The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-3/W12-2020, 2020
2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS 2020), 22–26 March 2020, Santiago, Chile

AERIAL IMAGE SEGMENTATION IN URBAN ENVIRONMENT FOR VEGETATION MONITORING

José Martins[1],*Diego André Sant'Ana[2,3], José Marcato Junior[1], Hemerson Pistori[3], Wesley Nunes Gonçalves[1,4]

[1] Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brazil
(jose.a,jose.marcato)@ufms.br
[2] Federal Institute of Mato Grosso do Sul, Aquidauana, Brazil
diego.santana@ifms.edu.br
[3] Dom Bosco Catholic University, Department of Local Development, Inovisão, Campo Grande, Mato Grosso do Sul, Brazil
pistori@ucdb.br
[4] Faculty of Computer Science, Federal University of Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brazil
wesley.goncalves@ufms.br

ICWG

KEY WORDS: SLIC, Aerial Image, Computer Vision, Classifiers, Geoscience, under-sampling , Machine Learning

ABSTRACT:

Urban forests are crucial for the population well-being and improvement of the quality of life. For example, they contribute to the rain damping and to the improvement of the local climate. Therefore a correct and accurate mapping of this resource is fundamental for its correct management. We investigated a method that combines machine learning and SLIC superpixel techniques using different Superpixels (k) number to map trees in the metropolitan region of the municipality of Campo Grande-MS, Brazil with aerial orthoimages with GSD (Ground Sample Distance) of 10 cm. The combination of superpixels and machine learning algorithms were checked out with a set of weka classifiers and achieved good results i.e. F-1 %98.2, MCC %88.4 and Accuracy of % 96.8, supporting that this method is efficient when used for urban trees mapping.

1. INTRODUCTION

Urban forests, provide a variety of ecosystem services to city-dwellers, such as air purification and humidification, good environment for recreation, better drainage and many others (Gómez-Baggethun, Barton, 2013, Davies et al., 2011, Baró et al., 2014). In towns and cities, the abundance and diversity of trees perform a significant role in determining the structure and composition of fauna and flora(Evans, 2013, Stagoll et al., 2012). For the good preservation and maintenance of this important environment, urban and regional planners require the nearly continuous acquisition of data to formulate governmental policies and programs, this needed data acquisition and evaluation happens because of the dynamic characteristic of the urban environment (Seto et al., 2012, Gaston, 2000). With this goal, an extensive and continuous data acquisition of the environment necessary, and with nowadays computational capacity and data acquisition tools, this task is made possible, with the historycally comproved capacity of discerning land cover types (Nichol, Lee, 2005, Small, 2001, Zhang et al., 2010), remote sensing has been widely used for vegetation mapping in diverse environments. Methods for monitoring vegetation in urban areas using satellite and aerial imagery are a hot topic in remote sensing for a considerable time (Nichol, Lee, 2005, Small, 2001, Zhang et al., 2010, Tigges et al., 2013, Li, Shao, 2013, Fassnacht et al., 2016, dos Santos Ferreira et al., 2017, Georganos et al., 2018, Yan et al., 2018, Zhou et al., 2019, Luo et al., 2019, Tigges et al., 2013, Li, Shao, 2013).

Following this line, the objective of this paper is to investigate the use of machine learning (ML) methods to perform the mapping of tree canopies in a dense urban environment. For this task, we used the Pynovisão software (dos Santos Ferreira et al., 2017), which uses the simple linear iterative clustering (SLIC) superpixel technique (Achanta et al., 2012) associated to ML methods. In our study case, high resolution aerial orthoimages obtained in a dense urban environment were used. Finally, we performed the classification using a set of classifiers, that historically provide good results for image classification task, such as: SVM (Platt, others, 1999); Random forest (Ho, 1995) and; IBK (Aha et al., 1991).

*Corresponding author

2. METHODS

2.1 Image Acquisition

The dataset used in this study includes aerial orthophotos obtained from Campo Grande municipality, Mato Grosso do Sul, Brazil, located in the geographic coordinates Latitude: 20 ° 26 '37' 'South, Longitude: 54 ° 38' 52 " West. A total 1323 orthophotos composes the complete dataset. For this study, we used only one ortophoto. Each orthophoto has a range of 561.9 by 594.6 meters, so 33410,74 m2. The GSD (Ground Sample Distance) of the image is 10 cm. The total number of pixels per image is equal to 334105740. The orhophotos are from 2013 and were made available by the municipal government of Campo Grande . The map of the studied area can be seen in Figure 1.

2.2 Software

In this study we used only 3 softwares, Quantum GIS (Qgis)(QGIS Development Team, 2019), for the manual seg-

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-349-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165618

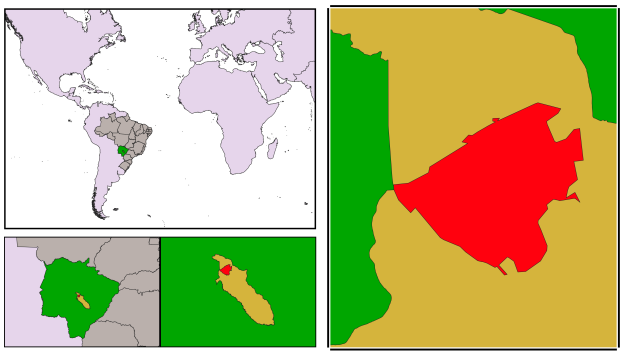


Figure 1. In the world map, we have to detach Brazil, in green, the state of Mato Grosso do Sul in dark yellow the province of Campo Grande and in red the metropolitan region of Campo Grande, where the image was obtained

metation. Pynovisão (dos Santos Ferreira et al., 2017), for the atributes extraction, Superpixels segmentation and classification. Api Rasterio (rasterio.readthedocs.io/en/stable), for the association between the manual image notation and the segmentation provided by Pynovisão.

2.3 Manual image annotation

The orthophoto was uploaded in QGIS software to manually annotate the tree canopies. After that, the archive was uploaded to the api rasterio to transform the file into a mask image. Thus, allowing the use of the mask in Pynovisão software to perform the selection of superpixels class.

2.4 Segmentation process and feature extraction

For the task of mapping the vegetation of urban areas, we adopt the robust segmentation algorithm: simple linear iterative clustering (SLIC), with different numbers of superpixels (k) configurations, to create the superpixels. The Pynovisão software is used in the majority of the process performed in this work, including the segmentation, feature extraction and classification.

In the process of segmentation, using the Pynovisão interface, we are able to search for the best parameters k, which is the number of superpixels. To choose the k values that best fit our task, we considered values equal to 3000, 6000 and 9000. Then, based on the mask, these k segments are divided into two classes background and Tree. For compactness and sigma, we considered the values of 10 and 5, respectively. We reach this number because they make a good representation of the image characteristics and have a multiplier factor between then, making the process a little more intuitive.

In the aerial image, the area of the orthophoto representing background was much bigger than the tree canopies area resulting in unbalanced dataset. Because of this difference, we performed an under-sampling (US) process to equalize the k number of Tree and Background. In this work, we used 3000, 6000, and 9000 (unbalanced) and also applied this configuration to under-sampled (UD.) (balanced) datasets.

The Pynovisão software performed the extraction of attributes with the extractors presented in Table 2.

Table 1. Extractors implemented in Pynovisão

Extractor	Citation
Color Statistics	(Swain, Ballard, 1991)
Gray-Level Co-Ocurrence	(Soh, Tsatsoulis, 1999)
Histogram of Oriented Gradients	(VAN KLAVEREN et al., 1999)
Hu Image Moments	(Ming-Kuei Hu, 1962)
Image Moments (Raw/Central)	(Ming-Kuei Hu, 1962)
Local Binary Patterns	(VAN KLAVEREN et al., 1999)
Gabor Filter Bank and	(Feichtinger, Zimmermann, 1998)
K-Curvature Angles	(Abu Bakar et al., 2015)

Table 2. Dataset used with classes Background and Tree

Superpixels Configuration (k)	Background segments	Tree segments
3000	2202	224
6000	4542	506
9000	6967	820
3000 US.	224	224
6000 US.	506	506
9000 US.	820	820

Together, these extractors add up to 405 attributes for each superpixel in the dataset. Then, we performed an under-sampling method in order to balance the k of background and tree classes. In this way, both classes will have the same k number.

2.5 Classifiers

In order to verify the quality of the segmentation processes, a set of Weka classifiers (Hall et al., 2009) was used to classify the superpixels. For this task we used the classifiers Support Vector Machine (SVM)(Keerthi et al., 2001), K-nearest neighbours classifier (IBK)(Aha et al., 1991) and Random Forest (RF)(Breiman, 2001) with a 10-fold Cross-Validation. For the evaluation, we used the following metrics: F-measure (F1), Matthews correlation coefficient (MCC), and accuracy. They allow a more detailed analysis of the results presented in the Confusion matrix. These will be called derivations matrix.

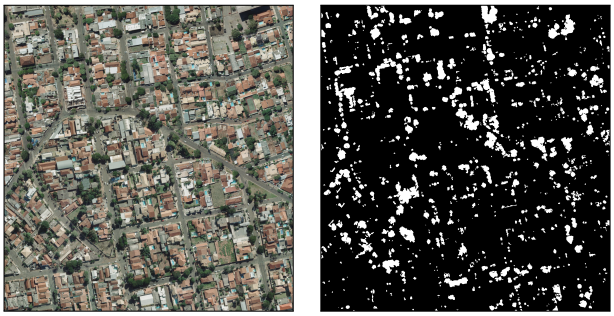
2.6 Hardware

The operating system used to process the images and segmentation was a Linux Ubuntu 18 and the hardware used for the SLIC segmentation, and the process of training the classifiers (IBK, SVM and Random Forest) had a configuration of motherboard Ryzen 7 (1800x); 16GB Ram; 240 GB SSD;Titan Xp (12GB) video card.

3. RESULTS

The original image used in the paper can be seen in Figure 2, we can also see the mask created using Rasterio and the QGIS tool. The effect in the image of the superpixels segmentation with SLIC with different k values can be seen in Figure 3. In this image, the k values used are 3000, 6000, and 9000.

In the confusion matrix presented in Figure 4, we can see the performance of the three algorithms and 6 configurations. At the end of the process, we observe that the classifiers with the



(a)Original Image (b)Mask Image

Figure 2. Original image(a) and Mask Image created using Qgis and Rasterio(b)

Figure 3. Original image with a detach for a particular region. The impact of different values for k can be verified in the amplified images.Superpixels are in yellow

		PREDICTED						
		BK	Tree	BK	Tree	BK	Tree	
REAL	BK	2174	28	2065	137	2181	21	3000
	Tree	50	174	59	165	90	134	
	BK	201	23	160	64	197	27	3000_UD
	Tree	11	213	11	213	12	212	
	BK	4466	76	4251	291	4495	47	6000
	Tree	99	407	125	381	169	337	
	BK	469	37	398	108	460	46	6000_UD
	Tree	22	484	19	487	23	483	
	BK	6856	111	6504	463	6908	59	9000
	Tree	159	661	166	654	260	560	
	BK	752	68	610	210	748	72	9000_UD
	Tree	37	783	31	789	43	777	
		SMO		IBK		RF		

Figure 4. The confusion matrix generated by the classifiers with the different k configurations and balance. The colors illustrate the results of the classification process with blue being the highest results and red the lowest. The colors are separated by the k number.

F1	**0,982**	0,955	0,975	3000	K Number and balance
	0,922	0,810	0,910	3000_UD	
	0,981	0,953	0,977	6000	
	0,941	0,862	0,930	6000_UD	
	0,981	0,954	0,977	9000	
	0,935	0,835	0,929	9000_UD	
	SMO	IBK	RF		
	CLASSIFIERS				

MCC	0,801	0,591	0,697	3000	K Number and balance
	0,849	0,685	0,828	3000_UD	
	0,804	0,609	0,743	6000	
	0,884	0,761	0,865	6000_UD	
	0,812	0,640	0,765	9000	
	0,873	0,724	0,860	9000_UD	
	SMO	IBK	RF		
	CLASSIFIERS				

ACCURACY	**0,968**	0,919	0,954	3000	K Number and balance
	0,924	0,833	0,913	3000_UD	
	0,965	0,918	0,957	6000	
	0,942	0,875	0,932	6000_UD	
	0,965	0,919	0,959	9000	
	0,936	0,853	0,930	9000_UD	
	SMO	IBK	RF		
	CLASSIFIERS				

Figure 5. This figure shows the results of the derivations of the confusion matrix generated in this paper. F1 is the harmonic mean of precision and sensitivity. The Matthews correlation coefficient (MCC) is used in machine learning as a measure of the quality of binary (two-class) classifications, and the accuracy refers to closeness of the measurements to a specific value

best results for accuracy were the unbalanced ones, with a result. The SVM achieved a rating of 96.8% for the 3000 unbalanced dataset, and the worst result was 83.3% for IBK the 6000 balanced dataset. The complete result of the classification showed in the form of a confusion matrix can be seen in Figure 4 and Figure 5.

The better results are in the unbalanced dataset with a 3000 superpixels and balanced with 6000 superpixels. This result is better to understand confusion matrix because it is possible to understand where the learning algorithms have many mistakes. In Figures 4 and 5 is possible to see the best results of the unbalanced and balanced experiment. In this case, both learning algorithms have the best results using an SVM.

When comparing the two classes background and tree with the metric f-measure, precision and Recall, we created the table 6, which presents the performance by class, where we can see in the F-measure results that the background in general had greater relevance than the tree class.

In order to evaluate the best technique, we can observe in Figure 6, the second step of classifier performance comparison using F-measure, precision and Recall metrics to show the difference between the techniques and to scrutinize better the results we present also the, usefulness (precision) and completeness (recall) of the results. Observing the results, we pointed to SVM as the best result with the unbalanced group and the balanced group. We have the three models using the unbalanced SVM (SVM (3000K), SVM (6000K) and SVM (6000k)) at the top, side by side, we have three more models with under-sampling (SVM_US (3000K), SVM_US (6000K) and SVM_US (6000k)). However, of these models, the best-unbalanced result was SVM

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-349-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165618

(3000k) and SVM_US (6000k) for the balanced.

3000				
SVM	0,978	0,987	0,982	background
	0,861	0,777	0,817	tree
IBK	0,972	0,938	0,955	background
	0,546	0,737	0,627	tree
RF	0,96	0,99	0,975	background
	0,865	0,598	0,707	tree
3000_US				
SVM	0,948	0,897	0,922	background
	0,903	0,951	0,926	tree
IBK	0,936	0,714	0,81	background
	0,769	0,951	0,85	tree
RF	0,943	0,879	0,91	background
	0,887	0,946	0,916	tree
6000				
SVM	0,978	0,983	0,981	background
	0,843	0,804	0,823	tree
IBK	0,971	0,936	0,953	background
	0,567	0,753	0,647	tree
RF	0,964	0,99	0,977	background
	0,878	0,666	0,757	tree
6000_US				
SVM	0,955	0,927	0,941	background
	0,929	0,957	0,943	tree
IBK	0,954	0,787	0,862	background
	0,818	0,962	0,885	tree
RF	0,952	0,909	0,93	background
	0,913	0,955	0,933	tree
9000				
SVM	0,977	0,984	0,981	background
	0,856	0,806	0,83	tree
IBK	0,975	0,934	0,954	background
	0,585	0,798	0,675	tree
RF	0,964	0,992	0,977	background
	0,905	0,683	0,778	tree
9000_US				
SVM	0,953	0,917	0,935	background
	0,92	0,955	0,937	tree
IBK	0,952	0,744	0,835	background
	0,79	0,962	0,868	tree
RF	0,946	0,912	0,929	background
	0,915	0,948	0,931	tree
	PRECISION	RECALL	F-MEASURE	

Figure 6. Comparison of Precision, Recall and F-measure by number of Superpixels, Balanced and class.

4. CONCLUSIONS

In this paper, we explored the combination os SLIC superpixel and machine learning methods to segment trees in high resolution aerial imagery. The experiments showed that better results were achieved using SVM classifier with 3000 superpixels and unbalanced dataset. The mistakes to classify background is low because the learning algorithms are specialized in its recognition. Also, a reason that the balanced datasets performed worse than the unbalanced ones is that the under-sampling technique excludes relevant data and valuable information from the original dataset. One of our hypothesis for this is that the background is much more diverse in colors and shapes than the tree canopies. Consequently, more background superpixels are necessary in Pynovisão software.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, FUNDECT, Foundation to Support Development of Education, Science and Technology of the State of Mato Grosso do Sul, Brazil and CNPq (National Council for Scientific and Technological Development) through research grants (p. 433783/2018-4 and 314902/2018-0). The present work was also carried out with a research project financed by CAPES through the program Print-CAPES. Also, we would like to thanks NVIDIA Corporation for the donation of the GPU used in this research.

REFERENCES

Abu Bakar, M. Z., Samad, R., Pebrianti, D., Mustafa, M., Abdullah, N. R. H., 2015. Finger application using K-Curvature method and Kinect sensor in real-time. *2nd International Symposium on Technology Management and Emerging Technologies, ISTMET 2015 - Proceeding*, 218–222.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Su¨sstrunk, S., 2012. SLIC Superpixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Aha, D. W., Kibler, D., Albert, M. K., 1991. Instance-Based Learning Algorithms. *Machine Learning*.

Baró, F., Chaparro, L., Gómez-Baggethun, E., Langemeyer, J., Nowak, D. J., Terradas, J., 2014. Contribution of ecosystem services to air quality and climate change mitigation policies: The case of urban forests in Barcelona, Spain. *Ambio*.

Breiman, L., 2001. Randomforest2001. *Machine Learning*.

Davies, Z. G., Edmondson, J. L., Heinemeyer, A., Leake, J. R., Gaston, K. J., 2011. Mapping an urban ecosystem service: Quantifying above-ground carbon storage at a city-wide scale. *Journal of Applied Ecology*.

dos Santos Ferreira, A., Matte Freitas, D., Gonçalves da Silva, G., Pistori, H., Theophilo Folhes, M., 2017. Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, 143, 314–324.

Evans, K. L., 2013. Individual species and urbanisation. *Urban Ecology*.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data.

Feichtinger, H. G., Zimmermann, G., 1998. A Banach space of test functions for Gabor analysis. *Gabor Analysis and Algorithms*, Birkhäuser Boston, Boston, MA, 123–170.

Gaston, K. J., 2000. Global patterns in biodiversity.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., Wolff, E., 2018. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience and Remote Sensing*.

Gómez-Baggethun, E., Barton, D. N., 2013. Classifying and valuing ecosystem services for urban planning. *Ecological Economics*.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., 2009. The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*.

Ho, T. K., 1995. Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*.

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., Murthy, K. R. K., 2001. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation*.

Li, X., Shao, G., 2013. Object-based urban vegetation mapping with high-resolution aerial photography as a single data source. *International Journal of Remote Sensing*.

Luo, N., Wan, T., Hao, H., Lu, Q., 2019. Fusing high-spatial-resolution remotely sensed imagery and OpenStreetMap data for land cover classification over urban areas. *Remote Sensing*.

Ming-Kuei Hu, 1962. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, 8(2), 179–187. http://ieeexplore.ieee.org/document/1057692/.

Nichol, J., Lee, C. M., 2005. Urban vegetation monitoring in Hong Kong using high resolution multispectral images. *International Journal of Remote Sensing*.

Platt, J., others, 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*.

QGIS Development Team, 2019. QGIS Geographic Information System.

Seto, K. C., Güneralp, B., Hutyra, L. R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences of the United States of America*.

Small, C., 2001. Estimation of urban vegetation abundace by spectral mixture analysis. *International Journal of Remote Sensing*.

Soh, L.-K., Tsatsoulis, C., 1999. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 780–795. http://ieeexplore.ieee.org/document/752194/.

Stagoll, K., Lindenmayer, D. B., Knight, E., Fischer, J., Manning, A. D., 2012. Large trees are keystone structures in urban parks. *Conservation Letters*.

Swain, M. J., Ballard, D. H., 1991. Color indexing. *International Journal of Computer Vision*, 7(1), 11–32. https://link.springer.com/content/pdf/10.1007

Tigges, J., Lakes, T., Hostert, P., 2013. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. *Remote Sensing of Environment*.

VAN KLAVEREN, E. P., MICHELS, J. P. J., SCHOUTEN, J. A., 1999. THE ORIENTATIONAL AND STRUCTURAL PROPERTIES OF N 2 and N 2 –AR SOLIDS AT HIGH PRESSURE. *International Journal of Modern Physics C*, 10(02n03), 445–453. https://www.worldscientific.com/doi/abs/10.1142/S0129183199000334.

Yan, J., Zhou, W., Han, L., Qian, Y., 2018. Mapping vegetation functional types in urban areas with WorldView-2 imagery: Integrating object-based classification with phenology. *Urban Forestry and Urban Greening*.

Zhang, X., Feng, X., Jiang, H., 2010. Object-oriented method for urban vegetation mapping using ikonos imagery. *International Journal of Remote Sensing*.

Zhou, X., Li, L., Chen, L., Liu, Y., Cui, Y., Zhang, Y., Zhang, T., 2019. Discriminating urban forest types from Sentinel-2A image data through linear spectral mixture analysis: A case study of Xuzhou, East China. *Forests*.

Revised November 2019