

Estudo da classificação de imagens a partir de sensoriamento remoto de Santa Catarina

Artur Ricardo Bizon
Universidade Regional de Blumenau
abizon@furb.br

Luciana Pereira de Araújo
Kohler
Universidade Regional de Blumenau
lpa@furb.br

Adilson Luiz Nicoletti
Universidade Regional de Blumenau
alnicoletti@furb.br

Fernanda Dal Bosco
Universidade Regional de Blumenau
fbosco@furb.br

Marlon Yuri Andrade
Universidade Regional de Blumenau
myandrade@furb.br

Murilo Schramm da Silva
Universidade Regional de Blumenau
muriloschramm@furb.br

Thales Bohn Pessatti
Universidade Regional de Blumenau
tbpessatti@furb.br

Julio Cesar Refosco
Universidade Regional de Blumenau
refosco@furb.br

Veraldo Liesenberg
Universidade do Estado de Santa
Catarina
veraldo.liesenberg@udesc.br

Alexander Christian Vibrans
Universidade Regional de Blumenau
acv@furb.br

ABSTRACT

The mapping of the use of the land in a certain region allows the identification of the available resources in it. In that way, the area of remote sensing searches for ways of processing large amounts of data with reliability in its results. This paper compares the automated process of image, along with a spacial filter, with the image edited manually. The overall accuracy for each categorized image resulted in 71,3% for simple category, 72,6% for filtered category and 94,5% for manually edited category.

KEYWORDS

Sensoriamento Remoto, Random Forest, Filtro majoritário

1 INTRODUÇÃO

Sensoriamento remoto é a prática de utilização de ferramentas para a coleta de dados, bem como o uso de satélites para a coleta de informações sobre o uso e cobertura da terra, sem ter contato direto com a área em estudo [1]. Desta forma, acontece a identificação de diferentes usos da terra, por exemplo, corpos de água, floresta e agricultura.

O sensoriamento remoto pode ser utilizado no processamento de imagens abrangendo áreas que podem ter dimensões locais como uma propriedade ou até alcançado grandes áreas de caráter nacional e/ou mundial. A prática de processamento de imagens de satélite vem crescendo com o passar do tempo. Entretanto, no início do sensoriamento remoto, os dados produzidos a partir de imagens de satélites, eram processados manualmente com o uso de ferramentas *Geographic Information System* (GIS).

Por definição, um sistema GIS é um sistema para captura, armazenamento e análise de dados geograficamente referenciados [2]. O sistema GIS se mostrou muito útil e com o aumento de interesse

sobre sua capacidade de processamento, este tipo de sistema se tornou mais acessível por ser possível realizar mais funcionalidades, sem ter muita experiência com o uso da ferramenta [2].

Com a evolução do sistema GIS, a criação de mapas temáticos passou a ser mais fácil, entretanto ainda de forma manual. Como o mapeamento manual é trabalhoso, formas automatizadas de mapeamento estão sendo desenvolvidas e aprimoradas, desta forma sendo possível gerar mapas de determinada região com mais frequência ou mapas de áreas maiores.

Mapas temáticos são mapas que mapeiam determinados atributos de uma área [3], como o tipo de sua cobertura. Como este é um trabalho em andamento, o objetivo deste estudo é realizar a comparação da classificação de imagens com o algoritmo *Random Forest* com e sem a utilização do Filtro Majoritário.

Esta comparação será realizada com o intuito de identificar possíveis melhorias no processamento automatizado de imagens, para substituir ou ao menos diminuir a necessidade do uso de técnicas de classificação por interpretação visual.

2 SOLUÇÃO PROPOSTA

Os mapas temáticos, são gerados utilizando imagens do satélite *Landsat 8* da NASA coleção 1 *Level 2* (*Surface reflectance*). No pré-processamento das imagens, aplica-se o algoritmo de correção topográfica *Sun-Canopy-Sensor + C* (SCS + C) proposto por [4].

Após a correção topográfica, a imagem é classificada pelo classificador *Random Forest* (RF) disponibilizado pela API *Scikit-learn*. O RF é um classificador *ensemble learn* que se baseia em classificadores estruturados em formato de árvore [5], sendo um exemplo o classificador de árvore de decisão *Decision Tree*. A quantidade de árvores utilizadas na classificação foi de 500. Segundo [6], 500 árvores podem ser consideradas um número de árvores adequado para o classificador RF. Para o treinamento do classificador, foram

utilizadas amostras baseando-se em imagens de alta resolução e informações de campo.

Na classificação da imagem foi aplicado o filtro majoritário. O filtro majoritário é um filtro espacial que leva em consideração o valor dos vizinhos de determinado pixel, sendo que para cada pixel é atribuído determinado valor. Este valor foi retirado a partir da moda do valor do pixel e de seus vizinhos. Após a execução do filtro majoritário, uma edição manual dos erros grosseiros foi realizada sobre a classificação filtrada.

3 VALIDAÇÃO DA IMAGEM

Com a imagem devidamente processada, aplica-se a metodologia de testes definida por [7]. Inicialmente são sorteados aleatoriamente os pontos de teste conforme a representatividade de cada classe do uso da terra no mapa. Em seguida, um avaliador confere a verdade terreno de cada ponto de teste sorteado com uma imagem de melhor qualidade. Caso o ponto esteja incorreto, o avaliador define a verdade terreno do ponto que posteriormente é utilizado no cálculo da acurácia.

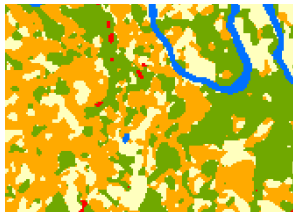


Figura 1: Classificação filtrada com o filtro majoritário

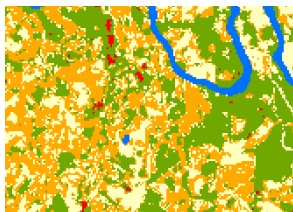


Figura 2: Classificação sem a utilização de filtro

Após todos os pontos de testes serem reavaliados gera-se uma matriz de confusão para posterior cálculos da acurácia global, do usuário e do produtor. Além da análise dos dados estatísticos retirados da matriz de confusão e dos cálculos de acurácias, também é feita uma avaliação visual no intuito de validar os resultados estatísticos visualmente.

4 COMPARAÇÃO DAS IMAGENS

A comparação foi realizada através dos critérios de avaliação apresentados em três diferentes classificações. A primeira foi a imagem apenas com a classificação utilizando o *Random Forest* com 500 árvores de decisão, a segunda com a aplicação do filtro majoritário sobre a classificação do RF e a terceira imagem é a classificação editada a partir de um sistema GIS.

5 RESULTADOS E DISCUSSÕES

Após a classificação, obteve-se como mapeamento da folha a compreensão das seguintes classes: “Agricultura”, sendo que a mesma compreende tanto agricultura perene, quanto atividades agrícolas intermitentes; “Floresta”, formações florestais em estágio médio e avançado de regeneração e estágio primário; “Pastagem”, compreende as pastagens destinadas a atividade pecuária, tanto quanto cobertura de terra por gramíneas; “Reflorestamento”, áreas de florestas plantadas e/ou plantio comerciais; “Áreas Construídas”, qualquer construção independente de sua inclusão no perímetro urbano ou rural; e “Corpos de água”, qualquer acúmulo de água continental.

Vale ressaltar que áreas como estradas e rodovias não são mapeadas, pois dada a resolução do pixel das imagens utilizadas, torna-se inviável o mapeamento de estradas. Isso porque normalmente elas possuem menos do que 30 metros de largura, sendo que 30 metros é a área de detecção de cada pixel das imagens usadas neste estudo. Com relação as áreas área urbanas, optou-se por mascaras-las logo após a classificação, pois dado sua amplitude espectral, terá inúmeros erros de comissão nas outras classes.

Foram analisadas três formas diferentes de processamento de imagem: a primeira apenas a classificação do *Random Forest*, na segunda forma a classificação junto com a aplicação do filtro majoritário e a última uma edição manual da classificação. Os resultados de acurácia para as três formas diferentes de pós-processamento são apresentados na Figura 3 (acurácia global e acurácia de usuário).

Com base nas acurácias apresentadas pode-se observar que a aplicação do filtro majoritário, que tem como principal intuito de diminuir o efeito *salt and pepper* da classificação, contribuiu para o aumento da acurácia do mapa. Este efeito pode ser melhor observado quando comparado duas classificações: uma com o filtro e outra sem o filtro, exemplo Figura 2 e Figura 1.

Com base na tabela de acurácias apresentada é possível observar que para esta imagem classificada há a necessidade de edição manual do resultado da classificação. Entretanto, o processo de edição é lento, trabalhoso e oneroso, logo a quantidade de energia posta nesta atividade nem sempre pode ser recompensada. Na classificação que recebeu a edição manual houve um aumento de aproximadamente 23% na acurácia global em relação à classificação sem essa edição.

6 CONSIDERAÇÕES FINAIS

O artigo apresentou um estudo sobre a comparação da classificação de imagens com o algoritmo *Random Forest* com e sem a utilização do Filtro Majoritário. Pode-se observar que o uso do filtro melhorou a representação visual da imagem, além da acurácia. Contudo, a acurácia ainda pode ser melhorada por meio de outras técnicas.

Uma possível alternativa para a melhora da acurácia da classificação seria uma melhor seleção nos pontos de treino. Baseando-se nas acurácias de usuário obtidas, segundo [8] a acurácia de usuário e/ou erro de comissão ($1 - \text{acurácia de usuário}$) pode ser melhorado a partir de uma amostragem de treino proporcional ao tamanho de cada classe.

Outra alteração seria a utilização de um filtro espacial que elimina grupos de *pixels*, pois como pode ser percebido na Figura 1, ainda existem pequenos ruídos que por serem formados de mais de um *pixel*, o filtro majoritário não os eliminou por completo.

	Folha SG 22 Z C IV	Classificação		Classificação com filtro majoritário		Classificação com edição manual	
		Acurácia	Intervalo de confiança	Acurácia	Intervalo de confiança	Acurácia	Intervalo de confiança
		Acurácia global	0,7136	0,0307	0,7259	0,0304	0,9454
Acurácia de Usuário	1 - agricultura	0,5346	0,0665	0,5769	0,0673	0,9605	0,0311
	2 - floresta	0,7257	0,0569	0,7098	0,0558	0,9434	0,0312
	3 - pastagem	0,7733	0,0548	0,7788	0,0542	0,9283	0,0311
	4 - reflorestamento	0,8842	0,0647	0,9294	0,0548	0,9769	0,0259
	5 - área construída	0,9231	0,0731	0,9231	0,0731	0,9344	0,0626
	6 - corpos de água	0,9565	0,0485	0,9565	0,0485	0,9600	0,0446

Figura 3: Acurácias de usuário e acurácia global.

Dessa forma, como trabalhos futuros, pretende-se melhorar a acurácia da classificação das imagens de forma automática por meio do uso de outras técnicas atreladas ao *Random Forest*.

REFERÊNCIAS

- [1] John A. Richards. *Remote Sensing Digital Image Analysis: An Introduction*. Springer, 2012. URL <https://www.amazon.com/Remote-Sensing-Digital-Image-Analysis-ebook/dp/B00A9YH460?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimb05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=B00A9YH460>.
- [2] A. K. Deakin. *Evolution of geographic information and visualization systems*, volume II of *Environmental Monitoring*. Encyclopedia of Life Support Systems, 2007.
- [3] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls. Multimodal classification of remote sensing images: A review and future directions. *Proceedings of the IEEE*, 103(9):1560–1584, Sep. 2015. ISSN 0018-9219. doi: 10.1109/JPROC.2015.2449668.
- [4] S. A. Soenen, D. R. Peddle, and C. A. Coburn. Scs+c: a modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2148–2159, Sep. 2005. ISSN 0196-2892. doi: 10.1109/TGRS.2005.852480.
- [5] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001. ISSN 0885-6125. doi: 10.1023/A:1010933404324. URL <https://doi.org/10.1023/A:1010933404324>.
- [6] Mariana Belgiu and Lucian Drăguț. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, April 2016. doi: 10.1016/j.isprsjprs.2016.01.011. URL <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- [7] Pontus Olofsson, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148:42–57, May 2014. doi: 10.1016/j.rse.2014.02.015. URL <https://doi.org/10.1016/j.rse.2014.02.015>.
- [8] Huiran Jin, Stephen V. Stehman, and Giorgos Mountrakis. Assessing the impact of training sample selection on accuracy of an urban classification: a case study in denver, colorado. 2014.