

HqRF: Novos Descritores de Forma para Auxiliar no Diagnóstico de Câncer de Mama

Walter Alexandre A. de Oliveira, Denise Guliato

¹Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Av. João Naves de Ávila, 2121 – Uberlândia – MG – Brasil
INCT-MACC Medicina Assistida por Computação Científica

{walter,guliato}@facom.ufu.br

Abstract. *Malignant breast tumors and benign masses appear in breast cancer exams with different shape characteristics: the former usually have rough, spiculated, or microlobulated contours, whereas the latter commonly have smooth, round, oval, or macrolobulated contours. Visual features that characterize shape roughness can assist in distinguishing between tumor types. In this work we propose two new approaches to describe 2D and 3D shapes based on Hilbert curves. The descriptors are applied to content-based images/volumes retrieval. The methods were evaluated using a set of breast contours and a 3D object synthetic database. As the experimental evaluations show, we achieved a mean average precision of 1.00 (to 2D shapes) and 0.99 (to 3D shapes).*

Resumo. *Tumores malignos e lesões benignas aparecem em exames radiológicos com diferentes características: os primeiros geralmente apresentam contornos grosseiros, espiculados ou microlobulados, enquanto os últimos comumente têm contornos suaves, arredondados, ovalados ou macrolobulados. Características visuais que descrevem a aspereza da forma podem auxiliar na distinção entre os tipos de tumor. Neste trabalho são propostos novos métodos para descrever formas 2D e 3D, baseados na curva de Hilbert. Os descritores são aplicados para a recuperação de imagens/volumes por conteúdo. O método foi avaliado usando uma base de dados de contornos de mama e uma base de dados sintética com objetos 3D. A avaliação dos experimentos mostram uma precisão média de 1.00 (para formas 2D) e 0.99 (para formas 3D).*

1. Introdução

O câncer de mama é o tipo mais comum de câncer entre mulheres no mundo todo, chegando a cerca de 16% [WHO 2010]. A Organização Mundial de Saúde (OMS) estima que 460 mil mulheres foram vítimas de câncer de mama durante o ano de 2008. Ainda de acordo com a OMS, cerca de 70% dos casos de câncer de mama ocorrem em países pobres ou em desenvolvimento. Em regiões como a América do Norte, a taxa de sobrevivência chega a 80%, graças, principalmente, à detecção precoce da lesão de mama [WHO 2010]. No entanto esta modalidade de exame apresenta limitações para detectar tumores malignos com achados radiológicos ocultos em mamas densas ou em mamas com implante, devido à sobreposição de tecidos causada pela projeção bidimensional da mama. Exaustivos esforços têm sido feitos nos últimos 15 anos para incluir imagens de ressonância magnética (MRI) e ultrasonografia como modalidades complementares para melhorar a interpretação de imagens de mama [Turnbull et al. 2010, Chen et al. 2006].

Independente da modalidade, lesões benignas e malignas podem ser diferenciadas pelas características da sua forma. É bem estabelecido que a maioria das lesões benignas tem contornos arredondados ou ovais, textura interna relativamente homogênea e margens bem definidas. As lesões malignas tipicamente apresentam contornos mal definidos, rugosos ou espiculados e textura interna heterogênea. Entretanto, casos atípicos podem ocorrer, o que torna o diagnóstico de câncer de mama uma tarefa difícil e sujeita a erros [Gupta et al. 2006, American College of Radiology, Reston, VA 1998].

Com o objetivo de auxiliar o radiologista na tomada de decisão, vários sistemas de apoio ao diagnóstico (CADx) de câncer de mama têm sido propostos no últimos anos. Uma das tarefas realizadas por um sistema CADx é analisar, objetivamente, as características do contorno das lesões presentes no exame radiológico para classificá-las como lesões benignas ou malignas. Diversos trabalhos têm sido desenvolvidos para caracterizar os contornos de lesões de mama de forma automática. Rangayyan et al. em [Rangayyan et al. 2000] propuseram métodos para obtenção de índices de espículos (SI) e de concavidade (F_{cc}), baseados em modelos poligonais para caracterização do contorno da lesão. Guliato et al. em [Guliato et al. 2008b] propuseram métodos para obtenção de índice de espículo (SI_{TA}), índice de concavidade (CX_{TA}) e índice de dimensão fractal (FD_{TA}) baseados na assinatura da função tangente (*Turning Angle Function*) de um dado contorno. Os métodos foram aplicados à classificação de lesões de mama, tendo alcançado altas taxas de precisão ($AUC = 0.93$). Caballero et al. em [R.G.-Caballero et al. 2007] utilizaram a análise de componentes independentes (*Independent Component Analysis - ICA*) para obter um conjunto de coeficientes usados na classificação de mamogramas com adição de ruído. A classificação foi realizada usando redes neurais e imagens da DDSM (*Digital Database for Screening Mammography*) [Heath et al. 1998]. Weiqiang et al. em [Weiqiang et al. 2008] combinaram diversos índices para representar as características do contorno de lesões de mama. Os índices utilizado no trabalho incluem: índice de espículo, índice de lóbulo, dimensão fractal e circularidade do contorno. Os autores relataram ter alcançado bons resultados na classificação de 93 imagens obtidas na base de dados do MIAS (*Mammographic Image Analysis Society*) [Suckling et al. 1994]. Em [Guliato et al. 2010] os autores apresentam um novo descritor de silhueta do contorno, utilizando a curva de Hilbert, denominado *2Hq Feature*. A silhueta de um contorno é definida como uma imagem em que o interior do contorno é preenchido pela cor branca, e o fundo, pela cor preta. O *2Hq Feature* foi utilizado na classificação de 111 contornos de lesão de mama [Alto et al. 2005]. A classificação das lesões como benignas ou malignas foi realizada usando uma rede neural. Os experimentos mostraram alta taxa de precisão ($AUC = 0.99$). Apesar dos resultados obtidos, o *2Hq Feature* proposto por [Guliato et al. 2010] apresenta alta dimensionalidade e foi desenvolvido para caracterizar silhuetas bidimensionais.

Neste artigo propomos uma modificação do descritor proposto por [Guliato et al. 2010], de tal forma a reduzir a sua dimensionalidade e será referido neste trabalho como *2HqRF*. Propomos também estender o *2HqRF* para descrever silhuetas tridimensionais, denominado *3HqRF*. Uma vez obtidas as características que descrevem as silhuetas 2D ou 3D das lesões de mama, estas podem ser utilizadas em tarefas de classificação de padrões ou de recuperação de imagens por conteúdo. Neste trabalho exploraremos o uso dos descritores *2HqRF* e *3HqRF* (*HqRF* para simplificar) para tarefas de recuperação de imagem, utilizando o método KNN (*K-Nearest Neigh-*

bors). Os experimentos para a avaliação dos resultados obtidos pelo descritor $2HqRF$ foram realizados com a mesma base de dados utilizada em [Guliatto et al. 2008b]. Para avaliar os resultados obtidos pelo descritor $3HqRF$, criamos uma base de dados sintética que simula lesões em silhuetas arredondas ou espiculadas.

O restante deste artigo é organizado da seguinte forma: na Seção 2 são apresentados os conceitos teóricos necessários ao entendimento do trabalho; na Seção 3, o descritor proposto $2HqRF$ e sua extensão para o espaço 3D são apresentados; a Seção 4 descreve as bases de dados utilizadas; os experimentos e uma discussão dos resultados obtidos encontram-se na Seção 5; finalmente, a Seção 6 traz a conclusão do trabalho.

2. Conceitos Teóricos

Nesta seção são resumidos os conceitos básicos necessários para o entendimento deste trabalho, apontando suas características de maior interesse.

2.1. Curva de Preenchimento de Espaço de Hilbert

Uma curva de preenchimento de espaço é um caminho contínuo que visita todos os pontos em uma grade n -dimensional e nunca se cruza [Prusinkiewicz and Lindenmayer 2004]. Ela pode ser aplicada a imagens ou volumes (neste caso a matriz de *pixels* ou o volume de *voxels* são as grades que a curva irá percorrer).

Uma característica deste tipo de curva é a sua capacidade de preservar a localização da informação, produzindo uma alta correlação dos dados obtidos. Em outras palavras, diz-se que características próximas na imagem tendem, tanto quanto possível, a permanecer próximas na representação unidimensional obtida pela curva. Dentre as diversas curvas de preenchimento de espaço, a de Hilbert é a mais popular, uma vez que é também a que apresenta uma maior correlação dos dados [Ebrahim et al. 2008].

A Figura 1a mostra as curvas de Hilbert no espaço 2D com níveis 1, 2, 3 e 4. Para obter a curva de nível $(n + 1)$, quatro cópias da curva de nível n são devidamente transladadas, rotacionadas e conectadas [Ebrahim et al. 2008]. A curva de Hilbert pode ser estendida para o espaço tridimensional. Sua construção, similar à curva no espaço 2D, utiliza oito cópias da curva de nível n para obter a curva de nível seguinte $(n + 1)$ [Prusinkiewicz and Lindenmayer 2004, Armstrong et al. 2009]. A Figura 1b ilustra as curvas de Hilbert no espaço tridimensional para $n = 1$ e $n = 2$.

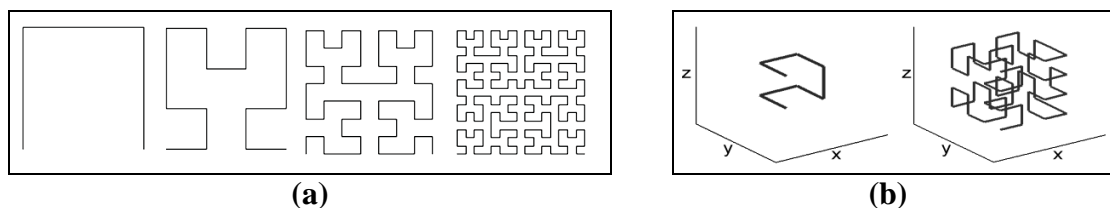


Figura 1. a) Curvas de preenchimento de espaço de Hilbert no espaço 2D (níveis 1 a 4). b) Curvas tridimensionais de Hilbert para os níveis 1 e 2.

2.2. Ganho Acumulativo Descontado

O ganho acumulativo descontado (do inglês *Discounted Cumulative Gain* - DCG) é uma medida de avaliação de desempenho que tem sido muito utilizada em aplicações

de recuperação de informação e em algoritmos de busca na *Web* [Croft et al. 2009]. A medida valoriza mais fortemente aqueles resultados corretos que ocorrem no início da lista de respostas em uma dada consulta (*ranking list*) [Manning et al. 2009], [Shilane et al. 2004].

Suponha-se que n candidatos sejam recuperados para uma consulta. O candidato na i -ésima posição da lista de respostas é associado a um valor rel_i , que indica a relevância deste candidato para a consulta. Os valores para o DCG e DCG normalizado para o intervalo $[0,1]$ ($NDCG$) são definidos nas Equações 1 e 2, respectivamente [Croft et al. 2009]:

$$DCG_k = rel_1 + \sum_{i=2}^k \frac{rel_i}{\log_2 i}, k \leq n \quad (1)$$

$$NDCG_k = \frac{DCG_k}{IDCG_k} \quad IDCG_k = 1 + \sum_{i=2}^{|C|} \frac{1}{\log_2 i} \quad (2)$$

em que $rel_i \in \{0, 1\}$, k indica o número de candidatos na lista de respostas, $IDCG$ corresponde ao valor ideal do DCG (obtido com uma lista de classificação perfeita) e $|C|$ é o número de elementos da classe do objeto de consulta.

2.3. O Descritor *2Hq Feature* Proposto em [Guliatto et al. 2010]

Nesta Seção, descrevemos sucintamente o descritor *2Hq Feature* conforme proposto por [Guliatto et al. 2010] para descrever silhuetas 2D usando a curva de Hilbert. A silhueta de um objeto é dividida em sub-regiões, de acordo com o nível de decomposição L definido *a priori*. Cada uma dessas sub-regiões é percorrida, independentemente, pela curva de Hilbert 2D com o objetivo de mapear a representação bidimensional para uma representação unidimensional, que será posteriormente subamostrada e normalizada para o intervalo $[0,1]$. Finalmente, o descritor *2Hq Feature* é obtido pela composição das representações normalizadas de cada sub-região.

Para um melhor entendimento, considere-se I uma silhueta e L o nível de decomposição da imagem. No nível L , a silhueta I é dividida em 4^L sub-regiões $r_i, i = 1, \dots, 4^L$. A representação normalizada, denominada $2Hq_i$ é obtida para cada sub-região r_i . O processo para obter o $2Hq_i$ é ilustrado na Figura 2.3. O descritor *2Hq Feature* que descreve I é composto pelo conjunto de todos os Hq_i , ou seja, $2HqFeature = \{2Hq_1, \dots, 2Hq_{4^L}\}$.

O processo para a obtenção do *2Hq Feature* é sumarizado a seguir:

1. *Redimensionamento*: nesta primeira etapa, as silhuetas são redimensionadas para as dimensões $2^N \times 2^N$, onde N é um inteiro positivo. Esta etapa evita a perda de informações, uma vez que a curva de Hilbert irá passar por todos os *pixels* de uma imagem com tamanho $2^N \times 2^N$.
2. *Decomposição*: cada imagem é dividida em quatro sub-regiões de tamanho igual. Cada sub-região pode ser novamente dividida em outras quatro sub-regiões e assim por diante, em diversos níveis.
3. *Aplicando a curva de Hilbert*: a curva de Hilbert é aplicada a cada sub-região. O resultado é uma representação unidimensional de cada sub-região, com a vantagem de manter a relação de localidade entre os *pixels* (uma característica da curva de Hilbert).

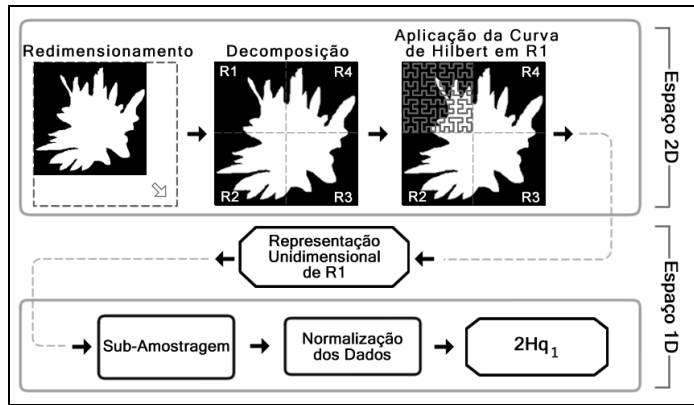


Figura 2. Fluxograma do processo empregado para obtenção do $2H_{q_1}$ de uma silhueta. Neste exemplo, é utilizada uma decomposição com $L = 1$.

4. *Subamostragem*: a representação obtida na etapa anterior é subamostrada, para um número predefinido q , pela aplicação sucessiva da transformada discreta de wavelet, considerando apenas o coeficiente de aproximação.
5. *Normalização*: os dados subamostrados para cada sub-região r_i são normalizados para o intervalo $[0, 1]$, produzindo o $2H_{q_i}$. É importante notar que, feita a decomposição, todas as sub-regiões são processadas de maneira independente, podendo ser processadas em paralelo.

3. Proposta dos descritores $HqRF$

3.1. Proposta para o Descritor de Silhuetas $2HqRF$

O descritor $2Hq$ Feature proposto por [Guliatto et al. 2010] tipicamente utiliza como parâmetros: $q = 256$ e $L = 1$. Neste caso, o $2Hq$ Feature é formado pela composição de quatro vetores de 256 elementos cada um. À medida que L aumenta, o número de vetores aumenta exponencialmente. Esta solução, tendo apresentado ótimos resultados em tarefas de classificação, seria praticável também na recuperação de imagens por conteúdo.

Neste trabalho propomos reduzir a dimensionalidade do $2Hq$ Feature utilizando uma rede neural artificial (RNA). O número de elementos que compõem o novo descritor, referido neste trabalho como $2HqRF$, depende do número de classes da aplicação. O uso da RNA para este fim foi inspirado pelo trabalho de [Barcelos et al. 2008].

O mapeamento do descritor $2Hq$ Feature para um vetor de característica de alto nível, aqui denominado por $2HqRF$, é feito por uma RNA *feed-forward*, multicamada com *back propagation*. A rede possui 4^L entradas, uma para cada $2H_{q_i}$ obtido, conforme Seção 2. A ilustração da RNA é apresentada na Figura 3.

3.2. Extensão do $2HqRF$ para descrever silhuetas tridimensionais

O método para a obtenção do descritor $2HqRF$, descrito anteriormente, pode ser estendido para obter descritores de forma no espaço tridimensional. Para isto, é necessário que:

- i) os objetos sejam definidos como volumes de *voxels*;
- ii) as etapas 1, 2 e 3 descritas na Seção 2.3 sejam ajustadas para o espaço 3D e;
- iii) a RNA seja ajustada para mapear os dados para obter o $3HqRF$.

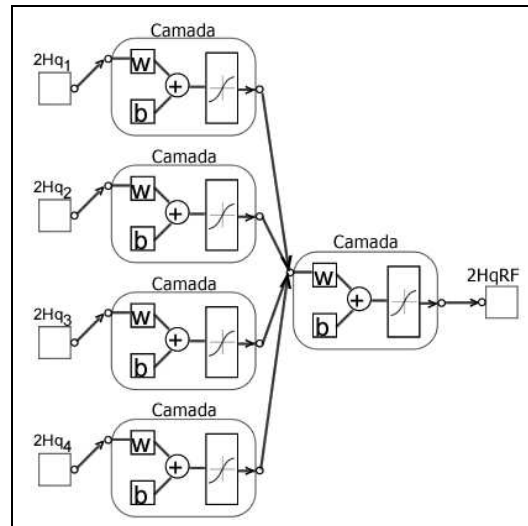


Figura 3. Arquitetura da RNA utilizada para obter o $2HqRF$ (nesta ilustração é considerada a decomposição com $L = 1$).

4. Bases de Dados

Para a realização dos experimentos apresentados na Seção 5, duas bases de dados foram utilizadas: a primeira contendo contornos de lesões em duas dimensões, e a segunda contendo objetos tridimensionais que simulam lesões tridimensionais da mama. Esta seção descreve ambas as bases de dados.

4.1. Base de Dados de Contornos 2D

Os contornos presentes nesta base têm origem em três fontes: *Screen Test: The Alberta Program For The Early Detection Of Breast Cancer* [Alto et al. 2005]; *Mammographic Image Analysis Society* (MIAS, UK); e *Foothills Hospital* de Calgary [Rangayyan et al. 2000, Rangayyan et al. 1997]. O contorno de cada lesão foi manualmente traçado por um radiologista especializado em mamografias. A base de dados combinada possui 111 contornos, incluindo casos típicos e atípicos de lesões benignas (65) e tumores malignos (46). A classificação como maligna ou benigna de cada contorno foi baseada em biópsia. Esta base de dados é aqui denominada BC-111. A Figura 4 exibe algumas silhuetas da base de dados BC-111.

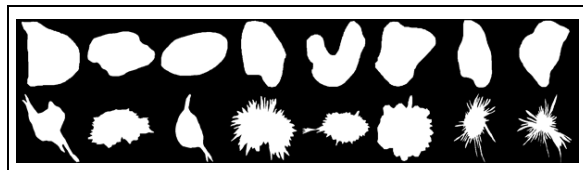


Figura 4. Exemplo de silhuetas de lesões de mama presentes na base BC-111.

4.2. Base de Dados de Volumes de Voxels

Com o objetivo de avaliar a potencialidade do descritor $3HqRF$ para caracterizar objetos tridimensionais arredondados ou espiculados, criamos a base de dados W3D-111¹. A Figura 5 ilustra alguns objetos desta base.

¹Disponível no endereço eletrônico <http://www.facom.ufu.br/~walter/hqrf>

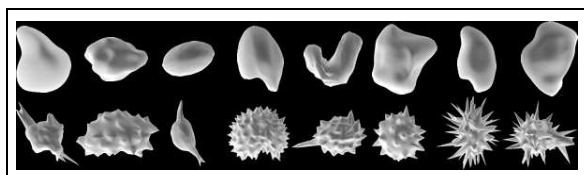


Figura 5. Exemplo de objetos da base de dados W3BC-111.

Os objetos são divididos em duas classes, espiculados (62 objetos) e arredondados (49 objetos), e são representados como volumes de *voxels* com dimensões $256 \times 256 \times 256$.

5. Experimentos e Discussão dos Resultados

Nesta seção, os descritores de forma $2HqRF$ e $3HqRF$ são avaliados por meio de experimentos de recuperação de formas. A RNA foi treinada utilizando a *10-fold cross-validation*. A similaridade entre dois descritores é medida por meio da distância Euclidiana. Os resultados são avaliados em termos da média das curvas de Precisão-Revocação (PR) e da média dos valores NDCG.

5.1. Experimentos Realizados no Espaço 2D

A fim de avaliar a qualidade dos resultados obtidos com o descritor de forma $2HqRF$, proposto neste artigo, é realizada uma comparação do método proposto com os fatores de forma SI_{TF} descrito em [Guliatto et al. 2008a], CX_{TA} e FD_{TA} descritos em [Guliatto et al. 2008b]. A base de dados utilizada neste experimento é a BC-111. Os parâmetros utilizados para a obtenção do $2HqRF$ são listados na Tabela 1. Na Figura 6 são exibidos os 20 primeiros resultados de consultas realizadas neste experimento (o primeiro objeto em cada linha é o objeto de consulta). A Figura 7a exibe as curvas PR média para o $2HqRF$ e para os métodos SI_{TF} , CX_{TA} e FD_{TA} . Os valores da média dos NDCG são apresentados na Figura 7b. Note que o método proposto obteve uma precisão média de 100% para todos os valores de revocação, superando os demais métodos utilizados neste experimento.

Tabela 1. Parâmetros utilizados na obtenção do descritor $2HqRF$.

Parâmetro	Valor
Nível de decomposição (L)	3
Família Wavelet	Haar
Num. de neurônios nas camadas escondidas	20
Tamanho do $2HqF_i(q)$	256

5.2. Experimentos Realizados no Espaço 3D

A base de dados W3BC-111 é utilizada para os experimentos no espaço tridimensional. Os parâmetros utilizados nestes experimentos são apresentados na Tabela 2. A Figura 8 exibe os 20 primeiros objetos recuperados em consultas realizadas (note que o objeto de consulta é o primeiro em cada linha). A curva PR média obtida neste experimento é ilustrada na Figura 9. O valor de NDCG médio obtido no mesmo experimento é de 0,99.

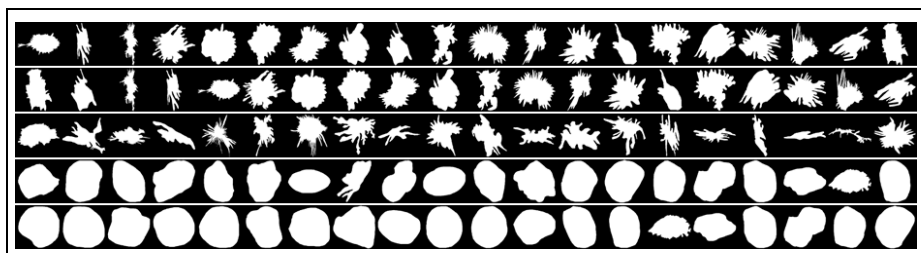
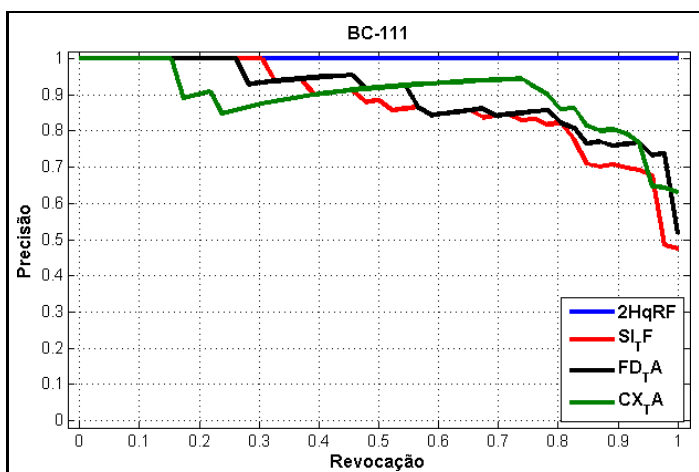


Figura 6. Consultas realizadas com a base de dados BC-111.



(a)

Descritor	NDCG
$2HqRF$	1.0000
SI_{TF}	0.8849
CX_{TA}	0.9656
FD_{TA}	0.9579

(b)

Figura 7. a) Curva precisão-revocação média obtida nos experimentos com a base BC-111. b) Valores de NDCG médio obtidos nos experimentos com a base de dados BC-111.

Tabela 2. Parâmetros utilizados na obtenção do descritor $3HqRF$.

Parâmetro	Valor
Nível de decomposição (L)	1
Família Wavelet	Haar
Num. de neurônios nas camadas escondidas	20
Tamanho do $2HqF_i$ (q)	256

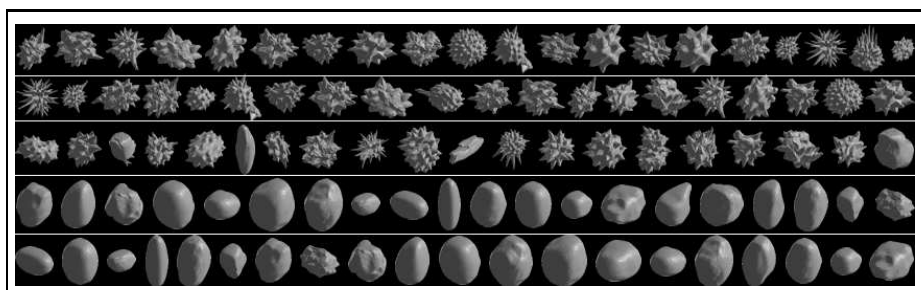


Figura 8. Consultas realizadas com a base de dados W3BC-111.

6. Conclusões

Neste artigo foram apresentados dois novos métodos para descrição de formas nos espaços 2D e 3D. Os descritores propostos foram aplicados à recuperação de tumores

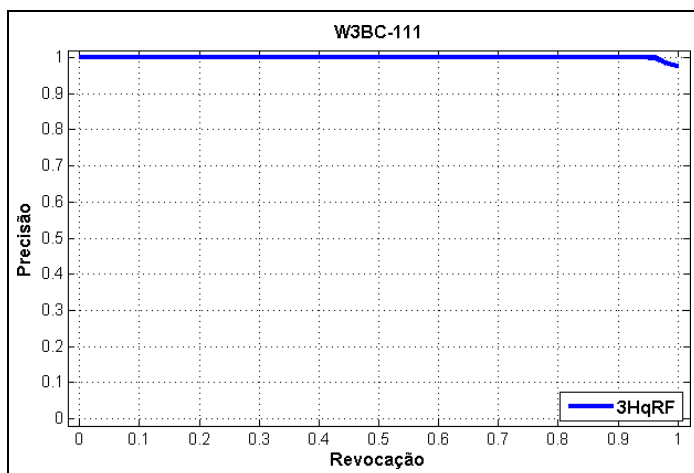


Figura 9. Curva precisão-revocação média obtida nos experimentos com a base W3BC-111.

de mama. A análise dos resultados mostrou uma superioridade do $2HqRF$, em termos dos valores médios de NDCG e de curvas PR, quando comparados a trabalhos relacionados. O $3HqRF$ mostrou potencialidade em discriminar formas arredondadas e espiculadas, que são características importantes para auxiliar o diagnóstico de câncer de mama em exames radiológicos tridimensionais, como ressonância magnética ou tomografia computadorizada. Esforços têm sido feitos para se construir uma base com lesões de mama conseguidas a partir de exames de ressonância magnética.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro.

Referências

- Alto, H., Rangayyan, R., and Desautels, J. (2005). Content-based retrieval and analysis of mammographic masses. *Journal of Electronic Imaging*.
- American College of Radiology, Reston, VA (1998). Illustrated breast imaging reporting and data system BI-RADSTM).
- Armstrong, J., Ahmed, M., and Chau, S.-C. (2009). A rotation-invariant approach to 2d shape representation using the Hilbert curve. *Image Analysis and Recognition*, 5627:594–603.
- Barcelos, C., Ribeiro, E., and Batista, M. (2008). Image characterization via multilayer neural networks. *International Conference on Tools with Artificial Intelligence*, pages 325–332.
- Chen, W., Giger, M., and Bick, U. (2006). A fuzzy c-means (fcm)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced mr images. *Acad. Radiol.*, 13(1):63–72.
- Croft, W., Metzler, D., and Strohman, T. (2009). *Search Engines: Information Retrieval in Practice*. Addison Wesley.

- Ebrahim, Y., Ahmed, M., Abdelsalam, W., and Chau, S.-C. (2008). Shape representation and description using Hilbert curve. *Pattern Recognition Letters*.
- Guliatto, D., de Oliveira, W., and Jr., C. T. (2010). A new feature descriptor derived from Hilbert space-filling curve to assist breast cancer classification. 1:303–308.
- Guliatto, D., Rangayyan, R., Carvalho, J., and Santiago, S. (2008a). Polygonal modeling of contours of breast tumors with the preservation of spicules. *IEEE Trans Biomed Eng*.
- Guliatto, D., Rangayyan, R., de Carvalho, J., and Santiago, S. (2008b). Feature extraction from the turning angle function for the classification of contours of breast tumors. *J. Digit Imaging*.
- Gupta, S., P.F., C., and Markey, M. (2006). Breast cancer cadx based on *BI-RADSTM* descriptors from two mammographic views. *Med. Phys.*, 33(6):1810–1817.
- Heath, M., Bowyer, K., and Kopans, D. (1998). Current status of the digital database for screening mammography.
- Manning, C., Raghavan, P., and Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University Press.
- Prusinkiewicz, P. and Lindenmayer, A. (2004). *The Algorithmic Beauty of Plants*. Springer-Verlag.
- Rangayyan, R., El-Faramawy, N., Desautels, J., and Alim, O. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*.
- Rangayyan, R., Mudigonda, N., and Desautels, J. (2000). Boundary modelling and shape analysis methods for classification of mammographic masses. *Medical and Biological Engineering and Computing*.
- R.G.-Caballero, C.J.G.-Orellana, H.M.G.-Velasco, and M.M.-Macías (2007). Independent component analysis applied to detection of early breast cancer signs. *Computational and Ambient Intelligence*.
- Shilane, P., Min, P., Kazhdan, M., and Funkhouser, T. (2004). The princeton shape benchmark. *Shape Modeling International*.
- Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., and Savage, J. (1994). The mammographic image analysis society digital mammogram database.
- Turnbull, L., Brown, S., Harvey, I., Olivier, C., Drew, P., Napp, V., Hanby, A., and Brown, J. (2010). Comparative effectiveness of mri in breast cancer (comice) trial: a randomised controlled trial. *The Lancet*, 375(9714):563–571.
- Weiqiang, Z., Xiangmin, X., and Wei, H. (2008). Shape and boundary analysis for classification of breast masses. *International Symposium on Computational Intelligence and Design*.
- WHO (2010). Breast cancer: prevention and control.