

1 Introdução

Esta pesquisa tem como objetivo melhorar a qualidade das análises realizadas pelos sistemas computacionais para o estudo de imagens biomédicas de alta resolução. O foco principal é direcionado para os sistemas que trabalham com a análise de imagens de secções microscópicas de tecidos. Analisar essas imagens possibilita o estudo de doenças em nível celular ou sub-celular. Estas análises podem auxiliar na caracterização da morfologia dos tecidos sub-celulares e podem facilitar o entendimento dos mecanismos de doenças e melhorar a avaliação da resposta aos tratamentos. Isto é possível, pois muitas informações sobre células e morfologia celular podem ser obtidas a partir das análises destas imagens [2].

Entretanto, capturar e analisar essas informações em imagens de larga escala mostra-se um desafio, por ser necessário a revisão humana subjetiva [1].

Uma plataforma disponível para realizar análise de imagens médicas é a desenvolvida pelo grupo do Prof George Teodoro denominada *Region Template Framework* (RTF) [3]. Esse sistema pode analisar imagens de tecidos de forma a extrair informações de elementos das imagens (como células ou núcleos celulares) e algumas de suas características, como tamanho, formato e aspectos da textura. Essas características são utilizadas para desenvolver modelos morfológicos que podem ser utilizados para o aprendizado de novos conhecimentos. Um exemplo da utilidade desta plataforma é na análise de gliomas difusos, os núcleos das células do tumor são de interesse significativo para a comunidade científica [4].

Um fluxo de trabalho padrão desta plataforma consiste nas seguintes etapas: normalização, segmentação, computação de características, refinamento das características e classificação. Uma vez que um fluxo de trabalho busca analisar imagens de tecidos, temos que este sistema é dependente das imagens analisadas e conseqüentemente o resultado depende dos parâmetros de entrada do sistema. Para melhorar a qualidade das análises, portanto, é necessário estudar e compreender o impacto desses parâmetros no resultados.

Para melhorarmos a qualidade das análises do RTF, iremos buscar estratégias que auxiliem na avaliação das características extraídas de uma imagem. Ela é uma etapa que influi diretamente na qualidade dos resultados das análises e ao ajustar os parâmetros buscando resultados com maior acurácia poderemos ter uma qualidade maior das análises.

Atualmente, na etapa de segmentação, algumas estratégias já foram utilizadas para avaliar o impacto dos parâmetros de entrada em um fluxo de trabalho. As etapas de segmentação e de avaliação de características apesar de serem etapas distintas, compartilham atributos similares. Logo, podemos utilizar dos conhecimentos obtidos da utilização dessas estratégias na etapa de segmentação para direcionar um ponto de partida do estudo que será realizado na etapa de avaliação de características. Também temos que ambas etapas dependem da revisão subjetiva de especialistas e dado o tamanho das imagens, essa revisão torna-se proibitiva. A partir das melhores estratégias estudadas e dos conhecimentos obtidos da etapa de segmentação, poderemos desenvolver modelos capazes de avaliar com alta precisão as características obtidas pelos métodos computacionais. Esses modelos podem ser utilizados na etapa de refinamento das características, pois seremos capazes de comparar os diversos resultados obtidos e assim extrapolar a necessidade da revisão de especialistas.

A etapa de refinamento de características pode ser separada em duas fases. A primeira, é a aplicação de métodos e técnicas de análise de sensibilidade (SA) e a segunda fase é o refinamento dos parâmetros. O objetivo final é encontrar dentro as possíveis combinações de resultados o conjunto de valores dos parâmetros que melhor influência no resultado final.

2 Justificativa

É imprescindível a busca de maneiras eficientes para se realizar processamento de alto desempenho para tarefas de análise de imagens. As análises de imagens de tecido de larga escala são problemas computacionalmente desafiadores. Um dos desafios é obter a escalabilidade dos vários fluxos de trabalho e/ou de grandes conjuntos de imagens com ganhos de qualidade.

O *National Institute of Health's The Cancer Genome Atlas* (TCGA) disponibiliza uma grande quantidade de dados multimodais contendo imagens de patologia, radiologia, genômica e dados clínicos para o glioblastoma (gliomas cancerígenos) e outros tipo de tumores [5]. Atualmente, o RTF contém um banco de dados com mais de 600 imagens cada com uma com uma média de 400 mil núcleos por secção. O processo de caracterizar as morfologias celulares produz cerca de 1.5GB/secção

de meta-dados que descrevem os parâmetros dos algoritmos, os limites dos objetos e as características das células [6].

Uma vez que a quantidade de dados gerada é elevada, não é humanamente possível avaliar todos resultados da caracterização das células e núcleos. Para isso, buscamos criar modelos capazes de avaliar o conjunto de características.

O desenvolvimento desses modelos é necessário para que as fases de refinamento de parâmetros e a análise de sensibilidade possam ser aplicados no contexto da computação de características. Após a implementação do modelo mais representativo capaz de avaliar a qualidade de um conjunto de características, podemos comparar os resultados obtidos entre os diversos conjuntos de parâmetros.

3 Objetivos

A partir do trabalho já realizado em [7], pretende-se melhorar a otimização de avaliação de parâmetros com foco direcionado para a etapa de classificação das características. Dessa forma espera-se incrementar a qualidade das análises realizadas pelo sistema e, se possível, com ganhamos computacionais de tempo.

3.1 Objetivos Específicos

1. Gerar e implementar modelos de avaliação das características das células e núcleos;
2. Atualizar os métodos e técnicas de análise de sensibilidade e refinamento de parâmetros disponíveis no RTF;
3. Otimizar a qualidade das análise das características das células e núcleo e consequentemente a do sistema como um todo.

4 Revisão da Literatura

Na Seção 1 comentamos que buscamos impactar a qualidade das análises do RTF ao aprimorar a avaliação das características extraídas de uma imagem. Esse aprimoramento é dividido em duas fases, a análise de sensibilidade (SA) e o refinamento de parâmetros. Para que a aplicação dessas fases sejam viáveis em um sistema de análise de imagens é necessário propor e utilizar eficientemente SA e métodos de refinamento.

Já estão presentes, no RTF, os métodos de SA Variance-based Decomposition (VBD) [8] [9] e o Morris One-At-A-Time design (MOAT) [10] [11]. Para o refinamento de parâmetros, os métodos existentes são o Nelder-Mead simplex (NM), Parallel Rank Order (PRO), e o Algoritmo Genético de auto-refinamento.

A maioria dos trabalhos recentes desenvolvidos para a otimização de parâmetros aplicam técnicas para modelos de segmentação específicos [12, 13, 14, 15]. O sistema Tuner [16] foca em algoritmos de segmentação em geral e cria modelos estatísticos que descrevem uma função de resposta da segmentação e utiliza de um modelo Gaussiano para explorar o espaço de busca. Já o sistema DAKOTA inclui métodos diversos de otimização de parâmetros, assim como métodos para modelagem surrogate, métodos de amostragem estatística de dados, estratégias de otimização baseada em surrogate e etc [17]. Algumas outras opções que também serão analisadas como os métodos de otimização Particle swarm, de otimização by branch-and-fit, de otimização Bayesiana, algoritmos de Hit-and-Run.

Estes métodos serão utilizados como ponto de partida, pois nenhum deles é voltado para a análise de imagens médicas de patologia, logo não são diretamente aplicáveis. Além disso, não existe uma ferramenta ou estudo que aborda o problema de refinamento de parâmetros e/ou de análise de parâmetros voltado para entendimento de características das células de uma imagem, que é o foco desse trabalho.

Para o trabalho proposto as ferramentas já disponíveis no RTF e os métodos pré-citados serão incorporados para análise de sensibilidade e refinamento de parâmetros. Os resultados obtidos e as influências observadas serão analisados, buscando a maximização a qualidade dos resultados gerados pelo RTF.

5 Metodologia

Inicialmente, será conduzida uma investigação mais detalhada da literatura sobre caracterização e previsão de custos de tarefas computacionais, sobre métodos de refinamento de parâmetros, com foco em algoritmos genéticos e os presentes na plataforma RTF, e sobre modelagem de características de células e núcleos celulares. Também será estudado mais profundamente o domínio de fluxos de trabalho de tratamento de imagens médicas. Em seguida, serão incorporados os modelos de características de células e núcleos celulares em conjunto com as técnicas de refinamento de parâmetros mais promissores da literatura e os já presentes na plataforma RTF.

6 Plano de Trabalho

Este trabalho será realizado em 5 etapas:

1. Implementação da modelagem das características de células e núcleos celulares;
2. Implementação de métodos de refinamento de parâmetros;
3. Análise do reuso dos métodos de refinamento de parâmetros do RTF;
4. Realização de testes para avaliar o desempenho obtido por cada otimização individualmente e em conjunto;
5. Análise de resultados de testes e escrita da dissertação e artigos científicos.

7 Cronograma

Atividade	2017		2018	
	1º	2º	1º	2º
Cursar Disciplinas de Mestrado	X			
Revisão Bibliográfica	X	X		
Implementação da modelagem das características		X	X	
Implementação dos métodos de refinamento		X	X	
Realização de testes		X	X	X
Escrita Da Dissertação			X	X

References

- [1] Jun Kong, Lee Cooper, Fusheng Wang, Candace Chisolm, Carlos Moreno, Tahsin Kurc, Patrick Widener, Daniel Brat, and Joel Saltz. A comprehensive framework for classification of nuclei in digital microscopy imaging: An application to diffuse gliomas. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 2128–2131. IEEE, 2011.
- [2] Jun Kong, Lee Cooper, Ashish Sharma, Tahsin Kurc, Daniel J Brat, and Joel H Saltz. Texture based image recognition in microscopy images of diffuse gliomas with multi-class gentle boosting mechanism. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 457–460. IEEE, 2010.
- [3] George Teodoro, Tony Pan, Tahsin Kurc, Jun Kong, Lee Cooper, Scott Klasky, and Joel Saltz. Region templates: Data representation and management for high-throughput image analysis. *Parallel computing*, 40(10):589–610, 2014.
- [4] Meenakshi Gupta, Azita Djalilvand, and Daniel J Brat. Clarifying the diffuse gliomas. *American journal of clinical pathology*, 124(5):755–768, 2005.
- [5] Roger McLendon, Allan Friedman, Darrell Bigner, Erwin G Van Meir, Daniel J Brat, Gena M Mastrogiannis, Jeffrey J Olson, Tom Mikkelsen, Norman Lehman, Ken Aldape, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068, 2008.

- [6] Lee AD Cooper, Jun Kong, Fusheng Wang, Tahsin Kurc, Carlos S Moreno, Daniel J Brat, and Joel H Saltz. Morphological signatures and genomic correlates in glioblastoma. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1624–1627. IEEE, 2011.
- [7] George Teodoro, Tahsin Kurc, Luís Rabelo, Alba C. M. A. Melo, Jun Kong, Yi Gao, and Joel Saltz. Efficient algorithm sensitivity analysis with parameter tuning on large imaging datasets. *IPDPS, 2016*, submitted.
- [8] V Gregory Weirs, James R Kamm, Laura P Swiler, Stefano Tarantola, Marco Ratto, Brian M Adams, William J Rider, and Michael S Eldred. Sensitivity analysis techniques applied to a system of hyperbolic conservation laws. *Reliability Engineering & System Safety*, 107:157–170, 2012.
- [9] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1):271–280, 2001.
- [10] Max D Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.
- [11] Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10):1509–1518, 2007.
- [12] Sanjiv Kumar and Martial Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1150–1157. IEEE, 2003.
- [13] Martin Szummer, Pushmeet Kohli, and Derek Hoiem. Learning crfs using graph cuts. In *European conference on computer vision*, pages 582–595. Springer, 2008.
- [14] Chris McIntosh and Ghassan Hamarneh. Is a single energy functional sufficient? adaptive energy functionals and automatic initialization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 503–510. Springer, 2007.
- [15] Thomas Schultz and Gordon L Kindlmann. Open-box spectral clustering: applications to medical image analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2100–2108, 2013.
- [16] Thomas Torsney-Weir, Ahmed Saad, Torsten Moller, Hans-Christian Hege, Britta Weber, Jean-Marc Verbavatz, and Steven Bergner. Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1892–1901, 2011.
- [17] Anthony A Giunta. Use of data sampling, surrogate models, and numerical optimization in engineering design. *AIAA paper*, 538:2002, 2002.