Characterizing Irregular Settlements Using Machine Learning and Satellite Imagery: Case Study of Bengaluru, Karnataka, India

May 2018

Duke SANFORD SCHOOL OF PUBLIC POLICY

Nikhil Kaza, Associate Professor, City & Regional Planning, UNC at Chapel Hilll, nkaza@unc.edu

Raju Vatsavai, Associate Professor, Computer Science, NC State University, rrvatsav@ncsu.edu

Motivation

About a third of the urban population in the world (~830 million) live in irregular/informal settlements/favelas/shanty towns/slums (UNDP, 2017). These settlements are at the core of many urban regions around the world, yet they are vastly undercounted or uncounted in official estimates. In urban areas, these types of settlements can constitute 30%-60% of the city surpassing, in some cases, the extent of formal neighborhoods. Identifying irregular settlements, mapping and monitoring them using traditional approaches is costly and labor intensive.

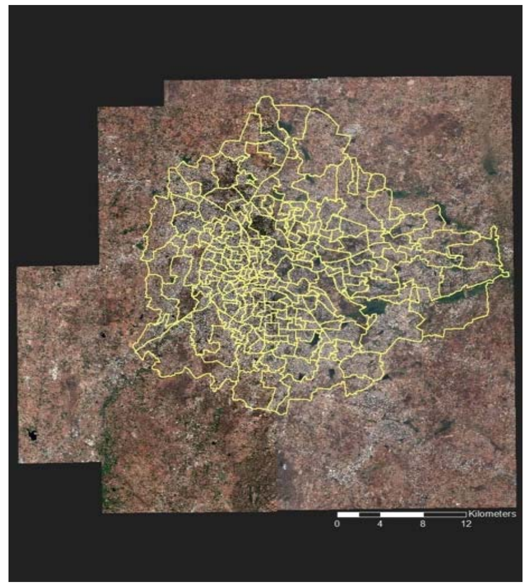

Figure 1 Size & Extent of Study Area, Bengaluru, India

Problem

Very high-resolution multi-spectral satellite imagery has proven to be highly useful settlement mapping; however, irregular settlement mapping is still an open challenge. In this study, we explored various feature extraction and machine learning approaches for automated identification of irregular settlements using 2m resolution data (2016) and their condition in 2002 using 30m resolution data. The area of the study was urban Bengaluru, as depicted in Figure 1. This research brief is a companion to the study done by Anirudh Krishna, Erik Wibbels and team at Duke University which describes that informal settlements come in many different types and built environments in these settlements vary along a continuum. The worst-off slums consist of temporary structures – four poles and a rough tarpaulin roof – but there are three-story concrete structures at the other end of the continuum

The authors are grateful, in addition to Omidyar Network, to the Jana Urban Foundation, Professor MS Sriram, Grady Lenkin, and IIM Bangalore. We thank the Indian Space Research Organisation for many helpful discussions. DigitalGlobe Foundation kindly donated the high resolution satellite imagery used in this project.

Characterizing Irregular Settlements Using Machine Learning and Satellite Imagery

Key Findings

Figure 2. Changes (left: ~2002 vs. right: 2016). Regions: Top left (Near Yeswanthpur Suburb II Stage), Middle (Near BEML Layout/ Thubarahalli Lake), and Bottom (Near Dollar Layout/JP Nagar Layout) Left images: Google Earth. Right images: Courtesy of the DigitalGlobe Foundation

- **The actual number of slums is far more than those stated in government records**: Between 4-17% of the city area and between 600-2000 neighborhoods are classified by the algorithms as irregular depending on the classification method. Classification evaluation on independent test data showed greater than 70% accuracy in correctly identifying formal and informal settlements. Our findings suggest that these settlements are lot more numerous and in different locations than the ones located by the Karnataka Slum Development Board.
- **A number of new settlements have evolved over last 15 years**: Change detection method showed that several irregular settlements were formed after 2002 (See Figure 2).
- **Algorithms are better at detecting slums at the lower end of the continuum with poorer infrastructure and housing quality**: We observed that certain type of settlements (e.g., temporary, semi -permanent) can be accurately discriminated as compared to others (e.g., multi-story). The slums with better housing quality tend to significantly overlap with formal settlements. Further study (both features and classification methods) is required to improve the accuracy of these algorithms to identify irregular settlement types.
- **Features such as vegetation index, built-up index and road-density help improve accuracy of the algorithms**: Spectral features alone are insufficient for irregular settlement mapping. Features that account for spatial context and autocorrelations are critical for identifying various settlement types. We observed almost 15-20% improvement in classification with additional features, such as Vegetation indices, Haralick textures, Built-up index, Morphological Building Index, road density etc.
- **Patch classification method showed greater promise than pixel-based classification**: Most machine learning approaches are designed to deal with single instance (or pixel). However, irregular settlement mapping requires reasoning with object sizes and their spatial arrangements at a bigger spatial footprint. Though adding structural features greatly improved accuracy, natural approach would be deal with image patches. Deep learning (CNNs) methods showed great promise (71% accuracy).

Research Implications

- The addition of ancillary datasets, such as roads, water bodies improve with the accuracy of the classification methods. However, the coverage, spatial and thematic accuracy of these data from available datasets is varied at best.
- The two trained human coders who visually interpreted the image to identify the boundaries of irregular settlements produced substantially different results compared to one another, demonstrating the significant open challenge of the problem of neighborhood classification. It is likely that visual interpretation by large number of coders are likely to achieve convergent results and accurate labels.
- The satellite image classification techniques can reduce and focus field efforts but won't completely substitute them. Accurate field verification in fact is crucial to providing labels for the neighborhoods, which in turn improves the classification methods.
- Irregular settlement mapping requires high-spatial resolution. Investments should be made in novel aerial data collection such as Doves and Drones. However, historical analysis will have to rely on medium resolution imagery such as Landsat (30m; 8 bands). At this resolution, coarse thematic classes can be delineated and can help with identifying land cover change rather than neighborhood types.
- Extending study sites to other cities and collecting large number of image patches for training can improve accuracy and help generalize the classification methodology.

Policy Implications

- Analysis of satellite imagery can be an effective way to track human settlement patterns. It will be important to invest in building technical capacity for geospatial analysis to support hyper-local urban planning: Understanding human settlement patterns in rapidly urbanizing cities with exploding population growth is important as it creates stress on civic resources and public utilities. Geospatial analysis can help identify stress zones, and allow civic authorities to focus their efforts in localized areas. It will be important to develop technical capacity within government and policy research institutions to analyze large volumes of geospatial data for improved urban planning, allowing authorities to plan and prioritize at a hyperlocal level.
- Robust, decentralized national level Spatial Data Infrastructure that will maintain and disseminate accurate thematic data is important for future efforts: A national level spatial data infrastructure, as available in many developed countries, will help build this technical capacity. Access to high quality data on roads, infrastructure etc. will improve the efficacy of different machine learning algorithms. Moreover, it is important for different government bodies e.g. housing, sanitation, roads, electricity etc. to share common geospatial databases.

Characterizing Irregular Settlements Using Machine Learning and Satellite Imagery

Methodology

We used the 8-band, 2m resolution data from DigitalGlobe from 2016 for a 2000 sq.km area. We also collected ancillary spatial data from several sources including governmental sources such as Bhuvan, CartoSat and OpenStreetMap to include datasets such as Digital Elevation Model, Airports, Bus stops, Cemeteries, Commercial property, Fire stations, Fuel stations, Golf courses, Greenspace, Hospitals, Industrial property, Libraries, Police stations, Rail networks, Road networks, Schools and Sports facilities. We merge information from these multiple sources to construct 1 ha neighborhood attributes. Most machine learning approaches deal with characterizing one pixel at a time; however, irregular settlement recognition requires reasoning with image patches in order to characterize different neighborhoods. To address this challenge, we explored two distinct approaches: (i) extract structural features that account for spatial autocorrelations followed by pixel-based classification methods (Naïve Bayes, Decision Trees, K Nearest Neighbors, Multilayer Perceptron, Boosted Trees, Random Forests), and (ii) image patch-based classification using deep learning (convolutional neural networks – CNN) algorithms. These approaches were used to classify 1 ha neighborhoods into 4 types of Irregular settlements; Temporary, Semi-Permanent, Single Story and Multi-Story We used 123 irregular neighborhoods identified in previous field surveys as a training sample for the classification algorithms. Post-processing is done to eliminate small areas. We have also designed a change detection scheme to identify the status of irregular settlements by classifying Landsat image (30m) from 2002. Due to limitations in spatial resolutions, we could only identify the status of settlements in terms of broad land-use changes. In addition to the machine learning algorithms, two human coders also employed the entire image by focusing on each ½ x ½ km grid. The coders were asked given few examples of the irregular settlements of different types and were asked to systematically scan the image for the settlements and demarcate them.