To access the published version, please go to 10.1109/IGARSS.2018.8518711

CORRECTING MISALIGNED RURAL BUILDING ANNOTATIONS IN OPEN STREET MAP USING CONVOLUTIONAL NEURAL NETWORKS EVIDENCE

John E. Vargas-Muñoz[1], *Diego Marcos*[2], *Sylvain Lobry*[2], *Jefersson A. dos Santos*[3], *Alexandre X. Falcão*[1], *Devis Tuia*[2]

[1]Institute of Computing - University of Campinas, Campinas, Brazil
[2]Laboratory of Geo-information Science and Remote Sensing, Wageningen University & Research, Netherlands
[3]Department of Computer Science - Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

ABSTRACT

Mapping rural buildings in developing countries is crucial to monitor and plan in those vulnerable areas. Despite the existence of some rural building annotations in OpenStreetMap (OSM), those are of insufficient quantity and quality to train models able to map large areas accurately. In particular, these annotations are very often misaligned with respect to the buildings that are present in updated aerial imagery. We propose a Markov Random Field (MRF) method to correct misaligned rural building annotations. To do so, our method uses i) the correlation between candidate aligned OSM annotations and buildings roughly detected on aerial images and ii) the local consistency of the alignment vectors.

***Index Terms*—** Rural buildings, Markov random fields, OpenStreetMap, Convolutional neural networks, Registration

1. INTRODUCTION

Nowadays, mapping information covering most existing human settlements in the world is available via Web services such as OpenStreetMap. However, most rural buildings are not mapped in any of the aforementioned systems. Rural building mapping is important to support demographic studies and to plan actions in response to crises that affect those vulnerable geographical areas.

Convolutional Neural Networks (CNNs) are widely used to obtain land cover/use maps from remote sensing imagery [1] and have been used to detect and delineate buildings in urban areas [2, 3, 4] . However, the bottleneck is that they require large amounts of labeled samples to be trainable. Although several Volunteered Geographic Information systems, such as OpenStreetMap (OSM), have gathered large amounts of building annotations data, the annotations they provide have two main problems [2]: *1)* due to an incomplete annotation or an imagery update, some objects that appear in the imagery might not be present on the annotation map; *2)* the location of the annotation of an object in the map might be inaccurate (misalignment errors). The authors in [2] propose new loss functions to reduce the effect of these two problems while training the CNN model for urban building classification. However, for the case of rural building classification, the problem of misalignment is more severe, especially since, when compared to urban buildings, rural buildings are smaller and scarcer in OSM [5]. Figure 1 shows an example of OSM building polygons overlayed on an image obtained from Bing maps. We can observe that some objects in the image and the annotations do not overlap, making these annotations, as they are, of little use for training a CNN.

Fig. 1. Misaligned OSM building annotations superimposed into imagery obtained from Bing maps. Some buildings in the imagery and their corresponding annotations do not overlap.

Three main tasks need to be performed to correct many of the existing rural building annotations in OSM:

1. Correct misalignments of existing annotations.
2. Remove annotations of buildings that no longer exist.
3. Add new annotations for buildings that appear in the updated imagery for the first time.

In this work, we propose a solution for the first two tasks, namely alignment correction and incorrect annotations re-

The authors acknowledge the FAPESP (grant 2016/14760-5, 2017/10086-0 and 2014/12236-1), the CNPq (grant 449638/2014-6) and the Swiss National Science Foundation (grant PP00P2-150593).

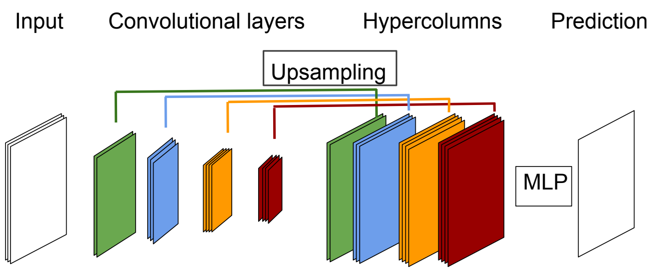

Fig. 2. CNN architecture used to predict building probability maps for correcting misaligned annotations.

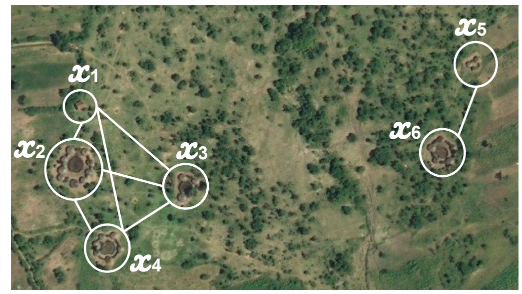

Fig. 3. Neighboring system of the proposed model.

moval. Our solution uses a Markov Random Field (MRF) [6] based registration method to solve the alignment problem. MRFs have been shown to work well in practice for nonrigid registration of images, both in the medical [7] and remote sensing domains [8]. In our case, the MRF maximizes the overlap between the misaligned annotations and the rough probability map generated by a CNN model (*unary term*), while encouraging nearby annotations to use a similar shift correction (*pairwise term*). Given that rural buildings are gathered in small groups in which the individual buildings often have the same misalignment error, we consider these small groups of rural buildings as nodes of the MRF and we compute just a single alignment vector for all the buildings in each group.

2. CORRECTING MISALIGNED ANNOTATIONS

In order to obtain the best alignment vector for each OSM rural building annotation, we need to measure how well a collection of displacement vectors performs. The correlation between the buildings probabilities provided by a CNN model and the aligned annotations is a good heuristic for that purpose.

We used a CNN model based on [9], trained on a small data set of manually annotated data. We applied four convolutional layers to the input, but instead of upsampling the activations using deconvolutions, as in [9], we used the concept of hypercolumns, as in [10]. This way, the model becomes more efficient to train while obtaining similar performance. The architecture is depicted in Figure 2: after applying the convolutional layers, all the output activations are upsampled to the original size of the image. Then, a per pixel Multi-Layer Perceptron (MLP) is applied to obtain the final predictions.

As pointed out above, rural buildings are concentrated in small groups, in which all the buildings tend to undergo the same alignment. Using this prior knowledge, we aligned groups of buildings (using the same alignment vector) with the building probability maps rather than individual buildings.

An additional observation is that close groups of buildings tend to undergo similar displacements as well. To find the best alignment complying with both CNN evidence and this smoothness prior, we used a MRF model solving the global inference problem of displacing all groups of buildings as nodes in a graph. MRF is a widely used framework to enforce spatial dependencies between correlated features. The problem of correcting misaligned annotation can be seen as seeking a realization $\mathbf{d} = \{d_0, d_1, \ldots, d_n\}$ of a MRF (where d_i is an alignment vector) to be applied on the original positions of the annotations $\mathbf{x}$ based on the CNN probability map $\mathbf{y}$. As stated above, we use groups of buildings, or *sites*, as the nodes in our graph. Sites i and j are defined to be neighbors (i.e. $j \in N_i$) if they are spatially close as shown in Figure 3.

The optimal alignment $\hat{\mathbf{d}}$ is then given by:

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}\in\mathcal{D}^{\mathbf{N}}} \sum_i U(d_i|x_i, y_i) \tag{1}$$

$$= \arg\min_{\mathbf{d}\in\mathcal{D}^{\mathbf{N}}} \sum_i -\log C(d_i(x_i), y_i) + \beta \sum_{j\in N_i} \frac{1}{Z}||d_i - d_j||_2,$$

where $\mathcal{D} = \{D_1, D_2, \ldots, D_m\}$ is the set of the m possible alignment vectors, β is the parameter that controls the spatial regularization, Z is a normalization factor (maximum possible distance between two vectors in $\mathcal{D}$) and $C(d_i(x_i), y_i)$ is the normalized correlation between the annotation after alignment $d_i(x_i)$ and the CNN building probability map y_i. This is equivalent to taking the average of the building probability predictions of the pixels contained in the annotations aligned with d_i. Finally, we used the Iterated Conditional Modes (ICM) [6] initialized with $\mathbf{d} = \{\arg\max_{d\in\mathcal{D}} C(d_i(x_i), y_i)\}_i$ to minimize the proposed energy.

In order to remove the annotations of buildings that no longer exist (and therefore addressing problem 2 raised in the Introduction), we compute the average CNN output probability values of the pixels contained in the aligned building annotations. If the computed value is less than 50%, we remove the corresponding building annotation. This simple heuristic works well, since probability values are generally very low if the building is absent (the CNN uses only the image as input and is not aware of existing OSM annotations). Algorithm 1 summarizes the proposed method.

Algorithm 1 – MRF-BASED ALIGNMENT ALGORITHM

INPUT: Map of misaligned annotations M.
OUTPUT: Alignment vectors $\mathbf{d}$.

```
Group the misaligned building annotations M according to their spatial distance from each other, obtaining the set x.
Define the neighbors N_i of each site i as spatially close sites.
Initialize each d_i to arg max_{d∈D} C(d_i(x_i), y_i)
Run Iterated Conditional Modes (ICM) for MaxIters iterations
For t = 1...MaxIters, do
    For each x_i ∈ x, do
        For each D ∈ D, do
            If U(D|x_i, y_i) < U(d_i|x_i, y_i), then
                d_i = D
Return d
```

3. EXPERIMENTS

3.1. Data and setup

The CNN model was trained with 3134 OSM rural buildings annotations, manually verified/corrected on a set of images acquired over the Geita, Mtwara and Manyara regions of Tanzania. We used Bing aerial imagery with a spatial resolution of 30cm. For the validation data set, we manually corrected 1094 misaligned building annotations close to the city of Mugumu in Tanzania, where we found annotations with different misalignment orientations. Note that the validation region has not been used to extract any training annotation. The number of filters in the convolutional layers were 32, 64, 128 and 128, and the corresponding kernel sizes used were 7×7, 5×5, 5×5 and 3×3. After every convolution operation, we applied spatial pooling with stride 2 and kernel size 3×3. We allow the alignment vectors $D = (D_x, D_y)$ to take values in the range $\{-30, -29, ...0, ..., 29, 30\}$ (in pixels) for both D_x and D_y and set $\beta = 2$ and $MaxIters = 10$ experimentally.

3.2. Results

Table 1 shows the processing times and performances of different alignment strategies and the performance of the original misaligned annotation in terms of pixel-based precision and recall. After performing the alignment, we applied our heuristic to remove buildings that no longer exist in the imagery (See section 2) for all methods. The first observation is that all registration methods drastically improve the results obtained without alignment, thus showing the need for an alignment of OSM data. Among them, those based on the proposed MRF with pairwise connections outperform the models based only on correlation alignment, showing that alignments are smooth and that such prior knowledge is helpful to reach the best results. Finally, proceeding per groups of buildings gives both numerical improvements and computational speedups over working at the single buildings level. Note that

Table 1. Pixel-based performance of alignment correction methods.

Methods	Precision	Recall	Time (sec)
Annotation without alignment	0.108	0.115	0
Correlation alignment ($\beta = 0$) per building	0.672	0.446	61.7
MRF alignment per building	0.745	0.634	67.8
Correlation alignment ($\beta = 0$) per groups	0.775	0.625	22.2
MRF alignment per groups	**0.802**	**0.638**	23.3

the recall measure is low for all the alignment methods, since many of the annotations in the updated ground truth (that we are using for evaluation) belong to new buildings that were not present when the OSM annotations were performed (Problem 3 in the Introduction). As we did not tackle this problem in this paper, these buildings simply cannot be recovered.

Figure 4 shows three different groups of buildings, their RGB images, building probability maps and their annotations aligned with different methods. The original misaligned annotations are presented with yellow circles. Figure 4c shows with green circles the result of aligning the annotations using only the correlations computed with groups of buildings (i.e. $\beta = 0$). Figure 4d shows a more precise correction using our MRF method (blue circles). Figure 4g shows with orange circles the alignment result of the MRF using individual buildings as nodes of the MRF graph. One of the buildings is not correctly aligned because the values of the probability map are high in a location where there are no buildings: this makes the algorithm select a wrong alignment vector. This behavior does not happen in the MRF applied over a graph of groups of buildings (blue circles in Figure 4h), since we apply the prior knowledge that buildings that are spatially very close should undergo nearly the same alignment. Finally, in Figure 4i we provide an example where the MRF method using groups of buildings obtains a wrong alignment vector, since the building probability map (Figure 4j), has high values in locations where there are no buildings. This also makes fail our heuristic for removing annotations. Although the MRF based on groups of buildings alleviates this issue, the quality of the CNN probability maps estimations is still one of the main factors to obtain accurate alignment vectors.

4. CONCLUSION

We presented an MRF-based method to correct misaligned rural building annotations from OpenStreetMap. The energy function to be minimized uses the correlation of the aligned annotations with a CNN output indicating building probabilities and considers the dependency between the alignment vectors of buildings that are spatially close. We observed high quality registration of existing vector annotations on newly acquired aerial images, thus obtaining directly instance level annotations. By letting the CNN lead a registration problem instead of trying to map directly, we both addressed the issues of geometric inaccuracy of the CNN buildings map and of the large human effort required to screen and correct existing OSM annotations of rural buildings when the imagery is updated.

5. REFERENCES

[1] X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: a review," *IEEE Geosci. Remote Sens. Mag.*, in press.

[2] V. Mnih and G. E. Hinton, "Learning to label aerial images from noisy data," in *Proc. ICML*, 2012, pp. 567–574.

[3] E. Maggiori, G. Charpiat, Y. Tarabalka, and P. Alliez, "Recurrent neural networks to correct satellite image classification maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4962–4971, Sept 2017.

[4] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb 2017.

[5] J. Chen and A. Zipf, "DeepVGI: Deep Learning with Volunteered Geographic Information," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 771–772.

[6] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc.*, vol. 68, pp. 259–302, 1986.

[7] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, "Deformable medical image registration: Setting the state of the art with discrete methods," *Annual Review of Biomedical Engineering*, vol. 12, pp. 219–244, 2011.

[8] D. Marcos, R. Hamid, and D. Tuia, "Geospatial correspondences for multimodal registration," in *Proc. CVPR*, 2016.

[9] M. Volpi and D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, 2017.

[10] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proc. CVPR*, 2015, pp. 447–456.

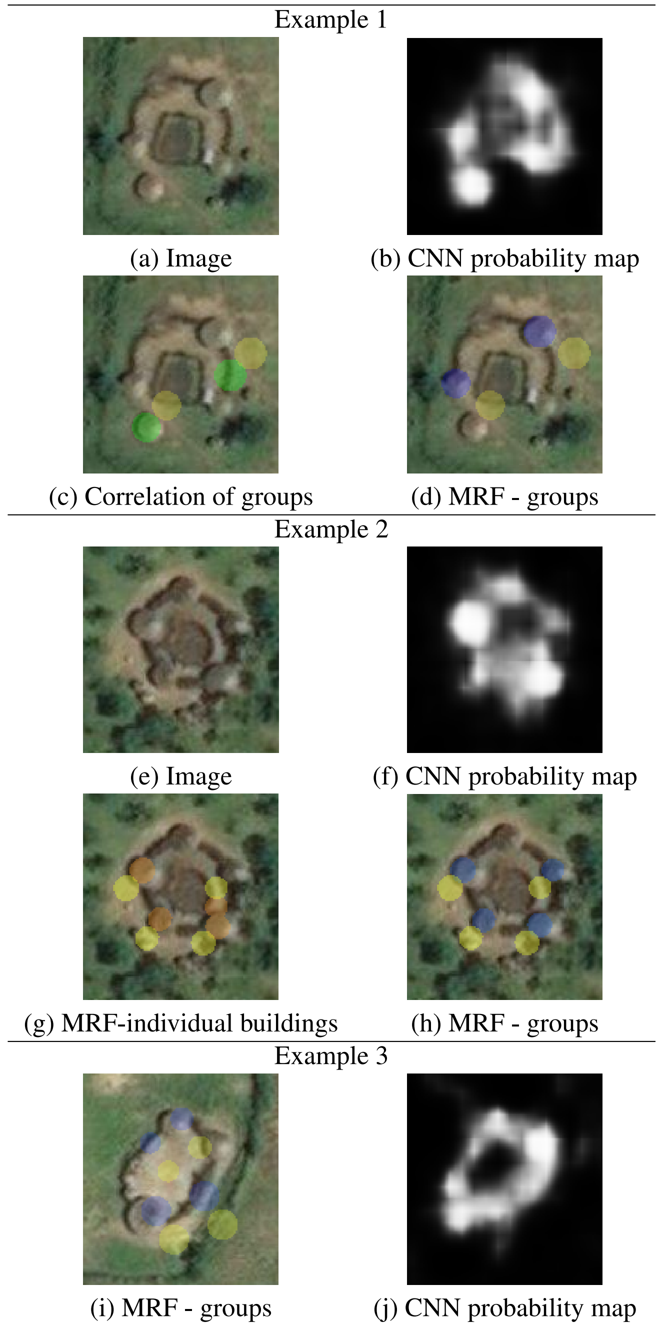

Fig. 4. Alignment result examples (the original misaligned annotations are presented as yellow circles). Example1: a-d) image (obtained from Bing maps), probability map, result using correlation of groups of buildings (green circles) and results using MRF with groups of buildings (blue circles). Example 2: e-h) image, probability map, result using MRF with individual buildings (orange circles) and result using MRF with groups of buildings. Example 3: i) A failure case where the MRF method, that uses groups of buildings, does not align well the annotations, j) probability map.