Gaussian Multiple Instance Learning Approach for Mapping the Slums of the World Using Very High Resolution Imagery*

Ranga Raju Vatsavai[†]
Computational Sciences and Engineering Division
Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
vatsavairr@ornl.gov

ABSTRACT

In this paper, we present a computationally efficient algorithm based on multiple instance learning for mapping informal settlements (slums) using very high-resolution remote sensing imagery. From remote sensing perspective, informal settlements share unique spatial characteristics that distinguish them from other urban structures like industrial, commercial, and formal residential settlements. However, regular pattern recognition and machine learning methods, which are predominantly single-instance or per-pixel classifiers, often fail to accurately map the informal settlements as they do not capture the complex spatial patterns. To overcome these limitations we employed a multiple instance based machine learning approach, where groups of contiguous pixels (image patches) are modeled as generated by a Gaussian distribution. We have conducted several experiments on very high-resolution satellite imagery, representing four unique geographic regions across the world. Our method showed consistent improvement in accurately identifying informal settlements.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining; I.5 [**Pattern Recognition**]: Models—*Statistical*

*Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT-Battelle, LLC for the U. S. Department of Energy under contract no. DEAC05-00OR22725.

†(Partial) support for this work was provided through Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research (and Basic Energy Sciences/Biological and Environmental Research/High Energy Physics/Fusion Energy Sciences/Nuclear Physics)

KDD'13, August 11–14, 2013, Chicago, Illinois, USA.

Keywords

Spatial Data Mining, Remote Sensing, MIL

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. On the other hand, very high resolution (VHR) imagery is useful in mapping complex patterns, such as formal and informal settlements. 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 (consisting 100's of adjacent pixels) in building a classification model.

In this work, we present a classification framework based on image window or patch (multi-instance) learning for mapping informal settlements using VHR images. From remote sensing perspective, informal settlements share unique spatial characteristics that distinguish them from other urban structures like industrial, commercial, and formal residential settlements [10]. To overcome the limitations posed by single-instance classifiers in modeling complex patterns, we developed a novel multi-instance based machine learning approach, which showed improvements in accurately identifying informal settlements. We have conducted several experiments on high-resolution satellite imagery, representing four unique geographic regions across the world.

1.1 Application Significance

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 (submeter) are emerging fast. Such imagery provides new opportunities to monitor and map both natural and man made structure across the globe. For past several years, we are engaged in developing several new approaches to efficiently process these imagery to support applications of national importance, such as biomass monitoring [6, 7], nuclear proliferation monitoring [22,24], and settlement mapping [10] at finer spatial and temporal scales. 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 20^{th} 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. This work is a step towards developing a computationally efficient and automated framework that is capable of detecting new settlements (especially slums) across the globe.

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). This process is shown in Figure 1. Typical classification involves: (i) collection of ground-truth (training/test) data at few locations (Figure 1(a)) in the image for simple thematic classes (e.g., urban, water, forest, agriculture) or finer classes (e.g., high-density urban, low-density urban, deep water, shallow water, hardwood forest, conifer forest, soybean, wheat, corn), (ii) build a classification model (e.g., Naive Bayes, decision trees, neural networks), and (iii) predict labels for entire image (Figure 1(b)). A review of these techniques can be found in [10, 25].

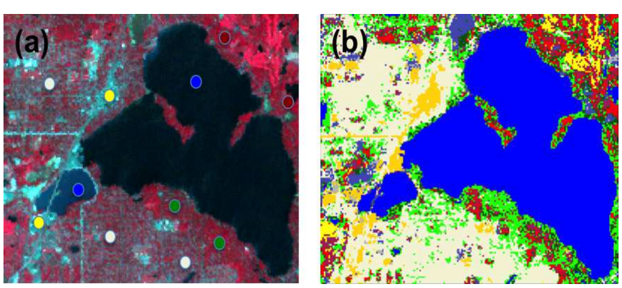

Figure 1: Raw satellite image and the corresponding classified image

An improvement over per-pixel classification schemes are the spatial classification schemes such as MRF [18]. Most classification schemes model the correlations in feature space and often ignore spatial correlations in spatial data, such as satellite images. In spatial classification schemes both spatial correlations (context) and feature correlations are modeled simultaneously, as a result the final classified image contains much smoother (spatially) class distribution and eliminates salt and pepper noise. Figure 2 compares (a) non-spatial (maximum likelihood) and (b) spatial (MRF) classification schemes. However, it should be noted that both of these schemes based on single instance learners.

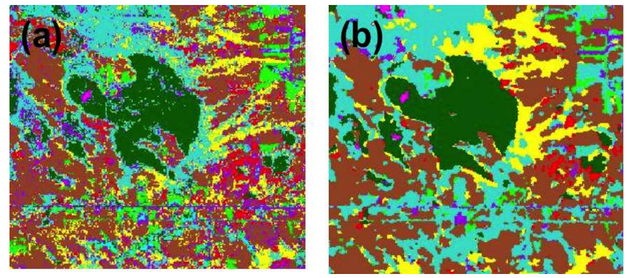

Figure 2: Classified images: (a) non-spatial, (b) spatial

Though single instance learning schemes are widely used in remote sensing image classification (approximately 30 meter spatial resolution), they are mostly applied for discriminating simple thematic classes (both aggregate and finer). However with increasing spatial resolution (sub-meter) current satellite images contains much more spatial heterogeneity (rich spatial information), see Figure 3. As result, it is possible to extract more complex classes, such as, informal (slums, shanty towns, burrows) settlements from these very high resolution images. Single instance learning (non-spatial or spatial) are ineffective in such cases due to the fact that the size of the pixel (less than one m^2) is much smaller than the size of the objects (for example, average building size in US is 250 m^2).

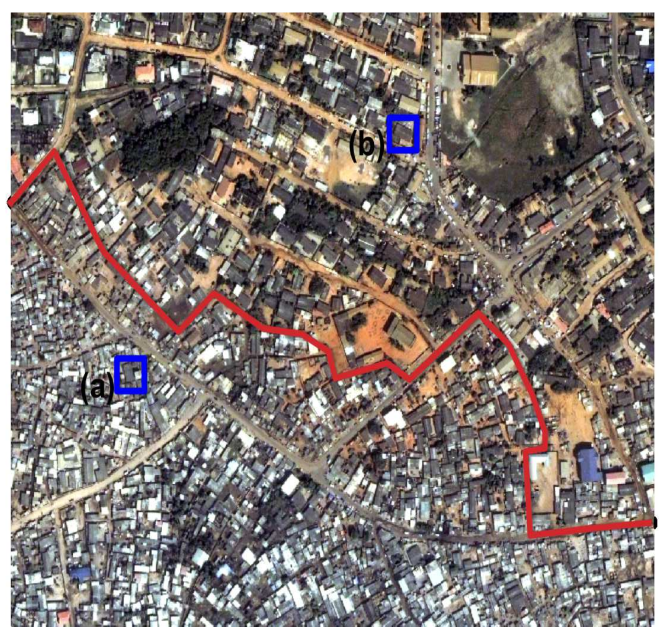

Figure 3: Accra (Ghana) city with two distinct settlement patterns speared by hand drawn boundary (red)

One way to overcome single instance limitation is to look at additional features beyond spectral features, because features that exploit spatial contextual information are highly useful in the classification of very high-resolution images. Recent studies [10,21,25] show the improved performance of single instance learners when the spectral features are combined with a broad set of extended features such as morphological, texture, and edge density. Although these studies showed that the extended features which exploit spatial contextual information resulted in improved accuracy, the classification schemes utilized are still single-instance learners. For complex object recognition we need to look at a bigger spatial region. Figure 3 illustrates this problem clearly. As can be seen in this figure, humans can easily identify two distinct settlement patterns, north of boundary is formal and south is informal. However, there is not much difference between spectral values if you look at individual pixels drawn for similar objects across the boundary, for example, buildings. To illustrate this point, we selected two small windows (blue windows) around buildings, window (a) is drawn from informal settlements and window (b) is drawn from the formal settlements. These windows are zoomed in Figure 4. As one can see, the difference between these building (at pixel level) is minimal, as a result per-pixel (single instance) classification schemes cannot predict them into formal vs informal classes, rather they can accurately predict both into buildings. On the other hand, when you a look at bigger patch as a whole, then it is easy to see the difference between informal and formal settlements, as these patches contains much richer spatial information.

Figure 4: Two small patches drawn from informal and formal settlements in Accra city

Object based classification schemes [3, 15] are one step in that direction. Typically, object based methods seek to segment the image into meaningful objects by exploiting spatial and spectral features. One can build a meta classifier on the features extracted from the objects, for example, area, perimeter, compactness, shape index, and fractal dimension. Or one can aggregate all feature vectors into a single feature vector and then apply any single instance learning algorithm. However, all these approaches loose important structural and spatial properties in the aggregation process.

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 seminal work of Dietterich et. al. [8], Diverse Density [14], and Citation-KNN [26]. Recently, MIL algorithms have also been applied to remote sensing image classification as well. For example, in [20] MIL approach is explored for sub-surface landmine detection using hyperspectral (HS) imagery. In [4], authors have developed MIL based binary classification scheme for identifying targets (landmines) in HS imagery. While each of these algorithms have advantages and disadvantages over per-pixel based classification schemes, in general they are shown to perform (accuracy) better than single instance learning schemes. Key idea behind multi-instance learning schemes is the utilization of all instances drawn from the image patches or windows. In multi-instance learning, the training data consists of many bags (windows) where each bag contains several examples (pixels). A bag is positively labeled if it contains at least one positive instance (e.g., informal settlement) and negative otherwise (e.g., formal settlement). This scheme is conceptually depicted in the Figure 5(a). As shown in this figure, the decision boundary is optimized such that positive and negative bags are separated using decision rule just described. On the other hand, Figure 5(b) shows traditional single instance learning schemes where the objective is minimize the number of misclassified (single) instances. 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).

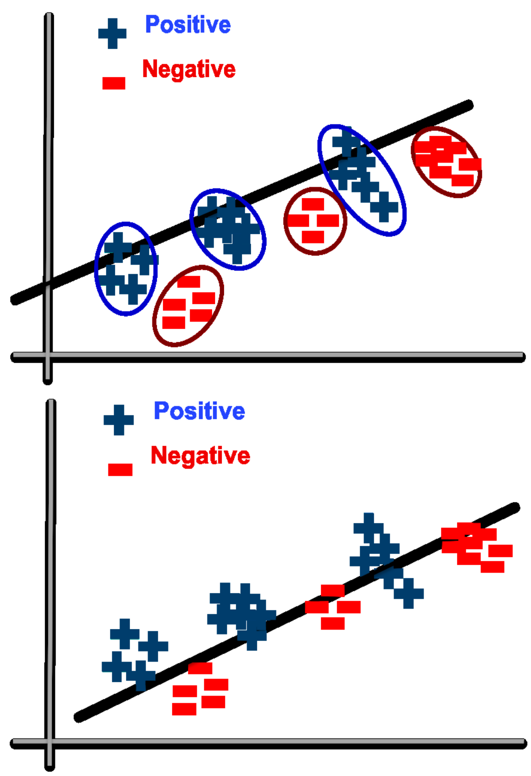

Figure 5: Decision boundaries resulting from: (a) multi-instance learning, and (b) single instance learning

In a recent feasibility study we successfully applied Citaion-KNN algorithm for complex settlement mapping [23]. However, the high computational cost of Citaion-KNN has led us to design a computationally efficient algorithm. This algorithm is very similar to the Citation-KNN approach in spirit. In Citation-KNN approach, the similarity between bags (patch in image parlance) is determined by minimizing the Hausdroff distance which is computationally expensive. As can be seen in the following sections, our proposed method is not only computationally efficient but is also consistently accurate than the Citation-KNN algorithm.

3. GAUSSIAN MULTIPLE INSTANCE LEARNING (GMIL)

The key idea behind the proposed GMIL algorithm is to model each bag as a Gaussian distribution. As a result we are simplifying the computation of similarity measure (e.g., Hausdroff distance) and at the same time capturing the heterogeneity through the covariance matrix. Parameters of this distribution can be estimated from the training data directly. Since we are abstracting each window (or patch) as a Gaussian distribution, the trained model essentially consists a bag of Gaussians (BoG) of size "N," where "N" is the number of training bags. We assume a bag representation instead of a set representation, because it may be possible that two Gaussian distributions may be extremely *close* (or highly overlapping), if not exactly the same. Once we have a BoG model generated from the training data, we can use this model to predict a class label for any new image patch or window. Therefore, the BoG algorithm can abstracted into following key steps.

Basic primitives of BoG algorithm are as follows:

- Divide the image into regular grids (or patches)
- A fast training acquisition system
- Construction of BoG model from the training data
- Match query bag with the bag of Gaussians
- Apply nearest neighbor based classifier to assign a class label to the query bag

Our objective is to design these steps such that accuracy is maximized while the computational complexity is minimized.

3.1 Image Grids or Patches

In the first step, we divide the image into regular grids, or blocks, or patches. A grid is essentially a square or rectangular block whose size (pixels x lines) determines the quality and computational cost of the algorithm. If the grid is too large, it may result in poor representation. For example, larger grids may contain more than a single object, therefore the Gaussian distribution fitted to the grid may not have a single peak. However, the computation time drastically reduces with the increase in grid size. On the contrary, if the grid size is too small it would increase the computational cost, and may also lead to errors in model parameter estimation and matrix inversion problems. The optimal size is typically dictated by the pixel resolution, typical object sizes found in the imagery, and the number of image bands (i.e., dimensions). Figure 6 shows the grids superimposed on a high-resolution satellite image. In the remaining sections, we refer to the term grid when referencing this first step.

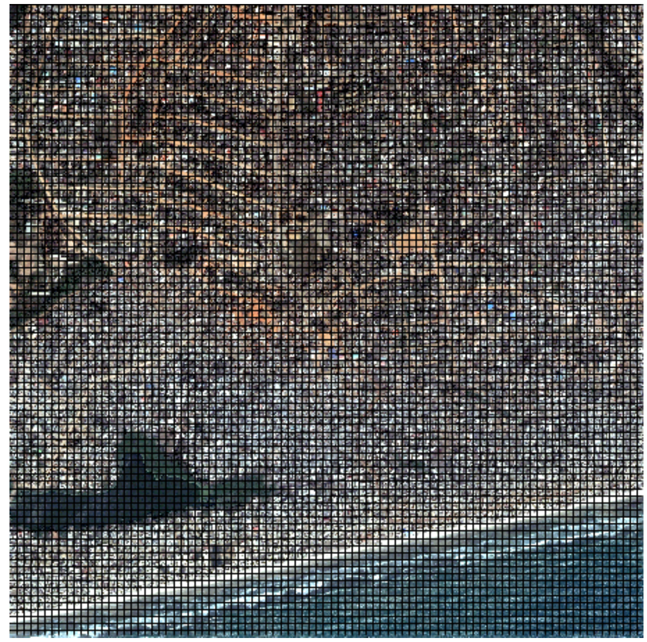

Figure 6: User defined grids superimposed on a high-resolution satellite image

3.2 Fast Training Acquisition

One of the bottlenecks in image classification is the acquisition of the training data. Often an analyst has to accurately digitize the object boundaries and label them. By the very design of our algorithm, analyst do not have to digitize at all. An analyst simply displays the image with grids overlaid and picks up a few representative grids by just clicking on the grids for each class (or thematic category). The training acquisition system is integrated with the popular open source QGIS [1]. Resulting training data is shown in Figure 7. Each colored grid represents a class label given by the analyst.

3.3 Constructing BoG model

We model the image data in each grid, that is, all multi-dimensional feature vectors from each pixel in the grid, are generated by a multi-variate Gaussian distribution described in the following equation.

$$p(x|y_j) = \frac{1}{\sqrt{(2\pi)^{-N}|\Sigma_j|}} e^{\frac{-1}{2}(x-\mu_j)^t|\Sigma_j|^{-1}(x-\mu_j)} \quad (1)$$

The standard multi-variate Gaussian distribution is described by the parameters mean (μ) and covariance matrix (Σ). These parameters are estimated for each grid separately from the corresponding image data, as shown in Figure 8.

3.4 Prediction

Once BoG model is constructed, then class labels can be predicted based on how similar the new image windows (or patches) are to the examples in the trained model. For each new window, the algorithm works as follows.

1. Compute the probabilistic distance between a given window (Gaussian distribution, P_i) and each Gaussian

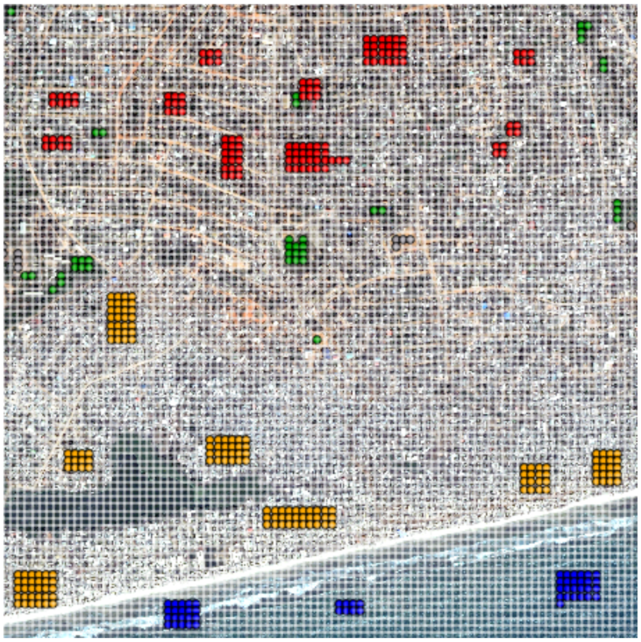

Figure 7: Ground-truth collection system (each color represents a unique class)

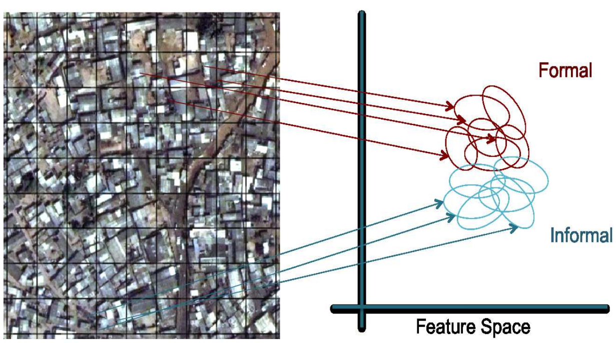

Figure 8: BoG Model Constructed From The Training Data

(Q_j) in the BoG model. There are various divergence and distance measures readily available from the literature. The most notable ones are the Kullback-Leibler (KL) divergence, Bhattacharyya distance, and Mahalanobis distance. We implemented all three measures, but experimentally found KL divergence to be slightly better than the other two. The KL divergence is a non-symmetric measure of the difference between two probability distributions P and Q, given by:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx \qquad (2)$$

For Gaussian distributions, the KL divergence is given by:

$$\begin{aligned} D_{KL}(P||Q) = \frac{1}{2}[log\frac{|\sigma_Q|}{|\Sigma_P|} + Tr(\Sigma_Q^{-1}\Sigma_P) \\ +(\mu_P - \mu_Q)^T\Sigma_Q^{-1}(\mu_P - \mu_Q)] \end{aligned} \qquad (3)$$

Unfortunately, KL divergence is not a distance metric and is not scaled between 0 and 1 (whereas Bhattacharyya and Mahalanobis are scaled). However, the symmetric version of KL divergence is easy to compute:

$$D_{KL}(P||Q) = \frac{1}{2}(D_{KL}(P||Q) + D_{KL}(Q||P) \qquad (4)$$

2. Rank the distances (similarity score)
3. Assign label to the window (or patch) based on the winner from top K nearest neighbors.

This prediction process is schematically represented in Figure 9. As shown in the figure, after ranking the computed similarity between a given image (or query) patch and all the training patches, "formal" class label is assigned to the query window as it got the maximum votes (3) in comparison to the "informal" class. Though the example is shown for two class problem, the actual algorithm is designed for multiple classes.

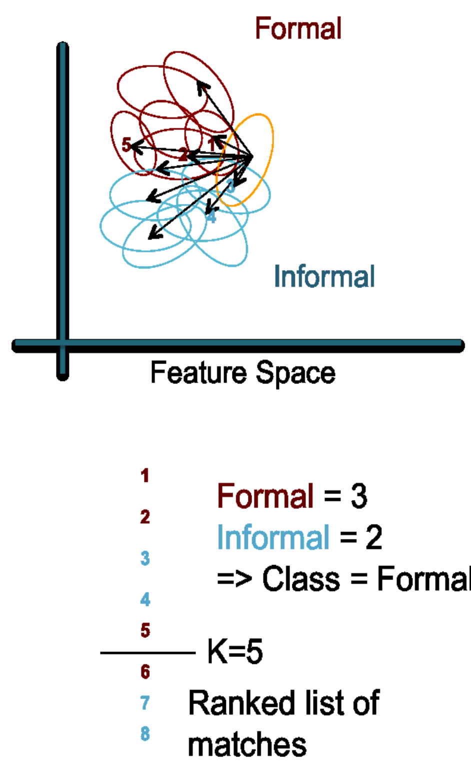

Figure 9: Prediction based on similarity (probabilistic distance) measure

This general framework can be adopted for Citation-KNN by simply replacing the modeling and prediction as following. In addition, single instance learning algorithms can also be applied by transforming the patch into a single feature vector (e.g., centroid, average).

3.5 Citation-KNN

The training dataset is collected as described previously (section 3.2). Citation-KNN is a simple extension of regular kNN algorithm. In its simplest form (k=1), the kNN algorithm assigns a given pixel (feature vector) to the same class label as the closest data point. Though computation of distance (Euclidian or probabilistic) between feature vectors is straight forward, computing distance between bags is not as straight forward. Citation-KNN algorithm uses Hausdorff distance as the distance between two bags. Let us assume that A and B are two given bags, and a_i, b_j are instances from the corresponding bags, then the distance between A and B is found by minimizing the following equation.

$$\begin{aligned} \mathrm{Dist(A,B)} &= \mathrm{Min}_{\substack{1\leq i\leq n \\ 1\leq j\leq n}} (Dist(a,b_j)) \\ &= \mathrm{Min}_{a\epsilon A} Min_{b\epsilon B} \|a-b\| \end{aligned} \quad (5)$$

This minimal Hausdorff distance allows regular kNN algorithm to be applied to the multi-instance learning. Once all the minimal Hausdorff distances have been computed between query bag and training bags, a simple majority voting is used to choose the label for the query bag. In addition, Citation-KNN uses the notion of reference, that is, a query bag is assigned not only based on its neighbor relationships but also by taking into account the bags that regard the query bag as its neighbor. This citation approach is shown to be more robust to the noise in the training data.

4. PIXEL (SINGLE INSTANCE) BASED CLASSIFICATION SCHEMES

We now briefly describe four major classification schemes used to evaluate the performance against the Citation-KNN and the GMIL classification schemes.

4.1 Logistic Regression

Given an $n-$vector $\mathbf{y}$ of observations and an $n \times m$ matrix $\underline{X}$ of explanatory data, classical linear regression models the relationship between y and $\underline{X}$ as $\mathbf{y} = \mathbf{X}\beta + \epsilon$. One simple way to extend regression for classification is to perform regression for each class. While this simple extension works for classification, it violates the basic assumptions of regression. That is, errors are statistically independent and normally distributed because the observations (y) take only 0 and 1 values. Logistic regression [9,13] does not suffer from these limitations. In the case of Logistic regression, the target variable (y) is transformed via the Logistic function and the dependent variable is interpreted as the probability of finding a given class.

4.2 Tree Based Classification (Random Forests)

Random forests is an ensemble method used to construct a series of decision trees. Each tree is constructed on a different training dataset of the same size generated by random sampling, with replacement from the original training dataset. Random forests retain many benefits of decision trees and avoid pruning. Random forests have also shown to generalize well, and accuracies are typically higher than a single tree. Uses of random forests for image classification can be found in [11].

4.3 Neural Networks

Artificial neural networks, which are non-parametric classifiers as opposed to Bayesian classifiers, are gaining popularity in remote sensing image classification. This popularity can be attributed to several factors: 1) previous studies [2,16] have shown that their performance is as good as MLC and in many cases more accurate, 2) they are non-parametric, so they are capable of classifying multi-source data, and 3) they have several desirable characteristics like nonlinearity, adaptability, and fault tolerance.

The use of neural networks in remote sensing data analysis has been somewhat limited until recent years because of the complexities associated with establishing suitable parameters for network training, the lack of knowledge about the internal workings of networks (especially how they divide the feature space), and lack of comparative studies. The previous "black box" view of neural networks – which limited its use – is now clear with the insights provided by recent studies [12], [19], [17]. Several recent studies [5], [2], [16] were also focused on comparing statistical and neural network classification of remote sensing data. We used the Multi-layer perceptrons (MLP) architecture in this experiment.

4.3.1 Naive Bayes Classification

One of the simplest Bayes classifiers is naive the Bayes (NB) classifier. In the general Bayesian setting, it is assumed that the samples in feature space are correlated, meaning greater emphasis is placed on an accurate estimation of the covariance matrix (Σ). One challenge in estimating the full covariance matrix is that one needs a large number of training samples. This assumption is relaxed in a NB classifier, where it is assumed that features are independent. That is, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is not dependent on the presence (or absence) of other features. As a result, the full covariance matrix does not have to be estimated. Instead, estimation of variance is sufficient to construct a NB classifier. With independent feature assumption, the Gaussian distribution simplifies to:

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (6)$$

5. EXPERIMENTS

Four cities were chosen for this study to thoroughly evaluate the performance of the proposed algorithm. The cities chosen are as follows: Accra (1), Caracas (2), La Paz (3), and Kandahar (4). The population estimate in 2010 for Caracas and La Paz was 3.098 million and 1.69 million, respectively. As of 2006 estimate Kandahar has population of 468,200. The imagery used for this study is came from the DigitalGlobe CityShere database. Spatial resolution is 0.6m and each image has 3 spectral bands.

We chose these four cities as they represent diversity in terms of different climates, cultures, and economies. Caracas, Kandahar, and La Paz reside in a tropical, dry, semi-arid, and sub-tropical highland climate, respectively. Caracas, which has one of the largest "mega-slums" on the planet, has an estimated 44% of its population living in informal settlements.

	F	I	V	B	W	Recall
F	105	7	0	0	0	93.75
I	2	74	0	2	0	94.87
V	0	0	7	0	0	100.00
B	0	0	0	22	0	100.00
W	0	0	0	0	35	100.00
Precision	98	91	100	92	100	**95.67**

Table 1: GMIL test accuracy (contingency table) for Accra city

City	C-KNN	Log.Reg.	RF	MLP	NB	**GMIL**
1	76.25	71.25	72.08	69.58	75.66	95.66
2	82.96	78.15	81.85	81.81	74.07	85.00
3	80.97	77.17	78.26	80.23	76.08	83.25
4	79.78	64.89	69.14	73.93	60.10	81.20

Table 2: Overall classification accuracy for each study site

For each study site, we have chosen roughly 4% of grids for training data. Couple of domain experts who have close knowledge about these cities then labeled these grids. The classes considered for these cities are: formal (F), informal (I), vegetation (V), bare soil (B), and water (W). This data is then divided into independent training (40%) and test (60%) datasets. Trained model is then used to predict labels for entire study site. The performance is evaluated by constructing contingency table. Table 1 shows the contingency table for GMIL classifier for Accra. Reporting full contingency table for each city and each classifier would be overwhelming, therefore for simplicity we are just reporting the overall accuracies. The overall classification results were summarized in the Table 2.

As can be seen from the table, GMIL performs better than all other classification schemes. It is interesting to note that Citation-KNN model is also performed better than single instance approaches. Figure 10 shows the GMIL classification output (overlaid on raw image) of the Accra city. As single instance learning algorithms can't be directly applied on the bags, we did post-processing by converting grids into a single class by applying majority vote (*Modal filtering*).

5.1 GMIL vs Citation-KNN

Though the performance of Citaiton-KNN is close to GMIL except in Accra, the computational complexity makes it infeasible for image classification. Let n be the average number of instances per bag, and N be the number of training bags, and d be the number of dimensions, then the computational complexity of Citation-KNN is $O(n^2Nd)$. As a result this algorithm can not be applied to large images. Computational complexity is greatly reduced for GMIL as we are computing the distance between two distributions rather than the n^2 pair-wise distances between instances across the bags. To measure the actual difference in cpu-time, we ran the experiment on a very high-resolution (1 m^2) multispectral (3 bands) satellite imagery. The image cropped to $1km^2$ consisting of 1,000 x 1,000 pixels. This image is then divided into 10,000 blocks where each block consists of 10 x 10 pixels. Of these 10,000 blocks, we have selected 380 blocks for training. This data is divided into independent training

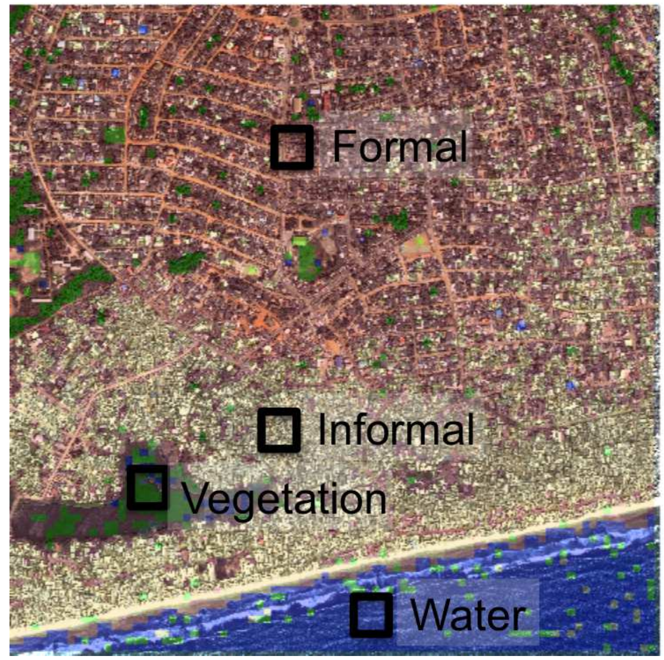

Figure 10: GMIL Classified Image Overlaid on Raw Image

(130 blocks) and test (250 blocks) datasets. We ran both GMIL and Citaion-KNN algorithms on a single node with two dual Xeon hex-core processors (3.46 GHz) with 48 GB of 1333 MHz ECC DDR3 memory. Even for this small image, Citation-KNN took about 27.8 hours, where as GMIL took 3.1 hours. GMIL is 9 times faster than Citation-kNN in addition to being more accurate on all study sites. Though GMIL is faster, it requires parallel implementation to make it operational.

6. CONCLUSIONS

In this study, we presented a computationally efficient multiple instance learning algorithm. Experimental studies showed that algorithm is capable of finding complex patterns in the high resolution satellite images. Comparative analysis showed our algorithm performed better than many standard per-pixels classification algorithms as well as the Citation-KNN algorithm. In addition, our algorithm is computationally efficient. We are using this algorithm for mapping the slums (informal settlements) across the globe. Analyzing very high resolution remote sensing images for complex pattern extraction is an emerging research topic. In addition, automated mapping of informal settlements fills a critical need of national governments and planning agencies. We are working on further improving the algorithm as we are classifying more and more cities. We are also working on parallel implementation of this framework on GPUs. As mentioned earlier, our ultimate goal is to build a computationally efficient and automated framework that is capable of detecting new settlements (especially slums) across the globe in a continuous manner. As our work matures, we hope that this framework becomes an important tool in complex settlement mapping using very high-resolution satellite imagery for a broader community.

7. ACKNOWLEDGMENTS

I would like to thank my collaborators B. Bhaduri, J. Graesser, A. Cheriyadat, E. Bright, V. Chandola, and C. Symons for their invaluable feedback into this research.

This manuscript has been authored by employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

8. REFERENCES

[1] Quantum GIS: open source geographic information system (gis). http://www.qgis.org/.

[2] A. Benediktsson, P. Swain, and O. Ersoy. Neural network approaches versus statistical methods in classificaion of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):550, 1990.

[3] T. Blaschke, S. Lang, and G. Hay. *Object-Based Image Analysis*. Springer, 2008.

[4] J. Bolton and P. Gader. Application of multiple-instance learning for hyperspectral image analysis. *Geoscience and Remote Sensing Letters, IEEE*, 8(5):889 –893, sept. 2011.

[5] L. Bruzzone, C. Consese, F. Masellit, and F. Roli. Multisource classification of complex rural areas by statistical and neural-network approaches. *Photogrammetric Engineering & Remote Sensing*, 63(5):523–533, May 1997.

[6] V. Chandola and R. R. Vatsavai. Scalable time series change detection for biomass monitoring using gaussian process. In *CIDU*, pages 69–82, 2010.

[7] V. Chandola and R. R. Vatsavai. A scalable gaussian process analysis algorithm for biomass monitoring. *Statistical Analysis and Data Mining*, 4(4):430–445, 2011.

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

[9] S. Dreiseitl and L. Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35(5-6):352 – 359, 2002.

[10] J. Graesser, A. Cheriyadat, R. Vatsavai, V. Chandola, J. Long, and E. Bright. 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, aug. 2012.

[11] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh. Investigation of the random forest framework for classification of hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(3):492 – 501, Mar. 2005.

[12] I. Kanellopoulos and G. G. Wilkinson. Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18(4):711–725, 1997.

[13] B. Krishnapuram, L. Carin, M. A. Figueiredo, and A. J. Hartemink. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):957 –968, June 2005.

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

[15] S. Nussbaum and G. Menz. *Object-Based Image Analysis and Treaty Verification*. Springer, 2008.

[16] J. Paola and R. Schowengerdt. A detailed comparison of backpropagation neural network and maximum likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4):981–996, 1995.

[17] J. Paola and R. Schowengerdt. The effect of neural-network structure on a multispectral land-use/land-cover classification. *Photogrammetric Engineering & Remote Sensing*, 63(5):535–544, 1997.

[18] S. Shekhar, P. Schrater, R. Vatsavai, W. Wu, and S. Chawla. Spatial contextual classification and prediction models for mining geospatial data. *IEEE Transaction on Multimedia*, 4(2):174–188, 2002.

[19] A. Skidmore, B. Turner, W. Brinkhof, and E. Knowles. Performance of a neural network: Mapping forest using gis and remorely sensed data. *Photogrammetric Engineering & Remote Sensing*, 63(5):501–514, May 1997.

[20] P. Torrione, C. Ratto, and L. Collins. Multiple instance and context dependent learning in hyperspectral data. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop on*, pages 1 –4, aug. 2009.

[21] R. R. Vatsavai. High-resolution urban image classification using extended features. In *ICDM Workshops*, pages 869–876, 2011.

[22] R. R. Vatsavai. A data mining framework for monitoring nuclear facilities. In *ICDM Workshops (Industry/Government Track)*, page 917, 2012.

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

[24] R. R. Vatsavai, B. L. Bhaduri, A. Cheriyadat, L. F. Arrowood, E. A. Bright, S. S. Gleason, C. Diegert, A. K. Katsaggelos, T. Pappas, R. Porter, J. Bollinger, B. Chen, and R. Hohimer. Geospatial image mining for nuclear proliferation detection: Challenges and new opportunities. In *IGARSS*, pages 48–51, 2010.

[25] R. R. Vatsavai, E. A. Bright, V. Chandola, B. L. Bhaduri, A. Cheriyadat, and J. Graesser. Machine learning approaches for high-resolution urban land cover classification: a comparative study. In *COM.Geo*, page 11, 2011.

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