ERDC/GRL TR-21-1

Geospatial Research Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Evaluation of Automated Feature Extraction Algorithms Using High-resolution Satellite Imagery Across a Rural-urban Gradient in Two Unique Cities in Developing Countries

Andrew Griffin, Sean Griffin, Kristofer Lasko, Megan Maloney, S. Bruce Blundell, Mike Collins, and Nicole Wayant

March 2021

ERDC/GRL TR-21-1
March 2021

Evaluation of Automated Feature Extraction Algorithms Using High-resolution Satellite Imagery Across a Rural-urban Gradient in Two Unique Cities in Developing Countries

Andrew Griffin, Sean Griffin, Kristofer Lasko, Megan Maloney,
S. Bruce Blundell, Mike Collins, and Nicole Wayant

Geospatial Research Laboratory
U.S. Army Engineer Research and Development Center
7701 Telegraph Rd
Alexandria, VA 22314

Final report

Approved for public release; distribution is unlimited.

Prepared for National Geospatial-Intelligence Agency
7500 GEOINT Dr
Springfield, VA 22150

Under NGA, MIPR NR28G19138GS15

Abstract

Feature extraction algorithms are routinely leveraged to extract building footprints and road networks into vector format. When used in conjunction with high resolution remotely sensed imagery, machine learning enables the automation of such feature extraction workflows. However, many of the feature extraction algorithms currently available have not been thoroughly evaluated in a scientific manner within complex terrain such as the cities of developing countries. This report details the performance of three automated feature extraction (AFE) datasets: Ecopia, Tier 1, and Tier 2, at extracting building footprints and roads from high resolution satellite imagery as compared to manual digitization of the same areas. To avoid environmental bias, this assessment was done in two different regions of the world: Maracay, Venezuela and Niamey, Niger. High, medium, and low urban density sites are compared between regions. We quantify the accuracy of the data and time needed to correct the three AFE datasets against hand digitized reference data across ninety tiles in each city, selected by stratified random sampling. Within each tile, the reference data was compared against the three AFE datasets, both before and after analyst editing, using the accuracy assessment metrics of Intersection over Union and F1 Score for buildings and roads, as well as Average Path Length Similarity (APLS) to measure road network connectivity. It was found that of the three AFE tested, the Ecopia data most frequently outperformed the other AFE in accuracy and reduced the time needed for editing.

Contents

Figures and Tables

Figures

Tables

Preface

This study was conducted for the National Geospatial Intelligence Agency (NGA) under Project Number 478420, "Automated Feature Extraction Evaluation." The technical monitors were Hilary McGown, Dr. Enrique Montano, and Shona Kiser of NGA.

The work was performed by the Information Generation and Management Branch (IGMB) of the U.S. Army Engineer Research and Development Center, Geospatial Research Laboratory (ERDC-GRL). At the time of publication, John Nedza was Branch Chief; Jeffrey Murphy was Division Chief; and Ritch Rodebaugh was the Technical Director for the Geospatial Research Enterprise. The Deputy Director of ERDC-GRL was Valerie Carney and the Director was Gary Blohm.

The Commander of ERDC was COL Teresa A. Schlosser and the Director was Dr. David W. Pittman.

1 Introduction

Background

Extracting clear and relevant geographic information from imagery sources is an ongoing effort to yield important information about an area. Historically, analysts review imagery manually to digitize important information such as building footprints and road networks. This proven method yields quality results but is time intensive and extracted features can vary from analyst to analyst based on personal interpretation. With the advent of automated feature extraction (AFE), where a computer digitizes features from imagery typically using a machine-learning algorithm, the possibility arises to use a uniform and quick computer driven process to supplement or replace the laborious traditional method of digitization. However, there is concern that the automated methods do not fully replicate the results of a trained analyst (Arcot Sowmya 2000).

Advances in building AFE leverage panchromatic or RGB imagery with building morphology and classified morphological features using a shallow neural network (Benediktsson et al. 2003). Recent advances have led to building footprint AFE using fully convolutional Neural Networks (CNNs) or other advanced machine learning algorithms. Some studies use a single RGB satellite image (Katartizis 2007) while others use satellite imagery in conjunction with a fused Normalized Digital Surface model (nDSM), which provides relative heights (Bittner et al. 2018; Chen et al. 2012; Lu et al. 2018). Some of these studies also perform shallow machine learning in conjunction with object based image analysis (OBIA).

A new AFE method, promoted by Ecopia Tech, uses a combination of artificial intelligence and analyst review to semi-automatically extract feature data from imagery. According to Ecopia Tech, their data has the same “quality of a trained GIS professional” (Ecopia Tech 2018). Organizations, such as the Bill and Melinda Gates Foundation and Sustainable Development Technology Canada, have used Ecopia data for their organization’s projects in remote areas (Toman 2020). Yet, beyond the claim of quality, a comprehensive assessment of the dataset has not been completed.

Objective

The goal of this research project was to measure, quantifiably, the effectiveness and completeness of the Ecopia product compared against two different AFE datasets, Tier 1 and Tier 2, as well as manually extracted reference data. All AFE datasets were compared against the level of effort required to create the feature reference data. Additionally, a series of accuracy metrics were applied against multiple types of AFE to measure the individual quality and completeness of the datasets. Lastly, a recommendation was made of the most accurate and effective feature extraction method.

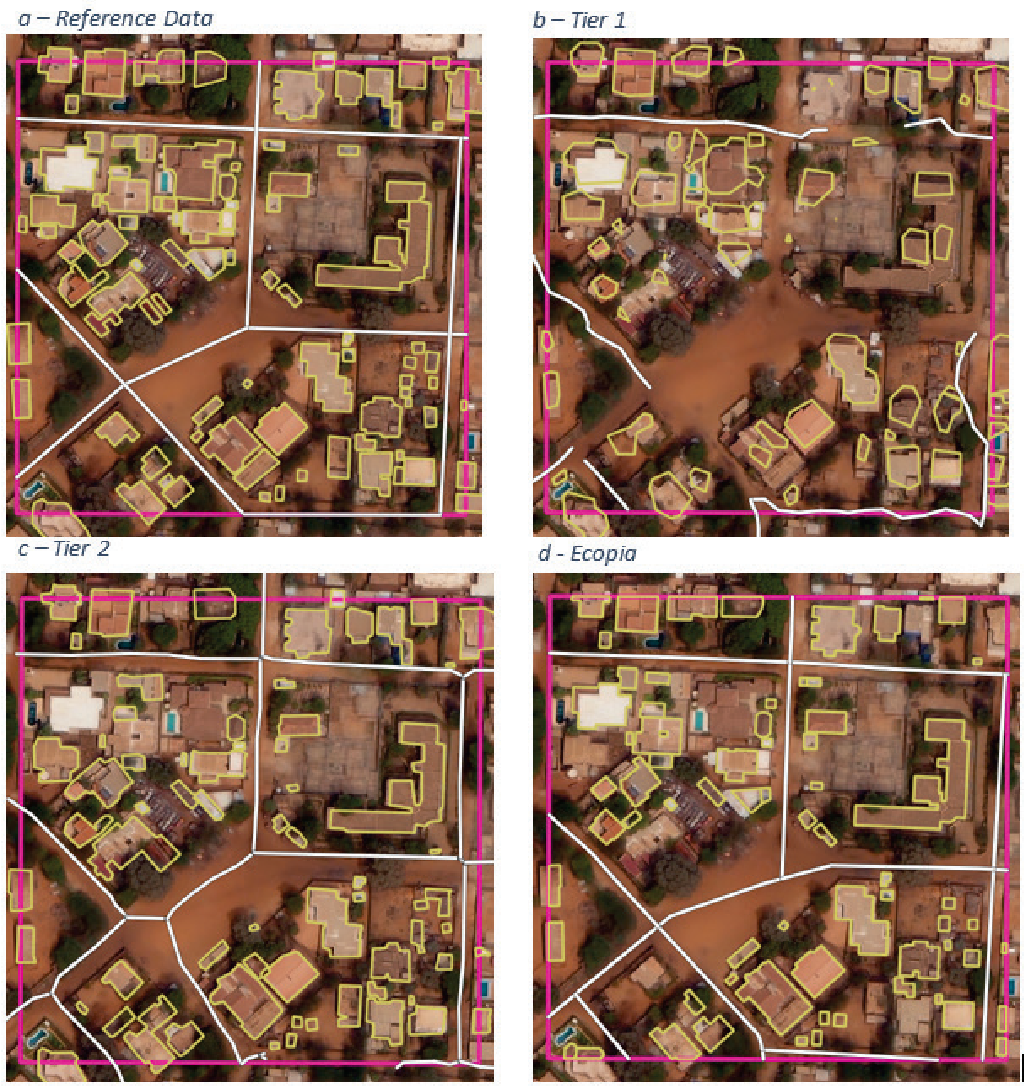
Approach

The different AFEE methods tested in this study were Tier 1, Tier 2, and Ecopia, which are shown in raw form in Figure 1. All data was provided by the National Geospatial Intelligence Agency (NGA), though it was generated independently by John Hopkins University's Applied Physics Laboratory. Tier 1 building data was created using the U-Net algorithm (Oktay et al. 2018) and used degraded Ecopia data to generate the road features. Tier 2 roads data was created with the City- scale Road Extraction from Satellite Imagery (CRESI) algorithm established by the SpaceNet challenge (Van Etten 2020) while the buildings were represented by downgraded Ecopia data. The Ecopia data was created by the private firm Ecopia Tech as part of their Global Feature Extraction product.

To judge the performance of the various AFE, a control feature dataset was manually digitized by a team of six analysts at the Geospatial Research Laboratory (GRL). The analysts that performed the manual digitization come from diverse educational and professional backgrounds, but all possess a wealth of geospatial experience. With between 6 and 35 years each of experience working with geospatial products, each analyst is highly qualified to independently work with and interpret geospatial data.

All AFE datasets were compared against the time required to digitize the feature reference data and the ground truth represented by the control dataset. All levels of feature extraction used the same Vricon True Ortho imagery over ninety test areas of interest (AOI) in each of two different regions: Maracay, Venezuela and Niamey, Niger. These environmentally and architecturally distinct regions, chosen by our partners at NGA, give the AFE datasets the opportunity to demonstrate performance in different conditions.

Figure 1. Examples of different tiers of unedited AFE and reference data over Niamey, Niger. The truest to imagery are the reference data and the Ecopia data, with outlines closely following the imagery. Tier 1 misses many features, and the features it does digitize only approximately capture the correct area and dimensions. Tier 2 improves on the performance of Tier 1, especially for roads, but still misses several buildings. Buildings are outlined in yellow and roads in white. a.) Reference Data, b.) Tier 1, c.) Tier 2, and d.) Ecopia.



2 Data

The two regions selected, Maracay (Figure 2) and Niamey (Figure 4), were chosen based on data availability, and represent different environments and architectural styles, which have historically been an obstacle to global accurate AFE (Yuan 2018). The regions were divided into ninety separate, randomly chosen AOI for testing and comparison using stratified random sampling with stratification based on low, medium, and high density areas using the Ecopia data to define density. This stratification ensured testing was not overly representative of any particular urban density, which may have skewed results to focus on performance in that level of density.

Figure 2. Imagery of Maracay, Venezuela region with test AOI marked with pink squares.

The Maracay region is a mix of dense urban structures, informal settlements, and suburban areas, with trees intermixed throughout the urban areas. Buildings in the urban sections typically abut one another and are difficult to distinguish from one another, as illustrated in Figure 3. Trees obscure building edges and corners, adding further difficulty to discerning the exact dimensions of features. Additionally, many areas use roofing material that appear similar to the surrounding pavement, adding barriers to determining where a roof ends and a courtyard or wall might begin. Though the imagery used for this region does allow identification and extraction of features, at larger scales there is a noticeable blur to features.

Figure 3. Imagery of a Maracay neighborhood indicative of the closely packed, abutting buildings.

Figure 4. Imagery of Niamey, Niger with test AOI marked with pink squares.

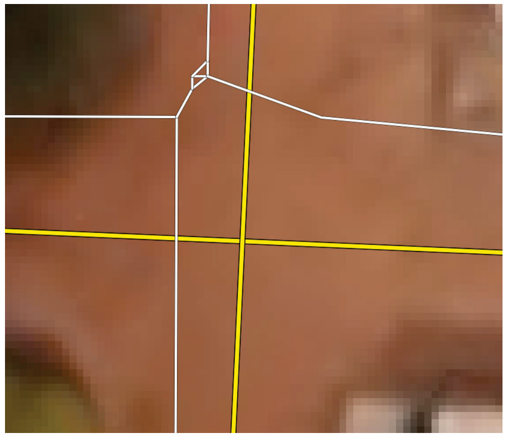
The Niamey region is a densely packed urban center that has abutting structures in built-up areas, but for the most part has distinct structures aiding in easier digitization. The imagery used in identifying and extracting features has an overall reddish tint, which along with some partially obscured objects, suggests some degree of dust cloud interference as demonstrated in Figure 5. This does not prevent feature extraction, but does make some features, especially those of a similar color to their surroundings, blend in with the soil. However, other than the dust interference, the imagery remains crisp at a higher resolution than that of Maracay. In a few cases, the provided imagery did not cover all of an AOI. Rather than use supplementary imagery in these areas, no feature extraction took place in areas missing True Ortho imagery.

Figure 5. A 1:500 scale comparison of the imagery used in Niamey for digitizing (left) and ArcMap Basemap (Esri n.d.) (right). While the images are not from the same date the difference between the images does reveal the distortion present in the digitizing image. Distortion is most evident along roadways where cars are barely visible as well as in the blurring around trees.

The different AFE techniques behaved consistently across both regions. The Tier 1 dataset was the coarsest level of feature data, capturing only the most basic of features, which provided a general sense of urbanization without accurately capturing building footprints. This AFE often, especially in areas of closely packed buildings, captured entire blocks as a single building footprint. Tier 1 AFE also only vaguely captures the road network, regularly failing to connect road segments, does not follow the roadbed closely, and missed several roads entirely.

Tier 2 data has a higher level of fidelity than Tier 1 data, capturing greater detail in buildings specifically. Along with the improved building performance, the road network gathered in Tier 2 also shows notable improvement, generally following the course of the road as shown in imagery. To accomplish this, many very small road segments are generated in the place of long continuous lines as shown in Figure 6, which increases the complexity of the network. Ecopia AFE data has the highest level of fidelity of the three AFE datasets assessed, regularly extracting features aligning with imagery, with little need for correction.

Figure 6. A 1:100 scale example of Tier 2 road layer generated excessive line segments. The white lines are Tier 2 generated road segments while yellow lines are manually digitized roads. Tier 2 AFE generated a series of small segments in a triangular pattern in place of a standard cross intersection.



3 Methods and Metrics

For each of the AOI, the AFE products were evaluated against the manually digitized reference data using the metrics listed below. All metrics were computed using the CosmiQ Solaris python library (CosmiQ Works 2020) and Jupyter Notebooks to compare both the raw and edited forms of data against the reference data. Unless otherwise noted, the metrics were used for both road and building layers.

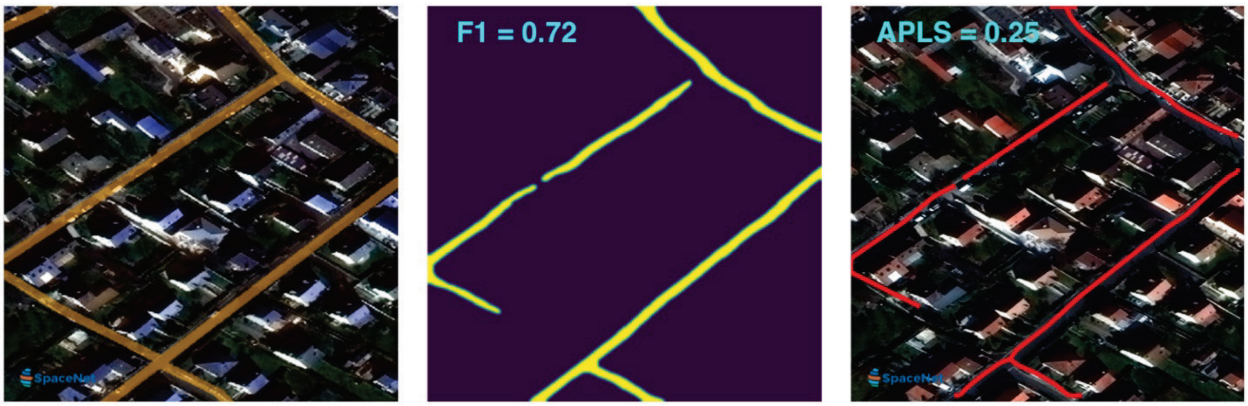
- Creation time of reference features from imagery or time required to clean up AFE data so it matches the reference data
- Performance of the raw and edited AFE features versus reference features:
  - Intersection Over Union (IoU) – ratio of intersected areas over merged areas. This metric measures correct placement of the extracted features. A value of 1.0 is a perfect score (Russakovsky et al. 2015) indicating all extracted and reference features are identical in placement and area.
  - F1 Score – A measurement of accuracy combining precision and recall representing accuracy and completeness in a single number. A value of 1.0 is a perfect score.
    - Precision – The number of correct extractions out of all extracted features. A score of 1.0 indicates all features extracted were correctly identified and actually a feature.
    - Recall – Correctly extracted features out of all the features actually present, this measures how complete the total extraction was. A score of 1.0 indicates correct extraction of all features present in an AOI.

Equation 1: The F1 calculation

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

  - Average Path Length Similarity (APLS) – the mean difference between the reference road lengths and the AFE road lengths. 1.0 is a perfect score (Van Etten, SpaceNet Road Detection and Routing Challenge — Part I 2017). This metric is unique to the roads layer. This score is based on graph theory and the connectivity of the collected road network. A lower score will indicate that there are gaps in the network as illustrated in Figure 7.

Figure 7. Graphic depiction of the different focus of F1 vs APLS and what can be discerned from each. The leftmost image shows a road ground truth. The high F1 score shown in the middle picture indicates that the pixels of the ground truth and compared dataset correctly overlapped accurately in most of the image. While the image on the right shows a significantly lower APLS score representing the poor connectivity of the comparison road network versus the ground truth. (Van Etten, SpaceNet Road Detection and Routing Challenge — Part I 2017).



Feature extraction was completed in local Geographic Information System (GIS) environments by six analysts using QGIS or ArcGIS. While some initial digitization was attempted in the research open mapping enclave (ROME), an online digitization collaboration portal, styled after OpenStreetMap (OSM), technical difficulties precluded continued use of that environment. The Maracay region was the first region processed, followed by the Niamey region. A random stratified sample of 90 AOI per region were selected to divide the regions into testing areas. Analysis was limited to the areas within AOI bounding boxes. During editing, analysts edited the features of the three AFE datasets to conform with the True Ortho imagery, limited to the areas within the AOI bounding boxes.

In the editing process, analysts used their experience to determine which features could be modified with a few edits to align to imagery and which should be deleted entirely and redrawn to save time. Given the unique differences between any two AOI, analysts used their best judgment and experience along with consultation with colleagues to accurately identify and separate out buildings from misidentified objects. To ensure consistent editing practices were followed, one team member reviewed 10% of all digitized or edited data for each AFE dataset edited and corrected errors.

4 Results and Discussion

As seen in Table 1, the results depicting the length of time to digitize or correct all features (buildings and roads) for the Maracay region show that Ecopia data took less time to clean than Tier 1 or Tier 2 AFE data. Ecopia took approximately a third less time than it took to digitize the reference data from imagery. Tier 1 took approximately two or more times longer than Tier 2, with the time to create reference data falling between Tier 1 and Tier 2. This pattern persisted when the sample was broken out into high, mid, and low urbanized AOI, which are shown in the last three rows of Table 1.

Table 1. Maracay data creation or average cleanup time.

Metric	Reference Data	Tier 1	Tier 2	Ecopia
Mean AOI Data Creation or Clean Up Time	45 min	61 min	27 min	16 min
Mean Highly Urbanized AOI Data Creation or Clean Up Time	100 min	131 min	53 min	33 min
Mean Mid-Urbanized AOI Data Creation or Clean Up Time	44 min	63 min	30 min	24 min
Mean Low Urbanized AOI Data Creation or Clean Up Time	20 min	23 min	12 min	4 min

However, this pattern did not persist in the Niamey region. As seen in Table 2, the time to digitize or clean data was relatively similar in all four datasets. Subsetting the AOI by urban density did not clarify any trends. The increased familiarity of the analysts with the process after editing Maracay may account for the shorter editing time for the second region.

Table 2. Niamey data creation or average cleanup time.

Metric	Reference Data	Tier 1	Tier 2	Ecopia
Mean AOI Data Creation or Clean Up Time	28 min	27 min	29 min	26 min
Mean Highly Urbanized AOI Data Creation or Clean Up Time	57 min	53 min	52 min	63 min
Mean Mid-Urbanized AOI Data Creation or Clean Up Time	27 min	31 min	30 min	28 min
Mean Low Urbanized AOI Data Creation or Clean Up Time	10 min	13 min	14 min	12 min

The accuracy metrics for the three AFE datasets are shown in Tables 3 - 6, first for buildings in Tables 3 and 4 and then for roads in Tables 5 and 6. Precision and recall are included for transparency but are both represented in the F1 score. All tables include metrics measured for the raw form of the data, before analyst editing, and for the edited form of the data, after an analyst has changed the polygons to align with the imagery. By comparing the scores for the raw versus edited data, the impact that an analyst's review has can be seen.

In both the Maracay and Niamey regions, Ecopia outperforms Tier 1 and Tier 2 data in mean time to correct (Table 1, Table 2). This aligns with analyst observations reporting that Ecopia data needed less editing overall to conform to reference data. Before editing, Ecopia AFE data scores higher than the other AFE in every category. After editing, the gap narrows considerably with other AFE occasionally outscoring Ecopia in some categories.

Notably, Tier 1 has remarkable improvement after editing that seems to signify some trait in the dataset that makes it stand out positively (Table 3, Table 4, Table 5, Table 6). After conferring with analysts, this positive performance is likely a side effect of how inaccurate the Tier 1 AFE is in its raw form, which led many analysts to erase the entire AOI and digitize from scratch, creating entirely new polygons that better matched the reference data. Thus, the edited accuracy ranking must be considered alongside the time metric, as Tier 1 displayed similar high accuracy to Ecopia at the cost of longer correction times and being essentially re-digitized by hand.

Table 3. Maracay building layer AFE performance compared to reference data.

Performance Metrics		Tier 1	Tier 2	Ecopia
IoU	Raw	0.75	0.72	0.77
	Edited	0.89	0.85	0.88
F1 Score	Raw	0.28	0.29	0.66
	Edited	0.80	0.64	0.80
Precision	Raw	0.28	0.43	0.66
	Edited	0.83	0.73	0.78
Recall	Raw	0.28	0.22	0.66
	Edited	0.77	0.58	0.83

Table 4. Niamey building layer AFE performance compared to reference data.

Performance Metrics		Tier 1	Tier 2	Ecopia
IoU	Raw	0.601	0.715	0.796
	Edited	0.962	0.957	0.866
F1 Score	Raw	0.098	0.463	0.677
	Edited	0.891	0.923	0.710
Precision	Raw	0.141	0.574	0.627
	Edited	0.859	0.910	0.610
Recall	Raw	0.075	0.388	0.736
	Edited	0.924	0.937	0.852

Across the two regions, the building layer scores (Tables 3, 4) reflect a general trend of improvement after editing, but diverge on the degree of improvement. In Maracay, the improvement is straightforward with the IoU scores before and after editing remaining within a few hundredths of one another for each AFE but showing consistent improvement (Table 3). The F1 score for Ecopia in the raw form is initially more than double that of the other AFE and after editing ties for the highest score overall with Tier 1 (Table 3). This suggests that Ecopia correctly extracts the most relevant features compared against the other AFE in this study.

In Niamey, the improvement is less stark. While Ecopia leads before editing in both IoU and F1, after editing, the other AFE datasets have better accuracy scores with Ecopia having reduced precision (Table 4). The minor decrease in precision for Ecopia indicates incorrect extractions when compared to the reference data, which would be a case of added features not aligning with a matching feature in the reference data. The seemingly closer alignment to the reference data post editing for Tiers 1 and 2 may be a case of Ecopia performing satisfactorily to capture building footprints while not aligning exactly with the reference data and analysts not editing it as heavily as the other AFE tiers as shown in Figure 8.

Figure 8 . Example of Ecopia (Yellow Hashed Line) and reference data (green line) representing the same structure.

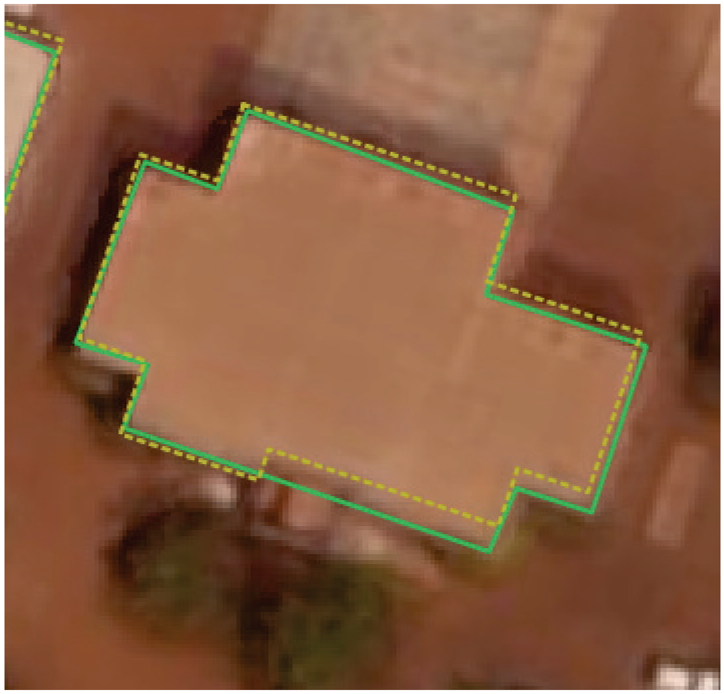


Table 5. Maracay roads AFE performance compared to reference data.

Performance Metrics		Tier 1	Tier 2	Ecopia
IoU	Raw	0.65	0.69	0.86
	Edited	0.63	0.69	0.89
F1 Score	Raw	0.71	0.80	0.87
	Edited	0.70	0.79	0.89
Precision	Raw	0.84	0.82	0.87
	Edited	0.82	0.80	0.88
Recall	Raw	0.65	0.80	0.89
	Edited	0.65	0.80	0.92
APLS	Raw	0.21	0.19	0.63
	Edited	0.42	0.49	0.81

Table 6. Niamey roads AFE performance compared to reference data.

Performance Metrics		Tier 1	Tier 2	Ecopia
IoU	Raw	0.11	0.32	0.70
	Edited	0.70	0.69	0.70
F1 Score	Raw	0.44	0.36	0.78
	Edited	0.94	0.92	0.87
Precision	Raw	0.61	0.38	0.84
	Edited	0.94	0.92	0.86
Recall	Raw	0.38	0.38	0.77
	Edited	0.95	0.93	0.89
APLS	Raw	0.20	0.22	0.50
	Edited	0.70	0.64	0.48

For roads, the IoU and F1 are useful indicators of pixel accuracy of the road features that were captured as shown previously in Figure 7. These scores demonstrate the correct extraction location and feature validity. However, APLS is a better indicator of how well connected the road network is and will be the focus of success in the road features. An example of the rationale can be found in the Maracay road feature layer, where the IoU and F1 for all AFE are all near 0.7 or higher, but the APLS score before editing hovers around 0.2 for Tier 1 and Tier 2, indicating a lack of connections in the extracted features (Table 5).

In both regions, the initial highest scorer in APLS was Ecopia. After editing, this remained true in the Maracay data (Table 5), but in the Niamey data both Tier 1 and Tier 2 overtook Ecopia after editing, while Ecopia had a minor drop of 0.02 in APLS (Table 6). This decrease means that network connectivity decreased while the F1 score increased, a situation illustrated in Figure 7. This suggests the addition of road segments that do not tie into the road network, potentially due to snapping errors during editing. The improvement of the Tier 1 and Tier 2 data is again likely attributable to analysts deleting the initial road features and re-digitizing in an effort to save time.

Given the consistent high initial scores in IoU, F1, and APLS, along with analyst impressions, the raw Ecopia data is the most accurate of the AFE tested in this study. The raw Ecopia data requires the least amount of time to correct to match the feature data of all the tested AFE. One consistent trend across both regions and performance metrics is the improvement of scores after analyst intervention. While AFE is a powerful tool, which can reduce the workload of an analyst, at this time it is not a replacement for an experienced analyst reviewing the output and making corrections. With the metrics used in this study, an analyst can get an appraisal of the quality of an AFE product without the need to visually review the entire dataset.

5 Summary

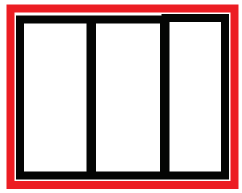
After reviewing the metric results for all raw and edited forms of the AFE datasets tested, Ecopia performs the best in the unedited stage and scores high after editing. While post-editing, Ecopia's metric scores did not consistently match the Tier 1 and Tier 2 scores, this was probably because of analysts' impressions that the raw form of Ecopia was sufficient to represent the features covered. Where available and practical to obtain, Ecopia data appears to be the best choice for feature representation when compared to the other AFE methods tested in this study. The results of Ecopia can be improved when enhanced by an analyst, but depending on the use case for the data, as well as time constraints, the raw form may be sufficient.

While analyst enhancements to Ecopia does add detail and missed features, it does not necessarily mean that the manually digitized reference data always outperformed Ecopia. In several AOI, when analysts compared Ecopia's initial results to the reference data, they were persuaded by the Ecopia data to revisit their judgement on some features and add them to the reference data. This is an example of the way that AI can supplement analyst judgment and lead to a superior end product.

Some limitations of the study and lessons learned are that a more robust foundational set of guidelines from the beginning of the study would yield a more consistent end product. While the analysts were well versed in geospatial products and data extraction, there exists room for different interpretations of the same guidelines. For instance, questions of what the minimum threshold size for a feature were not fully explored, which led some analysts to digitize small potentially temporary structures such as sheds, while others focused on larger permanent structures.

Additional uncertainty could be introduced by the difference in two methods for digitizing the same area. Illustrated in Figure 9, three theoretical abutting buildings could also be represented by one larger rectangle. Both options contain the correct area and approximate dimensions, but the larger option fails to capture all the detail of the situation. This situation was particularly likely in Maracay, due to the imagery quality previously described in the data section, which made discerning breaks in abutting buildings difficult. The tested AFE differed in their representation of buildings in this situation and due to disagreement with the reference data, may have had their performance score lower than would otherwise be the case.

Figure 9. Three abutting buildings (black) digitized as one building (red).



Future studies may benefit from testing more AFE techniques, including existing commercially available algorithms. Applying a more diverse selection of techniques to the same data would increase knowledge of the performance strengths and weaknesses of each AFE method. Using additional reference data to test AFE methods may also help to highlight any effect that different environments have on AFE performance. Existing analyst created datasets could potentially be leveraged to create additional reference data for further testing.

6 Supplemental Material

To access the code developed to test the AFE please go to https://gitlab.gs.mil/afee. The code was developed for use on any AFE output in addition to the ones specifically tested in this study. Access to the website does require a Common Access Card (CAC) and an invitation to view. Please contact either Hilary.K.McGown@nga.mil or Enrique.L.Montano@nga.mil for access.

References

Arcot Sowmya, J. T. 2000. "Modelling and prepresentation issues in automated feature extraction from aerial and satellite images." *ISPRS Journal of Photogrammetry and Remote Sensing* 34-47.

Benediktsson, J. A., M. Pesaresi, and K. Amason. 2003. "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations." *IEEE Transactions on Geoscience and Remote Sensing 41(9)* 1940-1949.

Bittner, K., F. Adam, S. Cui, M. Kömer, and P. Reinartz. 2018. "Building footpring extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks." *IEEE Journal if Selected Topics in Applied Earth Observations and Remote Sensing 11(8)* 2615-2629.

Chen, L., S. Zhao, W. Han, and Y. Li. 2012. "Building Detection in an urban area using lidar data and QuickBird imagery." *International Journal of Remote Sensing 33(16)* 5135-5148.

CosmiQ Works. 2020. *Solaris 0.3.0 : An open source machine learning pipeline for geospatial imagery*. https://github.com/CosmiQ/solaris.

Ecopia Tech. 2018. *Ecopia - Global Feature Extraction*. Accessed 2020. https://www.ecopiatech.com/global-feature-extraction.

Esri. 2020. *World Imagery [basemap]*. Accessed July 20, 2020. https://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer.

Katartizis, A., and H. Sahli. 2007. "A stochastic framework for the identification of building rooftops using a single remote sensing image." *IEEE Transactions on Geoscience and Remote Sensing 46(1)* 259-271.

Lu, T., D. Ming, X. Lin, Z. Hong, X. Bai, and J. Fang. 2018. "Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network." *Remote Sensing 10(9)* 1496.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert. 2018. "Attention U-Net: Learning Where to Look for the Pancreas." *arXiv preprint arXiv: 1804.03999.*

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z Huang, A. Karpathy, A. Khosla, M. Berstein, A. C. Berg, L. Fei-Fei. 2015. "ImageNet Large Scale Visual Recognition Challenge." *Int J Comput Vis 115* 211-252.

Toman, S. 2020. *Waterloo Stories: Creating the first complete map of our world - in real time*. January 2. Accessed July 24, 2020. https://uwaterloo.ca/stories/global-impact/creating-first-complete-map-our-world-real-time.

Van Etten, A. 2020. "City-Scale Road Extraction from Satellite Imagery v2: Road Speeds and Travel Times." *2020 IEEE Winter Conference on Applications of Computer Vision.*

Van Etten, Adam. 2017. *SpaceNet Road Detection and Routing Challenge — Part I.* The DownLinQ, CosmiQ Works.Yuan, J. 2018. "Learning Building Extraction in Aerial Scenes with Convolutional Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 40, no. 11* 2793-2798.

Acronyms

GIS – Geographic Information System

NGA – National Geospatial Intelligence Agency

IoU – Intersection over Union

CRESI - City- scale Road Extraction from Satellite Imagery

AFE – Automated Feature Extraction

APLS – Average Path Length Similarity

AOI – Area(s) of Interest

RGB – Red Green Blue

CNN – Convolutional Neural Network

nDSM – Normalized Digital Surface Model

OBIA – Object Based Image Analysis

OSM – Open Street Map

ROME – Research Open Mapping Enclave

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
March 2021	Final	

4. TITLE AND SUBTITLE

Evaluation of Automated Feature Extraction Algorithms Using High-resolution Satellite Imagery Across a Rural-urban Gradient in Two Unique Cities in Developing Countries

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Andrew Griffin, Sean Griffin, Kristofer Lasko, Megan Maloney, S. Bruce Blundell, Mike Collins, and Nicole Wayant

5d. PROJECT NUMBER
478420

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Geospatial Research Laboratory
U.S. Army Engineer Research and Development Center
7701 Telegraph Rd
Alexandria, VA 22314

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC/GRL TR-21-1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Geospatial-Intelligence Agency
7500 GEOINT Dr
Springfield, VA 22150

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

MIPR NR28G19138GS15

14. ABSTRACT

Feature extraction algorithms are routinely leveraged to extract building footprints and road networks into vector format. When used in conjunction with high resolution remotely sensed imagery, machine learning enables the automation of such feature extraction workflows. However, many of the feature extraction algorithms currently available have not been thoroughly evaluated in a scientific manner within complex terrain such as the cities of developing countries. This report details the performance of three automated feature extraction (AFE) datasets: Ecopia, Tier 1, and Tier 2, at extracting building footprints and roads from high resolution satellite imagery as compared to manual digitization of the same areas. To avoid environmental bias, this assessment was done in two different regions of the world: Maracay, Venezuela and Niamey, Niger. High, medium, and low urban density sites are compared between regions. We quantify the accuracy of the data and time needed to correct the three AFE datasets against hand digitized reference data across ninety tiles in each city, selected by stratified random sampling. Within each tile, the reference data was compared against the three AFE datasets, both before and after analyst editing, using the accuracy assessment metrics of Intersection over Union and F1 Score for buildings and roads, as well as Average Path Length Similarity (APLS) to measure road network connectivity. It was found that of the three AFE tested, the Ecopia data most frequently outperformed the other AFE in accuracy and reduced the time needed for editing.

15. SUBJECT TERMS

Algorithms – Evaluation
Automation
Pattern recognition systems
Machine learning
Remote-sensing images
Geospatial data
Cities and towns

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	SAR	28	