Complex Settlement Pattern Extraction With Multi-instance Learning

Ranga Raju Vatsavai, Budhendra Bhaduri
Computational Sciences and Engineering Division
Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
Email: {vatsavairr,bhaduribl}@ornl.gov

Jordan Graesser
Department of Geography
McGill University, Montreal, Quebec, Canada H3A 0B9
Email: jordan.graesser@mail.mcgill.ca

Abstract— **Per-pixel (or single instance) based classification schemes which have proven to be very useful in thematic classification have shown to be inadequate when it comes to analyzing very high resolution remote sensing imagery. The main problem being that the pixel size (less than a meter) is too small as compared to the typical object size (100s of meters) and contains too little contextual information to accurately distinguish complex settlement types. One way to alleviate this problem is to consider a bigger window or patch/segment consisting a group of adjacent pixels which offers better spatial context than a single pixel. Unfortunately, this makes per-pixel based classification schemes ineffective. In this work, we look at a new class of machine learning approaches, called multi-instance learning, where instead of assigning class labels to individual instances (pixels), a label is assigned to the bag (all pixels in a window or segment). We applied this multi-instance learning approach for identifying two important urban patterns, namely formal and informal settlements. Experimental evaluation shows the better performance of multi-instance learning over several well-known single-instance classification schemes.**

I. Introduction

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.

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 were designed for coarse 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 simple objects such as buildings and roads. What we are pointing at is that the single-instance learning algorithms are inadequate in modeling complex patterns which requires simultaneously looking at multiple instances which gives better spatial context for the classifier. The same limitations are also applicable to spatial contextual classifiers (e.g, *Markov Random Fields*), as these classifiers look at the 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. To overcome the limitations posed by single-instance classifiers in modeling complex patterns, we adopted a novel multi-instance based machine learning approach, which showed improvements in accurately identifying informal settlements. We have conducted several experiments on high-resolutions satellite imagery, representing four unique geographic regions across the world.

II. Related work and our contributions

Most of the existing classification approaches work with spectral features extracted from each pixel (spatial location). This is very ineffective 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 1, illustrates this problem clearly. As shown in this figure, single pixels drawn from bare soil and rooftop are very similar (spectrally). On the other hand, two small windows centered around the same pixel locations, shows better discrimination power as these windows (grids) contains much richer spatial (e.g., building edges) information.

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

978-1-4799-0211-8/13/$31.00 ©2013 IEEE

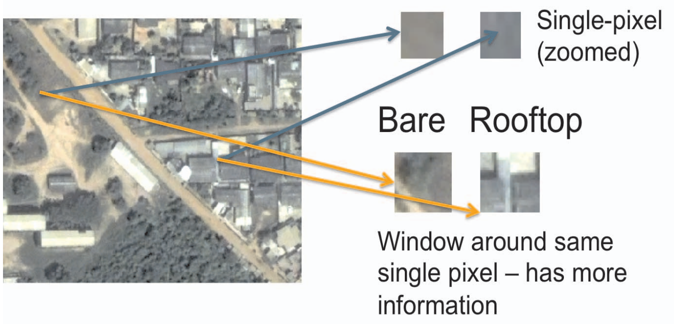

Fig. 1. Figure shows problem with single instance learners

studies [6], [15], [16] show the comparative performance of a broad set of widely used classification methods on extended features which included textures, edge density, and morphological features. 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.

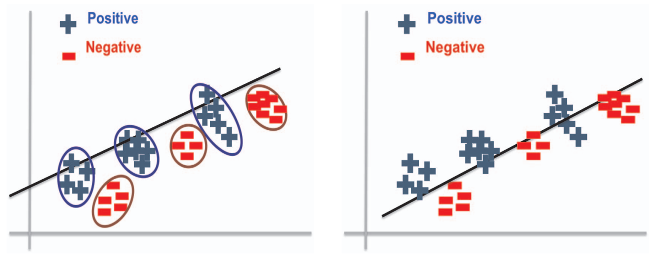

Fig. 2. Figure shows decision boundaries resulting from: (a) multi-instance learning, and (b) single instance learning

Multi-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. [4], Diverse Density [10], and Citation-KNN [17]. Recently, MIL algorithms have also been applied to remote sensing image classification. For example, in [14] MIL approach is explored for sub-surface landmine detection using hyperspectral (HS) imagery. In [2], 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 2(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 2(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).

III. Multi-Instance Learning

We now describe a multi-instance learning algorithm based on Citation kNN [17] which is adopted for classification of complex settlement patterns in very high-resolution remote sensing imagery. This algorithm can abstracted into following key steps.

- Divide the image into regular grids (or patches)
- A fast training acquisition system
- Assign query bag (window) to one of the training bag labels using Citation kNN
- Accuracy assessment

A. Image Grids or Patches

In the first step, we divide the image into regular grids, or blocks, or patches, or bags. 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 contain many irrelevant objects resulting in poor representation of the pattern. The optimal size is typically dictated by the pixel resolution and typical object sizes found in the imagery. Figure 3(a) shows the grids superimposed on a high-resolution satellite image. In the remaining sections, we refer to the term grid (or bag) when referencing this first step.

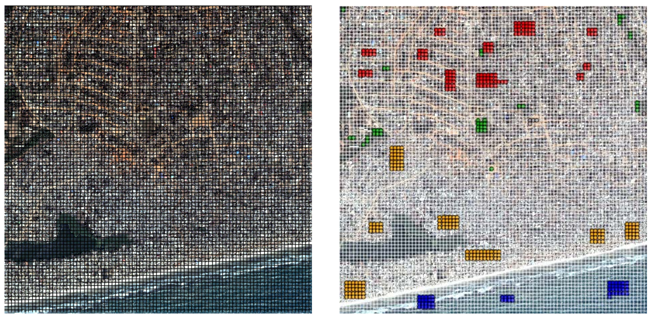

Fig. 3. Patch-based Traning: (a) User Defined Grids Superimposed on a High-resolution Satellite Image, (b) Training sites selected are highlighted using different colors

B. 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 complex pattern). Resulting training data is shown in Figure 3(b). Each colored grid represents a class label given by the analyst.

C. Citation-kNN

Given the training dataset collected as described in previous section, our objective is to predict a class label for a given new window (or bag). We call this new window as query bag. Citation-kNN is one such multi-instance classification system which is an 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 Hausdroff 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.

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

This minimal Hausdroff distance allows regular kNN algorithm to be applied to the multi-instance learning. Once all the minimal Hausdroff distances have been computed between query bag and training bags, a simple majority voting is used to chose 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.

IV. PIXEL (SINGLE INSTANCE) BASED CLASSIFICATION SCHEMES

We now briefly describe four major classification schemes used to evaluate the classification performance against the Citation kNN Model.

A. 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 $\mathbf{y} = \mathbf{X}\beta + \epsilon$. Here $\mathbf{X} = [1, \underline{X}]$ and $\beta = (\beta_0, \ldots, \beta_m)^t$. The standard assumption on the error vector ϵ is that each component is generated from an independent, identical, zero-mean and normal distribution, i.e., $\epsilon_i = N(0, \sigma^2)$. Though regression is widely used in the context of prediction (that is, y is a continuous attribute), linear regression can also be used for classification when the dependent variable is binary. One simple way to extend regression for classification is to perform regression for each class. The output is binary, where the output of regression is one for training instances that belong to a given class and zero otherwise. 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 [5], [9] 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.

B. 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 [7].

C. 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 [1], [11] 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 [8], [13], [12]. Several recent studies [3], [1], [11] 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.

V. EXPERIMENTS

Four cities were chosen for this study to thoroughly evaluate the performance of the proposed algorithm. The cities chosen are as follows: Caracas (001), Venezuela (002), La Paz (003), Bolivia (004), and Kandahar (005). 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 five cities as they represent diversity in terms of different climates, cultures, and economies. Caracas,

City	C-kNN	Regression	RF	MLP	NB
001	76.25	71.25	72.08	69.58	75.66
002	82.96	78.15	81.85	81.81	74.07
003	80.97	77.17	78.26	80.23	76.08
004	79.78	64.89	69.14	73.93	60.10
005	81.69	77.18	80.58	81.14	74.55

TABLE I
OVERALL CLASSIFICATION ACCURACY RESULTS

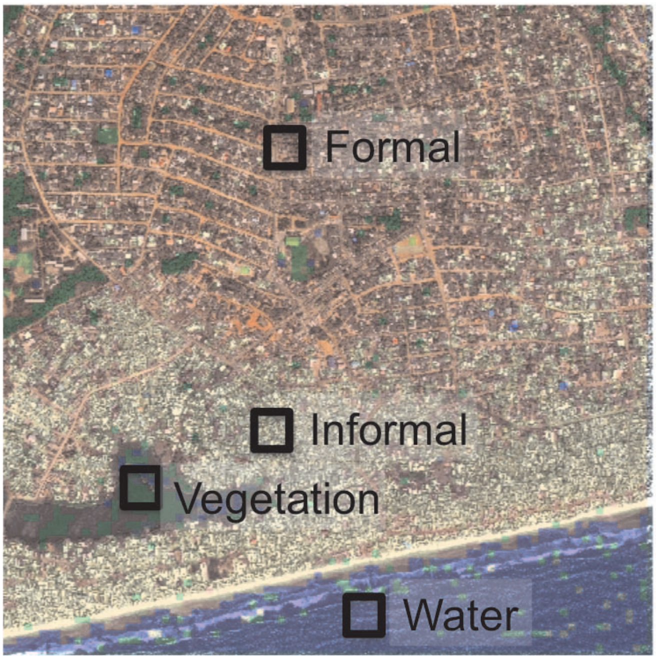

Fig. 4. MIL (Citation-kNN) Classified Image

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. The patterns considered for these cities are: formal, informal, vegetation, bare soil, and water. The overall classification results were summarized in the Table I.

As can be seen from the table, Citation-kNN model performed consistently well as compared to the other approaches. Figure 4 shows classification output (overlaid on FCC image) of Citation-kNN algorithm. 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 a modal filter.

VI. CONCLUSIONS

In this study, we presented a Citaion-kNN based multi-instance learning model for classifying complex urban patters. From experimental studies over 5 cities, we found that the Citation-kNN based multi-instance learning approach is very effective in discriminating complex urban patterns. However, Citation-KNN is computationally expensive. We are working on parallel implementation of this framework on GPUs. We hope that this approach becomes an important tool in complex settlement mapping using very high-resolution satellite imagery.

VII. ACKNOWLEDGMENTS

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.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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, May 1997.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.