

Segmentação de imagens de túbulos seminíferos no testículo de camundongos

Francisco de Assis Zampirolli
Centro de Matemática,
Computação e Cognição
Universidade Federal do ABC
Santo André, Brasil
fzampirolli@ufabc.edu.br

Carlos Alberto da Silva
Centro de Ciências Naturais e
Humanas
Universidade Federal do ABC
Santo André, Brasil
carlos.asilva@ufabc.edu.br

Glauber Rajiv de Andrade
Universidade Federal do ABC
Santo André, Brasil
glauber.Rajiv@aluno.ufabc.edu.br

Resumo — O artigo tem como objetivo o desenvolvimento de métodos automáticos para a classificação de imagens de túbulos seminíferos de camundongos, através do uso de técnicas de processamento de imagem e mineração de dados. Os dados apresentados indicam que a estratégia adotada apresentou 92,8% de acerto durante a classificação dos túbulos seminíferos e abrem perspectivas para a possível aplicação desta técnica nas análises histopatológica automáticas ou semiautomáticas no testículo de mamíferos.

Palavras-chave – *segmentação, túbulos seminíferos, espermatogênese, camundongos, classificação.*

I. INTRODUÇÃO

A espermatogênese é um processo cíclico altamente organizado e complexo que permite a formação de espermatozoides no interior dos túbulos seminíferos presentes nos testículos de mamíferos. Em um corte histológico do testículo observa-se que os túbulos seminíferos são revestidos por um epitélio altamente especializado formado por células germinativas e células de Sertoli e, na região central do túbulo observa-se a presença de um lúmen [1]. A análise morfológica dos túbulos seminíferos pode representar uma estratégia para avaliar a toxicidade de substâncias naturais ou sintéticas na espermatogênese, permitindo que os túbulos sejam classificados em normais, alterados, atrofiados, hipoespermatogênicos ou necrosados. Essas análises são complexas, pois dependem da experiência de um histopatologista para a identificação dos túbulos seminíferos e um grande número de amostras a serem analisadas. Com isso, a automação do processo de análise de imagens para a classificação dos túbulos seminíferos aperfeiçoaria a análise histopatológica no testículo. Sendo assim, o trabalho tem como objetivo a segmentação das imagens dos túbulos seminíferos normais e diferentes alterações morfológicas para posteriormente classificá-los com o auxílio.

Os túbulos seminíferos foram distribuídos em cinco categorias, nomeados como alterados, atrofiados, hipoespermatogênicos, necrosados e normais. A classificação foi estabelecida com base nas mudanças morfológicas dos túbulos seminíferos que refletem o grau de comprometimento da espermatogênese [2]. Cada grupo foi segmentado em secção externa, interna e intermediária; com exceção do atrofiado, por não possuir região de lúmen dos túbulos. Os dados morfológicos das secções foram coletados para gerar um meio de classificação.

II. PRELIMINARES

A. Processo de Segmentação de imagens

Em processamento de imagem a segmentação é o processo de divisão de uma imagem em segmentos permitindo assim uma análise computacional e matemática a fim de se extrair dados [3]. O resultado desse processo é uma imagem binária (preto – nível de cinza com valor zero, e branco – nível de cinza com valor 1) que segmenta o objeto desejado.

Para realizar o processo de segmentação, foi utilizada a linguagem e software matricial MATLAB®, a biblioteca *SDC Morphology Toolbox for MATLAB*, além de outras internas. A escolha da ferramenta se deve pelos elementos de linguagens de alto nível, como matriz de tamanho dinâmico [4], facilitando o trabalho com imagens. Além da grande disponibilidade de bibliotecas de processamento de imagem.

B. Processo de mineração de dados

Para a mineração de dados, utilizou-se o software de aprendizagem de máquina Weka, *Waikato Environment for Knowledge Analysis*, desenvolvido pela Universidade de Waikato. O software suporta uma grande quantidade de tarefas de mineração de dados. Utilizou-se o algoritmo SMO, *Sequential minimal optimization*, que utiliza métodos de classificação por vetores de suporte [5].

III. RESULTADOS

A. Ajuste de imagem

O processamento foi iniciado pela adequação e otimização das imagens dos túbulos seminíferos. Os algoritmos usados foram executados tendo como entrada imagens em escala de cinza, com 256 tonalidades. Como a imagem primária se encontra em RGB, foi necessário obter o valor correspondente de cada canal. Usando testes empíricos, cada canal de cor teve um peso correspondente na conversão para cinza, definido como 30% para o vermelho, 59% para o verde e 11% para o azul [6].

As imagens precisam ter uma distribuição equalizada de seus histogramas, para obter máxima variância [6] e melhor contraste. A Figura 1 ilustra a comparação dos histogramas. Na Figura 2, podemos perceber que um limiar preto, antes

escuro, agora é visível. Como segmentação se baseia em encontrar esses limiares, essa otimização é imprescindível.

B. Segmentação interna

Na segmentação da secção interna do túbulo, foi aplicado o algoritmo *Watershed* [7]. Esse algoritmo transforma uma imagem em picos e vales, usando como variável, a intensidade da escala de cinza. Logo após, há uma inundação nos vales e esses canais inundados podem ser interpretados como o limiar do lúmen. Antes da aplicação, foram usadas funções para filtrar objetos inadequados e eliminar buracos. Tais funções são encontradas na biblioteca *Image processing Toolbox* do MATLAB® e no *SDC Morphology Toolbox*, biblioteca para o MATLAB®. O conjunto de funções que compõem a solução para a segmentação interna foi compilada em uma função maior, onde algumas constantes, como área dos objetos descartados e erosão, foram definidas experimentalmente. A Figura 3 define a marcação do lúmen (segmento interno) em um túbulo hipoespermatozóico.

C. Segmentação externa

A segmentação externa foi mais complicada, pois a simples utilização do algoritmo *Watershed* foi ineficaz por não funcionar bem em bordas muito grossas, já que ele acaba traçando muitos limiares dentro da própria borda de segmentação. A solução foi usar inicialmente a técnica de *Threshold*, onde a amostragem da escala de cinza é dividida em dois intervalos, definidos pelo usuário, correspondendo ao 1 e ao 0 da imagem binária. A imagem resultante apresentou poucos detalhes e ficou em contato com os túbulos vizinhos. Esse problema foi corrigido pela técnica *mmdcells* [8] de picos, semelhante ao *Watershed*, mas que usa como variável a distância de cada ponto de valor 1 até o ponto nulo mais próximo, para a determinação dos picos. O limiar foi definido na mediana da distância entre dois picos. A Figura 4 exemplifica o processo. Desse modo, evitou-se que túbulos seminíferos vizinhos ficassem unidos com o estudado. Na Figura 5, há uma segmentação externa de um túbulo normal usando esta técnica.

D. Segmentação intermediária

A segmentação intermediária foi construída pela diferença entre a externa e a interna. A subtração é efetuada apenas de valores não nulos.

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Perceba que não ocorreram valores negativos.

E. Cálculo de medidas

Com os segmentos binários obtidos, foi possível obter medidas morfológicas. Foi utilizado 17 tipos de medidas, sendo 14 aplicadas ao segmento interno e no externo e 3 ao intermediário, totalizando 35 dados por túbulo seminífero.

Uma parte dos algoritmos de medidas foi encontrada na biblioteca de processamento de imagem do MATLAB®, outras foram desenvolvidas. A lista de medidas usadas e as descrições [9] foram: Perímetro: comprimento de contorno dos segmentos; Área: quantidade de pixels nos segmentos; Comprimento do maior eixo: o maior eixo da elipse que pode ser desenhada dentro do segmento; Comprimento do menor eixo: o menor eixo da elipse que pode ser desenhada dentro do segmento; Orientação: ângulo formado entre o menor e o maior eixo da elipse; Área convexa: a menor área convexa que consegue contornar o objeto; Excentricidade: razão entre a distância dos focos da elipse com o maior eixo; Diâmetro equivalente: diâmetro da circunferência com área igual ao do segmento; Solidity: $\frac{\text{área}}{\text{área convexa}}$; AreaBox: menor equilátero que consiga cobrir todo o segmento; Densidade: $\frac{\text{Área}}{\text{AreaBox}}$; Altura de borda: a maior distância perpendicular do segmento interno ao externo; Desvio padrão da altura de borda. Outras medidas encontradas em [10] foram usadas: FormFactor: $\frac{4 * \pi * \text{Área}}{\text{Perímetro}}$; Roundness: $4 * \frac{\text{Área}}{\pi * \text{Maior eixo}^2}$; AspectRatio: $\frac{\text{Comprimento do maior eixo}}{\text{Comprimento do menor Eixo}}$; Solidity2: $\frac{\text{Área convexa} - \text{Área}}{\sum_{i=1}^n (\text{Área convexa}_i - \text{Área}_i) / N}$.

F. Mineração de dados

O processo de mineração de dados iniciou-se pela criação de um arquivo de texto no formato arff, *Attribute-Relation File Format*. Um gerador foi criado através do MATLAB®, que pode ser lido pelo Weka.

Para a base de dados, foram utilizados 42 imagens de túbulos seminíferos e o algoritmo de classificação SMO, configurado com 10 dobras em validação cruzada. Nesse modo, os objetos foram divididos igualmente (dobras) e cada dobra é usada tanto para teste como para treinamento.

Os resultados finais mostram que 39 imagens foram classificadas corretamente e 3 incorretamente, tendo então 92,8% de acerto. A tabela de confusão mostra quais foram os erros cometidos:

	Normal	Alterado	Atrofiado	Hipoes.	Necros.
Normal	10	0	0	0	0
Alterado	1	8	0	0	0
Atrofiado	1	0	6	1	0
Hipoespermatozóico	1	0	0	5	0
Necrosado	0	0	0	0	10

IV. CONCLUSÕES

Os túbulos seminíferos foram distribuídos em cinco categorias, nomeados como alterados, atrofiados, hipoespermatozóicos, necrosados e normais. Como estas categorias não apresentarem características morfológicas semelhantes em cada tipo de túbulo, as segmentações das

imagens foi semiautomática, tendo parâmetros de Threshold e Watershed ajustados manualmente. Para cada categoria, também foram extraídas características interna, externa e intermediária. Ainda existem medidas dos filamentos do segmento interno que podem ser extraídas, como o comprimento, a espessura, a taxa de organização e a densidade de filamentos na borda, que poderiam elevar a taxa correta de classificação. Mesmo assim, conseguimos um excelente resultado de 92.8% de acerto na classificação das cinco categorias.

REFERÊNCIAS

- [1] D. L. Russel, A. R. Ettlin, S. P. A. Hikim, D. E. Clegg - Histological and Histopathological Evaluation of the Testis. In: Staining for the Laboratory Mouse. Cache River Press, p.119-61, 1991.
- [2] John R. Wenz, Rex A. Hess - Characterization of stage-specific tyrosinated α -tubulin immunoperoxidase staining patterns in Sertoli cells of rat seminiferous tubules by light microscopic image analysis. - publicado em Tissue & Cell, em março de 1998.
- [3] R. C. Gonzales and R.E. Woods. "Digital Image Processing". Addison-Wesley Pub. Company, 3rd, pp. 976, 2007.
- [4] MATLAB technical documentation – Mathworks.com. Acessado em 20 de maio de 2012.
- [5] J. Platt – Machines using Sequential Minimal Optimization - 1998.
- [6] William K. Pratt – Digital Image Processing. New York:.
- [7] Vincent, L.; Soille, P. - Watersheds in digital spaces: an efficient algorithm based on immersion simulations.
- [8] SDC Information Systems - *SDC Morphology Toolbox for Matlab* - 5, 2000.
- [9] F. A. Zampiroli, B. Stransky, A. C. Lorena, F. L. Melo Paulon - Segmentation and classification of histological images - application of graph analysis and machine learning methods – SIBGRABI(2010)
- [10] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," Breast Cancer Research and Treatment, vol. 89, pp. 179–18, 2005.

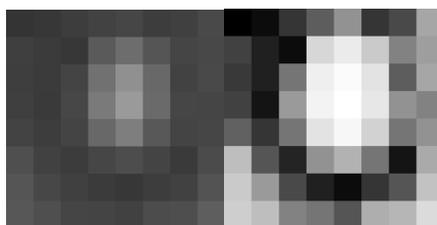


Figura 1. Imagem original (à esquerda) e otimizada por equalização global do histograma (à direita).

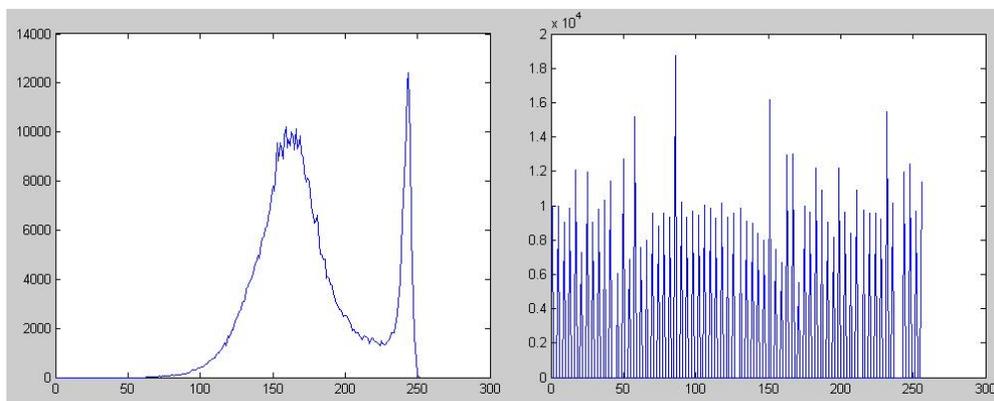


Figura 2. Exemplo de histograma (à esquerda) e equalizado (à direita).

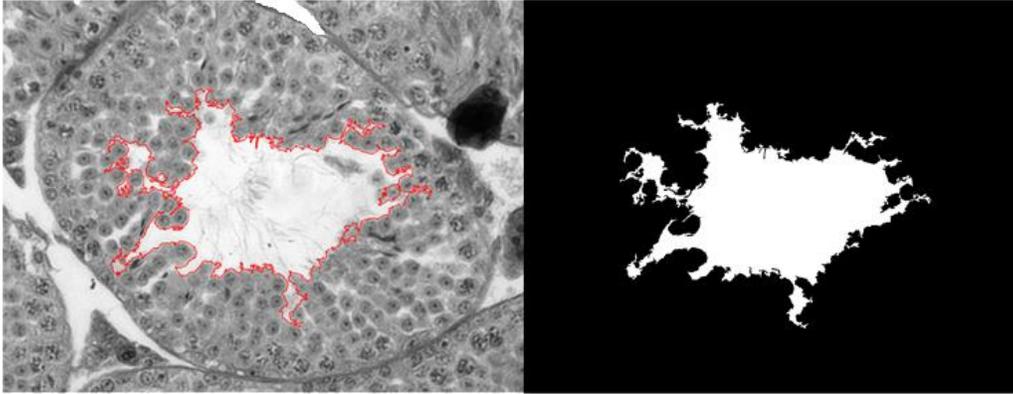


Figura 3. Túbulo hipoespermatogênico segmentado internamente.

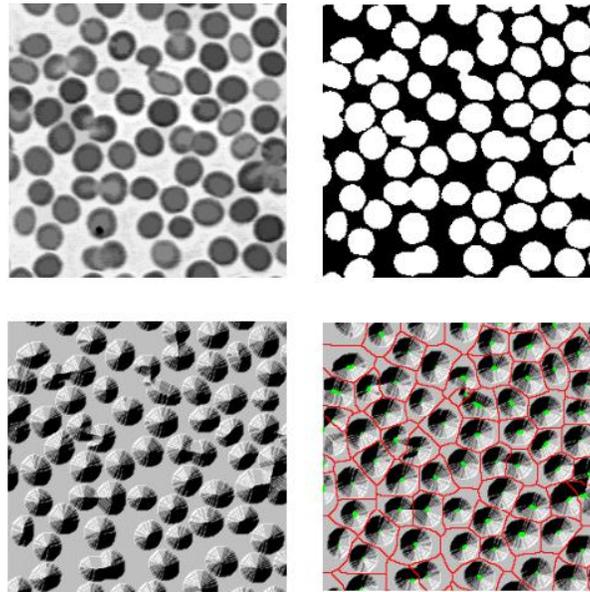


Figura 4. As etapas da técnica mmdcells. Em ordem: imagem original, binária com Threshold, em picos e a divisão pela mediana da distância dos picos [7].

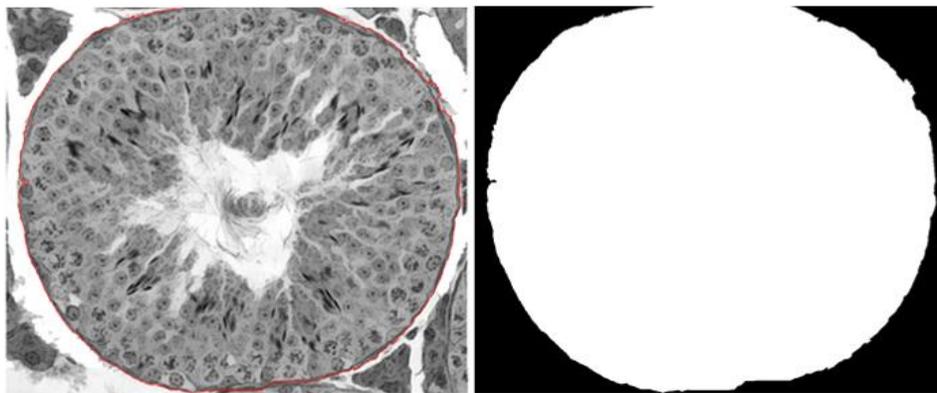


Figura 5. Segmentação externa.