Computationally Efficient Algorithms for Settlement Mapping Using Very High-resolution Satellite Imagery

Ranga Raju Vatsavai[1,2]

[1]North Carolina State University, 890 Oval Drive, Campus Box 8206- Raleigh, NC 27695-8206
[2]Oak Ridge National Laboratory, One Bethel Valley Rd., Oak Ridge, TN, 37831, USA
919-515-6019
Email: rvatsavai@ncsu.edu

Abstract

There is a great need for identifying and characterizing human settlements at global scale. Though very high-resolution (VHR) imagery has proven to be highly useful in identifying human settlements, the algorithms and computational approaches have proven to be inadequate and very slow. Existing per-pixel based classification approaches are shown to be inadequate for characterizing urban neighborhoods in VHR imagery. In this paper, we present a computationally efficient and parallel approach for mapping human settlements using VHR imagery.

Keywords: human settlement mapping, very high-resolution imagery, multiple instance learning, complex object based image analysis.

1. Introduction

Multi-spectral remote sensing imagery is widely used in mapping settlements, forests, crops and other natural and man-made objects on the Earth. With the recent launch of satellites by private companies such as Digital Globe (e.g., WorldView-2 in late 2009), applications around very high-resolution (VHR) imagery (sub-meter) are emerging fast. Such imagery provides new opportunities to monitor and map both natural and man made structure across the globe. Despite the great efforts by research community across the globe, neighborhood mapping is a challenging task. First of all, neighborhoods are not well defined, which is reflected in the quote by Galster (2001) – “Urban social scientists have treated ‘neighborhood’ in much the same way as courts of law have treated pornography: a term that is hard to define precisely, but everyone knows it when they see it.” There is no consistent nomenclature across the countries about neighborhoods and no consistent ground-truth, making it very difficult to build machine learning models for global scale problems. In addition, most of the neighborhoods are made up of complex objects (consisting of different types of objects, not just buildings and roads), which make it very difficult to obtain ground-truth data from images.

Mapping informal settlements is an important task both from national security and as well as humanitarian grounds. The high rate of urbanization, political conflicts and ensuing internal displacement of population, and increased poverty in the 20th century has resulted in rapid increase of informal settlements. These unplanned, unauthorized, and/or unstructured homes, known as informal settlements, shantytowns, barrios, or slums, pose several challenges to the nations, as these settlements are often located in most hazardous regions and lack basic services. Though several World Bank and United Nations sponsored studies stress the importance of poverty maps in designing better policies and interventions, mapping slums of the world is a daunting and challenging task. In this work, we present a computationally efficient and automated framework that is capable of identifying different neighborhoods (including informal settlements or slums).

VHR image classification poses several challenges because the typical object size is much larger than the pixel resolution. Any given pixel (spectral features at that location) by itself is not a good indicator of the object it belongs to without looking at the broader spatial footprint. However, existing per-pixel (single instance) based thematic classification schemes are designed for moderate spatial resolution (10 meters and above). This is not to say that well-known single instance learning algorithms are not applicable in classifying VHR images, in fact they are highly effective in identifying primitive objects such as buildings, roads, forest, and water. However, what we are pointing at is that the single-instance learning algorithms are inadequate in modeling complex (spatial) patterns. The same limitations are also applicable to spatial contextual classifiers (e.g, Markov Random Fields), as these classifiers look at the immediate neighboring pixels to modify the label of a single instance. Therefore, there is a great need for newer approaches, which looks at a bigger window or image patch or segment (consisting 100's of adjacent pixels) in building a classification model. These concepts were illustrated in fig 1. Image shows a small region from VFR image, and sub-images (a) and (b) shows pixels from two different categories, bare soil and rooftop respectively. Individual pixels from these locations are very similar, therefore hard to discriminate (using pixel-based (or single instance) classifiers). However, if we look at a bigger spatial footprint (e.g., small window around the pixel), then these two categories can be easily discriminated (using multiple-instance classifiers).

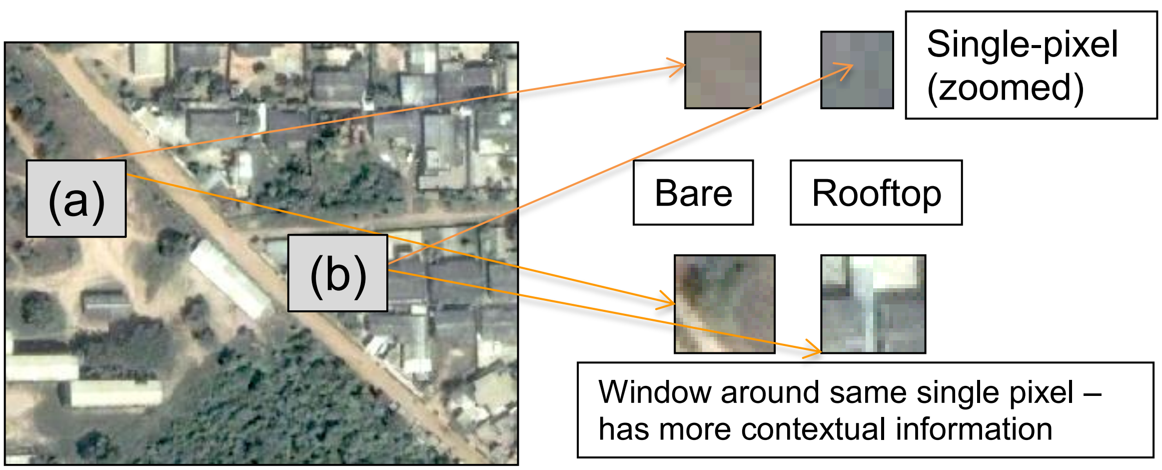
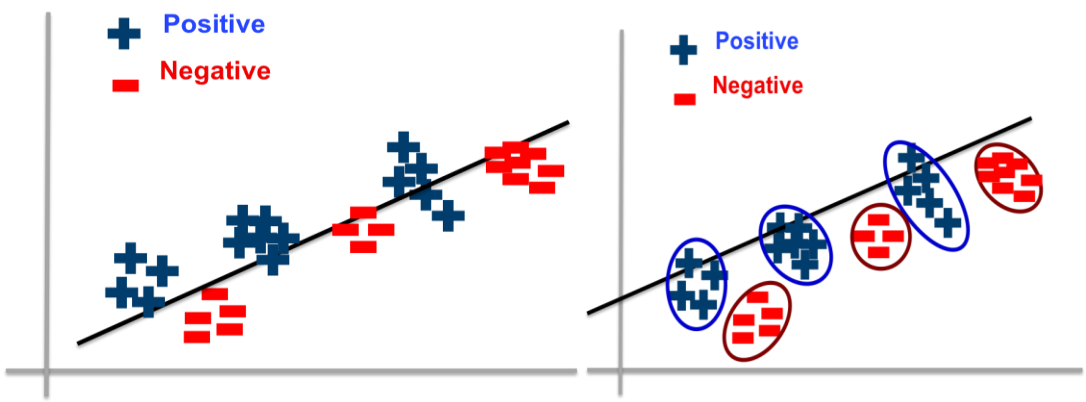


Figure 1. VHR Image. (a) Bare soil, and (b) Rooftop

2. Related Work and Limitations

Most of the existing classification approaches work with spectral features (e.g., blue, green, red, thermal infrared) and derived features (e.g., texture, band ratios like Normalized Difference Vegetation Index (NDVI), Histogram of Oriented Gradients (HOG)), extracted from each pixel (spatial location). These features were then used to learn a classification (single-instance or pixel-based) model using ground-truth data, and the model is then applied to the entire image (study regions) to classify each pixel into one of the predefined thematic class. A review of these techniques can be found in (Graesser et al. 2011, Vatsavai R. R. 2012). Global scale settlement mapping is also gaining momentum in recent years (Pesaresi et al. 2013). These methods were predominantly single-instance or pixel-based classifiers. Though incorporating features (e.g., texture) might provide some contextual information, still these methods cannot take the benefit offered by larger window, patch or segments. Patch-based and object-based approaches are also becoming quite popular (Vatsavai 2013, Blaschke 2010, Lang, 2008), but these methods were good at identifying/classifying individual objects (e.g., buildings, trees) but not the complex patterns (e.g., slums). Recently, we developed multiple-instance learning based algorithms (Vatsavai et al. 2013, Vatsavai 2013) which showed improved performance over single-instance learning algorithms for settlement mapping, however these approaches are computationally expensive. In this work, we present computationally efficient and shared memory-based parallel implementation results.



(a) Single Instance Learning (SIL)   (b) Multiple-instance Learning (MIL)

Figure 2. SIL vs. MIL

3. Complex Object Based Image Analysis

Though object based image analysis (OBIA) is becoming popular, OBIA methods require segmenting the image first. Image features are then extracted for objects (or segments), for example, area, shape, and size. Single instance learning algorithms like decision trees, neural networks, or support vector machines are then used to classify (label) objects (segments). However, image segmentation is still a challenging task in itself (Geoffrey 2008), and it is very difficult to obtain good segmentation in informal settlements. To alleviate these problems, we proposed a complex object based image analysis (COBIA) framework based on multiple instance learning (Vatsavai 2013), where segments (regular grids) consist of many full or partial objects (see fig. 3). Based on composition of objects, multiple instance learning scheme allows to label different segments into various neighborhoods.

Multi-instance (or Multiple instance) learning (MIL) methods have been developed to overcome some of the limitations of single instance learning schemes. Notable approaches include the axis-parallel rectangles (Dietterich et. al., 1997), Diverse Density

(O. Maron, et. al., 1998), and Citation-KNN (J. Wang, 2000). Basic distinction between single-instance learning (SIL) and multiple-instance learning (MIL) is shown in Figure 2. In SIL, objective is to minimize classification error, that is, minimize the number of pixels misclassified (Figure 2(a)). On the other hand, in MIL, objective is find a decision boundary such that entire window (segment or bag) is correctly classified. As shown in Figure 2(b), the decision boundary is optimized such that positive and negative bags are separated using decision rule just described. Key point to note here is that in multi-instance learning entire bag is assigned a single label, where as in single instance learning a single bag may have both positive and negative instances. Therefore, single instance learning algorithms are appropriate for thematic classification (e.g., roads, buildings), whereas multi-instance learning algorithms are designed for recognizing complex patterns (e.g., informal and formal settlements). We now describe our MIL framework briefly, more details can be found in (Vatsavai, et. al. 2013; Vatsavai, 2013). This framework admits both Citation-KNN (J. Wang, 2000), and Gaussian-MIL (Vatsavai, 2013) classifiers. Basic difference is that GMIL model each bag (window) as Gaussian distribution whereas Citation-KNN treats each winds as bag of pixels (feature vectors).

Basic primitives of framework are as follows:

1. Divide the image into regular grids (or patches) (see fig. 3)
2. A fast training acquisition system
3. Construction of learning model from the training data
4. Match query bag with the bag of Gaussians
5. Apply nearest neighbor based classifier to assign a class label to the query bag

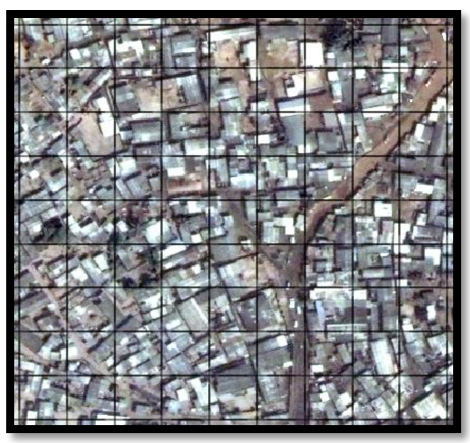


Figure 3: Square Grid Based Segments 1

This learning scheme (steps 3-5) is summarized in fig. 4. The first 2 steps are straightforward. Step (3) shows the learning model. Step (4) shows difference between Citation-KNN and GMIL. As shown in figure, in GMIL each window (bag) is treated as a Gaussian distribution and Citation-KNN is window is treated as a bag of feature vectors (points in multi-dimensional feature space). Classification (step 5) is done by: (i) computing the similarity between each window and all training windows, (ii) ranking based on similarity, and (ii) assign label based on majority. Citation-KNN uses Hausdorff distance for computing similarity between bags and GMIL uses the KL Divergence measure.

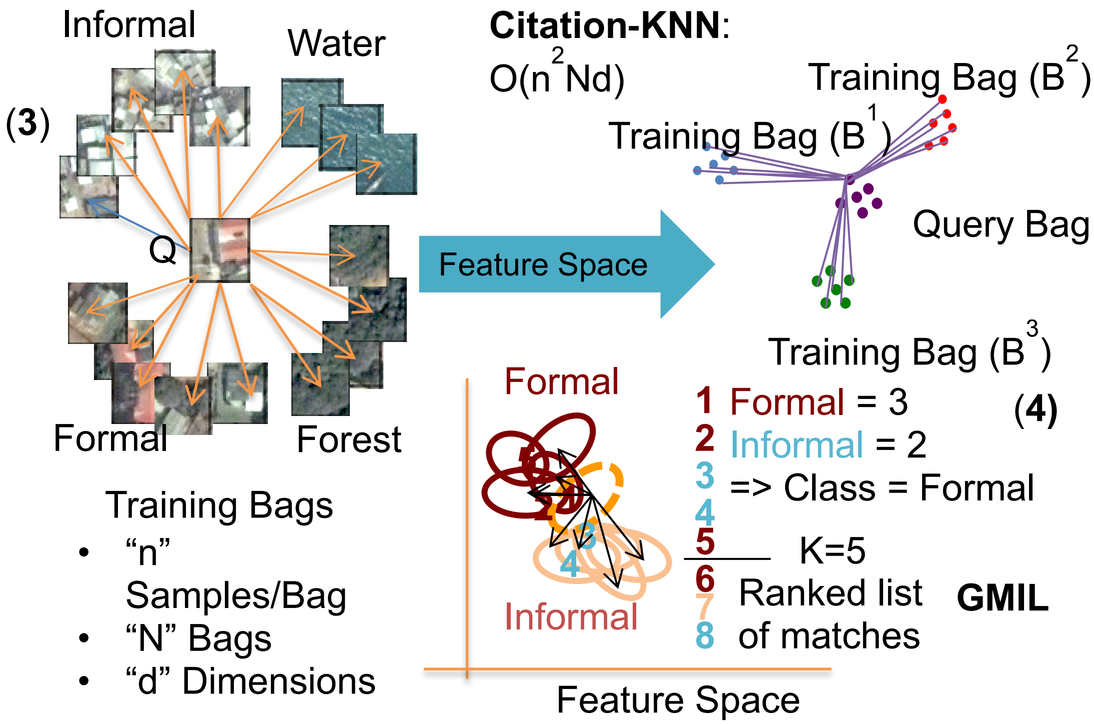


Figure 4. GMIL Classification Based COBIA Framework

3.1 Computationally Efficient Implementation

Citation-KNN and GMIL are both computationally expensive, though GMIL is order of magnitude faster than Citation-KNN. The computational complexity of Citation-KNN is $O(n^2Nd)$, where "n" is the average number of instances per bag, N is the number of training bags, and "d" is the number of features (dimensions). The "n^2" complexity comes from the fact that one has to compute pair-wise distance between "n" instances in training bag and "n" instances from the each bag that needs to be classified. This complexity is reduced for GMIL as the bag is treated as a Gaussian distribution and similarity is computed as KL divergence between two bags (Gaussians). Basic parallelization strategy is shown in fig. 5. We implemented a divide-and-conquer like parallelization strategy. As shown in fig. 5 we recursively breakdown the image into smaller grids (square segments). The size of final grid is determined two parameters; uniformity and user defined minimum grid-size. Each thread is initialized with training bags. Each bag (patch) that needs to be classified is assigned to the available thread in the pool. Each thread then matches the assigned patch to the training patches by computing similarity and raking the matches (steps 3 and 4 in fig. 4). The final output map is generated by assembling all the labelled patches (see fig. 6).

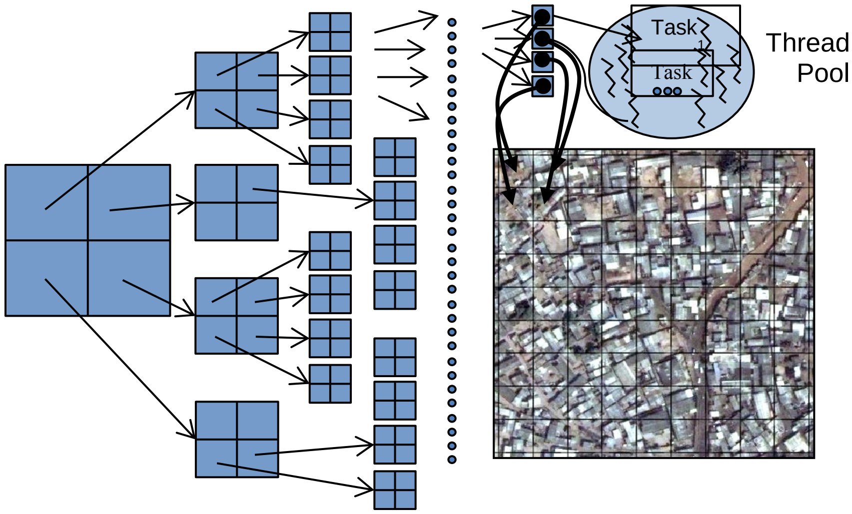


Figure 5: Parallelization Architecture 1

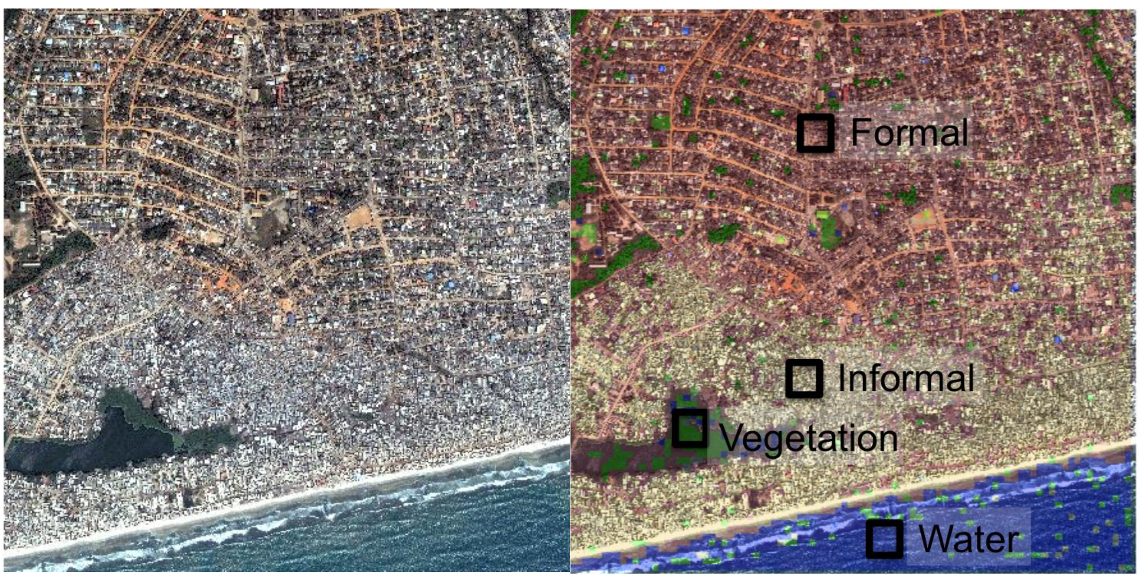


Figure 6. Raw (RGB) and Classified (GMIL) Images.

4. Results and Analysis

Detailed classification accuracy results were presented in (Vatsavai, et. al. 2013; Vatsavai, 2013). Table 1 shows comparison between MIL and leading SIL algorithms. In summary, GMIL consistently performed well over several SIL classifiers (Logistic Regression, Random Forests, Multilayer Perceptrons, and Naïve Bayes) and as well as Citation-KNN.

Parallelization results (accuracy) of GMIL are same as the sequential implementation. For parallelization, all the experiments were preformed on 1 km^2 image (1 meter pixels), with 10 x 10 bag (minimum patch) size. There are 10,000 bags, out of which 380 blocks were used for training. All the experiments were conducted on a single-node consisting of dual Xeon hex-core processors (3.46 GHz) with 48GB RAM. Sequential version of Citation-KNN took 27.8 hours and GMIL took 3.1 hours. GMIL is not only computationally efficient but also accurate (see table 1). Parallel implementation of GMIL took 20 minutes.

City	Citation-KNN	Regression	RF	MLP	NB	GMIL Model
Accra	76.25	71.25	72.08	69.58	75.66	95.66
Caracas	82.96	78.15	81.85	81.81	74.07	85
La Paz	80.97	77.17	78.26	80.23	76.08	83.25
Kandahar	79.78	64.89	69.14	73.93	60.1	81.2

Table 1. Comparison of MIL vs. SIL classification performance

5. Conclusions

In this paper, we presented a scalable implement of multiple instance learning (GMIL) based classification scheme for human settlement mapping. MIL approaches though perform better than traditional SIL algorithms (see table 1), they are computationally expensive. We presented a divide-and-conquer based parallel implementation on shared memory system, which showed 9x performance improvement for GMIL classification scheme. We are working on GPU implementation, which we are hoping to bring the classification of 1 km^2 image to less than a minute.

6. Acknowledgements

The preliminary work was carried out while the author was at the Oak Ridge National Laboratory. We would like to thank our collaborators B. Bhaduri, J. Grasser, E. Bright, and A. Chariyadat for their invaluable feedback, and anonymous reviews whose comments helped us to improve the technical quality of this manuscript.

7. References

Blaschke T, 2010, “Object based image analysis for remote sensing,” ISPRS J. Photogramm. Remote Sens., vol. 65, pp. 2–16, Jan. 2010.

Dietterich T. G, Lathrop R. H, Lozano-Perez T and Pharmaceutical A, 1997, Solving the multiple-instance problem with axis-parallel rectangles. Artificial Intelligence, 89:31–71.

Galster G, 2001, On the nature of neighbourhood. Urban Studies, 38(12):2111–2124.

Graesser J, Cheriyadat A, Vatsavai R. R, Chandola V, Long J and Bright E, 2012. Image based characterization of formal and informal neighborhoods in an urban landscape. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 5(4):1164 –1176.

Lang S, 2008, Object-based image analysis for remote sensing applications: modeling reality–dealing with complexity. Object-Based Image Anal., pp. 3–27.

Maron O and Lozano-Parez T, 1998, A framework for multiple-instance learning. In ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, pages 570–576. MIT Press.

Pesaresi, M.; Guo Huadong; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Linlin Lu; Marin-Herrera, M.A.; Ouzounis, G.K.; Scavazzon, M.; Soille, P.; Syrris, V.; Zanchetta, L., "A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results," Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, vol.6, no.5, pp.2102,2131, Oct. 2013.

Smith G, and Morton D, 2008, Segmentation: The Achilles' heel of object–based image analysis?. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(4).

Vatsavai R. R, 2011, High-resolution urban image classification using extended features. In IEEE ICDM Workshop on Spatial and Spatiotemporal Data Mining (SSTDM), pages 869–876.

Vatsavai R. R, 2013, Object based image classification: state of the art and computational challenges. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, pp. 73–80, ACM.

Vatsavai R. R, Bhaduri B, and Graesser J, 2013, Complex settlement pattern extraction with multi-instance learning. In Joint Urban Remote Sensing Event, pages 1 –4.

Wang J, 2000, Solving the multiple-instance problem: A lazy learning approach. In In Proc. 17th International Conf. on Machine Learning, pages 1119–1125. Morgan Kaufmann.