TOWARDS AUTOMATED DELINEATION OF SMALLHOLDER FARM FIELDS FROM VHR IMAGES USING CONVOLUTIONAL NETWORKS

Claudio Persello, Valentyn Tolpekin, John Ray Bergado, Rolf de By

Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

ABSTRACT

Automated delineation of smallholder farm fields is difficult because of their small size, irregular shape and the use of mixed-cropping systems. Edges between smallholder plots are often indistinct in satellite imagery and contours have to be identified by considering the transition of the complex textural patterns of the fields. We introduce a strategy to delineate field boundaries using a fully convolutional network in combination with a globalization and grouping algorithm to produce a hierarchical segmentation of the fields. We carry out an experimental analysis in a study area in Kofa, Nigeria, using a WorldView-3 image, comparing several state-of-the-art contour detection algorithms. The proposed strategy outperforms state-of-the-art computer vision methods and shows promising results by automatically delineating field boundaries with an accuracy close to human level photo-interpretation.

***Index Terms*—**Field boundary extraction, convolutional neural networks, deep learning, agriculture & food security, smallholder farming, remote sensing

1. INTRODUCTION

Improving the capability to map the spatial distribution of agricultural fields is crucial for increasing the agricultural production and ensuring food security in many parts of the world [1]. In Sub-Saharan Africa (SSA), agriculture is dominated by smallholder farms, characterized by rain-fed production for predominantly household consumption. Accurate and updated maps of agricultural fields are needed for decision making at both policy level, as well as at the level of daily field management [2]. A satellite-based approach can drastically reduce costs compared to traditional field surveys and can improve efficiency. Satellite imaging sensors provide the opportunity to acquire large volumes of data with increasing spatial, spectral and temporal resolution. Very High spatial Resolution (VHR) images can be used for mapping large geographical areas. Nevertheless, accurately mapping agricultural resources in Africa is a challenging task because of the characteristics of smallholder farms: (i) small plot size (< 2 ha); (ii) irregularly shaped fields with often indistinct boundaries; (iii) strong seasonal variations in surface reflectance; (iv) predominantly rain-fed practices that naturally coincide with high incidence of clouds; (v) high spatiotemporal dynamics.

In this paper, we focus on the segmentation of smallholder farm fields from VHR satellite imagery. Field segments are important because they enable to estimate cropland area size, to aggregate statistics and yield information at the field level [3]. Moreover, they facilitate the extraction of land tenure boundaries for cadastral systems [4]. Previous research on field boundary delineation from remote sensing data has mainly focussed on areas characterised by large field plots using medium resolution images [3]. Automatic delineation of smallholder farms in SSA is extremely challenging since field boundaries are not characterised by clearly visible edges, but need to be extracted by detecting changes in the textural patterns of different cultivations. In these circumstances, standard techniques for edge detection (e.g., Sobel, Prewitt, Canny operators) typically fail in achieving the required accuracy.

By taking advantage of the recent advancements in deep learning and contour detection, this paper introduces a novel strategy based on a Fully Convolutional Network (FCN) and a grouping algorithm to delineate farm field boundaries from satellite VHR images. The FCN can learn complex hierarchical spatial-contextual cues in the image space and accurately detect sparse (i.e., not connected) high-level object contours. The detected contours are then combined with local cues based on multiscale oriented gradients and spectral clustering to generate a hierarchy of closed segments based on contour strength.

2. STUDY AREA

Our study site is an area of intensive, small-scale, rain-fed agricultural production in the Sudano-Sahelian savanna region of northern Nigeria, around the city of Kofa. This area can be characterized as having small fields (average 0.22 ha), with only 5% pure crops, and more than 50%

having three or more crops at any moment in time in the crop season. The farm field landscape is further characterized by many scattered trees. Important crops in this area are sorghum, rice, millet, maize and groundnut. The site is under study by the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) Nigeria, ICRISAT Mali and the ITC Faculty.

Field boundary data, comprising over 5,000 field polygons, were obtained for the year 2015 by ICRISAT Nigeria through an intensive field campaign, using GPS-enabled smartphones and tablets. Using a WorldView-3 image, acquired through satellite tasking over the study site on September 25th 2015, we subsequently corrected that original dataset by human photo-interpretation and expanded it to over 5,700 field boundaries, using visual clues from the pan-sharpened image product. The WorldView-3 data contains a panchromatic (PAN) channel at 0.5 m resolution and eight multispectral (MS) bands at 2 m. The product is atmospherically corrected, orthorectified, co-registered and processed through the STARS project image workflow [6]. Four tiles of 1000×1000 pixels were selected for our experimental analysis (see one training tile in Figure 1). Three tiles are used for training and one for testing.

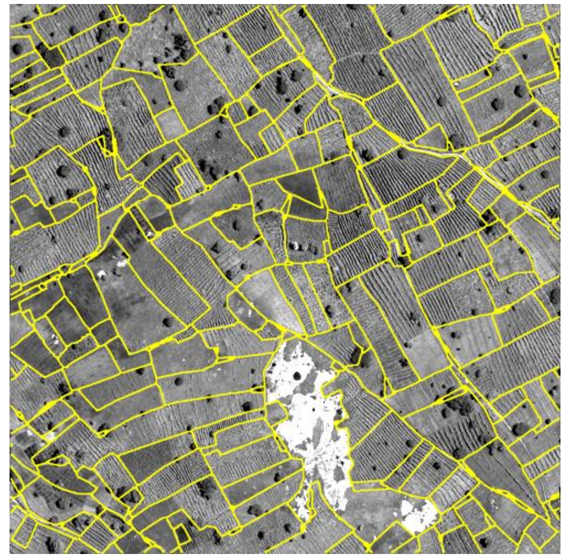

Figure 1. One of the training tiles used in the experimental analysis. Field reference boundaries are superimposed over the panchromatic channel.

3. PROPOSED METHOD

The proposed strategy takes advantage of the recent success of FCNs for pixel-wise classification [7]. In contrast to traditional CNNs, which predict one class label per input image, FCNs are designed to infer pixel-wise predictions directly, independently from the size of the input image. In these networks, the fully connected layers of standard CNNs are usually substituted by deconvolution or unpooling layers that upsample the feature maps learned by the convolutional layers to the resolution of the input image [8].

3.1. FCN for boundary detection

We formulate the boundary detection as a supervised pixel-wise image classification problem to distinguish "boundary" from "non-boundary" pixels, respectively. To this aim, we adopt the SegNet architecture [8], which consists of a deep encoder-decoder FCN for pixel-wise labelling. The encoder part of the network is topologically identical to the convolutional layers of the VGG-16 network [9], including 13 convolutional layers followed by batch normalization and Rectified Linear Units (RELU), and five max-pooling layers, each of them down-sampling the spatial resolution of the input feature maps by a factor two. The decoder is used to map the low-resolution feature maps learned by the encoder to the full resolution of the input image. Instead of using deconvolution or transposed convolutions, the decoder of SegNet uses pooling indices computed in the corresponding max-pooling layers of the encoder to perform non-linear upsampling. The obtained upsampled maps are sparse and are then convolved with trainable convolutional filters to produce dense feature maps. This procedure eliminates the need for learning to upsample, reducing the number of trainable parameters and improving the accuracy of boundary delineation. The abovementioned characteristics make SegNet well suited for the considered contour detection problem.

We initialise the encoder with the pre-trained VGG-16 net and the decoder with random weights. We then train SegNet for boundary detection on our VHR data set. From the training tiles we randomly extract 6000 training patches of 96×96 pixels from the panchromatic band, and the corresponding labelled patched from the reference data. Since the "boundary" and "non-boundary" pixels are extremely unbalanced, we set the penalty for misclassifying the "boundary" class to be 10 times higher than for the "non-boundary".

3.2. Hierarchical Segmentation

The contour detector based on the binary SegNet classification results in fragmented contours, which do not partition the image into closed segments. To recover a segmentation from the network output, one could apply a grouping algorithm directly to the network-derived boundary strength, e.g., the two-steps machinery developed in [10]. However, we notice that the boundary strength signal predicted by SegNet can capture strong boundaries between fields, while it is unable to detect fine-grained edges within fields or between fields and other classes (e.g., roads or bare soil). To address this problem, we linearly combine the SegNet-based contour detector with the *global Pb* (gPb) detector [10], which can capture small and detailed edges in the image, offering complementary information with respect to the SegNet output. The combined local cues are then globalised using normalized cuts adopting the fast eigenvector computation introduced in [11]; we finally combine global and local cues linearly and construct an *Ultrametric Contour Map* (UCM) based on the mean contour strength using the *Oriented Watershed Transform* (OWT).

4. EXPERIMENTAL ANALYSIS

4.1. Experimental setup

In our experimental analysis, we compare the performance of several state-of-the-art contour detection and hierarchical segmentation techniques on the considered test tile. The shallow hierarchical segmentation techniques gPb-owt-ucm [10], SCG and MCG [11] are applied to the pan-sharpened multispectral bands 7, 5, 2 of the WorldView-3 image. Pansharpening is performed using the Gram-Schmidt algorithm. SCG and MCG use the pre-trained structured-forest contour detector [12] for the extraction of the low-level cues while gPb-owt-ucm uses hand-crafted multiscale local cues based on colour, brightness and texture. We applied SegNet, trained in one case with the original cross-entropy loss function and in a second case adopting a ten time higher penalty for the "boundary" class (SegNet-W). We then combined SegNet-W with SCG using the structured-forest detector (SegNet-W-SCG) and finally obtained the result of the proposed technique by combining SegNet-W with gPb as described in the previous section (Proposed).

4.2. Results

The accuracy is assessed using the precision-recall framework introduced in [13], which is generally applied to evaluate the accuracy on computer vision contour detection benchmarks, but is not common in remote sensing. Table 1 reports the obtained results on the test tile by fixing the thresholds based on highest F-score on the training tiles. The results of the proposed technique show a significant improvement compared to the alternative methods. The advantage is captured by a higher F-score and generally a large increase in the recall with respect to the other techniques. Among the three shallow segmentation algorithms (gPb-owt-ucm, SCG, MCG), gPb-owt-ucm results in the most accurate boundary detection, showing higher transferability of the carefully engineered intensity and textural features from natural images to satellite data over the pre-trained structured-forest detector. Despite a significantly larger computational cost, MCG does not provide more accurate results than SCG. The gPb local features prove effective when combined with SegNet-W: SegNet-W-gPb-SCG shows a significant improvement over alternative techniques.

Figure 2 shows the detection map obtained by SegNet and the proposed technique on the test tile. The proposed strategy allows to properly connect the fragmented contours extracted by the deep network, resulting in accurate closed regions representing field plots. Despite a certain level of under-segmentation, the obtained segmentation visually look close to human-level photo-interpretation. We think this is a particularly encouraging result, taking into account the complexity of the considered landscape. Interestingly, the proposed algorithm is not significantly affected by the presence of the many trees and their shadows in the considered area. In several cases, the field boundary passing close or in correspondence with the tree crown is correctly identified.

Table 1. F-score (F), Precision (P), and Recall (R) on the considered test tile.

METHOD	F	P	R
GPB-OWT-UCM	0.663	0.717	0.617
SCG	0.623	0.625	0.621
MCG	0.598	0.514	0.715
SEGNET-W	0.749	**0.859**	0.664
SEGNET-W-SCG	0.773	0.847	0.711
PROPOSED	**0.795**	0.846	**0.750**

5. CONCLUSION

This paper proposes a contour delineation technique based on a deep fully convolutional network and a grouping algorithm to produce a segmentation of smallholder agricultural fields. The experimental analysis conducted using a WorldView-3 image acquired over the area of Kofa, Nigeria, shows promising results. The proposed technique compares favourably against state-of-the-art computer vision contour detection algorithms in terms of the accuracy assessed through the precision-recall framework. A visual inspection of the obtained segmentation results allows us to observe an accurate field plot delineation which is close to human photo-interpretation level. These results show that the proposed automated field delineation method could facilitate the extraction of cadastral boundaries and be incorporated into an object-based image analysis framework for accurate crop type classification.

6. ACKNOWLEDGEMENT

This publication was, in part, made possible by the STARS project (https://www.stars-project.org), an integrated effort to improve our understanding of the use of remote sensing technology in monitoring smallholder farming, funded by the Bill & Melinda Gates Foundation, under grant #1094229.

Reference Boundareis

SegNet

Proposed

Figure 2. Subsets of the detection maps obtained by selected contour and segmentation algorithms on the test tile.

REFERENCES

[1] S. R. Debats, D. Luo, L. D. Estes, T. J. Fuchs, and K. K. Caylor, “A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes,” *Remote Sens. Environ.*, vol. 179, pp. 210–221, Jun. 2016.

[2] R. Aguilar, R. Zurita-Milla, E. Izquierdo-Verdiguier, and R. A. de By, “A Cloud-Based Multi-Temporal Ensemble Classifier to Map Smallholder Farming Systems,” *Remote Sens.*, vol. 10, no. 5, p. 729, May 2018.

[3] A. Rydberg and G. Borgefors, “Integrated Method for Boundary Delineation of Agricultural Fields in Multispectral Satellite Images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2514–2520, 2001.

[4] S. Crommelinck, R. Bennett, M. Gerke, F. Nex, M. Yang, and G. Vosselman, “Review of Automatic Feature Extraction from High-Resolution Optical Sensor Data for UAV-Based Cadastral Mapping,” *Remote Sens.*, vol. 8, no. 8, p. 689, Aug. 2016.

[5] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik, “Multiscale combinatorial grouping,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 500, pp. 328–335, 2014.

[6] D. Stratoulias *et al.*, “A Workflow for Automated Satellite Image Processing: from Raw VHSR Data to Object-Based Spectral Information for Smallholder Agriculture,” *Remote Sens.*, vol. 9, no. 10, p. 1048, Oct. 2017.

[7] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–14, 2017.

[9] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” in *International Conference on Learning Representations*, 2015.

[10] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011.

[11] J. Pont-Tuset, P. Arbelaez, J. T. B. Barron, F. Marques, and J. Malik, “Multiscale combinatorial grouping for image segmentation and object proposal generation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 128–140, 2017.

[12] P. Dollár and C. L. Zitnick, “Fast Edge Detection Using Structured Forests,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1558–1570, 2014.

[13] D. R. Martin, “An Empirical Approach to Grouping and Segmentation,” 2003.