

Segmentação Supervisionada de Imagens pela Funcional de Mumford-Shah Utilizando Métricas de Distância Não-lineares

Antonio Carlos Sobieranski¹, Eros Comunello¹, Aldo von Wangenheim¹

¹Departamento de Informática e Estatística (INE)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Federal de Santa Catarina (UFSC) – Florianópolis – Brasil

{asobieranski,eros}@cyclops.ufsc.br, awangenh@inf.ufsc.br

Abstract. *In this work a supervised region-merging image segmentation approach is presented. In our approach, a distance metric is learned from some features on the image, and then a function for the Mumford-Shah model is optimized by this metric. The global optimum of the approximation function is inductively achieved under high polynomial terms of the Mahalanobis distance, extracting the nonlinear features of the pattern distributions into topological maps. The results we obtained show a better discrimination of object boundaries and the location of regions when compared with the conventional Mumford-Shah algorithm and other image segmentation methods. A quantitative objective evaluation of the proposed approach was performed in order to compute the quality of the obtained results.*

Image segmentation, Mumford-Shah model, distance metric learning, global optimization.

Resumo. *Neste trabalho é apresentada uma abordagem supervisionada de segmentação de imagens por crescimento de regiões. Nesta abordagem uma métrica de distância é modelada a partir de algumas características da imagem, e então utilizada para otimizar o modelo de Mumford-Shah. A função de aproximação do modelo é indutivamente obtida sobre termos de alta ordem polinomial da distância de Mahalanobis, possibilitando extrair as características não-lineares das distribuições em mapas topológicos. Os resultados obtidos apresentam uma melhor discriminação das fronteiras dos objetos quando comparados com o modelo convencional e em relação a outros métodos de segmentação. Para verificar a qualidade dos resultados obtidos, métricas de avaliação de qualidade foram utilizadas.*

Segmentação de imagens, modelo de Mumford-Shah, aprendizado de métrica de distância, otimização global.

1. Contextualização

A segmentação de imagens é uma das mais importantes ferramentas em Visão Computacional (VC) e Processamento Digital de Imagens (PDI). Seu uso possui aplicação direta nas mais diferentes áreas do conhecimento, tais como a indústria médica, robótica, aeroespacial, dentre outras. A segmentação consiste em codificar uma imagem de entrada I sem qualquer informação geométrica ou semântica em uma imagem de segmentos, onde cada segmento corresponda na totalidade ou parcialidade a uma região homogênea e/ou

contínua na cena. Através destes segmentos, além da imagem tornar-se mais simples de ser analisada (de pixels para regiões), os objetos constituintes tornam-se passíveis de reconhecimento e interpretação, possibilitando a posterior tomada de decisão.

Dentre os conceitos e métodos de segmentação existentes¹, um dos mais interessantes pela sua genialidade é o modelo da energia funcional de *Mumford-Shah* [Mumford and Shah 1989]. Este modelo variacional particiona uma imagem I em segmentos representativos W_i , onde a idéia básica é agrupar regiões onde o sinal de I seja homogêneo e delimitado por um sistema de fronteiras qualitativas K . Esta energia funcional é descrita através de um termo qualificador e dois termos regularizadores sendo:

$$E(u, K) = \beta \int_W (u - I)^2 + \int_{W/K} |\nabla u|^2 dx + \lambda \int_K d\sigma, \quad (1)$$

onde:

- o primeiro termo força a aproximação de um segmento u para a imagem I .
- o segundo termo quantifica e controla a suavidade do segmento u .
- o terceiro termo controla o comprimento, a suavidade e a localização das fronteiras K que dividem os segmentos.

De uma maneira geral, o modelo de *Mumford-Shah* acima descrito alcançou uma completa formulação geral para o problema da segmentação de imagens sob a forma de um problema de otimização [Brox 2005] (pg.102): a minimização da funcional $E(u, K)$ possibilita se obter uma imagem em *cartoon* (de regiões, ou W) em relação a uma determinada imagem de entrada I . A idéia geral do modelo de *Mumford-Shah* é buscar pela minimização desta funcional de energia através da otimização dos 3 termos penalizados acima descritos. Os coeficientes β e λ tornam este modelo multi-escala. β pondera a aproximação de I por u , e λ é um parâmetro de contraste que limita o comprimento total de fronteiras na segmentação: quando λ é pequeno, a minimização força os resultados a apresentarem muitas fronteiras (super-segmentação), enquanto que grandes valores de λ produzem o efeito contrário (sub-segmentação, com poucos segmentos)².

No entanto, o processo de otimização dos 3 termos penalizadores; isto é, a forma em como minimizar a energia tem sido deixada em aberto nesta funcional [Brox 2005] (pg.102). Embora o artigo original de *Mumford-Shah* [Mumford and Shah 1989] compreenda quase 100 páginas, este não apresenta uma simples sugestão em como implementar um framework básico para a segmentação [Brox and Cremers 2009]. Só para se ter uma idéia, o modelo chega a ser discutido pela sua genialidade em uma revista de fisiologia [Petitot 2003], apresentando uma visão “pedagógica” do modelo para o leitor não-matemático. Nos modelos analíticos apresentados existe uma grande diferença entre a sua formulação teórica e eficientes meios de se encontrar minimizadores desta funcional na prática [Brox and Cremers 2009]. Na dissertação em anexo são apresentados alguns trabalhos inteiramente dedicados à interpretação do modelo de *Mumford-Shah* (seção 3.2.2, pg.42), e uma estratégia de implementação bem difundida na comunidade científica é o modelo multi-escala proposto por [Koepfler et al. 1994][Morel and Solimini 1995], onde a funcional é acoplada a um modelo de segmentação por crescimento de regiões. De qualquer maneira, um dos maiores desafios identificados nas interpretações e conjecturas do

¹vide documento de dissertação em anexo, seções 2.2 e 2.3.

²conforme demonstrado na imagem em *cartoon* da Figura 3.2 no documento de dissertação (pg.49).

modelo de *Mumford-Shah* é em desenvolver algoritmos eficientes para computar penalizadores de alta qualidade desta funcional [Pock et al. 2009].

2. Método Supervisionado de Segmentação de Imagens

Termos penalizadores podem ser computados através de funções de similaridade, as quais correspondem a um conceito chave para o pleno sucesso ou falha em um algoritmo de processamento de imagens. Por similaridade logo imagina-se proximidade, ou quão duas entidades se assemelham de acordo com suas características. No entanto, a computação de similaridade no contexto de segmentação de imagens está sujeito a muitos fatores, tais como diferentes espaços de cores (RGB, HSV, CIE-Lab, etc) e diferentes métricas de distância (lineares, não-lineares, multivariadas, etc). Em [Sobieranski et al. 2011], é discutido como a grande variabilidade destes fatores pode dificultar na escolha de uma boa métrica de similaridade. Geralmente muitos algoritmos são penalizados pela grande quantidade de parâmetros de execução utilizados para controlar ao mesmo tempo homogeneidade e continuidade dos segmentos.

Nesta dissertação de mestrado optou-se por uma abordagem supervisionada de segmentação de imagens, onde diferentemente das abordagens parametrizadas, a funcional de *Mumford-Shah* é minimizada através de uma função de similaridade não-linear (chamada aqui de função de aproximação), modelada através da inferência do observador externo. Na Figura 4.2 no documento de dissertação é apresentada uma visão geral da metodologia desenvolvida, e esta pode ser sumarizada em 3 etapas:

- **Aprendizado da métrica de distância:** o observador infere sobre a cena, determinando um conjunto de treinamento (denominado de S), através da extração de características do objeto de interesse. Este conjunto S corresponde a uma distribuição multivariada das cores do objeto selecionado, e o objetivo nesta etapa é modelar estatisticamente o comportamento não-linear que S apresenta no espaço de cores de origem. Através desta modelagem, uma topologia M é obtida para ser utilizada como métrica de similaridade na funcional de *Mumford-Shah*;
- **Função de aproximação:** a energia funcional de *Mumford-Shah* é adaptada de modo que a topologia de M atue diretamente com os termos penalizadores da equação. Esta alteração nos termos da energia funcional permite que o processo de minimização ocorra em função desta nova topologia não-linear, alterando a ordem dos agrupamentos das regiões;
- **A etapa final do processo é a execução do algoritmo através do agrupamento das regiões utilizando os novos termos penalizadores.** Nesta etapa o usuário determina o número de regiões esperadas que a funcional deve produzir. O algoritmo inicia em nível de pixel, onde cada pixel é uma região que é verificada em relação aos adjacentes. Regiões similares de acordo a nova $E(u, K)$ são agrupadas sucessivamente, até que o número desejado de regiões seja obtido, conforme pode ser visualizado na Figura 4.2 supracitada, onde a imagem de entrada em questão foi reduzida para 4 segmentos representativos.

Essa não-linearidade é capturada através da distância polinomial de Mahalanobis [Grudic and Mulligan 2006], uma variação da distância estatística onde os termos de S são projetados em alta ordem dimensional. Com isto, as topologias de um determinado conjunto de entrada S podem ser expressas não-linearmente, conforme demon-

strados nos mapas topológicos no documento de dissertação³, seção 4.2.3, ou em [Sobieranski et al. 2011]. Quanto a função de aproximação, esta teve seus termos penalizados alterados de modo que a topologia de M pudesse trabalhar harmonicamente e com regularidade conforme as regiões fossem agrupadas.



Figure 1. Dois conjuntos de treinamento utilizados (objeto e background), demonstrando 2 minimizações diferentes da energia, ambas de 80 para 2 regiões.

Quanto aos resultados obtidos, o método supervisionado proposto foi testado com base no banco de imagens de *Berkeley*⁴, um *dataset* onde cada imagem possui entre 5 a 7 segmentações realizadas manualmente por especialistas humanos. A vantagem da utilização deste *dataset* é a possibilidade de utilizar estas segmentações manuais (denominadas de *ground-truths* - *GT*) para comparar os resultados de diferentes algoritmos de segmentação. Neste contexto, foram utilizadas 4 métricas de avaliação de segmentos, as quais possibilitaram avaliar quão similar é um resultado de segmentação produzido por um algoritmo em relação a uma imagem de *GT*.

O ambiente de experimento foi o seguinte:

- 60 imagens de *Berkeley* foram selecionadas, onde para cada imagem foram selecionados um conjunto de treinamento com base em alguma característica da cena. Em algumas destas imagens o conjunto de treinamento foi variado, totalizando em 63 experimentações. O algoritmo supervisionado (chamado de SMS) processou 22 segmentações para cada uma destas 63 imagens, decrescendo⁵ o número de regiões de 250 para 2, gerando 1.386 resultados. Um exemplo deste decréscimo é apresentado na Figura 1, onde 80 regiões são reduzidas a somente 2;
- cada imagem resultante (das 1.386) foram comparadas com os respectivos *GT*'s (de 5 a 7 por imagem), correspondendo a 7.601 comparações realizadas, conforme demonstrado na Figura 2;
- cada comparação foi regida por 4 índices de avaliação de qualidade: Rand, Fowlkes-Mallows, Jacard, Dongen (referenciados na seção 2.4 da dissertação), totalizando em 30.404 índices de avaliação obtidos.
- várias estatísticas foram geradas a partir destes índices, conforme seção 5 em [Sobieranski et al. 2011].

³Figuras 4.5, 4.7 e 4.9.

⁴<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>

⁵250, 225, 180, ..., 4, 3, 2 regiões, em 22 intervalos.

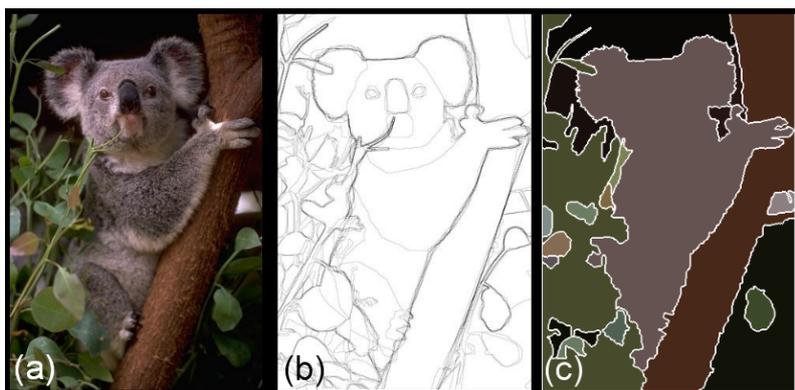


Figure 2. (a) Imagem de entrada, (b) imagem GT, (c) resultado SMS.

O experimento foi subdividido em 3 análises: (a) evolução da qualidade do agrupamento das regiões, com o objetivo de verificar a influência de diferentes conjuntos de treinamento na minimização da funcional (exemplificado na Figura 1); (b) comparação em relação a outros métodos de segmentação estado-da-arte, cujos índices médios de qualidade apresentaram uma vantagem para o método SMS; (c) comparação em relação ao modelo convencional de Mumford-Shah multi-escala.

Todos os experimentos descritos neste resumo estão detalhados no documento de dissertação (capítulo 5), no artigo publicado em [Sobieranski et al. 2011], e adicionalmente no site do método SMS: <http://www.lapix.ufsc.br/sms/>, onde todos os índices e resultados de segmentação podem ser visualizados.

3. Considerações Finais

O modelo supervisionado de segmentação de imagens proposto apresenta resultados superiores (visuais e quantitativos) em relação aos métodos comparados. Em alguns casos, a flexibilidade do método foi verificada através da utilização de diferentes conjuntos de treinamento sobre a mesma imagem. Quando comparado a outros algoritmos de segmentação, SMS consegue ao mesmo tempo descrever melhor os segmentos em um menor número de regiões, conforme demonstrado pelos índices médios de qualidade apresentados na seção 5.3.2 e 5.4 do documento de dissertação.

A grande quantidade de casos utilizados nos experimentos possibilitaram também avaliar indiretamente não somente a qualidade do *dataset* de *Berkeley*, mas também a representatividade dos índices de avaliação de qualidade de segmentações. Esta é uma tendência que vem sendo exigida na comunidade científica, e poucos trabalhos publicados vem realizando este tipo de validação de qualidade. Espera-se com este trabalho ressaltar esta cultura na comunidade científica, além de incentivar o desenvolvimento de novas métricas de qualidade. Quanto às contribuições científicas obtidas ao longo destes aproximadamente 3 anos de pesquisa, destacam-se:

- (preliminar, foco em imagem médica), artigo publicado e apresentado no 22nd IEEE CBMS – International Symposium on Computer-Based Medical Systems [Sobieranski et al. 2009b] – anexo A8 dissertação.
- (preliminar, foco em segmentação), Artigo publicado no periódico Pattern Recognition Letters (qualis A2) [Sobieranski et al. 2009a] – anexo A9 dissertação.

- (outro método de segmentação não relacionado diretamente ao SMS), Artigo publicado no periódico *Pattern Recognition Letters* (qualis A2) [Wangenheim et al. 2009] – anexo A10 dissertação.
- (resultado final dissertação), Artigo publicado no periódico *Computer Vision and Image Understanding* (qualis A1) [Sobieranski et al. 2011] – anexo A11 dissertação.
- site do método SMS (<http://www.lapix.ufsc.br/sms>), onde todas as imagens, tabelas de validação, métodos de avaliação de segmentos e experimentos realizados estão disponíveis para comunidade científica.

References

- Brox, T. (2005). *From pixels to regions: partial differential equations in image analysis*. PhD thesis, Faculty of Mathematics and Computer Science, Saarland University, Germany.
- Brox, T. and Cremers, D. (2009). On local region models and a statistical interpretation of the piecewise smooth Mumford-Shah functional. *Int. J. Comput. Vision*, 84(2):184–193.
- Grudic, G. Z. and Mulligan, J. (2006). Outdoor path labeling using polynomial mahalanobis distance. In *Robotics: Science and Systems*.
- Koepfler, G., Lopez, C., and Morel, J. M. (1994). A multiscale algorithm for image segmentation by variational method. *SIAM J. Numer. Anal.*, 31(1):282–299.
- Morel, J. M. and Solimini, S. (1995). *Variational methods in image segmentation*. Birkhauser Boston Inc., Cambridge, MA, USA.
- Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Pure Appl. Math.*
- Petitot, J. (2003). An introduction to the Mumford-Shah segmentation model. *Journal of Physiology-Paris*, 97(2-3):335 – 342. Neurogeometry and visual perception.
- Pock, T., Cremers, D., Bischof, H., and Chambolle, A. (2009). An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan.
- Sobieranski, A. C., Abdala, D. D., Comunello, E., and von Wangenheim, A. (2009a). Learning a color distance metric for region-based image segmentation. *Pattern Recognition Letters*.
- Sobieranski, A. C., Comunello, E., and von Wangenheim, A. (2011). Learning a nonlinear distance metric for supervised region-merging image segmentation. *Comput. Vis. Image Underst.*, 115:127–139.
- Sobieranski, A. C., Mantelli Neto, S. L., Coser, L., Comunello, E., and von Wangenheim, A. (2009b). Learning a nonlinear color distance metric for the identification of skin immunohistochemical staining. *IEEE Computer-based Medical Systems - CBMS 2009*.
- Wangenheim, A. V., Bertoldi, R. F., Abdala, D. D., Sobieranski, A., Coser, L., Jiang, X., Richter, M. M., Priese, L., and Schmitt, F. (2009). Color image segmentation using an enhanced Gradient Network Method. *Pattern Recogn. Lett.*, 30(15):1404–1412.