2014 IEEE International Conference on Data Mining Workshop

Learning A Spatial Ensemble of Classifiers for Raster Classification: A Summary of Results

Zhe Jiang*, Shashi Shekhar*, Azamat Kamzin*, Joseph Knight†

*Department of Computer Science & Engineering, University of Minnesota, Twin Cities, Minneapolis, USA.
Email: {zhe, shekhar, kamzin}@cs.umn.edu
†Department of Forest Resources, University of Minnesota, Twin Cities, Saint Paul, USA.
Email: jknight@umn.edu

Abstract*—Given a spatial raster framework *F*, a set of explanatory feature maps, training and test samples with class labels on *F*, as well as a base classifier type, the problem of ensemble learning in raster classification aims to learn a collection of base classifiers to minimize classification errors. The problem has important societal applications such as land cover classification but is challenging due to existence of class ambiguity from spatial heterogeneity, i.e., samples with the same feature values may have distinct class labels in different areas. Many existing approaches are non-spatial ensembles (e.g., bagging, boosting, random forest), which assume that learning samples follow an identical distribution. Some spatial ensemble approaches also exist, which simply partition the raster framework into several regular sub-blocks and combine classification results on each sub-block. However, these existing approaches can not address the class ambiguity issue among pixels. In contrast, this paper proposes a new spatial ensemble approach, which partitions the spatial framework into several spatial footprints to minimize class ambiguity of training samples and then learns a base classifier for each footprint. Experimental evaluations on a real world remote sensing dataset show that the proposed spatial ensemble approach outperforms existing approaches when strong class ambiguity exists.

***Keywords*-spatial ensemble, raster classification, class ambiguity, spatial heterogeneity, remote sensing**

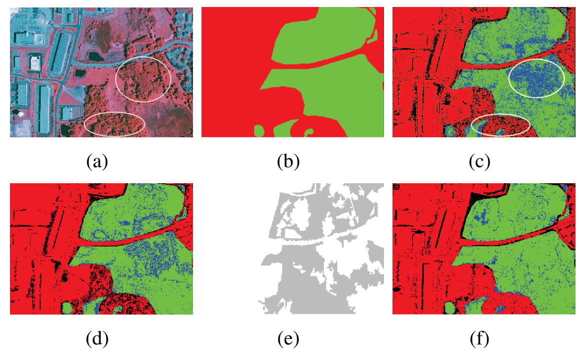


Fig. 1. Real world problem example (best viewed in color): (a) an input feature image; (b) ground truth classes: red color represents dry land, green color represents wetland; (c) predictions of a global decision tree, black and blue are prediction errors; (d) predictions of a random forest; (e) 2 partitions of the spatial framework (shown in white and gray colors); (f) predictions of a spatial ensemble on the two partitions;

I. Introduction

Given a spatial raster framework *F*, a set of explanatory feature maps, training and test samples with class labels on *F*, as well as a base classifier type, the problem of ensemble learning in raster classification aims to learn a collection of base classifiers to minimize prediction errors on test samples. Figure 1 shows a real world example in remote sensing. Comparing feature maps (Figure 1(a)) and a ground truth class map (Figure 1(b)), we can find class ambiguity in the white circles, i.e., their feature characteristics are very close, but their class labels are different. Given this input, the predictions of a single decision tree and a random forest are shown in Figure 1(c) and (d), both of which have lots of errors in the white circles. In contrast, a spatial ensemble of two decision trees is also learned, whose spatial footprints are shown in white and grey colors in Figure 1(e). Its prediction as shown in Figure 1(f) has much less errors.

Applications: The raster classification problem has many important societal applications. For example, it can be used to classify remote sensing images into different land cover types, such as wetland and dry land [1]. Wetland mapping is important for understanding climate change [2], natural resource management [3], and disaster management [4], etc.

Challenges: The raster classification problem is challenging due to the existence of class ambiguity from spatial heterogeneity, i.e., samples with the same feature values may have distinct class labels in different areas. For example, the two forest patches in the white circles of Figure 1(a) have similar feature characteristics but different class labels. Ignoring this effect, as by decision trees or random forests, can result in very poor classification performance (Figure 1(c)(d)).

Related work: As shown in Figure 2, existing ensemble learning techniques in raster classification can generally be grouped into two categories: non-spatial ensembles and spatial ensembles. The non-spatial ensemble approaches (e.g., bagging [5], boosting [6], and random forest [7]) have greatly boosted the performance of base classifiers in many applications. However, these methods assume that learning samples follow an identical distribution for the entire spatial framework, and thus may produce poor prediction accuracy when class ambiguity across subareas exists. There is also a spatial ensemble approach, which simply partitions an image into

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.166

regular sub-blocks and combines classification results of each sub-block [8] [9]. This approach is originally proposed for the scene classification problem, which aims to classify an entire image into one class (e.g., indoor or outdoor). Though combining relatively poorer individual sub-block classifiers can help boost accuracy of the class label for an entire image, it may not boost the accuracy of individual pixels.

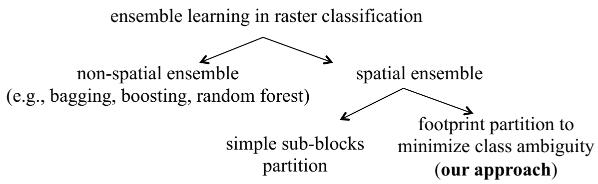


Fig. 2. Categorization of raster classification approaches

To address these limitations, this paper proposes a new spatial ensemble approach for raster classification. Feature images are first preprocessed to get relatively homogeneous and spatially contiguous clusters called islands. A grouping algorithm is designed to group these islands into k sets called footprints, to minimize class ambiguity among training pixels in each footprint. After this, training and prediction of a base classifier can be done in each footprint separately.

Contributions: This paper proposes a spatial ensemble learning approach for the raster classification. We make the following contributions:

- We propose a spatial ensemble approach, which decomposes the spatial raster framework into several footprints to minimize class ambiguity among training samples in each footprint, and learns a base classifier on each footprint.
- We conducted experimental evaluations on a real world remote sensing dataset. Results showed that our proposed approach outperforms the current approaches when class ambiguity exists.

Scope: This paper focuses on spatial ensemble learning to address the issue of class ambiguity. It considers a pixel as the minimum classification unit. Approaches that consider an object as the minimum classification unit (e.g., GEOBIA [10]) is out of the scope of the work.

Outline: The paper is organized as follows: Section II formalizes the raster classification problem; Section III introduces proposed approach. Experimental evaluations are in Section IV. Section V discusses some other relevant techniques. Section VI concludes the paper with future work.

II. Problem Statement

We formally define the problem of ensemble learning in raster classification as follows:

Given:
- a spatial raster framework F
- feature maps on F
- training and test samples on F
- a base classifier type

Find: a collection of base classifiers

Objective:
- minimize classifier error on test samples

Constraint:
- spatial autocorrelation exists on F
- spatial heterogeneity exists on F

III. Proposed Approach

This section discusses our proposed spatial ensemble approach. It consists of two main steps: a preprocessing step and a footprint construction step. After these two main steps, training and prediction of a base classifier can be done within each footprint separately.

A. Preprocessing

The preprocessing step decomposes the raster framework into many small homogeneous and spatially contiguous clusters called *islands*. This can be considered as an image segmentation task [11]. We use self-organizing map to cluster pixel locations based on their feature vectors, and merge feature clusters if Ripley's cross-K function [12] of their pixel locations is high. Then each spatially connected component from the feature cluster map is an island. Mod filter is used to remove small holes from final islands. According to the spatial autocorrelation effect, we assume that each island has no class ambiguity.

B. Footprint construction

Footprint construction step aims to group the homogeneous clusters (islands) into several disjoint sets (footprints) such that the training samples within each footprint have less class ambiguity and good class balance. Before introducing our footprint construction algorithm, we define some key concepts.

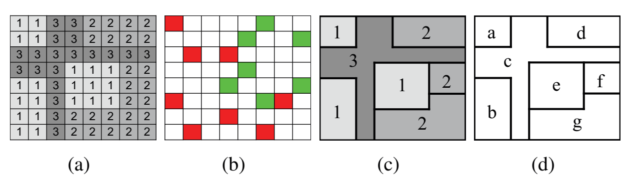


Fig. 3. Examples of key concepts. (a) a feature map; (b) training labels (R: red; G: green); (c) archipelagos, each with a unique id, e.g., 1, 2, 3; (d) islands

An *island* is a small homogeneous and spatially contiguous cluster (Figure 3(d)). A collection of islands with the same feature characteristics is an *archipelago* (Figure 3(c)). Each island may have a *dominant class*, which is the majority of training sample class labels. *Ambiguous islands* are islands from the same archipelago (with the same feature characteristics) but with different dominant classes (e.g., island d and g in Figure 3(d)). A *homogeneous island group* is a group of non-ambiguous islands from the same archipelago (Table I).

TABLE I
AN EXAMPLE OF HOMOGENEOUS ISLAND GROUPS (R: RED, G: GREEN)

island group	archipelago id	dominant class	training samples
$\{a, b\}$	1	R	(R:3, G:0)
$\{e\}$	1	G	(R:0, G:3)
$\{d, f\}$	2	G	(R:0, G:4)
$\{g\}$	2	R	(R:3, G:0)
$\{c\}$	3	R	(R:2, G:0)

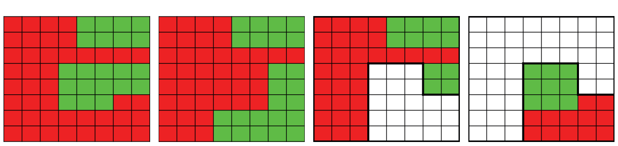
Algorithm 1 Footprint construction

Input:
- IM: an island map
- k: the number of footprints

Output:
- $(P_1, P_2, ..., P_k)$: a set of footprints

generate homogeneous island groups
for each k-partition of island groups $(P_1, P_2, ..., P_k)$ **do**
    for each P_i with $i = 1, 2, ...k$ **do**
        check **no-ambiguous-island constraint** on P_i
        check **class-balance constraint** on P_i
        if any constraint not satisfied **then**
            jump to next iteration of the loop
    return $(P_1, P_2, ..., P_k)$
return "no answer"

The footprint construction algorithm is shown in Algorithm 1, which takes a map of islands and a number k as inputs, and produces a partition of k footprints. The algorithm first generates homogeneous island groups from archipelagos (e.g., Table I from Figure 3(c)(d)). It then enumerates through all possible k disjoint partitions of all island groups. For each partition, it checks if ambiguous islands exist and if the training class labels are balanced (i.e., the entropy of the training sample class labels in the partition should not be lower than a threshold). If a partition fails to satisfy any constraint, the algorithm jump to next enumeration of k disjoint partitions. Otherwise, it will return the current k disjoint partitions as k footprints. Given the example of Table I where $k = 2$ and class balance constraint entropy threshold is 0.5, one feasible solution is $\{a, b, c, d, f\}$, $\{e, g\}$. Based on these two footprints, a spatial ensemble of two decision trees are learned, and their predictions are shown in Figure 4(c)(d), which are better than the prediction of a decision tree learned from the entire image (Figure 4(b)).



(a) (b) (c) (d)

Fig. 4. Prediction of proposed spatial ensemble; (a) ground truth classes; (b) prediction of a decision tree learned from the entire image; (c) and (d) prediction of two decision trees learned from footprints $\{a, b, c, d, f\}$ and $\{e, g\}$ respectively

IV. EXPERIMENTAL EVALUATION

The goal of the experimental evaluation was to investigate if the proposed spatial ensemble approach can achieve better classification accuracy compared with non-spatial ensemble methods. Due to space limit, we did not add existing simple sub-block partitioning approach (discussed in Section I) in the comparison. We will add it in future work.

A. Experiment Setup

Experiment design: The experiment design is shown in Figure 5. To evaluate classification performance, we compared the prediction accuracy (i.e., precision, recall, F-score) of single global classifiers, non-spatial ensembles (e.g., boosting, random forest [7]) and the proposed spatial ensemble framework using a tree classifier family (i.e., decision tree, spatial decision tree [13]).

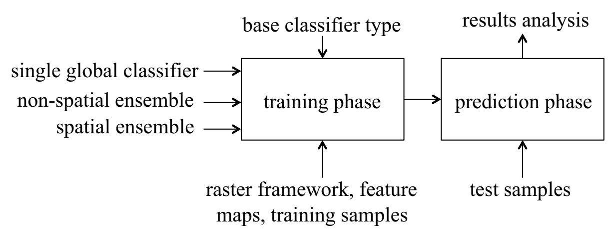


Fig. 5. Experiment Design

Dataset description: We used high resolution (3m*3m) remote sensing imagery collected from the city of Chanhassen, MN, by the National Agricultural Imagery Program and Markhurd Incorporation. There were 9 continuous explanatory features including multi-temporal (for the years 2005 and 2008) spectral information (R, G, B, NIR) and Normalized Difference Vegetation Index (NDVI). Class labels (wet land and dry land) were created by a field crew and photo interpreters between 2004 and 2005. We select a scene from the city with 243 by 339 pixels, among which we drew a training set with 993 in dry land class and 893 in wetland class.

Evaluation metric: We evaluated the classification performance with confusion matrices, the precision and recall, as well as F-score. We calculated the precision, recall, and F-score on the wetland class.

B. Classification Performance of Spatial Ensemble

Configuration of parameters: The base classifier type is set as the tree family (decision tree DT, spatial decision tree SDT). The number of partitions $k = 2$ (i.e., two tree classifiers are learned in spatial ensemble approach). The minimum tree node size is 10. The maximum neighborhood size in spatial decision tree is 5 by 5. The number of trees in random forest is 500. The number of trees in decision tree boosting is 100. Note "SE" represents our spatial ensembles.

Result analysis: Classification results is summarized in Table II. The first category summarizes results of a single classifier without ensemble. The prediction of a decision tree (DT) has an F-score of 0.74, while a spatial decision tree (SDT) increases the F-score from 0.74 to 0.78 due to ultilizing focal spatial contexts. The second category summarizes results of non-spatial ensembles including random forest and boosting

TABLE II
A SUMMARY OF CLASSIFICATION PERFORMANCE OF DIFFERENT CLASSIFICATION TECHNIQUES (DATASET 1)

Categories	Models	Confusion Matrix		Precision	Recall	F score	number of trees
single global model	single DT	37778	7109	0.78	0.71	0.74	1
		10382	25222				
	single SDT	38180	6707	0.80	0.76	0.78	1
		8536	27068				
non-spatial ensemble	DT Random Forest	38256	6631	0.81	0.79	0.80	500
		7436	28168				
	DT boosting	37183	7704	0.79	0.82	0.81	100
		6252	29352				
spatial ensemble (SE)	SE of DT	39266	5621	0.85	0.89	0.87	2
		3914	31690				
	SE of SDT	40023	4864	0.87	0.94	0.90	2
		2276	33328				

of decision trees, whose F-scores (around 0.80) are much higher than a single DT (0.74) and slightly higher than a single SDT (0.78). The results of spatial ensemble on both DT and SDT are in the third category. Spatial ensemble dramatically decreased the number of false positives and false negatives compared with a single DT or SDT and even compared with random forest and DT boosting on the dataset.

V. DISCUSSION

There are several other relevant work on ensemble learnings [14] [15] [16]. The difference between these studies and our spatial ensemble approach is that their partitioning of input space is based on features only, while we focus on constructing spatial footprints of base classifiers to explicitly minimize class ambiguity due to spatial heterogeneity. In scene classification problem, approaches are proposed to classify regular sub-blocks separately, and combine these classification results by majority voting, a neural network, or the "mixture of experts" method [8] [9]. Since scene classification does not require accurately classifying every pixel of the image, combining relatively poor individual sub-block classifiers can still boost the accuracy of classes of entire images. But the same is not true for our problem, which aims to accurately classify all pixels in one image.

VI. CONCLUSION AND FUTURE WORK

This paper investigates the problem of ensemble learning in raster classification. The problem is important in many applications such as land cover classification in remote sensing and lesion classification in medical image processing. However, the problem is challenging due to the effect of class ambiguity from spatial heterogeneity. To address this challenge, we proposed a novel spatial ensemble framework, which partitions a raster framework into different spatial footprints to minimize class ambiguity of training samples, and learns a base classifier for each footprint. Experimental evaluations on a real world dataset show that our approach outperforms existing approaches when class ambiguity exists.

In future work, we plan to evaluate the proposed approach on more datasets from different study areas. We also plan to analyze the theoretical properties of the spatial ensemble framework, e.g., quantifying class ambiguity degree and its relationship with classification accuracy. We may also investigate more efficient footprint construction algorithms.

REFERENCES

[1] Jennifer M Corcoran, Joseph F Knight, and Alisa L Gallant. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern minnesota. *Remote Sensing*, 5(7):3212–3238, 2013.

[2] Bryan Walsh. How wetlands worsen climate change. http://www.time.com/time/health/article/0,8599,1953751,00.html, January 2010.

[3] A. Deschamps, D. Greenlee, TJ Pultz, and R. Saper. Geospatial data integration for applications in flood prediction and management in the red river basin. In *International Geoscience and Remote Sensing Symposium, Toronto, Canada. Symposium, Geomatics in the Era of RADARSAT (GER'97), Ottawa, Canada*, 2002.

[4] CJ Van Westen. Remote sensing for natural disaster management. *International Archives of Photogrammetry and Remote Sensing*, 33(B7/4; PART 7):1609–1617, 2000.

[5] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[6] Yoav Freund, Robert Schapire, and N Abe. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612, 1999.

[7] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[8] Anna Bosch, Xavier Muñoz, and Robert Martí. A review: Which is the best way to organize and classify images by content. *Image and vision computing*, 25(6):778–791, 2007.

[9] Martin Szummer and Rosalind W Picard. Indoor-outdoor image classification. In *Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop on*, pages 42–51. IEEE, 1998.

[10] GJ Hay and G Castilla. Geographic object-based image analysis (geobia): A new name for a new discipline. In *Object-based image analysis*, pages 75–89. Springer, 2008.

[11] Robert M Haralick and Linda G Shapiro. Image segmentation techniques. In *1985 Technical Symposium East*, pages 2–9. International Society for Optics and Photonics, 1985.

[12] Philip M Dixon. Ripley's k function. *Encyclopedia of environmetrics*, 2002.

[13] Zhe Jiang, Shashi Shekhar, Xun Zhou, Joseph Knight, and Jennifer Corcoran. Focal-test-based spatial decision tree learning: a summary of result. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013.

[14] Anuj Karpatne, Ankush Khandelwal, Shyam Boriah, Snigdhansu Chatterjee, and Vipin Kumar. Predictive learning in the presence of heterogeneity and limited training data. In *SDM*. SIAM, 2014.

[15] Nitesh V Chawla, Thomas E Moore, Lawrence O Hall, Kevin W Bowyer, W Philip Kegelmeyer, and Clayton Springer. Distributed learning with bagging-like performance. *Pattern recognition letters*, 24(1):455–471, 2003.

[16] Robert E Banfield, Lawrence O Hall, Kevin W Bowyer, and W Philip Kegelmeyer. Ensembles of classifiers from spatially disjoint data. In *Multiple Classifier Systems*, pages 196–205. Springer, 2005.