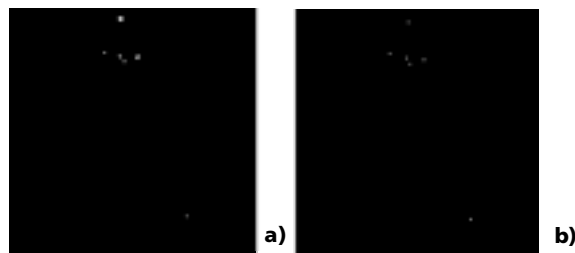


Figura 9. (a) Imagem binarizada, resultante do processo de segmentação; (b) imagem resultante após a aplicação da transformação área-ponto, na qual cada estrutura é transformada em um único ponto. **Figure 9.** (a) Binary image, after the segmentation procedure; (b) image resulting from the area-point transformation application, in which each structure is converted in a unique pixel.

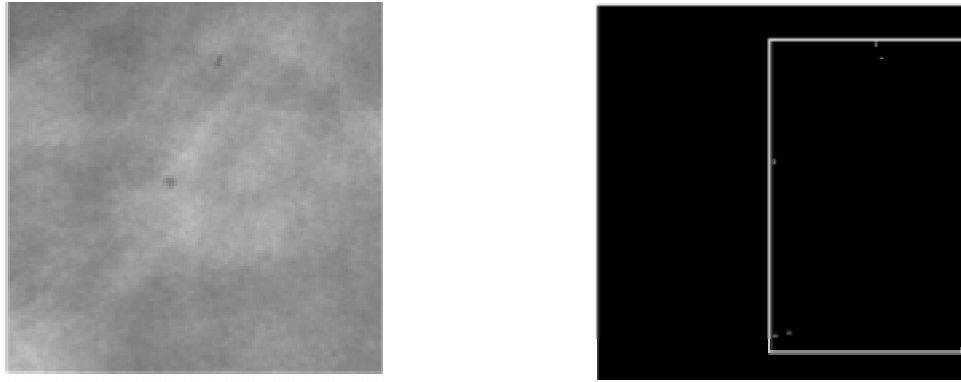


Figura 10. Exemplo de “cluster” identificado e demarcado em uma região de interesse. **Figure 10.** Example of identified cluster which was marked inside a region of interest.

detecção, procedem-se três novas etapas: (a) atribuição de cor aos sinais; (b) localização dos “clusters”; e (c) demarcação dos agrupamentos. No primeiro caso, aos sinais de cada “cluster” identificado é atribuída uma cor diferente a fim de diferenciá-los dos demais “clusters”. No segundo, com os dados armazenados durante o processo de identificação das regiões de interesse, cada agrupamento detectado é identificado em um dos quatro quadrantes da imagem: superior esquerdo, superior direito, inferior esquerdo ou inferior direito. Esta localização permite comparar também o resultado do processamento com o laudo médico. E, finalmente, em termos da terceira etapa, em torno de cada “cluster” é desenhado um retângulo que o contém, com a finalidade de facilitar a visualização. Um exemplo de um “cluster” demarcado na imagem é exibido na Figura 10.

Durante todas as fases da detecção dos agrupamentos de microcalcificações, são armazenadas informações em um Banco de Dados projetado para abrigar os dados próprios do sistema, cuja finalidade é permitir ao interessado consultas posteriores para acompanhamento dos casos processados. Assim, para cada região de interesse extraída, são gravadas informações que tornam possível seu relacionamento com o mamograma original. Após o processamento de cada região, armazena-se a quantidade de sinais detectados. Ao final do processamento é possível identificar todos os agrupamentos encontrados em um mamograma e suas respectivas localizações, como pode ser observado na Figura 11.

Bases de Dados

A fim de verificar o desempenho do sistema desenvolvido foram utilizados dois conjuntos de imagens a partir de mamogramas reais: o primeiro consistia em

16 mamogramas obtidos no Hospital das Clínicas de Ribeirão Preto, todos acompanhados com laudo médico. Os filmes foram digitalizados em um digitalizador UMAX UC1260-Pro, com resolução de 600 dpi e 256 níveis de cinza. O segundo conjunto era composto por 16 imagens obtidas do National Expert and Training Centre for Breast Cancer Screening e Department of Radiology at the University of Nijmegen, na Holanda, através da Internet, sendo fornecidas as localizações e o tamanho dos “clusters” com precisão. Os valores de resolução espacial e de contraste para esse segundo conjunto de imagens eram os mesmos do primeiro.

O processamento das imagens foi executado em computadores do tipo IBM PC Pentium, 133 MHz, com 32 Mb de RAM. Os resultados foram gravados em arquivos durante a execução dos programas e, posteriormente, comparados às informações conhecidas previamente sobre os laudos dos mamogramas.

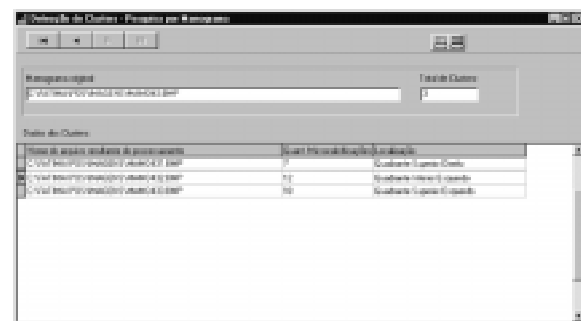


Figura 11. Exemplo da tela do banco de dados que permite identificar todos os agrupamentos encontrados em um mamograma e suas respectivas localizações.

Figure 11. Database screen example, allowing to determine all the clusters found in a mammogram as well as their respective locations.

Resultados e Discussões

Na Tabela 1 são apresentados os resultados do processamento do primeiro conjunto de imagens. Para tal, foram selecionadas regiões de interesse que poderiam conter agrupamentos de microcalcificações e outras que provavelmente não apresentavam “clusters”, a fim de verificar o desempenho do esquema na identificação de agrupamentos verdadeiramente existentes e na não identificação dos falsos. No total, foram processadas 64 regiões de interesse com 250 linhas x 250 colunas. Os testes do esquema foram executados variando-se o valor do parâmetro “desvio-padrão” durante a segmentação. Notou-se que, nesse caso, o aumento do valor desse parâmetro piorou o desempenho do sistema, aumentando o número de diagnósticos falsos-negativos. O melhor resultado observado para o desempenho do esquema foi uma taxa de 94% de acerto no diagnóstico de “clusters”, sem registros de qualquer falso-negativo, ao se utilizar um valor de 3,0 para o parâmetro desvio-padrão.

O desempenho obtido no processamento do segundo conjunto de mamogramas pode ser observado na Tabela 2. Foram processadas 50 regiões de interesse com 250 linhas x 250 colunas. O melhor resultado, nesse caso, foi obtido usando-se o valor de 3,3 para o desvio-padrão, quando foram detectados corretamente 90% dos “clusters” existentes (soma dos resultados verdadeiros-positivos e verdadeiros-negativos), também sem nenhum falso-negativo.

Alguns pontos importantes foram observados durante o processamento:

- **identificação das regiões de interesse** - algumas vezes o posicionamento da máscara para gravação de região de interesse ocorria na metade do “cluster” ou do trecho previamente identificado; a alteração dos parâmetros relativos ao tamanho da região e salto de linhas e colunas resolveu esse problema;
- **diferença da distribuição de intensidade** - este fator é um problema que traz grandes dificuldades durante o processamento. A distribuição de intensidade que chega no filme não é uniforme em todo o campo de radiação. Isso faz com que, algumas vezes, existam porções da imagem mais claras do que as partes onde estão as microcalcificações. Na detecção semi-automática de regiões de interesse, esse dado é um fator importante, pois pode ocorrer a não identificação de uma região em decorrência do fato da média de níveis de cinza daquela porção ser menor que a média de cinza da imagem. Para minimizar este problema, tornamos variável o parâmetro “porcentagem da média de cinza”, que indica qual a porcentagem da média geral deve ser utilizada para comparação e consequente detecção das regiões de interesse. Dessa forma, se uma determinada região não for identificada como sendo de interesse por estar mais escura do que o restante da imagem, o processamento pode ser refeito, atribuindo-se um valor menor a esse parâmetro.
- **presença de objetos** - algumas vezes, no exame mamográfico, são utilizados objetos, geralmente

Tabela 1. Resultado da detecção de “clusters” do primeiro conjunto de imagens. **Table 1:** Result from clusters detection for the first images set.

Desvio-padrão da segmentação	Total de imagens processadas	Total de diagnósticos verdadeiros-positivos		Total de diagnósticos falsos-positivos		Total de diagnósticos verdadeiros-negativos		Total de diagnósticos falsos-negativos	
		Quant.	%	Quant.	%	Quant.	%	Quant.	%
3,0	16	15	94%	1	6%	0	0%	0	0%
3,5	16	11	69%	1	6%	0	0%	4	25%

Tabela 2. Resultado da detecção de “clusters” do segundo conjunto de imagens. **Table 2:** Result from the clusters detection for the second images set.

Desvio-padrão da Segmentação	Total de regiões processadas	Total de diagnósticos verdadeiros-positivos		Total de diagnósticos falsos-positivos		Total de diagnósticos verdadeiros-negativos		Total de diagnósticos falsos-negativos	
		Quant.	%	Quant.	%	Quant.	%	Quant.	%
3,5	50	30	60%	5	10%	11	22%	4	8%
3,3	50	36	72%	5	10%	9	18%	0	0%

radiopacos (marcadores, em geral), com a finalidade de orientar o observador. Na maioria das vezes, a imagem desses objetos apresenta níveis de cinza maiores que no restante do filme (tendendo para o branco), o que influencia no cálculo da média geral de níveis de cinza, prejudicando a identificação das regiões de interesse. Com isso, as regiões de interesse são identificadas somente se for utilizado o valor 0 de desvio padrão. Isto causa incômodo, pois o usuário deve desprezar regiões sequencialmente até ser atingida aquela desejada. Para eliminar esse problema, deve-se identificar previamente as fronteiras da mama e processar somente a porção que faça parte dela na imagem.

- **contraste da imagem** - o contraste corresponde à diferença dos níveis de absorção de raios-X pelos diversos tecidos que compõem a mama. A solução mais efetiva para melhorar no nível de contraste foi alterar alguns parâmetros durante o processo de digitalização – como, por exemplo, o nível de contraste e brilho –, o que permitiu a obtenção de imagens digitalizadas em condições melhores do que as condições originais de observação dos filmes mamográficos.
- **tempo de execução** - observou-se que o tempo de processamento para a detecção de “clusters” aumenta consideravelmente com o aumento do tamanho das regiões de interesse, sendo que o processo de segmentação é o que mais contribui para este fato. Para diminuir este problema, procuramos trabalhar com o menor tamanho possível para as regiões de interesse.
- **transformação área-ponto** - o problema encontrado durante essa fase foi quando houve transformação de um único sinal em mais de um pixel, sendo que esse fato ocorre quando a forma da microcalcificação é próxima de um “Y” ou “V”. Se houver mais de três pixels entre os dois braços do Y ou V, o sinal será transformado em dois pixels e não em um só. Mas tal fato não constitui necessariamente uma desvantagem, visto que microcalcificações com esses formatos provavelmente constituem indício de tumor maligno e, portanto, é melhor mantê-las do que eliminá-las.

Durante a realização dos testes com os dois métodos de detecção de “clusters” (“mascaramento” e “crescimento de região”) notou-se que o algoritmo de “crescimento de região” apresentou uma tendência maior em dividir um “cluster” em dois ou mais, dependendo do posicionamento dos sinais. Em consequência disso, em algumas regiões de interesse

houve a detecção de dois “clusters” através do método de “crescimento de região”, enquanto o método de “mascaramento” detectou apenas uma aglomeração. O método de “mascaramento” realiza a identificação de agrupamentos verificando se em uma determinada área está presente uma quantidade mínima de sinais. A máscara utilizada percorre a imagem horizontalmente e, ao chegar no final de cada trecho de linhas, executa um salto verticalmente, reiniciando a verificação em um novo trecho de linhas. A cada posicionamento da máscara, os sinais detectados são eliminados da imagem para que não sejam contabilizados repetidamente. Estes procedimentos explicam a tendência desse método em detectar, de uma só vez, o “cluster”: como o tamanho das regiões de interesse (250 linhas x 250 colunas) é próximo ao tamanho das áreas verificadas (236 linhas x 236 colunas), geralmente o “cluster” é detectado na primeira iteração; como todos os sinais identificados como participantes do “cluster” são eliminados, pode acontecer que, devido ao tamanho da área, na próxima iteração não existam sinais suficientes para compor o “cluster”.

O método de crescimento de região percorre a imagem fazendo verificações para cada sinal de forma individual. Para cada um dos pixels existentes, verifica quais sinais estão localizados a uma distância máxima dele, englobando-os como participantes do mesmo agrupamento. A implementação do programa estabeleceu a fórmula da distância “City Block” (Gonzalez & Wintz, 1987) para essa verificação. Somente após a verificação da distância de todos os sinais a um determinado pixel e a eliminação dos pixels que atendem o critério de distância máxima, é que se parte para a próxima iteração, verificando-se o próximo sinal remanescente na imagem. Esse algoritmo faz com que sejam agrupados aqueles sinais mais próximos de um sinal considerado, no sentido diagonal. Os demais sinais são desprezados e poderão ser considerados como parte de outros “clusters” nas iterações posteriores. Com isso, pode ocorrer de dois sinais estarem bem próximos, mas serem alocados em “clusters” diferentes, simplesmente pela ordem de execução do algoritmo.

A diferença na forma de agrupamento dos sinais é o que produz resultados diferentes na identificação de “clusters”, fazendo com que eles variem de acordo com o posicionamento dos sinais. Na Figura 12 é apresentado um exemplo completo do processamento, incluindo a detecção dos “clusters” através dos dois métodos desenvolvidos. Nessas imagens é

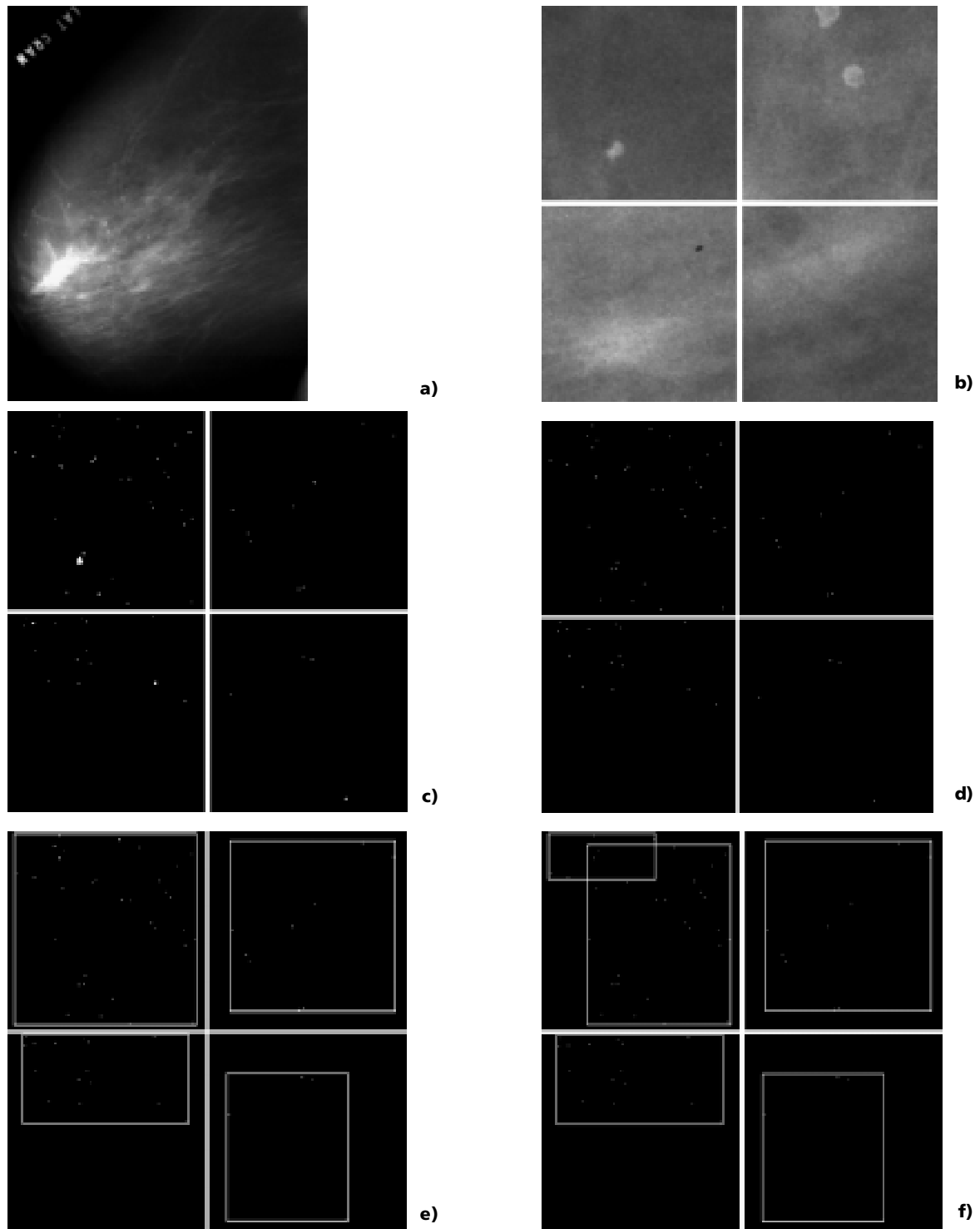


Figura 12. Exemplo completo do processamento, incluindo a detecção dos "clusters" através dos dois métodos desenvolvidos. **(a)** Mamograma original; **(b)** regiões de interesse selecionadas; **(c)** regiões de interesse após segmentação; **(d)** regiões de interesse após transformação área-ponto; **(e)** regiões de interesse após identificação de "clusters" pelo método de mascaramento; **(f)** regiões de interesse após identificação de "clusters" pelo método do crescimento de região. **Figure 12.** Complete example of the processing scheme, including the clusters detection from the both developed methods. **(a)** Original mammogram; **(b)** selected regions of interest; **(c)** regions of interest after segmentation; **(d)** regions of interest after area-point transformation; **(e)** regions of interest after clusters determination by the masking method; **(f)** regions of interest after clusters determination by the region growing method.

possível observar em quais casos os métodos apresentaram diagnósticos diferentes.

Conclusões

A utilização de técnicas de processamento de imagens para identificar agrupamentos de microcalcificações não se limita à aplicação de técnicas básicas. Ao contrário, exige adaptações específicas do processo e adequação individual em cada conjunto de imagens processado. Fatores como contraste baixo, não uniformidade na distribuição de intensidade da imagem e pequeno tamanho dos sinais são itens que dificultam o diagnóstico, exigindo grande atenção durante o desenvolvimento. Além disso, é difícil saber de antemão os melhores valores para os parâmetros ideais para cada grupo de imagens, uma vez que as características apresentam grandes variações.

Percebemos que o desempenho do sistema depende muito do conjunto de imagens de teste, como já havia sido observado por Nishikawa *et al.* (1994). No primeiro conjunto de imagens de mamogramas reais, o sistema atingiu 94% de eficiência, enquanto que no segundo conjunto a eficiência alcançou o índice de 90%. Os resultados obtidos foram contabilizados a partir da taxa de acerto (detecção ou não detecção correta) nas regiões de interesse. Apesar do número de mamogramas utilizados nos testes aqui citados ser pequeno – devido à dificuldade de obtenção de imagens adequadas – foram selecionadas várias regiões de interesse de cada mamograma a fim de que o esquema desenvolvido fosse submetido a um teste mais efetivo.

É importante que um sistema para auxílio no diagnóstico precoce do câncer de mama forneça um baixo índice de diagnósticos falsos-positivos e, principalmente, falsos-negativos. O primeiro pode levar a biópsias desnecessárias e o segundo pode induzir o médico a não dar a devida atenção a uma área onde pode existir um início de câncer. Neste sentido, o sistema aqui proposto teve um bom desempenho, visto que o tamanho do pixel (42 μm) permite identificar estruturas de até 0,13 mm (estruturas com tamanhos menores são eliminadas durante a segmentação, que exclui da imagem sinais com 3 ou menos pixels), que podem não ser percebidas a olho nu no mamograma, particularmente se se tratar de mama com densidade mais acentuada.

Como principais vantagens do sistema desenvolvido, podemos inferir: diminuição do tempo de processamento através da detecção semi-automática de regiões de interesse, parâmetros variáveis e

portabilidade. As desvantagens que pudemos registrar foram essencialmente: necessidade de conhecimento sobre o assunto para utilizar o sistema, considerando a grande quantidade de parâmetros variáveis, isto é, o usuário deve saber o significado dos parâmetros para que o sistema seja configurado adequadamente para o conjunto de imagens a ser processado e dificuldade em processar imagens muito grandes. Esses problemas, no entanto, devem ser tratados com treinamento e incremento de “hardware”, respectivamente.

Desenvolver um sistema que forneça um diagnóstico seguro e preciso para a detecção precoce do câncer de mama não é tarefa simples. As grandes variações nas características de cada imagem impedem o emprego de parâmetros genéricos. Esse fato faz com que os sistemas desenvolvidos para a área sejam complexos e, em muitos casos, de difícil manuseio. A solução, nesta pesquisa, foi o desenvolvimento de um sistema em que boa parte dos valores utilizados no processamento de imagens pode ser modificada, possibilitando a alteração dos parâmetros para cada imagem submetida ao processamento. Se forem processadas imagens pertencentes a um mesmo conjunto, isto é, com as mesmas características de aquisição, tais parâmetros podem ser aplicados ao conjunto todo e não somente a uma imagem. O fato de deixar valores passíveis de alterações pelo usuário não torna difícil a sua utilização, uma vez que todos os parâmetros apresentam valores iniciais já estabelecidos a partir de estudos prévios.

O esquema desenvolvido registrou um nível de eficiência médio de 92% para o processamento dos dois conjuntos de mamogramas reais testados. Acreditamos que esses resultados poderiam ser confirmados através de testes mais exaustivos com novos conjuntos de imagens. O sistema computacional descrito fornece informações com o objetivo de auxiliar os médicos na formação da opinião sobre a possível existência do câncer de mama no paciente e permite, através dos registros dos processamentos realizados, um acompanhamento de cada caso, quantificando os sinais identificados. Dessa forma, o esquema aqui apresentado pode-se constituir numa útil ferramenta em esquemas computadorizados de auxílio à detecção precoce do câncer de mama.

Agradecimentos

Os autores agradecem o apoio financeiro à FAPESP e ao Hospital das Clínicas de Ribeirão Preto, que possibilitou a utilização de um dos conjuntos de mamogramas para os testes.

Referências

- Chan, H-P.; Doi, K.; Galhotra, S.; Vyborny, C.J.; MacMahon, H.; Jokich, P.M. (1987). "Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography". *Med. Phys.*, v. 14, p. 538-548.
- Chan, H-P.; Doi, K.; Vyborny, C.J.; lam, K-L.; Schmidt, R.A. (1988a). "Computer-aided detection of microcalcifications in mammograms. Methodology and preliminary clinical study". *Original Investigations*.
- Chan, H-P.; Doi, K.; Lam, K-L.; Vyborny, C.J.; Schmidt, R.A.; Metz, C.E. (1988b) "Digital characterization of clinical mammographic microcalcifications. Application in computer-aided detection". *SPIE Med. Imaging II p. Image Formations Detection Processing and Interpretation*, v. 914, p. 591-593.
- Davies, D.H.; Dance, D.R. (1990) "Automatic computer detection of clustered calcification in digital mammograms". *Phys. Med. Biol.*, v. 35, p. 1111-1118.
- Dengler, J.; Behrens, S.; Desaga, J.F. (1993) "Segmentation of microcalcifications in mammograms". *IEEE MI*, v. 12, p. 634-642.
- Egan, R.L.; McSweeney, M.B.; Sewell, C.W. (1980) "Intramammary calcifications without an associated mass in benign and malignant diseases". *Radiology*, v. 137, p.1-7.
- Gonzalez, R.C.; Wintz, P. (1987) *Digital Image Processing*. Reading: Addison-Wesley.
- Jin, H-R.; Kobatake, H. (1993) "Extraction of microcalcifications from mammograms using morphological filter with multiple structuring elements". *Systems and Computers in Japan*, v. 24, n. 11, p. 66-74.
- Le gal, M.; Chavanne, G.; Pellier, D. (1984) "Valeur diagnostique des microcalcifications groupées découvertes par mammographies". *BULL Cancer*, v. 71, p. 57-64.
- Nishikawa, R.M.; Giger, M.L.; Doi, K.; Vyborny, C.J.; Schmidt, R.A. (1993) "Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms". *State of the Art in Digital Mammographic Image Analysis, World Scientific Publishing*.
- Nishikawa, R.M.; Giger, M.L.; Doi, K.; Metz, C.E.; Yin, F-F.; Vyborny, C.L.; Schmidt, R.A. (1994) "Effect of case selection on the performance of computer-aided detection schemes". *Medical Physics*, v. 21, n. 2, p. 265-269.
- Rangayyan, R.M. (1996) "Digital image processing techniques for computer-aided diagnosis of breast cancer". *Anais do III Fórum Nacional de Ciência e Tecnologia em Saúde*, v. 2, p. 831-843.
- Richardson JR., W.B. (1995) "Applying wavelets to mammograms". *IEEE Eng. Med. Biol.*, v. 14, p. 551-560.
- Sickles, E. A. (1986) "Breast calcifications: mammographic evaluation". *Radiology*, v. 160, p. 289-293.
- Wee, W.G.; Moskowitz, M.; Chang, N.C.; Ting, T.C.; Pemmeraju, S. (1975) "Evaluation of mammographic calcifications using a computer program". *Radiology*, v. 116, p.717-720.