This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Context-Based Filtering of Noisy Labels for Automatic Basemap Updating From UAV Data

Caroline M. Gevaert, Claudio Persello, *Senior Member, IEEE*, Sander Oude Elberink, George Vosselman, and Richard Sliuzas

***Abstract*—Unmanned aerial vehicles (UAVs) have the potential to obtain high-resolution aerial imagery at frequent intervals, making them a valuable tool for urban planners who require up-to-date basemaps. Supervised classification methods can be exploited to translate the UAV data into such basemaps. However, these methods require labeled training samples, the collection of which may be complex and time consuming. Existing spatial datasets can be exploited to provide the training labels, but these often contain errors due to differences in the date or resolution of the dataset from which these outdated labels were obtained. In this paper, we propose an approach for updating basemaps using global and local contextual cues to automatically remove unreliable samples from the training set, and thereby, improve the classification accuracy. Using UAV datasets over Kigali, Rwanda, and Dar es Salaam, Tanzania, we demonstrate how the amount of mislabeled training samples can be reduced by 44.1% and 35.5%, respectively, leading to a classification accuracy of 92.1% in Kigali and 91.3% in Dar es Salaam. To achieve the same accuracy in Dar es Salaam, between 50000 and 60000 manually labeled image segments would be needed. This demonstrates that the proposed approach of using outdated spatial data to provide labels and iteratively removing unreliable samples is aviable method for obtaining high classification accuracies whilereducing the costly step of acquiring labeled training samples.**

***Index Terms*—Basemap updating, image classification, informal settlements, label noise, random forests, unmanned aerial vehicles (UAVs), urban planning.**

I. Introduction

THE utilization of geospatial information to support urban planning is becoming common practice worldwide. A fundamental building block is the basemap (or topographic map), which provides information regarding the location of elemental objects of the urban fabric. As a foundation for many urban planning activities, it is imperative that this basemap provides accurate and up-to-date information. This is not always the case, as changes in an urban setting may occur more rapidly than the updating of such basemaps, which traditionally occurs through the manual digitization of satellite or airborne imagery. This is particularly relevant for informal settlements, which tend to be very dynamic urban environments.

Recent technological developments regarding data acquisition platforms such as unmanned aerial vehicles (UAVs), also known as remotely piloted aerial systems, display potential for quickly delivering high-quality spatial data for geomatics applications [1]. UAVs are capable of bringing imagery with a spatial resolution of mere centimeters and accurate 3-D information to urban planners at a low cost. However, the area that can be covered by a single flight is currently limited due to the technical characteristics of the type of UAVs commonly used for mapping activities [1] and national legislation often limits the maximum area that can be covered by a single flight [2]. UAVs are, therefore, especially suited for mapping tasks that require multiple acquisitions over a limited study area, such as incremental map updating.

In order to exploit the information contained in remotely sensed imagery, the images are usually translated into vector-based semantic information such as the basemaps mentioned previously. In many situations, basemap updating through UAV imagery is performed through manual digitization of new features [3], possibly even involving stakeholders through a participatory mapping approach [4]. An alternative strategy to digitization is the extraction of semantic information through supervised image classification methods.

Supervised classification methods make use of representative training samples to characterize the common characteristics and variability of objects pertaining to each semantic class. Based on the observed distributions of the samples in a defined feature space, a classification model is constructed. This model allows labels to be assigned to new, unlabeled samples. Supervised classification algorithms have been successfully applied in a number of studies to extract semantic information from UAV imagery of urban scenes. Some studies divide the orthoimagery into a grid of coarser resolution, and use a pretrained library to propose which urban objects may be present within each grid cell [5]. Other studies include 3-D features from the point cloud or digital surface model (DSM) to support the image-based features, e.g., [6] and [7]. The application of morphological filters of adaptive sizes on the DSM result in useful features for identifying urban objects with differing scales and can complement the information contained in the imagery [8].

One of the main difficulties in classifying subdecimeter resolution imagery obtained through UAV platforms is the spectral and spatial variability of urban objects (e.g., [5] and [8]). It is

Manuscript received June 8, 2017; revised September 1, 2017; accepted October 8, 2017. *(Corresponding author: Caroline M. Gevaert.)*

The authors are with the Faculty of Geo-Information Science and Earth Observation, University of Twente, 7500 AE Enschede, The Netherlands (e-mail: c.m.gevaert@utwente.nl; c.persello@utwente.nl; s.j.oudeelberink@utwente.nl; george.vosselman@utwente.nl; r.sliuzas@utwente.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2017.2762905

1939-1404 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

possible to address the spectral variability by clustering all the pixels in an unsupervised manner and using a majority voting of the reference labels per cluster to label the pixels [9]. Unfortunately, the collection of these reference labels is expensive, time consuming, and requires a relatively high level of knowledge to ensure that they are representative of the class distributions.

Rather than manually labeling training samples, it is also feasible to use existing spatial datasets to provide the labels. For example, vector data from existing basemaps or sources such as OpenStreetMap could be used [10], [11]. However, there are likely to be changes in the scene if there is a time lapse between the collection of the vector data at t_0 and the newly acquired UAV imagery at t_1. Furthermore, the existing vector information may have been digitized over imagery of a lower spatial resolution, causing misalignments when superimposed over the UAV imagery. Therefore, if existing spatial data are utilized to provide the training samples, it must be taken into account that a number of the training labels are likely to be incorrect.

Various strategies have been developed to deal with such errors in the training sample labels, i.e., *label noise*. A recent overview of the effect of label noise on classification algorithms [12] observed that there are following three main strategies to address this issue:

1) utilizing noise-robust classification algorithms;
2) data cleansing to remove potentially noisy labels from the training data;
3) explicitly modeling label noise.

Other strategies have been developed to specifically combat label noise for remote sensing applications. For example, by modeling label noise by combining noise robust logistic regression and conditional tandom fields (CRFs) for updating geospatial databases [13]; or by using the contextual information in a semisupervised setting in order to assess the reliability of training samples and obtain a classification algorithm that is more robust to mislabeled training samples [14].

In this paper, we utilize a data cleansing strategy that exploits context to identify samples, which are likely to have an incorrect label. Research from the fields of computer vision [15] and the human visual system [16] indicate that both global and local contextual cues are important for object recognition. Global context has to do with the statistics of the image as a whole. The underlying idea is that similar but disjoint objects in a single study area will have similar variations. So, by using existing labels over the entire scene, the classifier uses global statistics of the study area to identify the common variations of these objects in a feature space. Objects that are mislabeled are likely to fall outside of these common variations, causing the classifier to be uncertain about the output label. Local context has to do with the similarity of neighboring samples. Generally speaking, neighboring pixels or segments that have similar characteristics could be expected to belong to the same semantic class [17]. This forms the basis of the contrast-sensitive Potts model, which is commonly used in image analysis techniques such as CRFs to ensure a smooth labeling (e.g., [18]).

Object- or segment-based labeling, as opposed to pixel-based labelling, gives a single label to a group of pixels. Such a contiguous group of pixels with similar characteristics, is known as an *image segment*. This technique forms the basis of object-based image analysis (OBIA) [19], and could also be interpreted as a way to ensure smooth labels [17]. It has been advocated that OBIA is especially suitable for remote sensing applications where the object of interest is larger than the spatial resolution of the image [19]. It has been one of the most common techniques for slum identification from high resolution satellite imagery, though parameter tuning is important to avoid over- or under-segmentation [20]. In light of the difficulty of tuning image segmentation parameters, superpixels could be used [21]. These are in essence an oversegmentation of the image. Superpixel-based image analysis lowers the data redundancy and can speed up classification tasks compared to pixel-based strategies, while avoiding errors due to undetected object boundaries in cases of undersegmentation.

The main motivation behind this study is to combine both the *global contextual uncertainty* (i.e., class representation and object variability within the scene) with the *local contextual consistency* (i.e., similar neighbors having similar classes) to automatically identify and remove noisy labels from training data. By iteratively training supervised classifiers and removing potentially mislabeled samples after each iteration, the training sample set is iteratively cleaned and the accuracy of the classification model is improved. This allows existing vector outlines to be exploited as labels for newly acquired UAV imagery, thereby, reducing the need for the collection of costly training samples and speeding up the basemap updating workflow. The adopted methodology combines various aspects of the state-of-the-art in remote sensing of urban areas and image processing, such as OBIA, the integration of 2-D and 3-D features, and the inclusion of contextual information to improve classification accuracies.

The proposed technique is demonstrated through two applications. The first uses recently acquired UAV imagery and outdated building outlines of an informal settlement in Kigali, Rwanda. The second application demonstrates how the same method can be applied to improve the accuracy of crowd-sourced data. More specifically, to verify the building outlines of an informal settlement in Dar es Salaam, Tanzania, which were digitized by community members using OpenStreetMap. Various experimental setups demonstrate the necessity of using both global and local contextual cues, the sensitivity of the proposed method to the proportion of training labels that are incorrect, and approximate the number of training samples that would need to be manually labeled in order to obtain the same classification accuracy as the automated workflow.

II. Proposed Method

The proposed method takes image segments with descriptive features from the newly acquired dataset and an initial class label obtained from the outdated basemap data as input. Then, it applies three steps to identify samples with unreliable labels and remove these from the training set. These three steps (steps 4–6 in Fig. 1) are: prefiltering the image segments based on the uniformity of noisy labels acquired at t_0 for each segment from the images at t_1, performing a supervised classification, and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

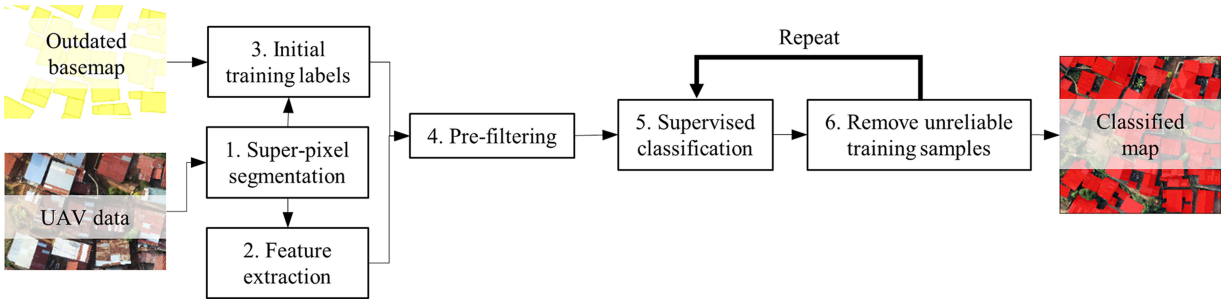

Fig. 1. Workflow of the proposed method for automatically identifying unreliable labels when using existing spatial data to provide training labels for the classification of UAV data.

Algorithm 1: *iterRF-LG.*

Inputs: $S = \{s_1,\ s_2,\ \ldots,\ s_n\}$ image segments with features from t_1, $R \subseteq S$ subset of segments which are used as training samples, $C^k = \{c_1^k, c_2^k, \ldots,\ c_n^k\}$ segment labels at iteration *k,* user-defined number of iterations $k_{\max}$, local contextual consistency threshold $\psi_{\min}$, global contextual uncertainty threshold $\theta_{\min}$.

Procedure:

1. Set $R = S$ and initialize C^0 with noisy training labels from t_0
2. If uniformity (s_i) < minimum uniformity criterion, remove s_i from R

For $k = 1 : k_{\max}$.

3. Train a random forest classifier using segments in R and labels from C^{k-1}
4. Apply the classification model to S and update C^k.
5. If $c_i^k \neq c_i^{k-1}$ OR $\psi_i < \psi_{\min}$ OR $\theta_i < \theta_{\min}$, remove s_i from R

Outputs: improved segment labels $C^{k_{\max}}$.

finally, removing unreliable training samples. Here, uniformity refers to the percentage of pixels in an image segment, which are assigned the label of the most prominent class within that segment. Label reliability is based on the *label consistency*, *local contextual consistency*, and *global contextual uncertainty*. The last two steps (classification and removing unreliable training samples) are repeated iteratively to improve the classification model. This improved classification model can then be used to assign a class label to each image segment and obtain a classified map.

In the following section, we explain how the proposed method works. We use the notation $S = \{s_1,\ s_2,\ \ldots s_n\}$ for the n segments in the image acquired at t_1 and $R \subseteq S$ to the set of segments which are used to train the classifier. Each image segment s_i has an area A_i, a feature vector $\mathbf{x}_i$ obtained from the image dataset at t_1, and a class label c_i^k where *k* refers to the iteration. For example, c_i^0 indicates the class label of s_i according noisy labels acquired at t_0, and c_i^1 refers to the label assigned to s_i after the first iteration of the algorithm. E_i refers to the set of all image segments adjacent to s_i and $l_{i,j}$ refers to the length of the shared border between s_i and $s_j \in E_i$. Pseudocode for the algorithm is provided in Algorithm 1. Please note that this section describes the general workflow of the proposed method, whereas the exact implementation employed for our UAV datasets (including image segmentation and feature extraction) is described in Section III-B. Experimental Analysis.

The reasoning behind the prefiltering step (i.e., step 4 in Fig. 1) is that the building outlines at t_0 may not always align with the image segments obtained from the imagery at t_1, causing these image segments to contain conflicting labels. Therefore, as an initial simple filtering mechanism, only "pure" segments where the percentage of labels from a single class meets a user-defined threshold are selected for the training set R used to develop the classification model. The segments that do not meet this purity criterion are only incorporated at the end of the workflow, when the final classification model is used to classify the entire image and obtain the final classification map.

For the supervised classification step, we propose to use random forests [22] as they have been demonstrated to be more robust to label noise compared to other classification methods [12], [23] and they can easily deal with large numbers of training samples, which is useful as all the segments are labeled in this application of map updating. Furthermore, it is intuitive to derive a confidence measure for the prediction, which is needed for the global contextual uncertainty criterion.

Then, the *label consistency*, *local contextual consistency*, and *global contextual uncertainty* are used to remove unreliable training samples. *Label consistency* implies that the label of a training sample is consistent with the label assigned at the previous iteration, i.e., $c_i^k = c_i^{k-1}$. For example, if one segment represents a building at t_1, it may be nonbuilding according to the outdated basemap labels at t_0. However, as the features of the segment are likely to be similar to other buildings in the area, it could feasibly be classified as building in the second iteration. Therefore, segments where a label is inconsistent causes it to be removed from the training set. This strategy has been previously employed for data cleansing techniques [24], [25], but may be dangerous when used on its own as it may also remove potentially informative samples [26], [27].

The underlying idea of the second criterion, *local contextual consistency*, is that if there are misalignments in the object boundaries at t_0 and at t_1, then the correctly labeled parts of the object may be used to identify neighboring mislabeled segments [see the example in Fig. 2(a)]. This is implemented by

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

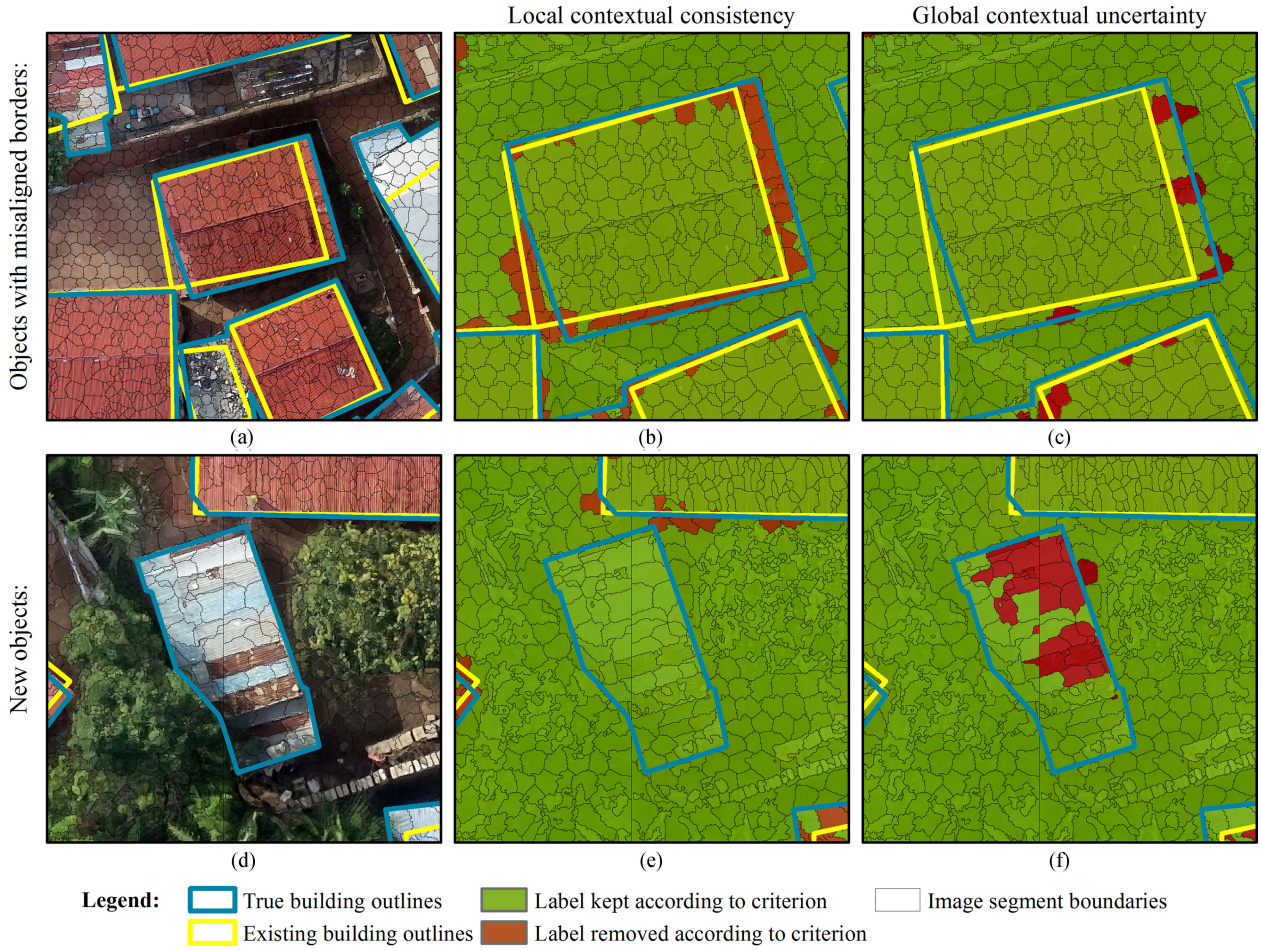

Fig. 2. Illustrative examples from the Kigali dataset showing the interplay between the local contextual consistency (b), (e) and global contextual uncertainty criteria (c), (f). The local contextual consistency is especially useful for updating object boundaries (a)–(c), whereas the global contextual uncertainty is required to capture new objects (d)–(f).

comparing the labels of neighboring pixels or image segments, and introducing a penalty for neighbors that have different labels but similar feature vectors. We exploit the idea of edge potentials commonly adopted in CRFs, and define our contextual consistency criterion using a contrast-sensitive Potts model [28]

$$\phi\left(c_i^k, c_j^k, \mathbf{x}_i, \mathbf{x}_j\right) = \begin{cases} \exp\left[-\beta \mathbf{x}_i - \mathbf{x}_j{}^2\right], & \text{if } c_i^k \neq c_j^k \\ 1, & \text{if } c_i^k = c_j^k \end{cases} \quad (1)$$

where c_i^k and c_j^k indicate the class labels and $\mathbf{x}_i$ and $\mathbf{x}_j$ the feature vectors of two neighboring image segments s_i and s_j. This assigns a value of 1 to edges between neighbors of the same class, and $\exp[-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2]$ to edges between neighbors adhering to different classes, where β equals the average square gradient between all neighboring segments as in [18]. The local contextual consistency index of segment s_i (ψ_i) is the weighted sum of (1) for all neighboring segments

$$\psi_i = \sum_{(i,j)\in E_i} w_{i,j} \cdot \phi(c_i^k, c_j^k, \mathbf{x}_i, \mathbf{x}_j) \quad (2)$$

where $w_{i,j}$ is the relative weight of the neighboring segment s_j. The relative weight ($w_{i,j}$) of each neighbors' edge potential is normalized by border length and relative size of the neighbors [29] as follows:

$$w_{i,j} = \frac{l_{i,j} \cdot A_j}{\sum_{(i,j)\in E_i} l_{i,j} \cdot A_j}. \quad (3)$$

This increases the influence of neighboring segments that share a longer border and larger neighboring segments, as larger segments are presumed to provide more stable feature values. Note that ψ_i penalizes similar segments with different labels, but it does not indicate which of the two neighbors is likely to be correct and which is likely to have the noisy label. Therefore, we only remove the samples for which both the local contextual consistency and the classifier uncertainty fall below a defined threshold. Furthermore, it is important to note that (1) only looks at neighboring segments that have different labels. This is useful for updating building outlines, but not in detecting new objects that are isolated. For example, if a building appears at t_1 in the middle of an area, which was entirely labelled as nonbuilding at t_0, ψ_i will not be suitable for identifying mislabeled training samples [i.e., Fig. 2(e)].

In such situations, mislabeled training samples can be identified by taking into account the *global contextual uncertainty*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

of segment s_i (θ_i). That is to say that the classification model describes the statistical attributes of the objects outlined by the outdated vector data at t_0 in the feature space derived from the imagery at t_1. If a sample is mislabeled by the provided labels, then it is likely to lie closer to its true label in the feature space and may, therefore, cause the classifier to be uncertain about the assignation of the label. If a group of neighboring segments consistently have a high uncertainty according to the classification model, they are likely to be mislabeled. Therefore, we calculate the weighted average of the classifier uncertainty over all neighboring segments

$$\theta_i = \sum_{(i,j)\in E_j} w_{i,j} \cdot u_j^k \tag{4}$$

where $w_{i,j}$ is again the relative weight between neighboring segments s_i and s_j as defined in (2) (i.e., the same weights are used for both the local contextual consistency and global contextual uncertainty) and u_j^k is the classifier uncertainty for segment s_j at iteration *k*. Note that although both ψ_i and θ_i change at each iteration, we omit the superscript *k* in order to simplify the notation.

We propose utilize random forests for the supervised classification task. Random forests consist of a group of classification trees, where each tree is trained using a random subset of the training samples and each decision node depends on a random subset of the features [22]. In the testing stage, a sample passes through each tree and each tree casts a vote for the label of the most prominent class of training samples, which ended up at that leaf. The final label of the sample is determined through a majority voting of the results of each tree in the forest. The uncertainty can easily be calculated as the fraction of reference samples of the leaf that have the most prominent label multiplied for each tree in the forest. For binary classification problems, u_j^k ranges from 0.5 to 1 with higher values representing more confident predictions.

In the proposed method, the steps of supervised classification and removing the unreliable training samples using the three consistency criteria are repeated iteratively. The number of iterations could be fixed by the user. Alternatively, the user could automatically stop the iterations by tracking the number of samples removed from the training set or the number of samples that have are assigned different labels compared to the previous iteration. The accepted classification model can then be used to classify the entire image.

III. Experimental Analysis

A. Datasets

1) Kigali, Rwanda: The first study area concerns an informal settlement in Kigali, Rwanda [see Fig. 3(a)]. In 2015, a DJI Phantom 2 Vision+ UAV was flown over the study area. The images were processed with Pix4Dmapper that provided a Digital Surface Model (DSM) and a true-color orthomosaic with a spatial resolution of 3 cm and point cloud with a density of up to 1014 points/m². Further details regarding this dataset can be found in [7]. A subset of 150 m × 150 m was selected for the present analysis. The outdated building outlines were provided by the local government as vector data, which was initially digitized over a 2008 orthomosaic of 22 cm and partially updated using 2014 Pléiades satellite imagery resampled to 50-cm pixels [30]. True reference data were obtained by manually digitizing the building outlines of the UAV orthomosaic. Around 11% of the segment labels provided by the existing outlines are incorrect according to this reference data.

2) Dar es Salaam, Tanzania: The second dataset was obtained by the Dar Ramani Huria project with the support of the World Bank.[1] This project mobilizes community members and university students to map flood-prone areas of the city. The results are used to support disaster response and are made available to the public through OpenStreetMap. To support these mapping activities, UAV flights were undertaken with a SenseFly eBee mounted with a 14-MP Canon Powershot RGB camera. The images were again processed with Pix4Dmapper to obtain a point cloud with an average density of 50 points/m², and a 5-cm DSM and orthomosaic. A 300 m × 300 m subset located in the Tandale ward was used for the present analysis [see Fig. 3(b)].

For this dataset, the "noisy" labels consist of the building outlines that were digitized over the 2015 imagery within the Ramani Huria project. Similar to the Kigali dataset, the true reference data were manually digitized over the subset for the purposes of the present study. Although most objects were correctly digitized by the Ramani Huria project, a label noise of about 10% is observed.

B. Experimental Setup

1) Image Segmentation: To segment the orthomosaic, the SLIC superpixel algorithm was used [21]. SLIC first defines a regular grid over the image, where the grid interval is based on a user-defined target super-pixel size. These samples are used to initialize a *k*-means clustering, where each pixel in the image is assigned to the nearest cluster center. The proximity to cluster centers is calculated in a 5-D space that consists of spectral (*L*, *a*, *b*, values of the pixel in CIELAB colorspace) and spatial (the *x*, *y* image coordinates) components. After all pixels are assigned to a superpixel, the cluster centers are updated by averaging the Labxy values of all pixels within the superpixel, and the process is repeated. In our experimental analysis, the target superpixel size was set to approximately 0.5 m² (i.e., 555 pixels for the Kigali dataset and 200 pixels for Dar es Salaam). However, it was noted that a number of very small segments were still present, which could unnecessarily slow down the image processing workflow. Therefore, all segments with an area less than 0.05 m² (i.e., 55 pixels for the Kigali dataset and 20 for Dar es Salaam) were merged with the most similar neighboring segment larger than 0.05 m². This segmentation strategy resulted in a total of 59 812 image segments for the Kigali dataset and 103 227 segments for Dar es Salaam.

2) Feature Extraction: For each segment, the average R, G, B color values, normalized r., g., b., and the ExG(2) vegetation index [31] were calculated, as well as a normalized histogram displaying the relative frequency of local binary pattern (LBP)

[1] http://ramanihuria.org/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

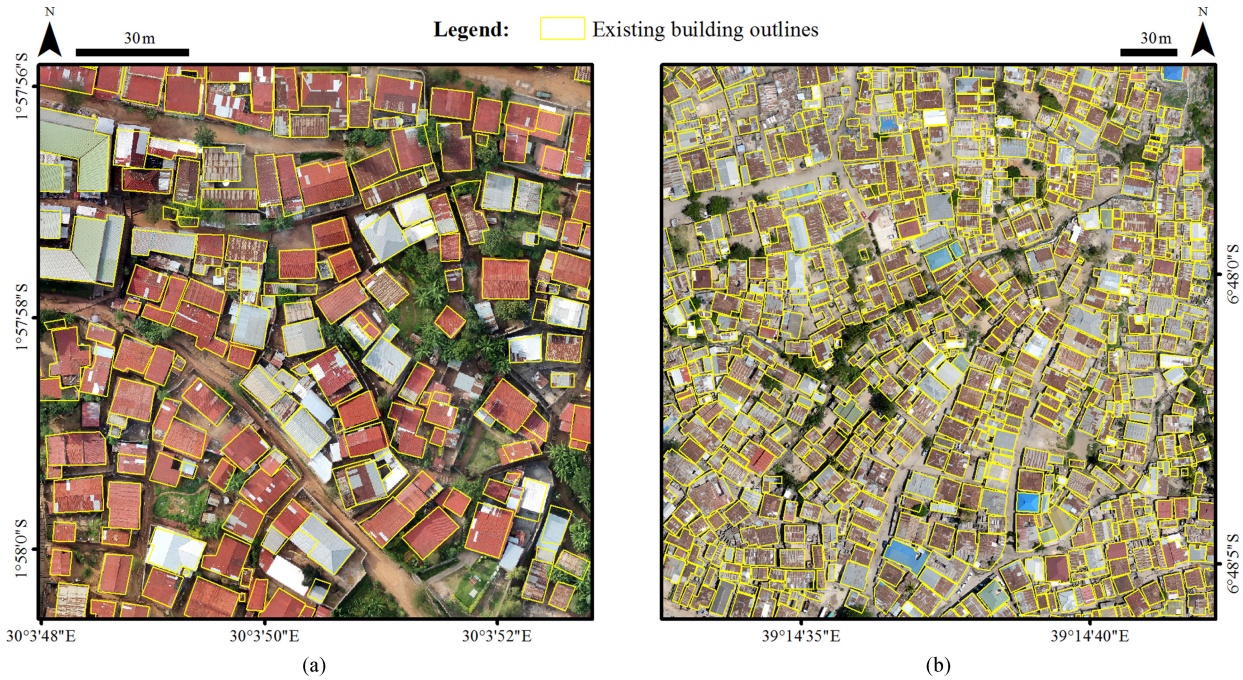

Fig. 3. (a) Kigali and (b) Dar es Salaam datasets used in this study. The building outlines (i.e., noisy labels) from t_0 are displayed in yellow over the images acquired at t_1.

texture patterns [32] within each segment. Features from the point cloud were also included in the classification: The number of points falling into the spatial extent of each pixel, as well as the range and standard deviation in the elevation values of these points. Planar segment and local neighborhood features of the highest point per pixel were also assigned to each image pixel. Planar segments were obtained through a surface-growing algorithm [33] and the number of points, average residual, inclination angle, and maximum height difference of the plane above neighboring points from different segments were included as features. The local neighborhood features were calculated according to the framework proposed by [34]. The values of each 3-D feature per pixel was averaged over the image segments. These image-based and point-cloud based features together form the set of features used for the classification task. A more comprehensive overview of the utilized features is provided in [35].

3) Initial Training Labels: The vector data from the existing basemap are first rasterized using the same grid as the UAV orthomosaic. Next, a majority voting is used to assign a binary label (building versus nonbuilding) to each SLIC segment, which represents the outdated label at t_0. The true labels at t_1, which are used for the performance assessment, were obtained by manually digitizing the building outlines in the UAV orthomosaic, and assigning them to the image segments in the same way.

4) Prefiltering: The previous processing steps yield: The input feature data consisting of segments that are described by radiometric, textural and geometric features derived from the UAV data, the outdated building versus nonbuilding labels at t_0, and true reference labels from the UAV data at t_1. In the prefiltering step, only segments for which at least 60% of the segment had the same label were retained. Experimental results using a threshold of 0% (i.e., not applying the prefiltering) and 100% (i.e., only including "pure" segments in the classification) are also provided in the results section to indicate the importance of this pre-filtering step.

5) Iterative Supervised Classification: The next step performs the iterative classification with a random forest classifier. The number of trees is optimized through cross validation, by randomly selecting 500 training samples, training forests with up to 200 trees, and selecting the number of trees with the lowest cross validation error. This optimal number of trees was then used to train the random forest classifier using all the training samples that was subsequently used classify the entire dataset.

6) Removal of Unreliable Training Samples: Four strategies are applied to illustrate the importance of combining both local and global contextual cues for identifying unreliable labels. The first method, iterRF, does not take any contextual criteria into account and simply uses the label consistency criterion. The second method, iterRF-L, removes samples for which the local contextual consistency index ψ_i is lower than the threshold value of 0.7, whereas iterRF-G only employs the global contextual uncertainty index θ_i and the same threshold value. Finally, the proposed method iterRF-LG uses both local contextual consistency and global contextual uncertainty criteria. For all four methods, 15 iterations (of steps 5 and 6 in Fig. 1) were performed.

7) Assessment: The four strategies are compared through the overall accuracy (OA) of all the segments, the OA of only the originally mislabeled segments, the number of false positives and false negatives in the training set, and the percentage

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
ACCURACY MEASURES OF THE PROPOSED ITERATIVE STRATEGIES AFTER 15 ITERATIONS

Data Cleansing Strategy	OA (%) All Segments	OA (%) Mislabeled Segments	False Positives (%)	False Negatives (%)	Percentage of Mislabeled Samples in Training Set
Kigali Dataset					
Using noisy labels	88.2	6.6	3.27	8.54	11.1
iterRF	88.9	7.3	2.94	8.15	11.0
iterRF-L	91.2	32.2	1.49	7.30	8.3
iterRF-G	90.1	23.6	2.14	7.79	9.1
iterRF-LG (proposed method)	*92.1*	*47.9*	*1.00*	*6.91*	*6.2*
iterRF-LG (no pre-filtering)	91.2	46.8	1.08	6.74	5.9
iterRF-LG (only uniform segments)	92.2	54.0	0.84	6.96	4.3
Dar es Salaam Dataset					
Using noisy labels	89.0	8.5	3.73	7.28	10.0
iterRF	89.6	8.6	3.45	6.92	10.0
iterRF-L	90.4	24.6	2.85	6.78	8.45
iterRF-G	90.4	24.7	3.33	6.31	8.23
iterRF-LG (proposed method)	*91.3*	*41.1*	*2.82*	*5.92*	*6.45*
iterRF-LG (no pre-filtering)	91.1	37.2	2.95	5.98	6.52
iterRF-LG (only uniform segments)	90.3	52.0	2.86	6.83	3.45

of mislabeled samples in the training set. Note that the underlying idea of the proposed method is to eliminate the need of collecting labeled training data by exploiting existing geospatial information. Therefore, in order to compare the proposed method to the traditional method of manually labeling training samples for the classifier, we provide an experiment that indicates how many (correct) training sample labels would need to be collected in order to obtain the same classification accuracy. This is done by randomly selecting tenfolds of a set number of reference samples at t_1, constructing a random forest classifier, and obtaining the OA. Finally, we also present the results of a sensitivity analysis. This was performed by taking the true labels of the Kigali dataset, and randomly changing training sample labels to induce a noise level of 0%, 5%, 10%, 20%, 30%, 40%, or 50%. Using these labels, iterRF-LG was again performed for 15 iterations. The average OA over three trials is reported for each iteration and noise level.

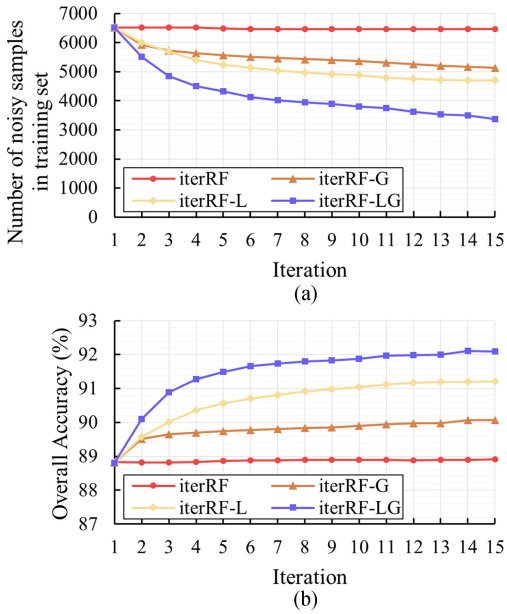

Fig. 4. Number of noisy training samples remaining in the set of samples used to train the classifier after each iteration (a) and the resulting overall accuracy for the Kigali dataset using the four different methods for filtering the training labels.

IV. RESULTS AND DISCUSSION

Table I provides the OA of the four different strategies for removing label noise from the training set after 15 iterations. The *iterRF* strategy, which only takes label consistency into account, does not improve the results significantly. The number of mislabeled and the classification accuracy is relatively stable after the first 15 iterations (see Fig. 4). The local contextual consistency (*iterRF-L*) and global contextual uncertainty (*iterRF-G*) achieve a similar accuracy for the Dar es Salaam dataset, correctly classifying about 90.4% of the image segments. Although a comparable number of noisy labels remain in the training set after 15 iterations (see Fig. 4), *iterRF-L* (91.2%) outperforms *iterRF-G* (90.1%) for the Kigali dataset. However, it is clear that the proposed method that combines all three criteria, *iterRF-LG*, obtains the best performance. For both datasets, a McNemar test with continuity correction [36] indicates that the results between *iterRF-LG* and the three other methods are statistically significant (p-value of < 0.001). The proposed method correctly classifies 92.1% of the segments for the Kigali dataset, corresponding to an improvement of 3.3% compared to using the initial, noisy training labels. This improvement was 1.7% for the Dar es Salaam dataset. The improvement is more visible when we consider only the segments which were mislabeled in the noisy training labels: The proposed method increased the accuracy of these segments from 6.6% to 47.9% in the Kigali dataset, and 8.5% to 41.1% in the Dar es Salaam dataset. Finally, the success of the method in removing unreliable labels is visible through the reduction of the fraction of mislabeled samples in the training data. This was effectively reduced from 11.1% to 6.2% in the Kigali dataset (effectively removing 44.1% of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

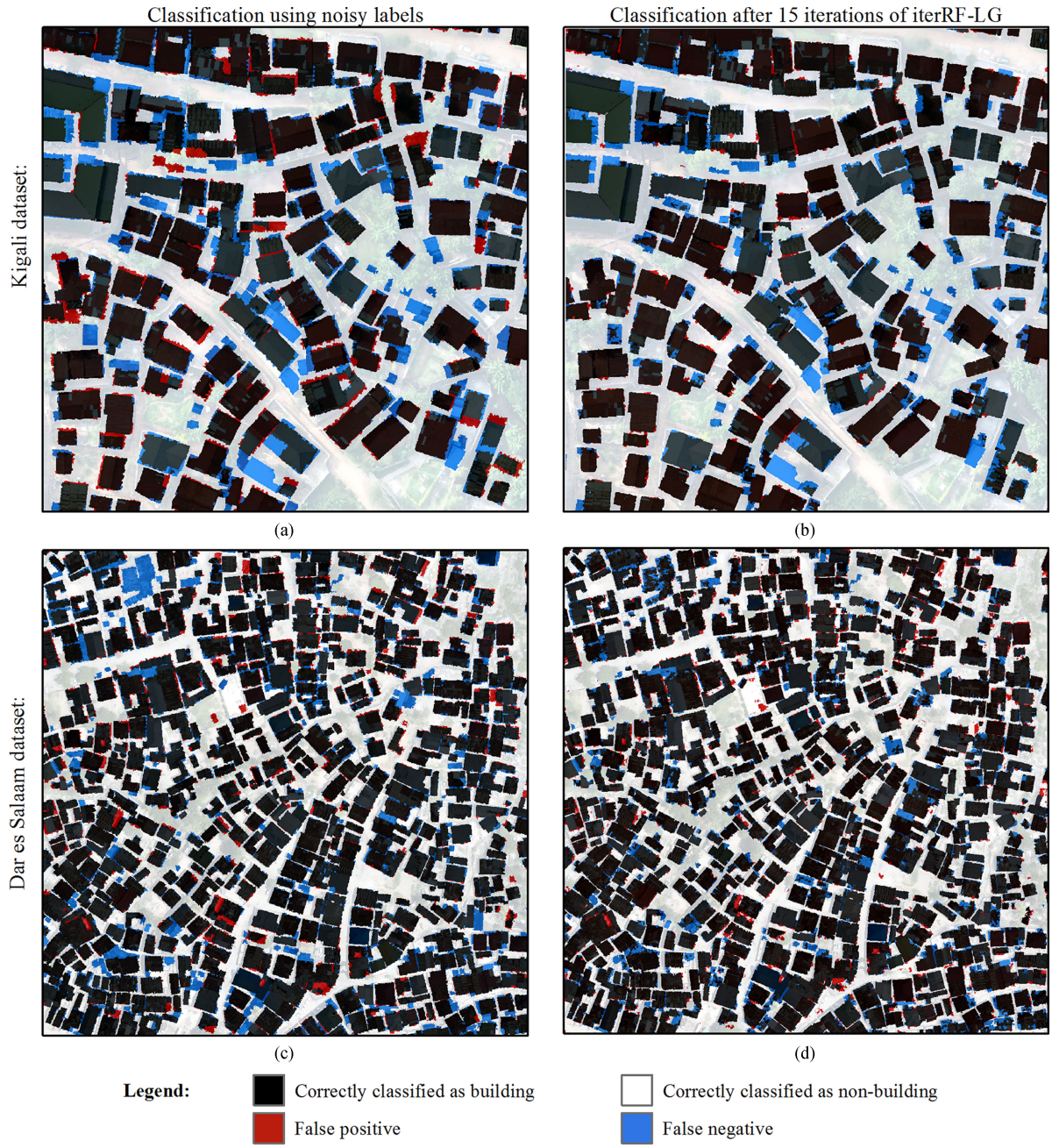

Fig. 5. Results of the classification using the noisy labels (a), (c) and after the fifteenth iteration of *iterRF-LG* (b), (d) for the Kigali (a), (b) and Dar es Salaam (c), (d) datasets.

the mislabeled samples from the training set), and from 10.0% to 6.4% in the Dar es Salaam dataset (removing 36.0% of the mislabeled samples).

The influence of the prefiltering step on the results of *iterRF-LG* is also visible in Table I. Increasing the uniformity criterion results in a decrease in the number of mislabeled segments in the training data and a more accurate classification of the mislabeled segments for both datasets. Using only pure segments (i.e., a uniformity of 99%) in the prefiltering stage improves the OA of the entire Kigali dataset by 0.1%, but decreases the accuracy of the Dar es Salaam dataset by 1.0%. The results, therefore, suggest that increasing the uniformity criterion in the prefiltering stage reduces the number of noisy labels in the training set and improves the classification accuracy of mislabeled samples. However, using a strict uniformity criterion may decrease the classification accuracy of the entire dataset, perhaps due to the exclusion of informative training samples.

The improved results of *iterRF-LG* compared to the other three methods is also visible in the output classification maps (see Fig. 5). For example, there is a notable reduction in false positives in the Kigali dataset. A building missed in the manual delineation of the buildings in Dar es Salaam (see Fig. 5(c), top left), is correctly identified through the *iterRF-LG* method [see (Fig 5(d)]. In the Kigali dataset, a number of roofs are

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

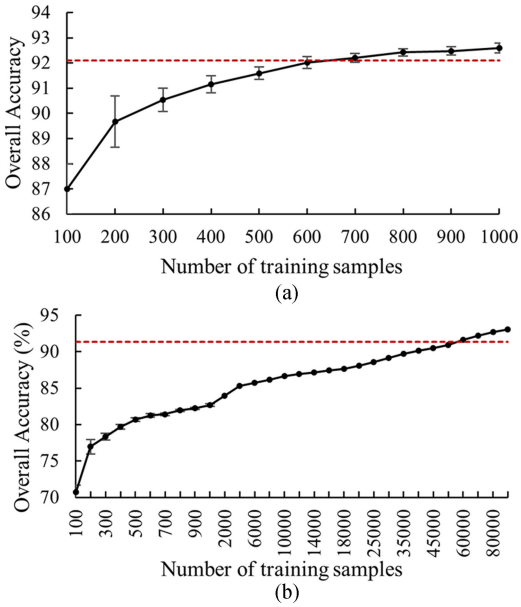

Fig. 6. Comparison between the overall accuracy achieved through *iterRF-LG* after 15 iterations (red dashed line) and the mean overall accuracy achieved by randomly selecting a set number of training samples with true labels (black line) for the (a) Kigali and (b) Dar es Salaam datasets.

still not recognized by the classification model, remaining false negatives in the *iterRF-LG* method [see Fig. 5(b)]. A visual analysis of the image indicates that many of these errors are in locations where the building extensions have been covered with a different roofing material than the (correctly labelled) adjacent construction. This causes a difference in the feature vectors of neighboring segments, and may, therefore, mislead the local contextual consistency criterion. If this is coupled to a consistent change in the representation of objects between t_0 and t_1—for example, if building extensions consist of a new type of roofing material that is not well represented by the existing labels—then the global contextual uncertainty may also fail. Note that in (1), the all segment features are weighted equally when determining the similarity between neighboring segments. Further research could consider incorporating more advanced techniques to select or weight the different features as previous research indicated that considering 3-D and 2-D features separately may improve image classification results [35].

Another set of experiments compared the proposed workflow with a traditional workflow, where image segments must be labeled manually. Experimental analysis indicates that approximately 600 correctly labeled training samples would be needed in the Kigali dataset to obtain the same accuracy as *iterRF-LG* after 15 iterations [see Fig. 6(a)]. For the Dar es Salaam dataset, this is much higher, and between 50 000 and 60 000 training samples would be needed [see Fig. 6(b)]. This could be due to the spectral similarity of building roofs and ground in the Dar es Salaam dataset. Furthermore, the slightly lower spatial resolution of the Dar es Salaam dataset makes it difficult to capture the texture of the corrugated iron roofs, which proved to be an important distinguishing attribute for the Kigali dataset [7].

The large number of training samples required for the Dar es Salaam dataset can easily be dealt with by a random

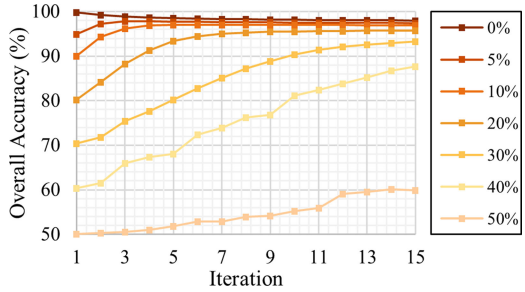

Fig. 7. Overall accuracy of *iterRF-LG* for the Kigali dataset after 15 iterations with initial label noise levels ranging from 0% to 50%.

forest classifier. Other supervised classification methods, such as support vector machine (SVM) also achieve high accuracies in remote sensing applications [37]. Future investigations regarding the use of SVM instead of random forests for the proposed *iter-LG* method would require two adaptations. First, the number of training samples would need to be reduced by sampling or using an SVM variant that is capable of dealing with large numbers of training samples such as DC-SVM [38]. Second, a classifier uncertainty measure (u_j^k) would need to be assigned to each training sample to calculate the global contextual uncertainty. An SVM does not directly provides a probability for the classification output, although strategies exist that use proxies to indicate the classification certainty, e.g., [39].

Finally, the results of the sensitivity analysis for the Kigali dataset are presented in Fig. 7. The results indicate that after 15 iterations the classification accuracy is above 93% for noise levels of up to 30%. In these experiments, the noise is introduced by switching the labels of randomly selected training labels. It is possible that the label noise in practical applications is more systematic (e.g., new constructions make use of a different roofing material, or a concentration of adjacent mislabeled samples), which may have a more significant impact on the results of the proposed method.

V. Conclusion

In this paper, we utilize two datasets to demonstrate how existing spatial data may be exploited to obtain labeled training samples for the application of supervised classification algorithms to UAV data. Considering that a number of labels provided by this outdated spatial data will be erroneous, local, and global image cues are used to filter out unreliable training samples. The local contextual criterion encourages neighboring image segments to have consistent labels. At the same time, a global contextual criterion uses the entire scene to capture the distribution of the semantic classes in the feature space, and is suitable for identifying isolated new objects. Sensitivity analyses show that classification accuracies of 93% or more are achieved, even in the presence of up to 30% erroneous training samples. There are two main implications of these results. The first is that the proposed method may lead to a considerable speed-up in the implementation of supervised classification methods for base-map updating by reducing the need of manually labeling image segments to train the classifier. Second, the interaction

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

between the local and global cues emphasizes that the inclusion of spatial contextual information is beneficial for data cleansing techniques in geomatics applications.

The proposed method may also be used for a number of other applications. For example, it could be used in a quality control application to verify the accuracy of volunteered geographical information such as OpenStreetMap. Furthermore, it could be used in a domain adaptation application, where the training labels are obtained from a classification model trained on a certain study area could be applied to a similar study area for which no data are available, rather than outdated spatial data. The main caveat of this method is that it assumes that the noisy data labels provided by the outdated spatial data cover all the representations of the semantic classes in the new UAV imagery. Therefore, if an entirely new variation of an object appears between t_0 and t_1, for example if an alternative type of roof material is only used in new constructions, the mislabeled segments will not be filtered by the proposed method. Further developments could explore active learning methods [40] to target such segments and potentially improve the classification accuracy, though this would require (limited) manual labeling.

REFERENCES

[1] F. Nex and F. Remondino, "UAV for 3D mapping applications: A review," *Appl. Geomatics*, vol. 6, no. 1, pp. 1–15, 2014.

[2] C. Stöcker, R. Bennett, F. Nex, M. Gerke, and J. Zevenbergen, "Review of the current state of UAV regulations," *Remote Sens.*, vol. 9, no. 5, pp. 1–26, 2017.

[3] M. Koeva, M. Muneza, C. Gevaert, M. Gerke, and F. Nex, "Using UAVs for map creation and updating. A case study in Rwanda," *Surv. Rev.*, pp. 1–14, 2016.

[4] R. Huria, "The Atlas of Flood Resilience in Dar es Salaam," *Dar es Salaam*, 2016, 126 pages.

[5] T. Moranduzzo, F. Melgani, M. L. Mekhalfi, Y. Bazi, and N. Alajlan, "Multiclass coarse analysis for UAV imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6394–6406, Dec. 2015.

[6] B. Chen, Z. Chen, L. Deng, Y. Duan, and J. Zhou, "Building change detection with RGB-D map generated from UAV images," *Neurocomputing*, vol. 208, pp. 350–364, 2016.

[7] C. M. Gevaert, C. Persello, R. Sliuzas, and G. Vosselman, "Informal settlement classification using point-cloud and image-based features from UAV data," *ISPRS J. Photogramm. Remote Sens.*, vol. 125, pp. 225–236., 2017.

[8] Q. Zhang, R. Qin, X. Huang, Y. Fang, and L. Liu, "Classification of ultra-high resolution orthophotos combined with DSM using a dual morphological top hat profile," *Remote Sens*., vol. 7, no. 12, pp. 16422–16440, 2015.

[9] J. Senthilnath, M. Kandukuri, A. Dokania, and K. N. Ramesh, "Application of UAV imaging platform for vegetation analysis based on spectral-spatial methods," *Comput. Electron. Agric.*, vol. 140, pp. 8–24, 2017.

[10] V. Mnih and G. E. Hinton, "Learning to label aerial images from noisy data," in *Proc. 29th Int. Conf. Mach. Learn.*, 2012, pp. 567–574.

[11] J. Chen and A. Zipf, "DeepVGI: Deep learning with volunteered geographic information," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 771–772.

[12] B. Frenay and M. Verleysen, "Classification in the presence of label noise: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 845–869, May 2014.

[13] A. Maas, F. Rottensteiner, and C. Heipke, "Using label noise robust logistic regression for automated updating of topographic geospatial databases," in *Proc. 23rd ISPRS Congr.*, Commission, Göttingen: Copernicus GmbH., vol. 3, No. 7, pp. 133–140, 2016.

[14] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009.

[15] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey," *Comput. Vis. image Underst.*, vol. 114, no. 6, pp. 712–722, 2010.

[16] S. ten Oever, V. Romei, N. van Atteveldt, S. Soto-Faraco, M. M. Murray, and P. J. Matusz, "The COGs (context, object, and goals) in multisensory processing," *Exp. Brain Res.*, vol. 234, no. 5, pp. 1307–1323, 2016.

[17] K. Schindler, "An overview and comparison of smooth labeling methods for land-cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4534–4545, Nov. 2012.

[18] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *Int. J. Comput. Vis.*, vol. 81, no. 1, pp. 2–23, 2009.

[19] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[20] M. Kuffer, K. Pfeffer, and R. Sliuzas, "Slums from space—15 years of slum mapping using remote sensing," *Remote Sens*., vol. 8, no. 6, p. 455–484, May 2016.

[21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[22] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[23] A. Folleco, T. M. Khoshgoftaar, J. Van Hulse, and L. Bullard, "Software quality modeling: The impact of class noise on the random forest classifier," in *Proc. IEEE Congr. Evol. Comput.*, 2008, pp. 3853–3859.

[24] J. Thongkam, G. Xu, Y. Zhang, and F. Huang, "Support vector machine for outlier detection in breast cancer survivability prediction," in *Proc. Asia-Pacific Web Conf.*, 2008, pp. 99–109.

[25] P. Jeatrakul, K. W. Wong, and C. C. Fung, "Data cleaning for classification using misclassification analysis," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 14, no. 3, pp. 297–302, 2010.

[26] N. Matic, I. Guyon, L. Bottou, J. Denker, and V. Vapnik, "Computer aided cleaning of large databases for character recognition," in *Proc. 11th IAPR Int. Conf. Pattern Recog.*, 1992, pp. 330–333.

[27] I. Guyon, N. Matic, and V. Vapnik, "Discovering informative patterns and data cleaning" Menlo Park, California, Tech. Rep. WS-94-03, 1996.

[28] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in *Proc. 8th IEEE Int. Comput. Vision*, 2001, vol. 1, pp. 105–112.

[29] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *Proc. 11th IEEE Int. Comput. Vision*, 2009, pp. 1–8.

[30] F. Bachofer, "Assessment of building heights from pléiades satellite imagery for the Nyarugenge sector, Kigali, Rwanda," *Rwanda J.*, vol. 1, no. 1S, 2016.

[31] D. M. Woebbecke, G. E. Meyer, K. Von Bargen, and D. A. Mortensen, "Color indices for weed identification under various soil, residue, and lighting conditions," *Trans. ASAE*, vol. 38, no. 1, pp. 259–269, 1995.

[32] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[33] G. Vosselman, "Automated planimetric quality control in high accuracy airborne laser scanning surveys," *ISPRS J. Photogramm. Remote Sens.*, vol. 74, pp. 90–100, Nov. 2012.

[34] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 286–304, Jul. 2015.

[35] C. M. Gevaert, C. Persello, and G. Vosselman, "Optimizing multiple kernel learning for the classification of UAV data," *Remote Sens.*, vol. 8, no. 12, p. 1025–1047, 2016.

[36] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.

[37] L. Bruzzone and C. Persello, "Approaches based on support vector machines to classification of remote sensing data," in *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific, 2010, pp. 329–352.

[38] C.-J. Hsieh, S. Si, and I. Dhillon, "A divide-and-conquer solver for kernel support vector machines," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 566–574.

[39] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.

[40] C. Persello, "Interactive domain adaptation for the classification of remote sensing images using active learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 736–740, Jul. 2013.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Caroline M. Gevaert received the B.Sc. degree in international land and water management from the University of Wageningen, Wageningen, The Netherlands, in 2011, the M.Sc. degree in remote sensing from the University of Valencia, Valencia, Spain, in 2013, and the M.Sc. degree in geographical information science from Lund University, Lind, Sweden, in 2014. Since 2014, she has been working toward the Ph.D. degree with the The Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, Netherlands.

The main topic of her research is the utility of Unmanned Aerial Vehicles for informal settlement mapping. She has published various papers in leading remote sensing journals.

Ms. Gevaert was nominated for the Student Paper Award at the Joint Urban Remote Sensing Event (JURSE) 2017 in Dubai.

Claudio Persello (S'07–M'11–SM'17) received the Laurea (B.S.) and Laurea Specialistica (M.S.) degrees in telecommunications engineering and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2003, 2005, and 2010, respectively.

He is currently an Assistant Professor with the Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands. In 2011–2014, he was a Marie Curie Research Fellow, conducting research activity at the Max Planck Institute for Intelligent Systems and the Remote Sensing Laboratory, University of Trento. His main research interests include the analysis of remote sensing data, machine learning, image classification, pattern recognition, and unmanned aerial vehicles.

Dr. Persello is a Referee for multiple journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Remote Sensing*, and *Pattern Recognition Letters*. He served on the Scientific Committee of the Sixth International Workshop on the Analysis of Multi-temporal Remote-Sensing Images (MultiTemp 2011). His Ph.D. thesis was awarded with the prize for the best Ph.D. thesis on Pattern Recognition published between 2010 and 2012 by the GIRPR, i.e., the Italian branch of the International Association for Pattern Recognition.

Sander Oude Elberink received the M.Sc. degree in geodetic engineering from the Delft University of Technology, Delft, The Netherlands, in 2000 and the Ph.D. degree in geoinformatics from the Faculty of Geo-Information and Earth Observation, University of Twente, Enschede, The Netherlands, in 2010.

From September 2009, he is an Assistant Professor with the Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

Dr. Oude Elberink's Ph.D. research was part of the project "3D Topography," which received the RGI Innovation Award in the category science in 2007. He received a young author's award for best papers at the ISPRS congress in Beijing, China in 2008. In 2009, he received the ITC Research Award for a journal paper on 3-D road reconstruction, which was coauthored by G. Vosselman and published in the Photogrammetric Record. In 2016, he received the ISPRS Guiseppi Inghilleri award for his high quality and innovative research in 3-D landscape modeling that has successfully been transferred to practice to serve the society.

George Vosselman received the M.Sc. degree in geodetic engineering from the Delft University of Technology, Delft, The Netherlands, in 1986 and the Ph.D. degree in photogrammetry from the University of Bonn, Bonn, Germany, in 1991.

From 1987 to 1992, he was a Researcher with the University of Stuttgart, Stuttgart, Germany. After a year, as a Visiting Scientist with the University of Washington, he was an appointed Professor in photogrammetry and remote sensing with the Delft University of Technology in 1993. Since 2004, he has been a Professor in geo-information extraction with Sensor Systems, University of Twente, Enschede, The Netherlands. His research interests include information extraction from point clouds and imagery, 3-D building and landscape modeling, quality analysis of point clouds, and sensor integration and fusion.

Prof. Vosselman is Board Member of the Netherlands Geodetic Commission and corresponding member of the German Geodetic Commission. He was an Editor-in-Chief of the *ISPRS Journal of Photogrammetry and Remote Sensing* from 2004 to 2012 and received the Hansa Luftbild Award, the ISPRS Otto von Gruber Award, the Schwidefsky Medal, the Karl Kraus Medal, and the ASPRS Photogrammetric (Fairchild) Award.

Richard Sliuzas received the Ph.D. degree in geographical sciences from Utrecht University, Utrecht, Netherlands, in 2004, for his research dealing with the management of informal settlements in Dar es Salaam, Tanzania, using geographic information technology.

He is an Australian/Dutch urban planner, specialized in the use of geospatial technologies for urban planning and management. From 1981 to 1983, he worked as a Planning Consultant and local government planner in South Australia. Since joining the Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands, in December 1983, he has worked in planning education and research. His research interests and activities are focused on the use of geo-spatial technologies in spatial planning for sustainable urban development with an emphasis on issues related to urban informality, urban poverty alleviation and the relationship between spatial planning and disasters. He is currently a committee member of the GEO Human Planet Initiative.