FAST BUILDING EXTRACTION BY MULTISCALE ANALYSIS OF DIGITAL SURFACE MODELS

François Merciol, Sébastien Lefèvre

Univ. Bretagne-Sud, UMR 6074, IRISA, F-56000 Vannes, France
{firstname.lastname}@irisa.fr

ABSTRACT

Building detection is a challenging issue that requires efficient solutions in many operational contexts. Hierarchical image representation through tree structures is known for its compliance with such requirements for fast methods in remote sensing. In this paper, we address the building detection problem using an underlying hierarchical image model, and rely solely on height information coming from digital surface models (DSM). The proposed method is thus very efficient and allows for interactive rule-based detection of buildings and other land use / land cover classes. Preliminary results on a standard dataset from ISPRS are provided and illustrate the relevance of such an approach.

***Index Terms*—** Hierarchical image representation, α-tree, DSM, building detection

1. INTRODUCTION

Nowadays, Earth Observation may be achieved with airborne and spatial sensors of very high spatial resolution (e.g., a few centimeters). While the level of details visible in the images allows for more accurate and precise analysis, it also results in very large datasets that can be processed only with scalable techniques. Among those, hierarchical modeling of images has recently appeared as a relevant alternative to pixel-wise analysis in order to tackle the massive amounts of remote sensing data through a multiscale representation. Various models based on tree structures have been successfully applied to problems faced in Earth Observation: α-tree for damage assessment (rubbish detection) [1] or land cover mapping [2], max-tree for interactive exploration of images in the context of human crisis management [3], binary partition tree for speckle noise reduction or object-based classification [4]).

In this paper, we propose to analyse DSM information with a hierarchical model in order to perform fast building extraction. The proposed methodology relies on a multiscale model built from the hierarchy of connected components (Sec. 2), that allows for analyzing of both local and global measurements of the height information (Sec. 3). Experiments conducted on a standard dataset lead to promising results (Sec. 4) and call for further exploration of this hierarchical strategy to efficiently process DSM data.

2. MULTISCALE IMAGE ANALYSIS WITH CONNECTED COMPONENTS

An α-tree is a multiscale representation of an image through its α-connected components (or α-CCs). While it finds roots in early work in computer vision, it has been revisited only recently by Soille and Ouzounis [5, 1]. This paradigm is very related to the single linkage procedure used in data clustering. It provides a compact representation of the image that allows its real-time processing. Furthermore, efficient algorithms have been recently introduced to ensure fast computation of this representation from complex images [6].

The concept of α-CC is an extension of the connected component (or CC). We recall that the latter is defined as a set of adjacent pixels that share exactly the same value (either scalar for panchromatic images, or vectorial for multi- or hyperspectral ones). Representing an image by its CCs allows for higher-level analysis (similarly to computer vision techniques relying on superpixels). However, the possibly great number of CCs in an image prevents their practical use. Indeed, adjacent pixels may belong to the same structure but have slightly different values, thus belonging to different CCs. The concept of α-CC has been introduced to allow such slight variations, leading to the following definition: an α-CC is a set of adjacent pixels that share similar values i.e., values with a dissimilarity lower or equal to a threshold α. The α-CC of a pixel p will thus contain all pixels q that can be reached with a path over neighboring pixels p_i $\langle p_1 = p, \ldots, p_n = q\rangle$ from p to q such as $d(p_i, p_{i+1}) \leq \alpha$ where d is the dissimilarity measure in use. The complexity and number of α-CCs are directly related to α. It allows one to build a hierarchical representation of an image, and to perform subsequent multiscale analysis (e.g., in an object-oriented strategy). This representation is called an α-tree. Each level of the tree is indexed by an α value, and its nodes are the corresponding α-CCs. A leaf in the tree is a 0-CC i.e., a standard CC in the image. Increasing α leads to the connection of α-CCs, resulting in the creation of higher nodes in the tree, until the root that contains the whole image.

The chaining effect is a well-known issue for the single-linkage paradigm. Since the α-tree is built on this paradigm, it is naturally affected by this problem. To explain, let us consider a series of neighboring pixels with values $u_1, \ldots, u_n$. If $d(u_i, u_{i+1}) \leq \alpha$ for all pairs $(i, i+1)$, these pixels will belong to the same α-CC. This remark holds even for a small value of α, values of neighboring pixels being then always very similar. However, the difference between any pair of pixels within the α-CC (e.g., $d(u_1, u_n)$) might be much larger than α. In this case, pixels with very distinct values are gathered in a unique α-CC assumed to be of low complexity (i.e. low α value). Such a situation is frequently observed on transition regions between objects of different spectral signatures. To counter this chaining effect, Soille has introduced the paradigm of constrained connectivity [5]. It consists in imposing a set of constraints to be fulfilled by the α-CCs. The most representative example is to apply a threshold ω on the global range of the α-CC i.e., $d(p, q) \leq \omega$ for all pairs (p, q) of pixel values belonging to this α-CC. Such a criterion leads to the definition of the (α, ω)-CCs. Computing this global range is efficiently achieved in the greyscale case by storing the minimum and maximum values of each α-CC, and updating these extrema when merging two α-CCs in the α-tree construction process.

3. APPLICATION TO BUILDING EXTRACTION FROM DSM DATA

We propose here to build the α-tree model from a digital surface model (DSM), where each pixel is characterized by its height (instead of its spectral signature in multispectral optical images). An α-CC thus gathers neighboring pixels sharing locally similar height, the similarity level being defined by the α value. The leaves of the α-tree are the flat regions observed in the DSM, while their ancestors in the tree represent regions with increasing gaps between heights of adjacent pixels. It is thus possible to distinguish between flat regions (building with flat roots, roads on a flat land, etc.) and non-flat ones (e.g., buildings with pitched roofs, trees, etc.). However, due to the pure local behavior of the α-tree, regions with small local variations in the DSM can be misclassified in pitched roofs. This issue can be tackled by using the global range constraint ω that considers the maximal height difference within each α-CC. (α, ω)-CCs with respectively low, average, and high dissimilarity value will thus correspond to roads or flat roofs, pitched roofs, and finally trees or steeples.

When a slope occurs on the ground, a flat object such as a road may appear non-flat on the DSM. It is important to analyse the height information of each pixel (or by extension CC) w.r.t. their neighborhood. To do so and prevent misclassification, we add some ancillary data to each α-CC or node in the α-tree. More precisely, and given a predefined neighborhood size, we compute a grid that provides the lowest height in the neighborhood of each pixel. In the tree representation, each single node thus comes with its slope or local range α, its global range ω, its size, its height (measured in its center pixel) and the minimal elevation in the neighborhood.

Once the tree has been built and each node equipped with its various attributes, the classification process takes place. It consists in scanning the tree from the root to the leaves, using predefined sets of criteria (see Sec. 4). Each node is then compared with the different sets of criteria under consideration. If it fits a given set of criteria, the node (and all its children) is assigned to the corresponding class. Otherwise the tree is further explored by analyzing the children of the remaining nodes. This top-down and rule-based approach allows for both efficient and robust extraction of objects, since significant objects (i.e., large size and global range) are analyzed first. Furthermore, the classification process can be further optimized by considering several tree scans. Indeed, a first top-down analysis of the tree can be conducted to extract the large objects (e.g., ground, buildings), while the remaining nodes are scanned in a second pass whose goal is to extract small objects (e.g., cars, trees).

4. EXPERIMENTS

The proposed strategy has been evaluated on a standard dataset containing DSM data, namely the Vaihingen dataset provided by ISPRS WG III/4 within the scope of the 2D Semantic Labeling Contest (ISPRS benchmark on urban object detection and 3D building reconstruction [7]). The dataset contains 33 patches (image tiles), each made of both a true orthophoto (8 bit color image with near infrared, red and green bands) and a DSM (32 bit floating values). Among them, 16 are provided with some ground truth under the form of a reference map where each pixel is assigned a label among 6 classes (impervious surfaces, buildings, low vegetation, trees, cars, and clutter/background), but for one of them the ground truth is not matching the DSM file (so it is discarded here). Size of these tiles varies from 2336×1281 (ca. 3 MPixels) to 2818×2558 (ca. 7 MPixels). Figure 4 shows two sample tiles and the corresponding ground truth, using the following color code: impervious surfaces in white, buildings in blue, low vegetation in cyan, trees in green, cars in yellow, and clutter/background in red.

We consider here only the 15 tiles coming with a correct ground truth (i.e. 73 MPpixels in total) and allowing a straightforward evaluation of classification accuracy. Such an evaluation relies on the F1-score measured pixelwise, following the evaluation protocol proposed with the contest. We recall that the F1-score is the harmonic mean of recall and precision, accounting for the ratio between true positive (correct detections) and false positive and false negative (errors). Table 4 summarizes the results obtained with the proposed approach considering a straightforward parameter setting / rule definition step.

Conversely to previous experiments conducted with this

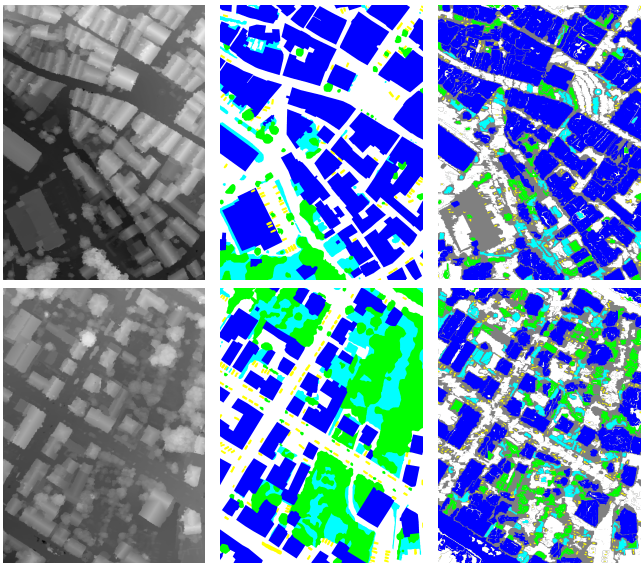

Fig. 1. Sample tiles *area 5* (top) and *area 7* (bottom). From left to right: greyscale DSM image, reference map (ground truth), and result provided by our method (unclassified pixels are shown in grey).

Tile No.	I	B	L	T	C
area 1	0.46	0.65	0.12	0.21	0.12
area 3	0.51	0.64	0.12	0.19	0.11
area 5	0.55	0.76	0.08	0.13	0.16
area 7	0.48	0.57	0.11	0.25	0.13
area 13	0.42	0.43	0.14	0.14	0.04
area 15	0.43	0.46	0.20	0.19	0.07
area 17	0.33	0.59	0.14	0.12	0.01
area 21	0.54	0.45	0.15	0.21	0.08
area 23	0.38	0.52	0.28	0.26	0.04
area 26	0.53	0.57	0.08	0.18	0.13
area 28	0.53	0.58	0.21	0.24	0.11
area 30	0.57	0.61	0.20	0.23	0.12
area 32	0.72	0.59	0.06	0.23	0.19
area 34	0.46	0.67	0.09	0.24	0.08
area 37	0.36	0.65	0.09	0.18	0.03
average	0.51	0.58	0.14	0.20	0.09

Table 1. Classification accuracy (F1-scores) of the proposed method for all evaluated tiles and classes: (I)mpervious surfaces, (B)uildings, (L)ow vegetation, (T)rees, (C)ars.

dataset (and reported on the contest website), we use here only the DSM file. Thus we do not exploit any color / spectral information, while such information is crucial (for instance to distinguish between tree / vegetation and other classes). It is then not surprising that the method performs poorly with some of the classes, as shown in Tab. 4 and the last column of Fig. 4. While the buildings (blue) are correctly detected, the low vegetation and trees cannot be handled satisfyingly. Let us note however that these presented results are preliminary, and the combination of DSM and TOP images within a unique hierarchical model is an objective of future research.

As described in Sec. 3, the classification procedure scans the tree and assigns each node to a class following some predefined rules. These rules simply consists in checking if the node properties fall into a predefined interval (one per class and per property). The full list of parameters used to process the dataset is given in Tab. 4. Object size is measured in dm^2. The grid size is used to identify pixels from which the lowest height is locally computed. Slope is measured in m/dm.

The role of parameter tuning is also crucial for the method to perform well. To illustrate, we provide in Fig. 4 another sample from the Vaihingen dataset (not from the 2D Semantic Labeling Contest). As we can see, the proposed method can lead to accurate classification between buildings, roads, trees, cars, and steeples even when relying only on the DSM data but considering several subclasses (or parameter sets). In this experiment, a postprocessing has been involved to assign a class to all unclassified pixels. Furthermore, this process is achieved very efficiently since the complexity of the whole process is linear w.r.t. the number of pixels. Indeed, computing the tree and its attributes from a 400 MPixels image requires less than one minute on a standard modern computer. The following steps are achieved even more efficiently thanks to the tree structure.

Efficiency of the proposed approach is indeed one of its main advantages. The most time-consuming step is the construction of the tree that can be built offline (40s for each tile *area X*). Computing height and ancillary properties for each node is relatively straightforward (5s) and can also be done offline. The last step, i.e. the two-pass top-down classification procedure based on predefined rules is very fast (3s) and can be conducted interactively (allowing the user to modify the rules by tuning the parameters before launching a new classification run). Furthermore, let us note that these CPU time include the image display cost, and have been obtained with a Java implementation using a standard laptop.

5. CONCLUSION

This paper aimed to explore how a hierarchical image representation (namely the α-tree model) can be used to derive

class	object size	grid size	minimal elevation	slope	height
Impervious surfaces	$[1000, +\infty]$	200	$[-0.2, 0.5]$	$[0, 1]$	$[0, 2]$
Buildings	$[3000, 200000]$	400	$[5, 25]$	$[0.03, 15]$	$[1, 25]$
Low vegetation	$[1000, 7000]$	400	$[0.5, 3]$	$[0, 1]$	$[0, 3]$
Trees	$[1000, 7000]$	200	$[1, 4]$	$[0, 10]$	$[0, 25]$
Cars	$[200, 1000]$	200	$[0.5, 0.25]$	$[0, 2]$	$[0, 1.5]$

Table 2. Parameter settings used with the ISPRS dataset.

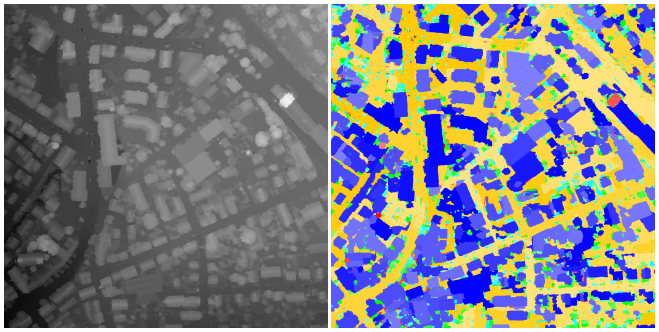

Fig. 2. Sample DSM image of 4000×4000 pixels (left) and corresponding classification map (right) obtained with the proposed method on a sample DSM from the Vaihingen dataset. The color correspondence table is the following: roads (yellow/orange), buildings (blue), steeples (red), cars (cyan), trees (green). Each shade corresponds to a set of selection criteria.

a multiscale analysis of DSM data. We illustrate the potential of this approach with experiments on building extraction from DSM, that can be achieved very efficiently using the proposed method. While promising, the preliminary results obtained with a challenging dataset made recently available call for several improvements.

First, DSM itself does not allow to perform accurate classification. It has to be coupled with color/spectral information from orthophoto. Combining such complementary information within a single tree structure remains challenging. Besides, more advanced learning techniques can be involved to improve recognition performance. Indeed, existing attempts to address the considered dataset rely on advanced strategies such as deep learning while our method simply consist in a set of rules (comparison of objects properties with predefined thresholds). A stronger reasoning process will definitely allow to achieve better results.

Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-13-JS02-0005-01 (Asterix project).

The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [7]: `http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html`. The authors would like to thanks Markus Gerke for fruitful discussions regarding the ISPRS 2D Semantic Labeling Contest.

6. REFERENCES

[1] G. K. Ouzounis and P. Soille, "Pattern spectra from partition pyramids and hierarchies," in *International Symposium on Mathematical Morphology*, Verbania-Intra, Italy, 2011, pp. 108–119.

[2] S. Lefèvre, L. Chapel, and F. Merciol, "Hyperspectral image classification from multiscale description with constrained connectivity and metric learning," in *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014.

[3] G. Ouzounis and L. Gueguen, "Interactive collection of training samples from the max-tree structure," in *IEEE International Conference on Image Processing*, 2011, pp. 1449–1452.

[4] A. Alonso-Gonzalez, S. Valero, J. Chanussot, C. Lopez-Martinez, and P. Salembier, "Processing multidimensional sar and hyperspectral images with binary partition tree," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 723–747, 2013.

[5] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1132–1145, 2008.

[6] J. Havel, F. Merciol, and S. Lefèvre, "Efficient schemes for computing α-tree representations," in *International Symposium on Mathematical Morphology*. 2013, vol. 7883 of *LNCS*, pp. 111–122, Springer.

[7] M. Cramer, "The dgpf test on digital aerial camera evaluation – overview and test design," *Photogrammetrie – Fernerkundung – Geoinformation*, vol. 2, pp. 73–82, 2010.