

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Um Método Não-Supervisionado de Detecção de Oclusões Textuais para Imagens de Sensoriamento Remoto

Dayara Pereira Basso¹

Departamento de Engenharia de Energia, UNESP, Rosana

Marilaine Colnago²

Departamento de Engenharia de Energia, UNESP, Rosana

Wallace Casaca³

Departamento de Engenharia de Energia, UNESP, Rosana

Resumo. Técnicas de Processamento Digital de Imagens (PDI) têm sido extensivamente empregadas em imagens de sensoriamento remoto. Nesse contexto, classificar alvos ou extrair características de interesse podem se tornar tarefas árduas devido à grande quantidade de oclusões que podem surgir nas imagens como, por exemplo, sombras de nuvens, textos sobrepostos, falhas na captura de dados por sensores, etc. Em vista disso, este trabalho apresenta uma nova metodologia para detecção automática de oclusões do tipo textual em imagens de sensoriamento remoto, a qual combina técnicas de segmentação morfológica, decomposição de imagens e estratégias de limiarização. A solução proposta considera filtros gaussianos para suavizar a imagem de forma a obter uma decomposição da mesma nas bandas “cartoon” e “texture” (padrões oscilatórios). Em seguida, essas bandas são processadas e binarizadas a partir do método de Otsu, gerando assim uma imagem auxiliar a qual é submetida a operações morfológicas, de modo a guiar a detecção textual na imagem original. Testes experimentais em imagens reais remotamente sensoriadas demonstram que o método proposto atinge elevada precisão na tarefa de detecção desse tipo de alvo, devolvendo imagens com áreas textuais segmentadas que podem ser utilizadas para aumentar a acurácia de outras aplicações clássicas da área como classificação e extração de feições de interesse.

Palavras-chave. Análise e processamento de imagens, Morfologia matemática, Detecção automática, Sensoriamento remoto, Filtros de suavização.

1 Introdução

Técnicas de *Processamento Digital de Imagens* (PDI) e de *Reconhecimento de Padrões* têm sido imprescindíveis quando se deseja segmentar uma imagem digital [4]. Essas técnicas tem auxiliado com sucesso problemas complexos que empregam a segmentação como etapa de pré-processamento como, por exemplo, restauração [7], classificação [8] e

¹dayara.pereira@unesp.br

²marilaine.colnago@unesp.br

³wallace.casaca@unesp.br

extração de feições em imagens aéreas [2]. O processo de detecção, tal como foi abordado nesse trabalho, objetiva particionar a imagem em segmentos ou subregiões, onde esses segmentos representam obstruções que poderiam, a posteriori, auxiliar na melhoria da qualidade visual das imagens, em particular, de sensoriamento remoto [1]. Nesta pesquisa, tem-se como foco as obstruções do tipo textual devido ao limitado volume de contribuições e baixa pluralidade de técnicas voltadas à detecção desse tipo mais específico de obstrução, principalmente na área de sensoriamento remoto.

De fato, na literatura geral de PDI, algumas abordagens foram propostas para detectar alvos textuais em fotos digitais. Dentre essas, podemos citar [11], a qual detecta degenerações em fotografias a partir de métodos de correlação, [9], que realiza a detecção de logotipos em imagens de documentos, e [6], que efetua reconhecimento e extração de textos *Bangla*⁴. Notadamente, diversas abordagens referem-se à detecção de regiões textuais em imagens digitais mais específicas como de idiomas e documentais. No entanto, a literatura ainda carece de uma diversidade maior de trabalhos e algoritmos destinados exclusivamente à tarefa de detecção textual em imagens de sensoriamento remoto, em especial, métodos inteiramente não-supervisionados tal como é o objetivo deste trabalho. Imagens de sensoriamento remoto tem sido alvos comuns de oclusões, contendo na sua composição cenas com nomes de ruas, topomínias, sombras de nuvens, erros de captura por sensores, entre outros artefatos indesejáveis não constituintes à cena, o que torna inválido o uso posterior da informação original que ali deveria estar presente.

Diante do exposto, esse trabalho propõe uma nova metodologia de detecção automática de oclusões textuais em imagens de sensoriamento remoto, na forma de um algoritmo integrado que, além de gerar bons resultados, procura empregar técnicas clássicas de PDI como segmentação morfológica e decomposição de imagens. A referida metodologia compreende as seguintes etapas básicas: inicialmente, a imagem é pré-suavizada de forma a obter as regiões de alta oscilação da mesma. Em seguida, as bandas originadas são binarizadas gerando uma máscara-guia. Por fim, uma sequência de operações morfológicas é aplicada na máscara a fim de guiar a detecção textual e obter o texto realçado. A saída (*output*) é uma imagem binária com os pixels do primeiro plano representando o texto.

2 Metodologia

A abordagem proposta visa detectar, de maneira automatizada, as regiões textuais de imagens de sensoriamento remoto através de uma combinação de técnicas de segmentação morfológica, filtragem gaussiana e de decomposição de imagens. A seguir, o algoritmo proposto é apresentado, e seus passos, discutidos em detalhes:

1. O método recebe como argumento de entrada uma imagem colorida I (Figura 1(a)), representada no sistema de cores RGB , sendo ela submetida a um filtro de suavização gaussiano (com $\sigma = 1$) a fim de gerar uma imagem suavizada I_{suave} (Figura 1(b)). Na sequência, a imagem componente I_{po} (Figura 1(c)) contendo os padrões oscilatórios da imagem original I é obtida pela equação $I_{po} = I - I_{suave}$, tal como realizado pela decomposição de imagens apresentada em [3].

⁴Língua indo-ariana falada pelas populações de Bangladesh e pelo estado indiano de Bengala Ocidental.

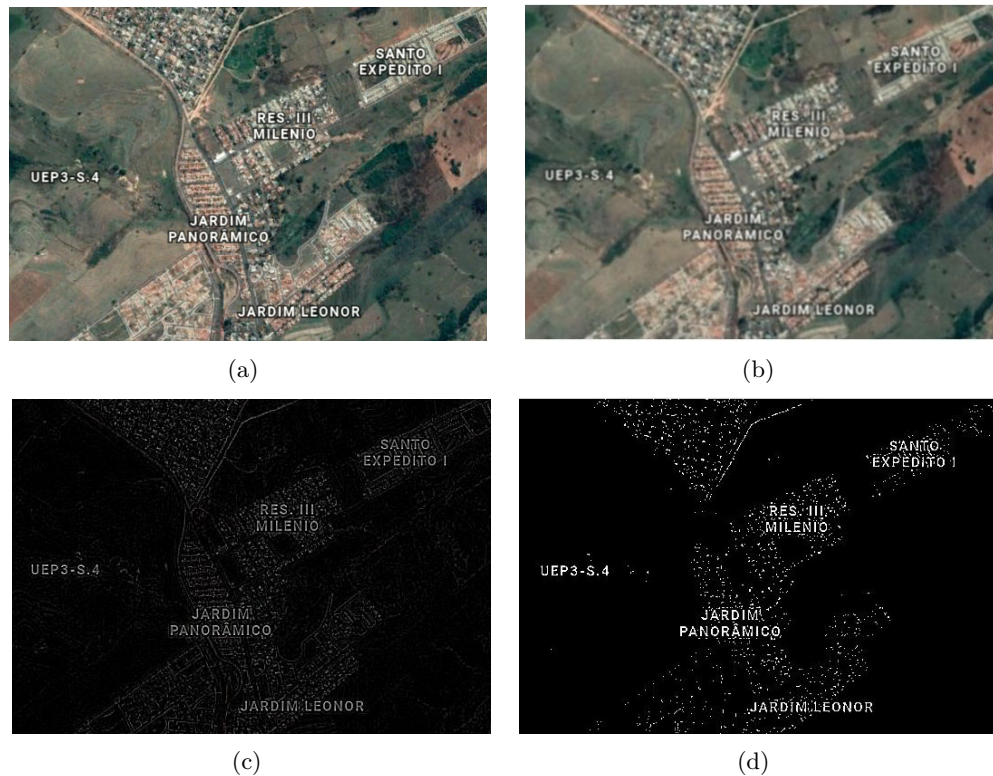


Figura 1: Principais etapas do processo da detecção textual da metodologia proposta. (a) Imagem original (fonte: *Google Maps*). (b) Imagem suavizada (*cartoon*). (c) Banda contendo parte dos textos e os padrões oscilatórios da imagem. (d) Imagem máscara M_1 .

2. Na sequência, o algoritmo calcula, por meio do método de Otsu [10], um valor base γ , o qual é para binarizar cada um dos canais de cores de I_{po} . Para isso, o método de Otsu utiliza a função *graythresh* do MATLAB, que determina um limiar para minimizar a variação intraclasse dos pixels da imagem, definido em termos de um valor de luminosidade escalar. Como saída, tem-se a imagem binária M_{po} .
3. Em seguida, através da função *bwconncomp*, são calculados os componentes conexos (CC) do binário M_{po} . Após isso, removem-se os CC que possuem valores inferiores ao CC médio observado para essa banda. A partir disso, tem-se uma nova imagem binária (M_1), composta por áreas de textos e resíduos de outros objetos da imagem (Figura 1(d)). M_1 atua como uma máscara nos próximos passos.
4. Como M_1 ainda apresenta elementos não-textuais, uma segunda sequência de passos é realizado, partindo-se agora da imagem original e da máscara M_1 . Inicialmente, dilata-se M_1 a partir de um disco de raio 1 (pixel) como elemento estruturante⁵. A partir do resultado dessa dilatação, aplica-se uma erosão, considerando o mesmo elemento estruturante anterior, o que resulta no binário M_2 . Por fim, calcula-se o valor médio dos pixels da imagem I mediante à região capturada na máscara M_2 .

⁵<https://www.mathworks.com/help/images/ref/strel.html>.

5. Sequencialmente, calcula-se a diferença entre a média obtida no passo anterior com cada pixel da imagem original, definindo-se como região textual todo pixel que obteve diferença menor que um dado limiar L , aqui escolhido empiricamente como $L = 200$ (valor obtido a partir de uma bateria de testes experimentais). Finalmente, a imagem resultante é dilatada com um elemento estruturante de um quadrado, de largura $w = 2$, no intuito de dar destaque ao texto extraído.

3 Resultados

A fim de avaliar o método proposto, foram considerados três estudos de caso. O algoritmo formulado, para fins de automação, assume como entrada somente a imagem original, e devolve uma imagem binária, onde os alvos textuais estão em primeiro plano, sendo que nenhum pós-ajuste foi realizado para os casos aqui descritos. Além disso, a acurácia (AC) [5] e a métrica F -score (F) [5] foram computadas para todos os exemplos.

O primeiro exemplo pode ser visualizado na Figura 2(a), onde tem-se uma imagem de sensoriamento remoto composta tanto por telhados de residências quanto por áreas verdes, reunindo diferentes tipos de padrões na mesma imagem. O texto sobreposto refere-se aos nomes dos bairros que correspondem à cidade de Presidente Prudente (SP). Na Figura 2(b), tem-se o resultado da detecção, onde nenhum dos caracteres foi omitido, tendo sido os mesmos detectados com o mesmo nível de precisão independentemente da região da imagem em que se encontravam. Para esse caso, obteve-se $AC = 99.96\%$ e $F = 98.90\%$.

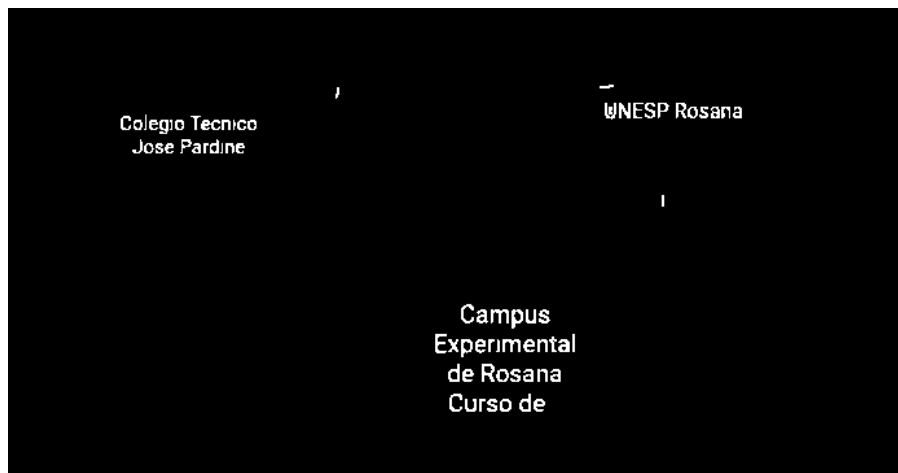
O próximo experimento também atesta a boa acurácia do método proposto na tarefa de detecção dos alvos textuais (vide Figura 3), resultando em uma medida de acurácia $AC = 99.97\%$ e de F -score $F = 98.78\%$. Neste exemplo, é possível observar que o texto sobreposto se mistura junto aos padrões homogêneos da imagem, principalmente na região dos telhados das construções, o que dificulta a distinção de padrões. Todavia, o método proposto conseguiu distinguir de forma satisfatória o texto dos demais elementos da imagem (vide Figura 3(b)).



Figura 2: (a) Imagem original (fonte: *Google Maps*). (b) Imagem final resultante.



(a)



(b)

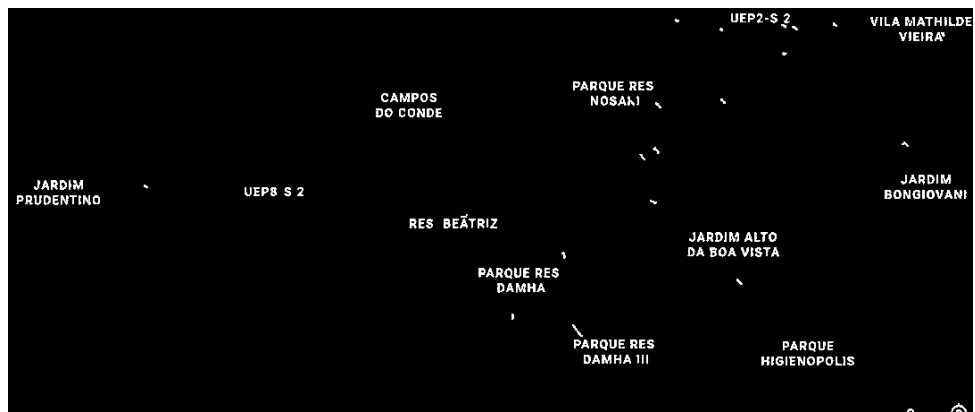
Figura 3: (a) Imagem original (fonte: *Google Maps*). (b) Imagem resultante.

Com o intuito de demonstrar a eficácia do método proposto, em imagens de elevado grau de complexidade, o mesmo foi aplicado à Figura 4(a), que além de possuir um grande número de alvos textuais, também apresenta um número bastante diversificado de elementos, incluindo vários agrupamentos de construções da região urbana de Presidente Prudente (SP).

O resultado obtido, após a aplicação do algoritmo proposto, é apresentado na Figura 4(b), onde foi gerada uma máscara textual com alta precisão de detecção dado o nível de detalhamento da imagem. Embora tenha-se a existência de alguns falsos positivos, o método foi meticuloso sem penalizar nenhum caractere em específico. Sendo assim, a abordagem proposta se manteve totalmente autônoma, isto é, sem que houvesse a intervenção do usuário e sem perder seu alto desempenho quanto à tarefa de detecção. Neste exemplo, obteve-se $AC = 99.90\%$ e $F = 97.08\%$.



(a)



(b)

Figura 4: (a) Imagem original (fonte: *Google Maps*). (b) Imagem resultante.

4 Conclusão

Este trabalho apresentou uma nova metodologia não-supervisionada para a tarefa de detecção de conteúdo textual em imagens de sensoriamento remoto. O método implementado obteve um nível de precisão bastante satisfatório na referida tarefa de identificação textual, onde para tal, foram utilizadas apenas técnicas clássicas de PDI como filtros de morfologia matemática e de suavização gaussiana, além de outras funções básicas disponíveis no MATLAB. Testes experimentais foram conduzidos com a metodologia proposta a partir de três imagens de satélite, onde foi possível constatar um comportamento linear do método para todos os casos estudados, tanto em termos boa acurácia como em nível de automatização.

Sendo assim, conclui-se que a detecção de alvos textuais foi realizada com sucesso pelo algoritmo proposto, o qual apresentou um desempenho robusto por intermédio das segmentações resultantes. Assim, essas imagens poderiam ser úteis se combinadas a outras aplicações-fins como, por exemplo, o processo de reconstrução de áreas textuais a fim de melhor classificar as regiões de interesse das imagens.

Agradecimentos

Os autores deste trabalho gostariam de agradecer à FAPESP (Proc. #2018/06756-3) pelo fomento à pesquisa.

Referências

- [1] S. Azevedo, G. Cardim, W. Casaca, E. Silva and R. Signh. Shadows Removal in High Resolution Remote Sensing Images using Local Inpainting Strategy, *Pecora 20: Observing a Changing Earth*, 2017.
- [2] J. A. Benediktsson, M. Pesaresi and K. Arnason. Classification and Feature Extraction for Remote Sensing Images From Urban Areas Based on Morphological Transformations, *IEEE Transactions on Geoscience and Remote Sensing*, 41:1940–1949, 2003. DOI:10.1109/tgrs.2003.814625.
- [3] W. Casaca and M. A. Boaventura. A Decomposition and Noise Removal Method Combining Diffusion Equation and Wave Atoms for Textured Images, *Mathematical Problems in Engineering*, 2010:1-21, 2010. DOI: 10.1155/2010/764639.
- [4] W. Casaca, L. G. Nonato and G. Taubin. Laplacian Coordinates for Seeded Image Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, p. 384-391, 2014. DOI: 10.1109/cvpr.2014.56.
- [5] T. Fawcett. An Introduction to ROC Analysis. *Pattern Recognition Letters*, 27:861–874, 2006. DOI: 10.1016/j.patrec.2005.10.010.
- [6] M. Al M. Hasan, M. A. A. Mehedi and M. W. Islam. A New Approach to Bangla Text Extraction and Recognition from Textual Image. *Proceedings of ICCIT*, 2005.
- [7] L. Lorenzi, F. Melgani and G. Mercier. Inpainting Strategies for Reconstruction of Missing Data in VHR Images, *IEEE Geoscience and Remote Sensing Letters*, 8:914-918, 2011. DOI:10.1109/lgrs.2011.2141112.
- [8] R. G. Negri, E. Silva and W. Casaca. Inducing Contextual Classifications With Kernel Functions Into Support Vector Machines, *IEEE Geoscience and Remote Sensing Letters*, 15:962-966, 2018. DOI:10.1109/lgrs.2018.2816460.
- [9] T. D. Pham. Unconstrained Logo Detection in Document Images. *Pattern Recognition*, 36:3023-3025, 2003. DOI: 10.1016/s0031-3203(03)00125-0.
- [10] M. A. B. Siddique, R. B. Arif and M. M. R. Khan. Digital Image Segmentation in Matlab: A Brief Study on Otsu’s Image Thresholding. *International Conference on Innovation in Engineering and Technology*, 2018. DOI: 10.1109/ciet.2018.8660942.
- [11] A. Toshiyuki. Correlation based Image Defect Detection. *IEEE International Conference on Pattern Recognition*, volume 1, p. 163-166, 2006. DOI: 10.1109/icpr.2006.419.