Watershed-Based Attribute Profiles With Semantic Prior Knowledge for Remote Sensing Image Analysis

Deise Santana Maia, Minh-Tan Pham, Sébastien Lefèvre

Watershed-Based Attribute Profiles With Semantic Prior Knowledge for Remote Sensing Image Analysis

Deise Santana Maia, Minh-Tan Pham, and Sébastien Lefèvre, *Senior Member, IEEE*

***Abstract*—In this article, we develop a novel feature extraction method that combines two well-established mathematical morphology concepts: watersheds and morphological attribute profiles (APs). In order to extract spatial-spectral features from remote sensing data, APs were originally defined as sequences of filtering operators on inclusion trees, i.e., the max- and min-trees, computed from the input image. In this study, we extend the AP paradigm to the more general framework of hierarchical watersheds. Moreover, we explore the semantic knowledge provided by labeled training pixels during different phases of the watershed-AP construction, namely within the construction of hierarchical watersheds from the raw image and later within the filtering of the resulting hierarchy. We illustrate the relevance of the proposed method with two applications including land cover classification and building extraction using optical remote sensing images. Experimental results show that the new profiles outperform various existing features using two public datasets (Zurich and Vaihingen), thus providing another high potential feature extraction method within the AP family.**

***Index Terms*—Attribute profiles (APs), building extraction, classification, remote sensing, watershed.**

I. Introduction

MATHEMATICAL morphology is an efficient tool that has a long history within the analysis and processing of remote sensing images, as attested by early surveys on this topic [1], [2]. Since the last decade, a special attention has been particularly given to a multilevel feature extraction method namely morphological attribute profiles (AP) [3], which appeared and mostly replaced the classical morphological profiles (MPs) [4] for the analysis of remote sensing images. Even though both of these methods are successful in conveying spatial-spectral features of those image data, APs have been proved to be more generalized and efficiently scalable to deal with large-scale data, thanks to their construction from tree-based hierarchical representation [5]. In this article, we study the relevance of hierarchical watersheds integrated in an AP processing framework for remote sensing applications. Watershed segmentation was proposed in the late 70's and, since then, this concept has been extended to several frameworks and implemented through a variety of algorithms. The intuition behind the various definitions of the watershed segmentation derives from the topographic definition of watersheds: dividing lines between catchment basins, which are, in their turn, areas where collected precipitation flows into the same regional minimum. These notions can be extended to gray-scale images and graphs, leading to different definitions of watershed segmentation. In this article, we focus on watershed-cuts and hierarchical watersheds defined in the context of edge-weighted graphs, as formalized in [6] and [7]. We present the notions of graphs, hierarchies of partitions, and hierarchical watersheds in Section III-A.

In addition to the construction of watershed-based APs, our second contribution in this work is to exploit semantic prior knowledge during different phases of the Watershed-AP construction so that the final extracted profiles could better encode and characterize the spatial-spectral multilevel features from the image. In the context of image segmentation, a widespread method to introduce prior knowledge in the results is to consider user-defined markers, which are subsets of image pixels indicating the locations of objects of interest. Such markers guide the segmentation algorithm and assure that the objects of interest are segmented into distinct regions. The notion of markers has been especially explored in watershed segmentation, in which catchment basins are grown from input markers instead of the regional minima of an image [8]. We provide more background notions and related studies of how semantic prior knowledge is exploited in watershed segmentation in Section II, as well as our proposed strategies to incorporate prior knowledge into watershed-based APs in Section IV-B.

It should be noted that this manuscript is an extension of our recent conference paper [9], which has provided preliminary results of watershed-APs applied to panchromatic remote sensing image classification. In the present article, we investigate other methods to compute watershed-APs with prior knowledge from training pixels and we present a more extensive evaluation of the proposed methods in land-cover classification. Moreover, considering that other methods in the literature, such as [10], [11], work well on binary pixel classification/segmentation of object/background, we decided to validate our method through another binary classification task, which is relevant for remote sensing imagery, namely building extraction. Both land-cover classification and building extraction are evaluated using multispectral images from two publicly well-known datasets: Zurich and Vaihingen.

Manuscript received October 2, 2021; revised February 5, 2022; accepted February 8, 2022. Date of publication February 23, 2022; date of current version April 6, 2022. This work was supported by the ANR Multiscale project under Grant ANR-18-CE23-0022. *(Corresponding authors: Deise Santana Maia; Minh-Tan Pham; Sébastien Lefèvre.)*

Deise Santana Maia is with the Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL), UMR 9189, Université de Lille, F-59000 Lille, France (e-mail: deise.santanamaia@univ-lille.fr).

Minh-Tan Pham and Sébastien Lefèvre are with the Institut de recherche en informatique et systèmes aléatoires (IRISA), UMR 6074, Université Bretagne Sud, F-56000 Vannes, France (e-mail: minh-tan.pham@irisa.fr; sebastien.lefevre@irisa.fr).

Digital Object Identifier 10.1109/JSTARS.2022.3153110

The rest of this article is organized as follows. We first review related works on the use of semantic prior knowledge in Section II. Then, we present some basic definitions of graphs, hierarchical watersheds, and APs in Section III. In Section IV, we introduce the proposed watershed-APs and our strategies to integrate semantic knowledge within its construction. Finally, experiments conducted on the two aforementioned remote sensing datasets are given in Section V with extensive analysis and discussion in Section VI. Finally, Section VII concludes this article.

II. Related Works

In this section, we review some related studies on the use of prior knowledge and markers in the field of image processing and analysis, with a focus on watershed segmentation. We present the main findings of each study and discuss how they relate to our proposed watershed-AP.

As already mentioned, the idea behind the watershed segmentation is to partition a surface/image into its different regional minima. To prevent oversegmentation, markers are often used to guide the computation of the watershed segmentation, in which catchment basins are grown from input markers instead of the image's regional minima. In this article, we show that the use of markers in a watershed segmentation can go beyond the introduction of new regional minima. The spectral-spatial information regarding the marked pixels can provide further knowledge about the objects we aim to segment. For instance, in [12], [13], the authors use the spectral signature of training samples in the construction of watershed segmentation of multispectral images. First, the spectral signature of training pixels is used to train a classifier, which is then applied on the whole image and used to obtain a probability map per class. Then, those maps are combined and used to obtain a single watershed segmentation. In the present work, we consider a similar approach to include prior-knowledge in the watershed-AP construction, with the main difference being the construction of a multilevel watershed segmentation instead of a single segmentation. Moreover, we go one step further to exploit such probability maps during the image reconstruction/filtering phase of the watershed-AP construction.

Another use of supervised classification for watersheds has been proposed in [10], where the watershed segmentation is computed from user-defined markers combined with probability maps computed for each targeted class. More precisely, catchment basins are grown from different markers, and the probability maps, combined with the original data, are used simultaneously in the process. In remote sensing, the later approach has been applied to the detection of buildings [14] and shorelines [15] in multispectral images. Similarly to [12] and [13], the method proposed in [10] deals with single level watershed segmentation. Finally, in [11], prior knowledge from markers is employed on several interactive image segmentation methods, including watersheds, in the framework of edge-weighted graphs. Edge weights are defined as a linear combination of the weights obtained from two sources: from the pixel values, and from the classification probability maps computed from the markers that are incrementally provided by the users. More generally, knowledge from markers can be used by other kinds of preprocessing methods beyond watershed segmentation. Namely, spectral signatures of training pixels have been used in [16] to optimize the data preprocessing with alternating sequential filters. Hence, training pixels are used for preprocessing the input data, as well as for the final pixel classification. A related approach is proposed in [17], where training pixels are used to optimize vector orderings for morphological operations applied to hyperspectral images.

In the context of hierarchical segmentation, prior knowledge can play a role in defining which regions should be highlighted at different levels of a hierarchy. In [18], a marker-based hierarchical segmentation is proposed for hyperspectral image classification. Labeled markers are derived from a probability classification map, which is obtained from training samples, as done in [10], [12], [13]. Then, those labeled markers guide the construction of a hierarchical segmentation by preventing regions of different classes to be merged, and by propagating the labeled markers to unlabeled regions. Another related approach, proposed in [19], uses prior knowledge to keep the regions of interest from being merged early in the hierarchy, *i.e.*, the details in the regions of interest are preserved at high levels of the hierarchy. This later idea is also explored in the watershed-AP framework, in which we aim to filter out the regions with low probability of belonging to a given ground-truth semantic class. Finally, in [20], the authors propose a knowledge-based hierarchical representation for hyperspectral images. In their approach, a dissimilarity measure learned from training pixels is employed in the construction of α-trees. This last approach share some common features with the one proposed in the present paper, with the main differences being the family of hierarchy under consideration as well as the learning algorithm used to explore prior knowledge from training pixels. Moreover, we also consider another way of using such prior knowledge which has not been considered in [20], namely in the filtering step rather than in he construction phase of a hierarchy.

Finally, following the current tendency of using deep learning for solving computer vision problems, various methods for coupling prior knowledge with deep learning based models have been explored in more recent works. For instance, in [21], edge information is combined with the output of a Fully convolutional network (FCN) in order to refine the segmentation results given by the later. And, in [22], crop classification in SAR time series are performed using autoencoders (AE), convolutional neural networks (CNN), and FCN, and then postprocessed using prior knowledge regarding crop dynamics, *i.e.* expert's knowledge about which crop transitions might or not occur over time in the same field. Though deep learning models perform well in computer vision tasks in general, including remote sensing imagery, there are advantages of using feature extraction methods based on hierarchical representations and morphological operators, such as the proposed watershed-APs. In particular, we can mention the interpretability of the method when compared to the black box parameters of deep learning models, and the low need for lots of ground-truth data, which makes those morphological methods well adapted to datasets with scarce annotated samples.

III. Background Notions

In this section, we present some basic notions of graphs and hierarchical watersheds. Then, we recall the definition of morphological attribute profiles and their extensions in the literature.

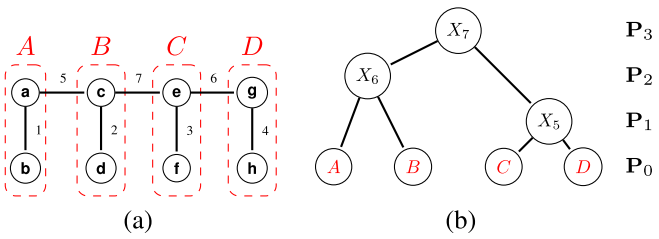

Fig. 1. (a) A weighted graph $\mathcal{G} = (V, E, w)$. (b) A tree representation of the hierarchical watershed $\mathcal{H}$ of $\mathcal{G}$ for the sequence (C, A, B, D) of minima of $\mathcal{G}$. (a) $\mathcal{G} = (V, E, w)$. (b) $\mathcal{H}$.

A. Graphs and Hierarchical Watersheds

A weighted graph is a triplet $\mathcal{G} = (V, E, w)$, where V is a finite set, E is a subset of $V \times V$, and w is a map from E into $\mathbb{R}$. The elements of V and E are called *vertices* and *edges* (of $\mathcal{G}$), respectively. Let $\mathcal{G} = (V, E, w)$ be a weighted graph and let $\mathcal{G}' = (V,' E,' w)$ be a graph such that $V' \subseteq V$ and $E' \subseteq E$. We say that $\mathcal{G}'$ *is a subgraph of* $\mathcal{G}$. A sequence $\pi = (x_0, \ldots, x_n)$ of vertices in V' is a *path (in* $\mathcal{G}'$*) from* x_0 *to* x_n if $\{x_{i-1}, x_i\}$ is an edge of $\mathcal{G}'$ for any $1 \leq i \leq n$. If $x_0 = x_n$ and if there are no repeated edges in π, we say that π *is a cycle* (in $\mathcal{G}'$). The subgraph $\mathcal{G}'$ of $\mathcal{G}$ is said to be *connected* if, for any x and x' in V', there exists a path from x to x'. Let $\mathcal{G}'$ be a connected subgraph of $\mathcal{G}$. We say that $\mathcal{G}'$ *is a connected component of* $\mathcal{G}$ if

1) for any x and x' in V', if $\{x, x'\} \in E$ then $\{x, x'\} \in E'$; and
2) there is no edge $e = \{y, y'\} \in E$ such that $y \in V \setminus V'$ and $y' \in V'$.

Let $\mathcal{G} = (V, E, w)$ be a graph and let $\mathcal{G}' = (V,' E,' w)$ be a connected subgraph of $\mathcal{G}$. If the weight of any edge in E' is equal to a constant k and if $w(e) > k$ for any edge $e = \{x, y\}$ such that $x \in V'$ and $y \in V \setminus V'$, then $\mathcal{G}'$ *is a (local) minimum of* $\mathcal{G}$. For instance, Fig. 1(a) illustrates a weighted graph with four minima delimited by the dashed lines. We note that in the remainder of this section, $\mathcal{G} = (V, E, w)$ denotes a connected weighted graph and n denotes the number of minima of $\mathcal{G}$.

Let $\mathcal{G}' = (V,' E,' w)$ be a subgraph of $\mathcal{G}$. A *Minimum Spanning Forest (MSF) of* $\mathcal{G}$ *rooted in* $\mathcal{G}'$ is a subgraph $\mathcal{G}'' = (V, E,'' w)$ of $\mathcal{G}$ such that

1) for every connected component X'' of $\mathcal{G}''$, there is exactly one connected component X' of $\mathcal{G}'$ such that X' is a subgraph of X'';
2) every cycle in $\mathcal{G}''$ is a cycle in $\mathcal{G}'$; and
3) $\sum_{e \in E''} w(e)$ is minimal among all graphs which satisfy conditions (1) and (2).

A *partition of* V is a set $\mathbf{P}$ of disjoint subsets of V such that the union of the elements in $\mathbf{P}$ is V. The *partition of* V *induced by a graph* $\mathcal{G}'$ is the partition $\mathbf{P}$ such that every element of $\mathbf{P}$ is the set of vertices of a connected component of $\mathcal{G}'$. A *hierarchy of partitions of* V is a sequence $\mathcal{H} = (\mathbf{P}_0, \ldots, \mathbf{P}_n)$ of partitions of V such that $\mathbf{P}_n = \{V\}$ and such that, for any $0 < i \leq n$, every element of $\mathbf{P}_i$ is the union of elements of $\mathbf{P}_{i-1}$. Any hierarchy of partitions $\mathcal{H}$ can be represented as a tree whose vertices correspond to the regions of $\mathcal{H}$ and whose edges link nested regions. For instance, Fig. 1(b) shows a tree representation of the hierarchy $\mathcal{H} = (\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$, where $\mathbf{P}_0 = \{\{a, b\}, \{c, d\}, \{e, f\}, \{g, h\}\}$, $\mathbf{P}_1 = \{\{a, b\}, \{c, d\}, \{e, f, g, h\}\}$, $\mathbf{P}_2 = \{\{a, b, c, d\}, \{e, f, g, h\}\}$, and $\mathbf{P}_3 = \{\{a, b, c, d, e, f, g, h\}\}$.

Let $\mathcal{S} = (\mathcal{M}_1, \ldots, \mathcal{M}_n)$ be a sequence of n distinct minima of $\mathcal{G}$ such that, for any $0 < i \leq n$, we have $\mathcal{M}_i = (V_i, E_i, w)$. The *hierarchy of MSFs of* $\mathcal{G}$ *for* $\mathcal{S}$, also known as *hierarchical watershed of* $\mathcal{G}$ *for* $\mathcal{S}$, is a hierarchy $\mathcal{H} = (\mathbf{P}_1, \ldots, \mathbf{P}_n)$ of partitions of V such that each partition $\mathbf{P}_i$ is the partition induced by the MSF of $\mathcal{G}$ rooted in the graph $(\bigcup_{j \geq i} V_j, \bigcup_{j \geq i} E_j, w)$.

A hierarchical watershed of the graph $\mathcal{G}$ of Fig. 1(a) for the sequence (C, A, B, D) of minima of $\mathcal{G}$ is illustrated in Fig. 1(b).

B. Attribute Profiles

In remote sensing image analysis, morphological APs [3] appears to be one of the most efficient multilevel feature extraction methods. To convey spatial-spectral features of remote sensing images, APs were initially defined as sequences of filtering operators on the max- and min-trees computed from the original data. Let $X : P \to \mathbb{Z}, P \subseteq \mathbb{Z}^2$ be a gray-scale image. The calculation of APs on X is achieved by applying a sequence of attribute filters based on a min-tree (i.e., attribute thickening operators $\{\phi_k^A\}_{k=1}^K$) and on a max-tree (i.e., attribute thinning operators $\{\gamma_k^A\}_{k=1}^K$) as follows:

$$AP(X) = \Big\{ \phi_K^A(X), \phi_{K-1}^A(X), \ldots, \phi_1^A(X), X \\ \gamma_1^A(X), \ldots, \gamma_{K-1}^A(X), \gamma_K^A(X) \Big\} \quad (1)$$

where ϕ_k^A and γ_k^A are, respectively, the thickening and thinning operators with respect to the attribute A and to the threshold k, and K is the number of selected thresholds. More precisely, the thickening $\phi_k^A(X)$ of X (resp. thinning $\gamma_k^A(X)$ of X) with respect to an attribute A and to a threshold k is obtained as follows: given the min-tree T (resp. max-tree T) of X, the A attribute values (e.g., area, circularity, and contrast) of the nodes of T are computed. If the attribute A is increasing, the nodes whose attribute values are inferior to k are pruned from the tree T; otherwise other pruning strategies can be adopted [5]. Finally, the resulting image is reconstructed by projecting the gray levels of the remaining nodes from the pruned tree into the pixels of X.

Since its appearance, the notion of APs has been extended to other hierarchical representations including tree-of-shapes and partition trees such as α-tree and ω-tree (see a comparative study of AP constructed from different trees in [23], and [24] for a more general survey on morphological trees). To obtain a profile from a partition tree instead of a component tree, some adaptations have to be made to the original definition of APs, as discussed in [25] and [26]. For instance, the nodes of a partition tree are not naturally associated to gray-level values, as it is the case of component trees. The strategy adopted in [25] is to represent each node as its level in the tree or as the maximum, minimum, or average gray-level of the leaf nodes (pixels) of this node. For

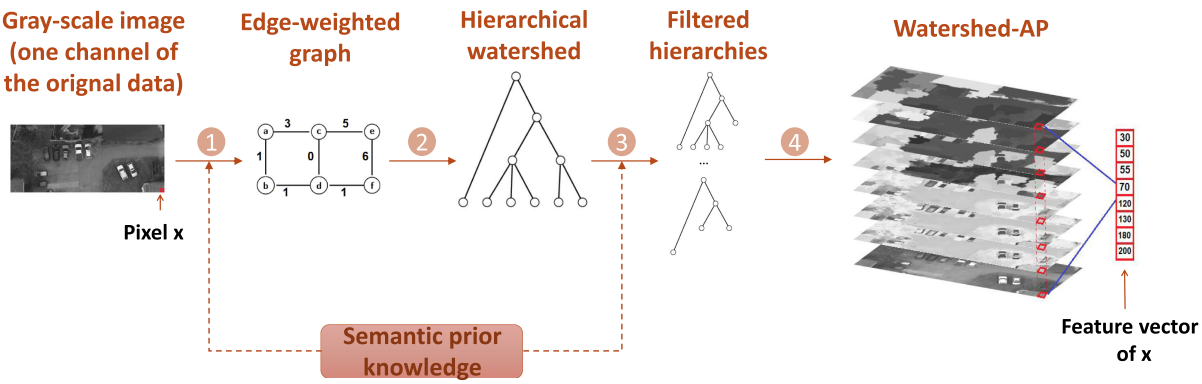

Fig. 2. General framework to compute Watershed-APs. Here, the solid arrows represent mandatory steps while the dashed arrows indicate optional steps to include prior knowledge. Details are provided in Section IV-B.

more details about APs' extensions, we invite readers to refer to a recent survey [5].

IV. Watershed AP

In this section, we present watershed-APs [9] and discuss different methods for introducing semantic prior knowledge during the generation and post-processing of hierarchical watersheds.

A. Watershed-Based APs

As mentioned in the previous section (see Section III-A), hierarchies of partitions, such as hierarchical watersheds, can be equally represented as a (partition) tree. Hence, the filtering strategy of watershed-APs is similar to the strategy described in [25] for the α- and ω-APs: if a node is filtered out, all of its descendants are also removed from the tree. As discussed in [25], image reconstruction from partition trees is not straightforward as it is from component trees. For node representation, we adopt one of the solutions proposed in [25] and already mentioned in Section III-B, in which a node is represented by the average gray-level of the pixels belonging to it. We highlight that, in the case of multichannel images, the average grey level computed on each band might lead to spectral values not present in the input image. However, in the context of APs used for pixel classification, our aim is not image filtering. Hence, the fact that new spectral values (a.k.a. false colors) are created is not a problem as long as they allow us to distinguish between different semantic classes.

Note that hierarchical watersheds are usually constructed from a gradient of the original image, which contains more information about the contours between salient regions than about the spectral signature of those regions. Hence, we consider the original pixel values to obtain the nodes representation instead of the image gradient.

Formally, let $X : P \to \mathbb{Z}$ be a gray-scale image and let $\mathcal{G} = (V, E, w)$ be a weighted graph, which represents a gradient of X, *i.e.*, $V = P$ and, for every edge $e = \{x, y\}$ in E, the weight $w(e)$ represents the dissimilarity between x and y (*e.g.*, $w(e) = |X(x) - X(y)|$). Let $\mathcal{S}$ be a sequence of minima of $\mathcal{G}$ ordered according to a given criterion C, and let $\mathcal{H}$ be the hierarchical watershed of $\mathcal{G}$ for the sequence $\mathcal{S}$. Given the tree representation $\mathcal{T}$ of $\mathcal{H}$, a *watershed-AP of X for the criterion C* is constructed as a sequence of image reconstructions from filtered versions of $\mathcal{T}$.

Fig. 2 summarizes the construction of the watershed-APs for a given gray-scale image I. The solid arrows represent mandatory steps for the watershed-AP construction, while the dashed arrows indicate optional steps to include prior knowledge, which will be discussed in the following section.

B. Watershed-APs With Semantic Prior Knowledge

As discussed in Section II, user-defined markers and prior knowledge from training pixels are valuable tools for optimizing the construction of various image representations, such as hierarchical segmentations, as well as for postprocessing such representations. In this article, we investigate the use of prior knowledge in the construction of watershed-APs. We present a general framework for including prior knowledge at different stages of the watershed-APs construction, followed by two instances of this framework that are later validated on multispectral remote sensing datasets.

AP and its variants are essentially unsupervised feature extraction methods, in which only the spectral values and the relative position between pixels are taken into consideration. During the filtering and reconstruction steps of APs, low levels of prior knowledge regarding the shape and size of the objects of interest may be taken into account, but semantic knowledge from training pixels remains little exploited. In this context, we have identified two ways to incorporate prior knowledge into watershed-APs.

1) Hierarchical Watersheds With Semantic Prior Knowledge: In general, a hierarchical segmentation is a satisfactory representation of an image for a given task when the lowest level of the hierarchy contains all regions of interest for this specific task, and when regions are merged in a meaningful way, *i.e.*, similar pairs of regions are merged before dissimilar pairs. In the context of feature extraction with watershed-APs, regions of interest are composed of neighbouring pixels belonging to the same semantic class. Due to the interclass similarity and intraclass variability of pixel in remote sensing images, hierarchical segmentations based only on pixel values often do not reflect

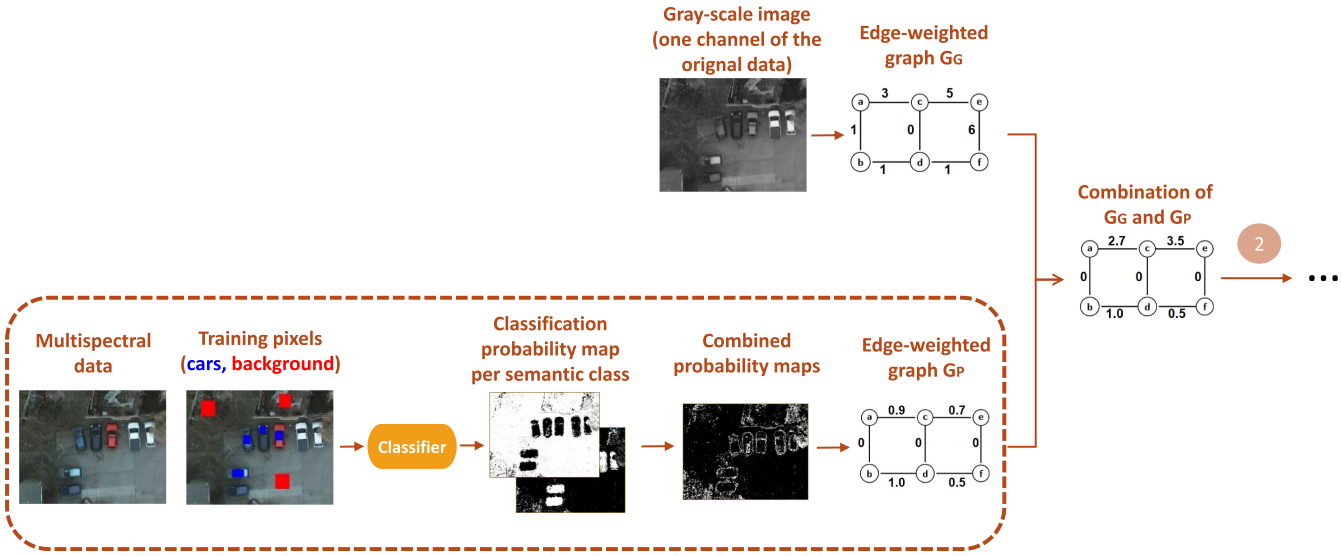

Fig. 3. Semantic prior knowledge combined with the image gradient before the construction of a hierarchical watershed. Given a multispectral image I, a classifier trained on a subset of pixels of I provides probability maps per semantic class. Then, those maps are combined into a single map which is used to obtain an edge weighted graph $\mathcal{G}_P$. The graph $\mathcal{G}_P$ is then combined with the image gradient $\mathcal{G}_G$ for each channel of I. The resulting graph is given to the step 2 of the watershed-AP pipeline presented in Fig. 2.

the inner semantic structure of the data. With that in mind, we rely on training pixels to obtain hierarchical watersheds whose regions partially reflect the semantic structure of the input data. In practice, we aim to *enforce* regional minima at the regions with high probability of belonging to any given ground-truth class. This is done through a combination of the methods proposed in [10], [12], [13], as described below.

Given a dataset I (*e.g.,* a panchromatic or an RGB image) and its training set composed of n classes, we compute its hierarchical watershed using prior knowledge as follows.

i) Training of a classifier using the training set of I and computation of perpixel class probabilities $(p_1, \ldots, p_n)$ for all pixels of I.
ii) Combination the class probabilities into a single map μ.
iii) Computation of a weighted graph $\mathcal{G}_P = (V, E, w_P)$ from μ.
iv) Computation of a weighted graph $\mathcal{G}_G = (V, E, w_G)$, which represents a gradient of I, *i.e.*, edge weights indicate the dissimilarity between neighboring pixels.
v) Combination of the weight maps w_P and w_G into a map w_{GP}.
vi) Computation of the hierarchical watershed of $\mathcal{G}_{GP} = (V, E, w_{GP})$ for a given sequence of minima of $\mathcal{G}_{GP}$.

Readers may note that each of the above steps can be implemented in many different ways. In our experiments, our choices have been adopted based on solutions present in the literature and on empirical results on the tested datasets.

In the first step of our method, we are aware that: 1) there might be sample pixels of a given class whose spectral values are not represented in the training set, and 2) there might be pixels in the training set with very similar spectral signatures but which belong to distinct classes. In those cases, we expect the classifier to assign low class probabilities to such pixels. This means that the watershed segmentation at those regions will be mostly guided by the original gray-levels of the image gradient. Then, in the step (ii), we combine the class probability maps into a single probability map μ. We expect this combination to provide flat zones of pixels with high probability of belonging to any given class, *i.e.*, subsets of pixels that should be merged *early* in the resulting hierarchical watershed. In the extreme case where the classifier assigns very high class probabilities to all pixels of I, we would have a single flat zone and, consequently, a hierarchy with a single segmentation level. In that case, we expect that the pixel features extracted from the resulting AP will have little influence in the final classification results. In other words, the final results will be similar to the ones obtained with the original pixel values. In the step (iii), a weighted graph (V, E, w_P) is obtained from the combined probability map μ. In our experiments, we chose to compute edge weights as the maximum between the probability values of neighboring pixels based on a few experiments with the datasets described in the conference version of this article [9]. In the steps (iv) and (v), a gradient (V, E, w_G) of I is computed and then combined with (V, E, w_P). In our experiments, this combination will be simply implemented as a multiplication of edge weights, similarly to [10]. We note that the proposed method is related to ones introduced in [19] and [20], the main difference being the type of hierarchy under consideration and how the original data is combined with the prior knowledge. Finally, in step (vi), we compute a hierarchical watershed $\mathcal{H}$ of $\mathcal{G}$ for a given sequence of minima of $\mathcal{G}$. Those minima are often ordered according to regional attributes (*e.g.,* area, volume) of the catchment basins associated to each minimum. Those regional attributes are known as *extinction values* [27], [28].

We now analyze the time complexity of the proposed method as described in Algorithm 1. Given an image I and a training set S of I, whose samples are labeled into n classes, we aim to compute a hierarchical watershed of I with prior knowledge

Algorithm 1: Computation of Hierarchical Watersheds with Semantic Knowledge from Training Pixels.

Input : Original image I, training set S of I composed of n classes
Output: Hierarchical watershed of I computed with prior knowledge

```
// Compute features of the pixels of I
// For simplicity, we often consider the
   raw pixel values as their features
1  F := compute_features(I)
// Train a classifier using the features
   F of the training set S
// For each pixel x of I, C(x) := (p_1^x, ..., p_n^x),
   where p_i^x is the probability of x to
   belong to the class c_i
2  C := train_classifier(I, S, F)
// Initialize an array μ that will store
   the class probabilities of each pixel
   of I provided by the classifier C
3  μ = array of length |I| initialized with zero values
4  foreach pixel x of I do
5  |  (p_1^x, ..., p_n^x) := C(x)
6  |  foreach semantic class c_i of I do
7  |  |  μ[x] := μ[x] + (p_i^x)^2
8  |  end
9  end
10 foreach pixel x of I do
11 |  μ[x] := 1 - sqrt(μ[x])
12 end
// Compute a connected graph from I
13 G := (V, E)
// Compute an edge weight map w_P : E → [0, 1]
   based on the probability map μ
14 foreach edge e = (x, y) of E do
15 |  w_P(e) := max(μ(x), μ(y))
16 end
// Compute an edge weight map w_G which
   corresponds to a gradient of I
17 foreach edge e = (x, y) of E do
18 |  w_G(e) := |I(x) - I(y)|
19 end
// Combine the edge weight maps w_P and w_G
20 foreach edge e = (x, y) of E do
21 |  w_PG(e) := w_P(e) × w_P(e)
22 end
// Compute a hierarchical watershed of the
   edge-weighted graph (V, E, w_PG)
23 H := compute_hierarchical_watershed(V, E, w_PG)
24 return H
```

from training pixels. Given that the pixel features are simply their spectral values or derived from a small window around each pixel, line 1 can be executed in linear time $O(|I|)$. Then, the time complexity to build and train the classifier $\mathbf{C}$ depends on the chosen method. For instance, if $\mathbf{C}$ is a Random Forest (RF) composed of m trees, obtained from t training samples with f features, the time complexity to build such a forest is $O(m \times t \times \log(t) \times f)$. Taking into account that f is a small constant in our experiments, the time complexity to build such a forest is $O(m \times t \times \log(t))$. In line 5, $\mathbf{C}(x)$ is thus computed in $O(m \times d)$, where d is the maximal depth of each tree. Considering that the number n of semantic classes is a small constant in most datasets, the *for* loops of lines $4-9$ are executed in time $O(|I| \times m \times d)$. The graph $\mathcal{G}$ can be constructed in time $O(|I|)$ given that each vertex in V corresponds to a pixel in I and that the number of edges increases linearly with the number of vertices in a 4 or 8-connected graphs. That being said, the three *for* loops in lines $14-21$ are executed in time $O(|I|)$. Finally, the hierarchical watershed $\mathcal{H}$ can be obtained in time $O(|I|\log|I|)$, as stated in [29]. Therefore, the overall time complexity of Algorithm 1 is $O(m \times t \times \log(t) + |I| \times m \times d + |I|\log|I|)$ if we consider the learning step of the classifier $\mathbf{C}$. However, since $\mathbf{C}$ is trained only once, we can simplify it to $O(|I| \times m \times d + |I|\log|I|)$, which is a function of the number m of trees of $\mathbf{C}$, their depths and the dimensions of the image I.

An illustration of the proposed method is given in Fig. 4. A crop of the image $zh20$ from the Zurich dataset [30], its ground-truth composed of six semantic classes and a class probability map (obtained as described in Algorithm 1) are shown in Fig. 4(a). Fig. 4(b) and (c) shows image reconstructions from filtered versions of two different hierarchical watersheds of $zh20$: in (b), we considered a hierarchical watershed obtained from the infrared channel of $zh20$ computed without any prior knowledge from training pixels and, in (c), we considered a hierarchical watershed with prior knowledge from the map μ, as described in Algorithm 1. When comparing both results, we observe that images in (c) preserve some of the boundaries between distinct semantic classes, as, for instance, between the regions belonging to classes *trees* (represented in dark green in the ground-truth) and *grass* (light green) in the lower right corner of the reconstructed images.

2) Hierarchy Filtering With Semantic Prior Knowledge: The illustration given in Fig. 4 shows that the hierarchical watersheds computed with prior knowledge from training pixels can indeed provide regions, which are more semantically coherent. On the other hand, considering that our objective is feature extraction at pixel level, including prior knowledge into the hierarchy construction may suppress part of the finer regions present in the original data. In order to preserve the information provided by those finer regions, we could delay the utilization of prior knowledge in the watershed-AP pipeline to the filtering step (step 3 of the pipeline given in Fig. 2). The idea is to replace the handcrafted criteria, which are usually employed at this step (*e.g.,* area and moment of inertia thresholds) by markers obtained from class probability maps per semantic class, as the ones given in Fig. 3. Given a dataset I and its training set composed of n classes, we propose the following pipeline.

1) Computation of a hierarchical watershed $\mathcal{H}$ of I.
2) Training of a classifier using the training set of I and computation of per-pixel class probabilities $(p_1, \ldots, p_n)$ for all pixels of I.
3) For each probability map p_i, computation of a list $L_i = (p_i^1, \ldots, p_i^k)$ of k binary thresholded versions of p_i such that, in each map p_i^m, white or one valued pixels are the ones whose class probabilities are larger than a given threshold.
4) Filtering out the nodes of regions of $\mathcal{H}$ which only contain black or zero-valued pixels.

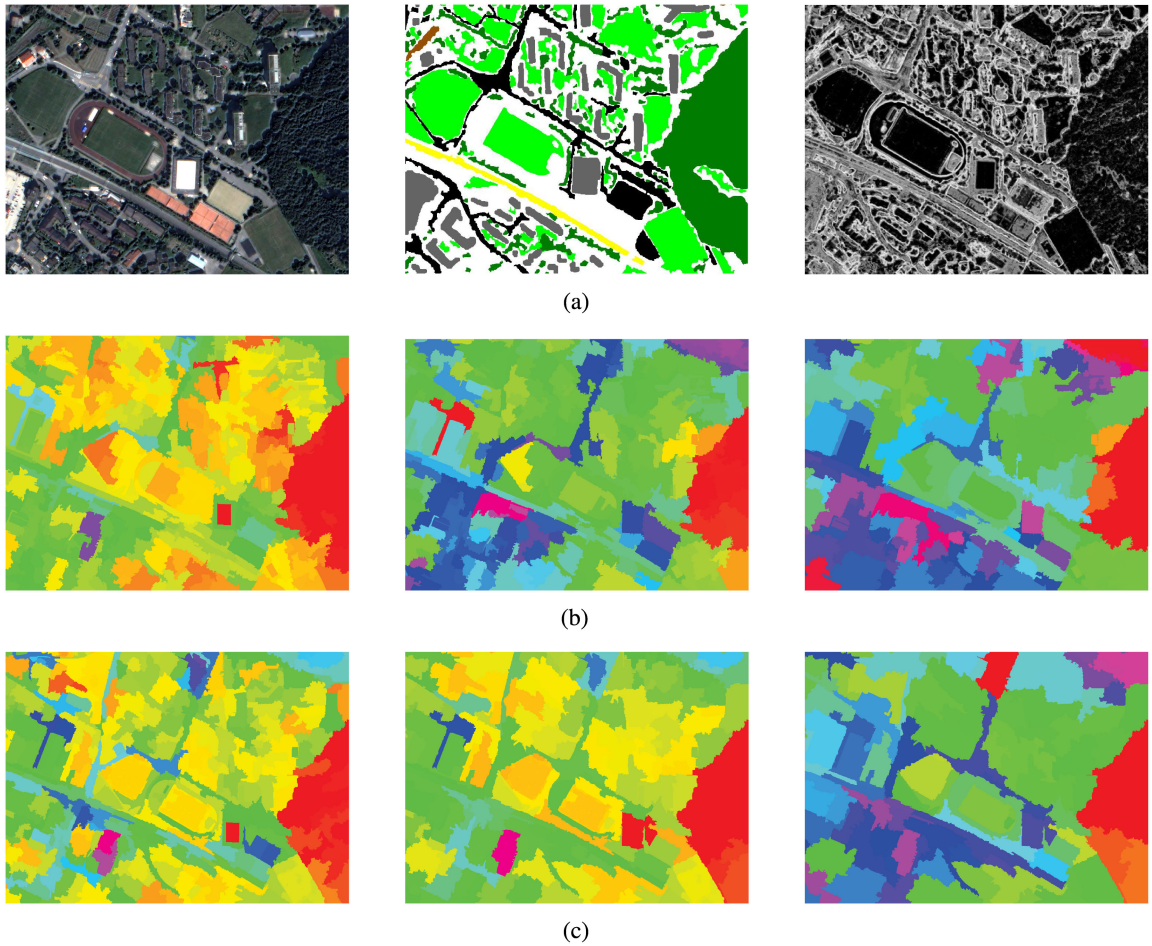

Fig. 4. Image reconstructions (represented in false colors) obtained from two hierarchical watersheds of the image $zh20$ (Zurich dataset V-A) computed without semantic prior knowledge (b) and with semantic prior knowledge given by the probability map μ (c). The reconstructions given in (b) and (c) are computed by filtering their respective hierarchical watersheds with the area attribute and the following thresholds: 5 k, 10 k, 20 k. (a) Cropped RGB image $zh20$ from the Zurich dataset, its ground-truth and the combined probability map μ. (b) Images reconstructed from filtered versions of a hierarchical watershed computed from a gradient of I. (c) Images reconstructed from filtered versions of a hierarchical watershed computed from a combination of a gradient of I with μ.

5) Reconstruction of an image from each of the filtered versions of $\mathcal{H}$ computed in the previous step.

The proposed pipeline is shown in Fig. 5 and described in Algorithm 2. In terms of time complexity, Algorithm 2 adds as much to the complexity of computing watershed-APs as does the method described in Algorithm 1. More precisely, both algorithms include the training of a classifier $\mathbf{C}$ and the computation of classification probability maps per semantic class, which consist of the most "time consuming" parts of the algorithms. Then, in the lines $7-17$ of Algorithm 2, we perform the filtering step of the watershed-AP. In practice, filtering the input hierarchy is indeed similar to the standard filtering step of APs using criteria such as area and moment of inertia: attribute values are propagated from the leaves to the root node, and nodes are removed if their attribute values are below a given threshold value. Therefore, we can conclude the proposed methods for including semantic prior knowledge into watershed-APs present equivalent time complexities.

Fig. 6 illustrates the method described in Algorithm 2 on a cropped patch of an image from the Vaihingen dataset [31]. The image and its ground truth regions are given in Fig. 6(a). In Fig. 6(b) and (c), we give thresholded versions of the probability map for the class *trees* (represented in green in the ground truth), along with the image reconstructions obtained from a hierarchical watershed filtered using each of those probability maps, as described in Algorithm 2. Similarly, Fig. 6(d) and (e) shows the results obtained for the class *cars* (represented in yellow in the ground-truth image). For a better visualization, all image reconstructions are represented in false colors. In Fig. 6(c), we observe that the finer regions belonging to the class *trees* are preserved in all reconstructions and the same is true for the class *cars* in Fig. 6(e).

V. Experimental Setup

In this section, we conduct experiments to evaluate the performance of the proposed watershed-AP (computed with and without prior knowledge) in the context of land-cover classification and building extraction of remote sensing images. We first describe the multispectral images considered in our study, as well as the experimental settings used for evaluation. We provide detailed analysis and show that oftentimes watershed-APs

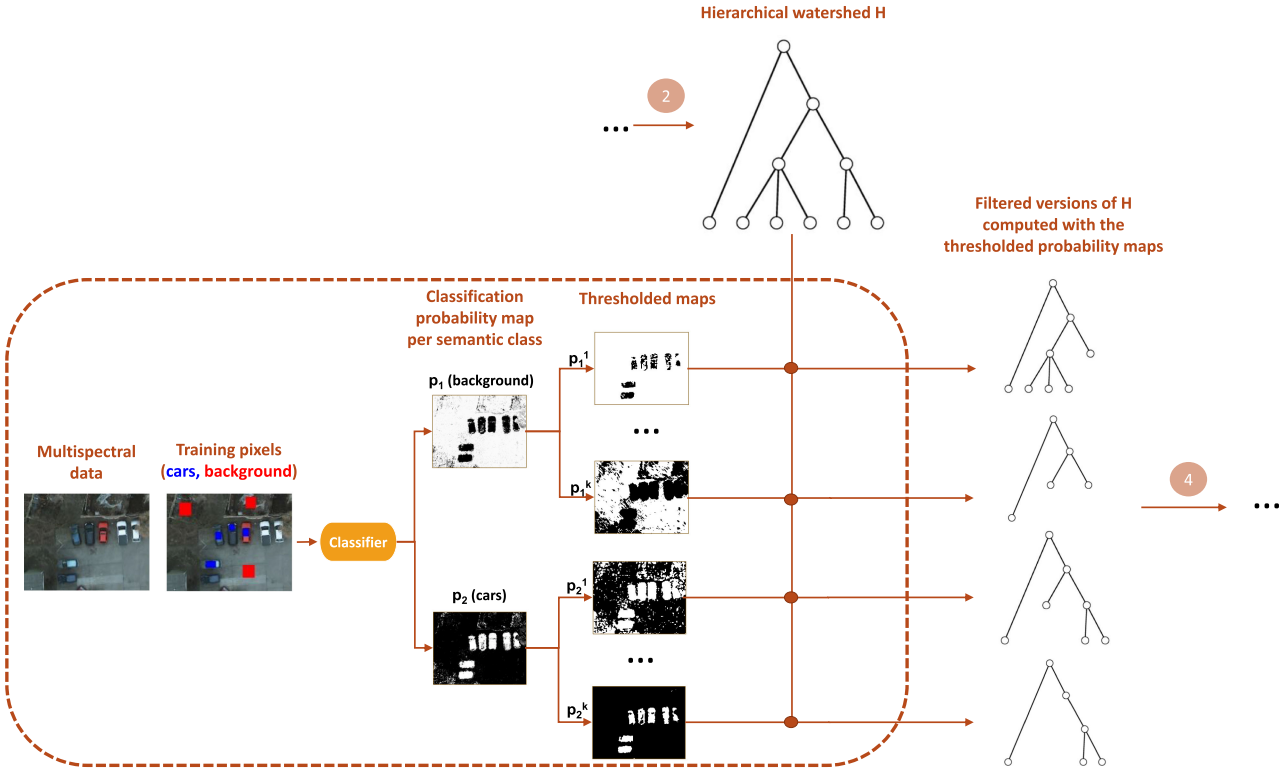

Fig. 5. Semantic prior knowledge employed at the filtering step of watershed-AP. Given a multispectral image I, a classifier trained on a subset of pixels of I provides probability maps per semantic class. Then, those maps are thresholded at many different levels, resulting in binary maps, whose white pixels have the highest probabilities of belonging to a given class. Finally, the threshold probability maps are used as markers to guide the filtering of the input hierarchical watershed.

outperform AP and its variants including SDAP [32], α-AP [25], and ω-AP [25], on both datasets in the following section.

Experiments were performed in Python[1] using the open source simple attribute profile (SAP)[2] and Higra [33] libraries.[3]

A. Datasets

Our experiments were conducted on the multispectral remote sensing Vaihingen [31] and Zurich [30] datasets.

The Vaihingen dataset [31] contains 38 images collected over the city of Vaihingen, Germany. Each image has a spatial resolution of 9 cm and is composed of four channels (near infrared, red and green and blue) plus a digital surface model (DSM). As done in [34], we only consider the first three channels, namely near infrared, red, and green in our experiments. Images width and height range in the intervals [1388,3816] and [1281,3313], respectively. Ground-truth annotations, which are provided for each of the 38 images, consist of at most six thematic classes: impervious surfaces, buildings, low vegetation, trees, cars, and background/clutter. Following previous works [34], we only consider the 16 images for which ground-truth annotations were provided during the ISPRS semantic labeling contest. Moreover, the background class is not used for land-cover classification. To perform building extraction, the ground-truth labels are divided into two classes: buildings and background (*i.e.,* the union of all remaining semantic classes). Training samples were extracted from eleven images (image IDs: 1, 3, 5, 7, 13, 17, 21, 23, 26, 32, 37), and the remaining five images (image IDs: 11, 15, 28, 30, 34) were used for evaluation. For each semantic class, training samples were composed of 1% of randomly selected pixels in the training images, leading to the number of training and test pixels per semantic class given in Table I. The test images of the Vaihingen dataset and their ground-truths are shown in Fig. 7.

The Zurich Summer dataset [30] is a collection of 20 images obtained from a QuickBird acquisition of the city of Zurich, Switzerland, in August 2002. The images in this dataset have various dimensions, with widths and heights in [622,1639] and [782,1830] ranges, respectively, and are composed of four channels (near infrared, red, green, and blue). The ground-truth provided for each image consists of at most nine thematic classes, namely: roads, buildings, trees, grass, bare soil, water, railways, swimming pools, and background. For the building extraction task, all classes except *buildings* are merged into a single *background* class. Following previous works [34], [35], training pixels were extracted from the first fifteen images of this dataset, and the remaining five images (zh16, zh17, zh18, zh19, and zh20) were used for evaluation. As for the Vaihingen dataset,

[1]Source codes are available in https://github.com/deisemaia/Watershed-attribute-profiles

[2]The documentation and source codes of the SAP package are provided in https://github.com/fguiotte/sap

[3]The documentation and source codes of the Higra package are provided in https://github.com/higra

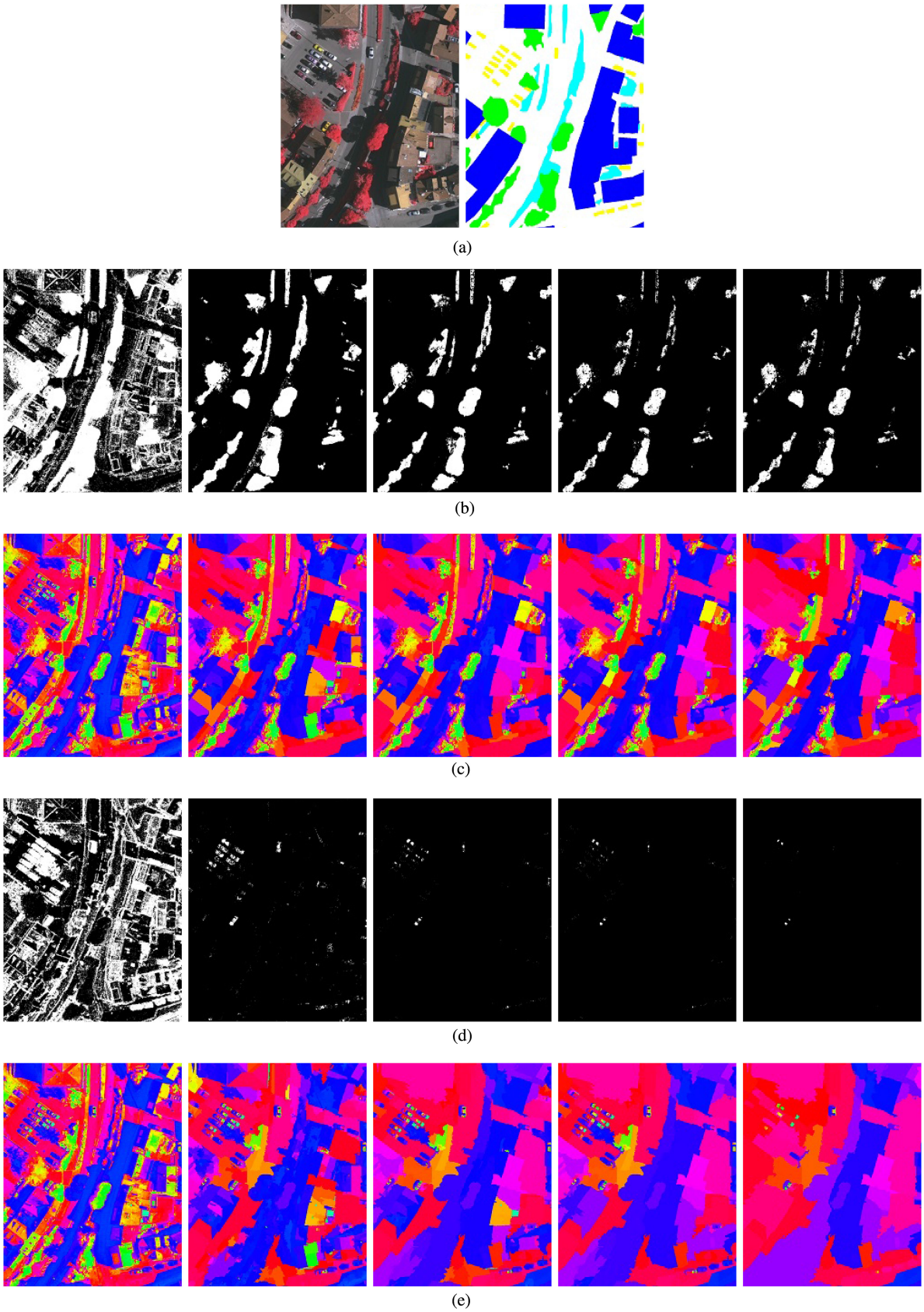

Fig. 6. Comparison between image reconstructions obtained with the filtering method illustrated in Fig. 5. To better distinguish between neighboring regions with similar gray levels, all image reconstructions are represented in false colors. (a) Cropped patch of the first image (denoted here as vn_1) from the Vaihingen dataset [31] and its ground-truth semantic classes. (b) Class probability map for the *tree* class (in green) thresholded at increasing values. (c) Image reconstructions (in false colors) obtained by filtering a hierarchical watershed of vn_1 using the maps given in (b). (d) Class probability map for the *car* class (in yellow) thresholded at increasing values. (e) Image reconstructions (in false colors) obtained by filtering a hierarchical watershed of vn_1 using the maps given in (d).

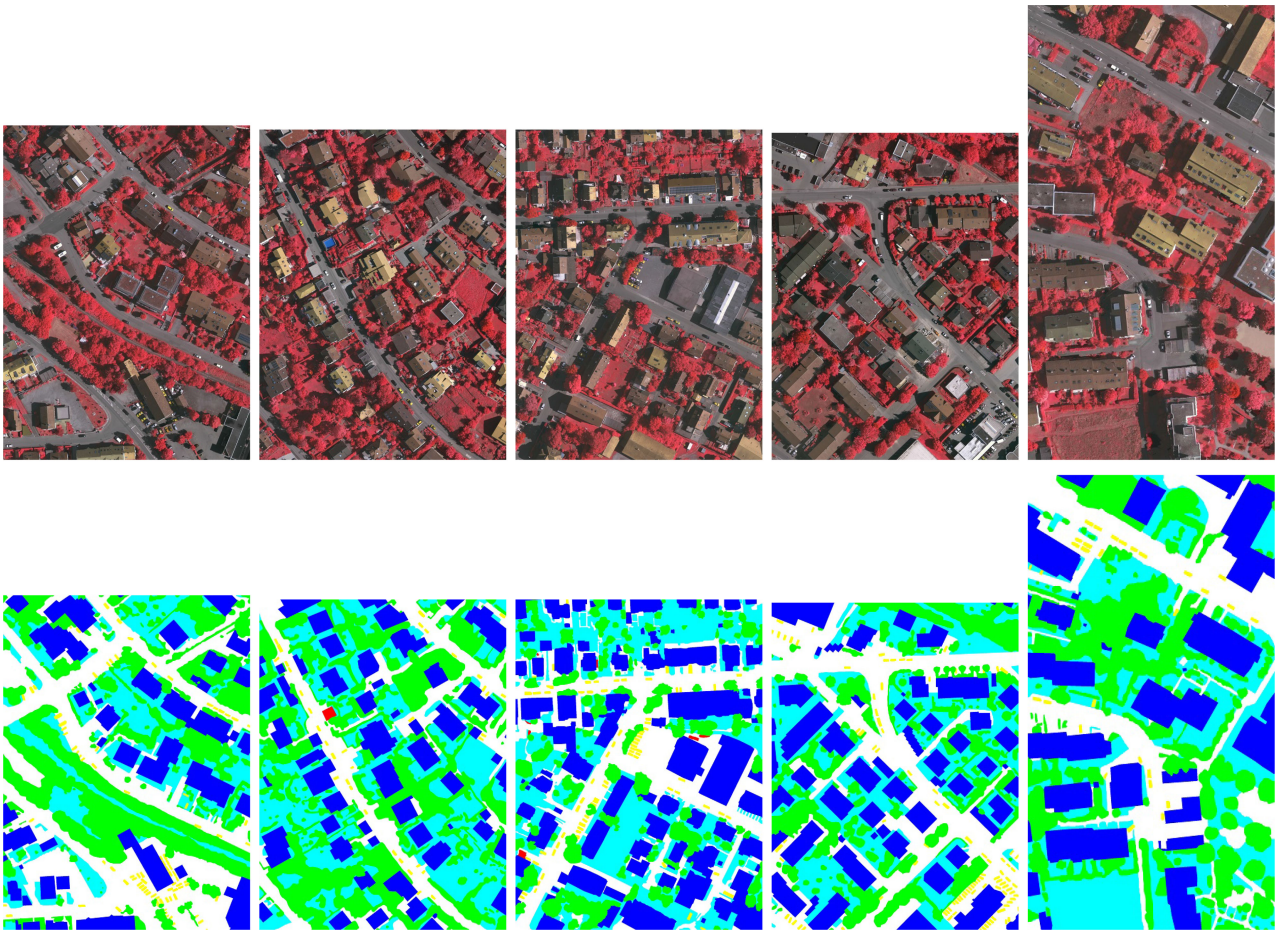

Fig. 7. The test images of the Vaihingen dataset. First line from left to right: NIR+RG images (IDs: 11, 15, 28, 30, and 34) represented as RGB images. Second line: ground-truth labels including six semantic classes: impervious surfaces, buildings, low vegetation, trees, cars, background.

the background class of the Zurich dataset is not considered for land-cover classification. In total, the training set is composed of 122,630 pixels, which corresponds to 1% of the labeled pixels randomly extracted for each class of each training image. The number of training and test pixels per semantic class is given in Table II. The test images and their ground-truths are given in Fig. 8.

B. Experimental Settings

Most works in the literature evaluate AP and its variants in the following way: training and test pixels come from the same remote sensing image, which allows training and test features to be extracted from the same hierarchical representation of the image under study. As discussed in [5], this setting does not generalize well to more realistic scenarios, in which we may have a dataset composed of several images without any annotation. For that reason, we extend the evaluation on the Zurich dataset already presented in our conference paper [9]. In the present article, training and test features are extracted from independent hierarchical representations computed from the training and test images.

AP and its variants, including watershed-APs, were computed with the usual area and moment of inertia (MoI) attributes. The following 10 area thresholds and four MoI thresholds were adopted for both datasets

$$\lambda_{\text{area}} = \{25, 100, 500, 1000, 5000, 10000, 20000, 50000, 100000, 150000\}$$

$$\lambda_{\text{moi}} = \{0.2, 0.3, 0.4, 0.5\}$$

The only exceptions are the watershed-APs filtered using semantic prior knowledge, as described in Section IV-B2. To compute those watershed-APs, the hard-coded criteria based on the area and MoI attributes are replaced by the nodes' probability of belonging to each ground-truth semantic class. In our experiments, the set T of threshold values used by Algorithm 2 is defined as an interval of evenly spaced numbers ranging from the minimum to the maximum values of each classification probability map. In order to obtain a similar number of features as the other APs, different numbers of threshold values were considered for each task and for each dataset (see Table III).

The APs and their extensions are first computed independently on each of the NIR+RGB bands (resp. NIR+RG) of the Zurich (resp. Vahingen) dataset and then concatenated. For

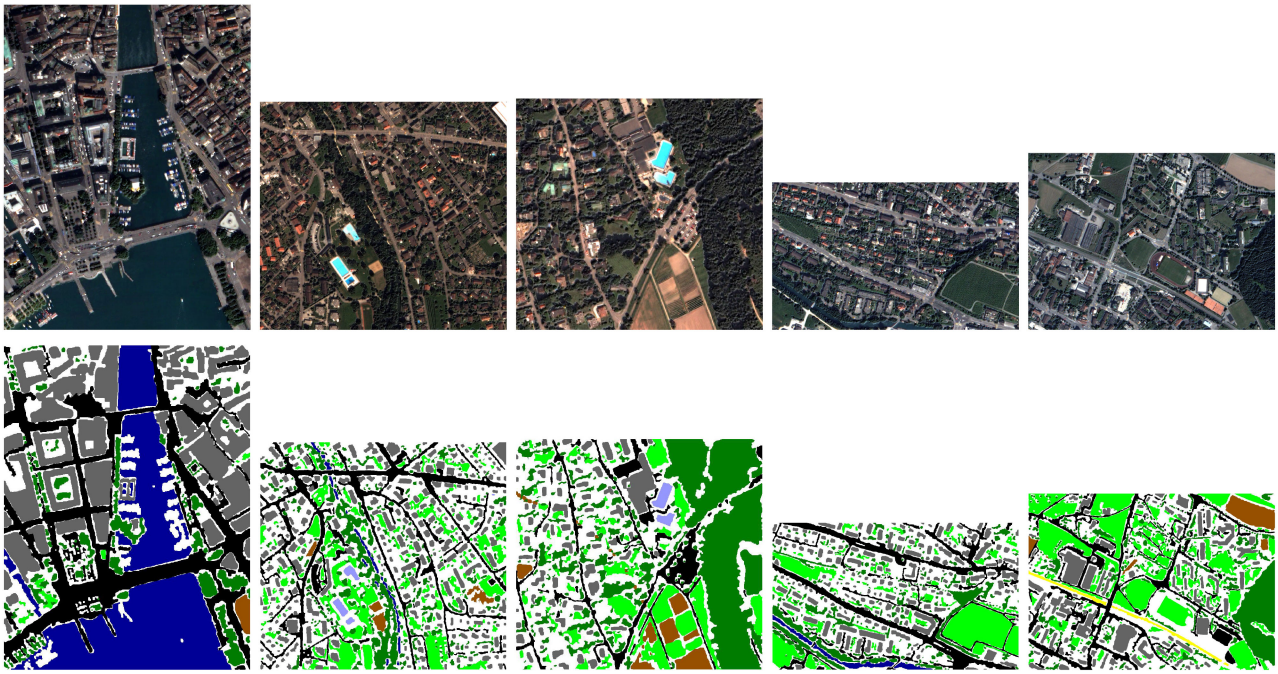

Fig. 8. Test images of the Zurich dataset. First line from left to right: RGB images with IDs from 16 to 20. Second line: ground truth labels including nine semantic classes: roads, buildings, trees, grass, bare soil, water, railways, swiming pools, background.

each image I in the Zurich and Vaihingen datasets, hierarchical watersheds were computed from two four-connected edge-weighted graphs: from the graph $\mathcal{G}_G = (V, E, w_G)$ obtained from a gradient of the original data I (without semantic prior knowledge from training pixels), and the second one computed from the combination of the graph $\mathcal{G}_G$ with the classification probability map obtained from the training set of I, as described in Section IV-A.

Supervised pixel classification was performed twice, once for obtaining the classification probability map and then to provide the final land-cover or building extraction classification. Both were performed using a RF classifier with 100 trees. The number of variables used for training was set to the square root of the feature vectors length. To optimize the construction and filtering of hierarchical watersheds using prior knowledge, the first RF, which is used to obtain the classification probability maps, is trained on the features extracted from a 5×5 window around each training pixel.

For land-cover classification, the different approaches are compared using the overall accuracy (OA), average accuracy per class (AA) and κ coefficient, as done in [3]. To evaluate the proposed methods on building extraction, the following standard measures were considered: OA, precision, recall, the F1 score, and the mean intersection over union (mIOU). For each tested method, we report the mean and standard deviation of the classification scores over ten runs in the form of $mean \pm standard\ deviation$.

As defined in Section III-A, hierarchical watersheds can be computed for any given ordering on the minima of a weighted graph. In our experiments, such orderings are obtained from *extinction values* [27], [28] based on the area, dynamics, and volume attributes.

VI. Results and Discussion

A. Land-Cover Classification

Tables IV–IX present the results of land-cover classification on the Vaihingen and Zurich datasets. We compare the performance of the following methods: the baseline, in which every pixel is represented by its NIR+RG or NIR+RGB values, AP-maxT, and AP-minT obtained by filtering the max- and min-tree, respectively; AP [3] obtained as a concatenation of AP-maxT and AP-minT; SDAP [32]; α-AP and ω-AP [25]; and the watershed-APs computed/filtered with and without prior knowledge. To simplify the notations, watershed-AP computed without prior knowledge is denoted as A-WS-AP, and the watershed-APs computed and filtered using prior knowledge are denoted as A-CPWS-AP and A-FPWS-AP, respectively, where A is the attribute used in the construction of the hierarchical watersheds, namely Area, Dynamics (Dyn) and Volume (Vol).

On the Vaihingen dataset, as shown in Tables IV and V, the Area-WS-AP and Volume-WS-AP, which are computed without any prior knowledge, outperform all other methods in the literature in terms of OA and κ scores. The same is true for Area-CPWS-AP and Vol-CPWS-AP, as well for all three FPWS-APs. Our best method, namely Vol-FPWS-AP, outperforms AP by 2.46% and by 3.29% in terms of OA and κ scores, respectively. Regarding the use of prior knowledge for WS-APs, all three FPWS-APs performed better than their respective CPWS-APs and WS-AP counterparts. The most meaningful improvements were observed for Dyn-FPWS-AP, which presented OA, AA,

Algorithm 2: Filtering of Hierarchical Watersheds with Semantic Knowledge from Training Pixels.

Input : Original image I, training set S labeled into n classes, hierarchical watershed $\mathcal{H}$, threshold list T
Output: List of filtered versions of the hierarchy $\mathcal{H}$

```
// Same steps 1-2 of Algorithm 1
```
$\mathbf{F} := compute\_features(I)$
$\mathbf{C} := train\_classifier(I, S, F)$
```
// Compute per-class and per-pixel
   classification probabilities
```
foreach *pixel* x *of* I **do**
  $p_1^x, \ldots, p_n^x := \mathbf{C}(x)$
end
```
// Initialize an empty list L that will
   store all filtered versions of H
```
$\mathcal{L} :=$ empty list
```
// For each threshold t in T and for each
   semantic class c_i, we filter out the
   nodes of H composed only of pixels
   whose class probability for the class
   c_i is inferior to t
```
foreach *threshold value* t *in* T **do**
  foreach *semantic class* c_i *of* I **do**
    ```
    // Create a copy of H
    ```
    $\mathcal{H}' :=$ copy($\mathcal{H}$)
    foreach *node* η *of* $\mathcal{H}'$ **do**
      if $p_i^x < t$ *for every pixel in* η **then**
        filter out the node η from $\mathcal{H}'$
      end
    end
    ```
    // Add H' to the list L
    ```
    $\mathcal{L} := \mathcal{L} \cup \mathcal{H}'$
  end
end
return L

TABLE I
NUMBER OF TRAINING AND TEST SAMPLES OF THE VAIHINGEN DATASET USED FOR LAND-COVER CLASSIFICATION AND BUILDING EXTRACTION

Class	Training	Test
Land-cover classification		
Impervious surfaces	159,323	5882512
Buildings	146,465	5,770,150
Low vegetation	110,075	5,264,832
Trees	121,183	5,991,642
Cars	6,661	279,069
Building extraction		
Buildings	146,465	5,770,150
Background	402,362	17,433,644

TABLE II
NUMBER OF TRAINING AND TEST SAMPLES OF THE ZURICH DATASET USED FOR LAND-COVER CLASSIFICATION AND BUILDING EXTRACTION

Class	Training	Test
Land-cover classification		
Roads	30,439	757,752
Buildings	46,343	739,365
Trees	18,789	652,037
Grass	12,129	650,969
Bare soil	1,552	79,039
Water	9,910	275,923
Railways	3,229	18,817
Swimming pools	229	10,457
Building extraction		
Buildings	46,343	739,365
Background	161,800	2,444,994

TABLE III
NUMBER OF THRESHOLD VALUES CONSIDERED IN THE COMPUTATION WATERSHED-APS FILTERED USING PRIOR KNOWLEDGE

Dataset	Task	#Thresholds
Zurich	Building extraction	7
Vaihingen	Land-cover classification	5
	Building extraction	7

TABLE IV
AVERAGE LAND-COVER CLASSIFICATION RESULTS FOR THE TEST IMAGES OF VAIHINGEN OVER TEN DIFFERENT RANDOM TRAINING SETS

Method	Dim.	Classification result		
		OA (%)	AA (%)	$\kappa \times 100$
NIR+RG	3	65.71 ± 0.04	54.60 ± 0.05	54.45 ± 0.06
AP-maxT [3]	48	69.86 ± 0.11	59.15 ± 0.11	59.92 ± 0.15
AP-minT [3]	48	70.18 ± 0.09	58.54 ± 0.15	60.31 ± 0.13
AP [3]	90	72.58 ± 0.08	61.92 ± 0.11	63.51 ± 0.11
SDAP [32]	48	72.27 ± 0.06	61.29 ± 0.13	63.11 ± 0.07
α-AP [25]	48	71.35 ± 0.07	58.85 ± 0.09	61.85 ± 0.10
ω-AP [25]	48	71.54 ± 0.11	58.99 ± 0.13	62.11 ± 0.14
Area-WS-AP	48	73.21 ± 0.09	61.26 ± 0.11	64.35 ± 0.11
Dyn-WS-AP	48	71.83 ± 0.13	59.19 ± 0.15	62.49 ± 0.18
Vol-WS-AP	48	73.46 ± 0.07	61.64 ± 0.10	64.69 ± 0.09
Watershed-AP constructed with prior knowledge:				
Area-CPWS-AP	48	73.15 ± 0.23	61.04 ± 0.26	64.27 ± 0.30
Dyn-CPWS-AP	48	72.49 ± 0.11	59.64 ± 0.15	63.36 ± 0.15
Vol-CPWS-AP	48	73.53 ± 0.15	61.66 ± 0.21	64.78 ± 0.20
Watershed-APs filtered with prior knowledge:				
Area-FPWS-AP	48	74.80 ± 0.07	62.18 ± 0.14	66.47 ± 0.10
Dyn-FPWS-AP	48	74.47 ± 0.12	62.57 ± 0.17	66.05 ± 0.16
Vol-FPWS-AP	48	$\mathbf{75.04 \pm 0.11}$	$\mathbf{62.61 \pm 0.16}$	$\mathbf{66.80 \pm 0.15}$
Deep learning models:				
FCN-Festa+dCFR [34]	-	77.99 ± 2.14	-	-
FCN [34]	-	86.51	-	-

and κ scores more than 2% higher than Dyn-WS-AP. Considering the classification result per semantic class given in Table VIII, we observe that the watershed-APs yielded the best results for the impervious surfaces, buildings, low vegetation and tree classes, but falls behind AP on the classification of cars.

To provide further insight into the performance of APs in the context of semantic segmentation, we compare one of our results with a recent deep learning method [34] (see Table VI). In [34], the authors propose an adaptation of the FCN to consider sparsely annotated training data. They trained their proposed model with different types of scribbled annotations, including points, lines, and polygons. On both datasets, their baseline is the FCN trained on the whole set of training pixels, and their best results were achieved by the FCN-Festa-dCRF method trained on scribbled lines. The FCN-Festa-dCFR and FCN methods were trained on 480,593 and 54,373,518 samples, respectively, while that our Vol-FPWS-AP was trained on 543,707 samples. On this dataset, Vol-FPWS-AP approached the scores of FCN-Festa-dCFR on four classes, namely impervious surfaces, buildings, low vegetation and trees, but presented much poorer results on the classification of cars. With respect to FCN, both Vol-FPWS-AP and FCN-Festa-dCFR performed worse in general, except for the *trees* class, for which Vol-FPWS-AP and FCN-Festa-dCFR outperformed FCN by

TABLE V
AVERAGE ACCURACY PER CLASS FOR THE TEST IMAGES OF VAIHINGEN OVER TEN DIFFERENT RANDOM TRAINING SETS

Method	Dim.	Classification result per class				
		Impervious surfaces	Buildings	Low vegetation	Trees	Cars
NIR+RG	3	73.03 ± 0.13	69.26 ± 0.14	49.20 ± 0.16	72.24 ± 0.17	9.26 ± 0.18
AP-maxT [3]	48	75.16 ± 0.41	72.82 ± 0.28	49.64 ± 0.30	82.09 ± 0.20	16.06 ± 0.28
AP-minT [3]	48	78.57 ± 0.24	73.07 ± 0.27	46.84 ± 0.38	82.40 ± 0.18	11.83 ± 0.46
AP [3]	90	80.70 ± 0.16	75.71 ± 0.18	49.57 ± 0.25	84.28 ± 0.12	**19.32 ± 0.47**
SDAP [32]	48	80.67 ± 0.21	73.60 ± 0.25	53.51 ± 0.21	81.80 ± 0.13	16.88 ± 0.72
α-AP [25]	48	76.21 ± 0.25	75.78 ± 0.30	47.98 ± 0.40	85.79 ± 0.15	8.49 ± 0.46
ω-AP [25]	48	75.55 ± 0.23	76.58 ± 0.27	49.39 ± 0.30	85.16 ± 0.18	8.26 ± 0.75
Area-WS-AP	48	79.53 ± 0.18	74.37 ± 0.13	54.47 ± 0.26	85.18 ± 0.09	12.74 ± 0.53
Dyn-WS-AP	48	76.85 ± 0.41	76.10 ± 0.43	48.88 ± 0.32	85.94 ± 0.16	8.18 ± 0.36
Vol-WS-AP	48	79.33 ± 0.17	75.20 ± 0.22	54.99 ± 0.19	85.03 ± 0.09	13.64 ± 0.37
Watershed-APs computed with prior knowledge:						
Area-PWS-AP	48	80.59 ± 0.29	72.14 ± 0.71	55.39 ± 0.73	85.28 ± 0.30	11.82 ± 0.85
Dyn-PWS-AP	48	77.45 ± 0.54	77.38 ± 0.41	48.50 ± 0.49	**87.00 ± 0.17**	7.85 ± 0.52
Vol-PWS-AP	48	80.68 ± 0.46	72.46 ± 0.46	**56.38 ± 0.58**	85.42 ± 0.22	13.35 ± 0.71
Watershed-APs filtered with prior knowledge:						
Area-PWS-AP	48	80.66 ± 0.29	78.19 ± 0.34	54.94 ± 0.34	86.22 ± 0.15	10.88 ± 0.51
Dyn-FPWS-AP	48	79.93 ± 0.41	79.15 ± 0.53	54.46 ± 0.25	84.99 ± 0.14	14.34 ± 0.46
Vol-FPWS-AP	48	**80.83 ± 0.44**	**79.34 ± 0.27**	54.68 ± 0.34	86.03 ± 0.13	12.19 ± 0.61

TABLE VI
COMPARISON OF PER-CLASS F1 SCORES OBTAINED WITH VOL-CPWS-AP FOR THE TEST IMAGES OF VAIHINGEN WITH THE RESULTS OF TWO DEEP LEARNING METHODS REPORTED IN [34]

Method	F1 scores per class					Mean F1
	Impervious surfaces	Buildings	Low vegetation	Trees	Cars	
Vol-FPWS-AP	78.36 ± 0.21	82.06 ± 0.16	60.44 ± 0.19	78.16 ± 0.03	18.5 ± 0.75	64.0
Deep learning models:						
FCN-Festa+dCFR [34]	80.06 ± 3.32	84.47 ± 2.23	64.35 ± 2.38	80.32 ± 0.92	43.72 ± 9.62	70.58
FCN [34]	88.67	92.83	76.32	74.21	86.67	83.74

TABLE VII
AVERAGE LAND-COVER CLASSIFICATION RESULTS FOR THE TEST IMAGES OF ZURICH OVER TEN DIFFERENT RANDOM TRAINING SETS

Method	Dim.	Classification result		
		OA (%)	AA (%)	$\kappa \times 100$
NIR+RGB	4	76.41 ± 0.18	65.16 ± 1.01	70.26 ± 0.24
AP-maxT [3]	64	81.29 ± 0.17	64.39 ± 0.30	76.34 ± 0.22
AP-minT [3]	64	76.64 ± 0.32	61.49 ± 0.32	70.45 ± 0.41
AP [3]	120	81.98 ± 0.14	64.55 ± 0.21	77.20 ± 0.18
SDAP [32]	64	81.98 ± 0.20	64.16 ± 0.89	77.20 ± 0.25
α-AP [25]	64	80.37 ± 0.18	63.35 ± 0.80	75.17 ± 0.24
ω-AP [25]	64	80.35 ± 0.35	63.41 ± 1.36	75.14 ± 0.46
Area-WS-AP	64	83.12 ± 0.18	65.28 ± 0.11	78.65 ± 0.23
Dyn-WS-AP	64	80.38 ± 0.12	62.84 ± 0.17	75.18 ± 0.15
Vol-WS-AP	64	83.36 ± 0.23	65.50 ± 0.13	78.96 ± 0.29
Watershed-APs computed with semantic prior knowledge (pipeline of Fig. 3)				
Area-CPWS-AP	64	85.25 ± 0.20	66.49 ± 0.18	81.34 ± 0.26
Dyn-CPWS-AP	64	83.46 ± 0.28	65.05 ± 0.32	79.07 ± 0.35
Vol-CPWS-AP	64	**85.27 ± 0.35**	**66.81 ± 0.24**	**81.37 ± 0.44**
Deep learning models:				
FCN-Festa+dcFR [34]	-	78.51 ± 2.21	-	-
FCN [34]	-	90.51	-	-

at least 3.95%. By comparing Vol-FPWS-AP and FCN-Festa-dCFR with FCN, we see a compromise between the number of training samples and the overall performance of each method.

On the Zurich dataset, as shown in Table VII, the Area-WS-AP and Vol-WS-AP, computed without any prior knowledge, as well as the CPWS-APs outperform all other methods in the literature in terms of OA and κ. Our best method, namely Vol-CPWS-AP, outperforms AP by 3.29%, 2.26%, and 4.17 in terms of OA, AA, and κ, respectively. Considering the classification result per semantic class given in Table VIII, we observe that the watershed-APs yielded the best results for most classes, namely roads, buildings, trees, grass, and water, but, for the remaining three classes, none of the APs outperformed the baseline NIR+RGB. Concerning the use of prior knowledge, it led to consistent improvements when used in the construction of hierarchical watersheds: all three CPWS-APs outperformed their respective WS-APs by at least 1% in terms of OA, AA, and κ. The most significant improvement was observed for the Dyn-CPWS-AP, which outperformed Dyn-WS-AP by 3.08%, 2.21%, and 3.89 in terms of OA, AA, and $\kappa \times 100$, respectively. However, regarding FPWS-APs, we concluded that this method is not well adapted for this dataset and, hence, we do not present land-cover evaluation scores of FPWS-APs on Zurich. The reason is that the Zurich test images do not contain ground truth pixels of all eight semantic classes and, different from CPWS-APs, the FPWS-APs are constructed by considering the classification probability maps of each class individually. For instance, the swimming pool class is present in only two among the five Zurich test images. Hence, the classification probability map for this semantic class would be meaningless for remaining three test images, leading to flat or redundant image reconstructions.

Table IX compares the F1 scores per class of our best method, namely Vol-FPWS-AP, with the results of two FCN based deep learning methods present in [34]. As for the Vaihingen dataset, their best results on Zurich were achieved by the FCN-Festa-dCRF method trained on scribbled lines, which are composed

TABLE VIII
AVERAGE ACCURACY PER CLASS FOR THE TEST IMAGES OF ZURICH OVER TEN DIFFERENT RANDOM TRAINING SETS

Method	Dim.	Classification result per class							
		Roads	Buildings	Trees	Grass	Bare Soil	Water	Railways	Swimming Pools
NIR+RGB	4	63.05 ± 0.32	79.17 ± 0.08	89.27 ± 0.10	76.82 ± 0.35	**25.29 ± 8.45**	93.46 ± 0.10	**2.27 ± 0.24**	**91.97 ± 0.58**
AP-maxT [3]	64	75.21 ± 0.50	84.94 ± 0.19	92.98 ± 0.16	79.51 ± 0.66	0.03 ± 0.03	93.36 ± 0.21	0.0 ± 0.0	89.08 ± 2.58
AP-minT [3]	64	67.13 ± 0.86	80.59 ± 0.22	92.99 ± 0.23	71.82 ± 1.64	0.13 ± 0.10	91.66 ± 1.14	0.0 ± 0.0	87.61 ± 1.09
AP [3]	120	77.93 ± 0.42	83.62 ± 0.36	94.01 ± 0.19	80.33 ± 0.76	0.06 ± 0.11	92.98 ± 0.81	0.0 ± 0.0	87.49 ± 1.71
SDAP [32]	64	75.06 ± 0.38	85.24 ± 0.36	94.10 ± 0.13	82.12 ± 0.89	0.34 ± 0.24	92.11 ± 0.38	0.0 ± 0.0	84.28 ± 6.80
α-AP [25]	64	73.49 ± 0.59	84.50 ± 0.31	95.03 ± 0.16	75.34 ± 0.75	2.54 ± 7.20	93.04 ± 0.43	0.0 ± 0.0	82.87 ± 0.96
ω-AP [25]	64	73.58 ± 0.89	84.03 ± 0.45	94.88 ± 0.11	75.77 ± 0.54	4.26 ± 11.11	92.70 ± 0.19	0.0 ± 0.0	82.04 ± 1.20
Area-WS-AP	64	74.36 ± 0.36	87.77 ± 0.28	95.12 ± 0.09	84.66 ± 0.69	0.06 ± 0.18	92.01 ± 0.47	0.0 ± 0.0	88.23 ± 0.71
Dyn-WS-AP	64	70.23 ± 0.83	**88.42 ± 0.41**	95.29 ± 0.09	74.87 ± 0.42	0.31 ± 0.81	92.85 ± 0.13	0.0 ± 0.0	80.74 ± 1.18
Vol-WS-AP	64	75.20 ± 0.36	88.28 ± 0.28	95.20 ± 0.13	83.81 ± 0.77	0.03 ± 0.03	92.94 ± 0.46	0.0 ± 0.0	88.57 ± 0.58
Watershed-APs computed with semantic prior knowledge:									
Area-CPWS-AP	64	79.86 ± 0.74	86.58 ± 0.57	96.03 ± 0.12	88.88 ± 0.50	0.07 ± 0.11	92.56 ± 1.33	0.01 ± 0.04	87.93 ± 1.51
Dyn-CPWS-AP	64	75.64 ± 0.82	87.02 ± 0.46	95.41 ± 0.08	85.17 ± 1.15	0.01 ± 0.04	92.65 ± 0.48	0.0 ± 0.0	84.48 ± 1.69
Vol-CPWS-AP	64	**81.62 ± 0.95**	84.22 ± 0.77	**96.09 ± 0.13**	**88.98 ± 0.68**	0.27 ± 0.36	**93.81 ± 0.61**	0.0 ± 0.0	89.46 ± 0.67

TABLE IX
COMPARISON OF PER-CLASS F1 SCORES OBTAINED WITH VOL-CPWS-AP FOR THE TEST IMAGES OF ZURICH WITH THE RESULTS OF TWO DEEP LEARNING METHODS REPORTED IN [34]

Method	F1 scores per class								Mean F1
	Roads	Buildings	Trees	Grass	Bare Soil	Water	Railways	Swimming Pools	
Vol-CPWS-AP	79.32 ± 0.71	82.06 ± 0.9	91.83 ± 0.26	91.26 ± 0.38	0.53 ± 0.7	96.53 ± 0.31	0.00 ± 0.00	94.43 ± 0.37	67.0
Deep learning models:									
FCN-Festa+dCFR [34]	71.74 ± 2.78	75.81 ± 4.18	81.20 ± 1.60	83.44 ± 1.51	66.49 ± 15.57	94.68 ± 0.52	0.00 ± 0.00	82.06 ± 6.80	69.43
FCN [34]	88.34	93.27	92.40	89.48	67.96	96.87	2.98	88.10	77.42

of 330,767 annotated samples from all semantic classes, except for background/clutter. As we can see in Tables VII and IX, our Vol-CPWS-AP outperformed FCN-Festa-dCRF in terms of OA and F1 scores for all semantic classes, except for the bare soil class. We note that our results were achieved by using less than a half of their number of training samples. On the other hand, in contrast to our experiments, the authors of [34] do not consider the blue channel of the Zurich dataset, which may explain the gap between both methods. Moreover, APs still falls behind the state-of-the-art FCN model trained on dense annotations (*i.e.,* trained on all pixels of the training set).

Fig. 9 illustrates the classification results on the test image *zh18* of the Zurich dataset. As already suggested by the scores given in Table VIII, we can see that none of the APs are able to improve the results for the bare soil class, but most of them provide better classification results for the roads (in black) and grass (in light green). When comparing the WS-APs and CPWS-APs, the most significant improvements are observed on the roads, trees, and grass classes, in the regions highlighted by the blue boxes in Fig. 9(m)–(o). The red boxes indicate a few regions in which classification results were worsened by the use of semantic prior knowledge.

B. Building Extraction

Taking into account that other methods in the literature, such as [10], [11] which also employ prior knowledge for image segmentation, perform well on binary classification/ segmentation of images into object and background regions, we aim to validate our watershed-APs on a binary classification task, which is meaningful for remote sensing imagery, namely building extraction. For evaluating our proposed methods on this task, we consider two semantic classes on the Zurich and Vaihingen datasets, namely *buildings* and *background*, such that the latter is the union of all remaining classes of each dataset, as mentioned in Section V-B. Evaluating our methods on this binary classification task might give us a clearer idea of their performance, since semantic prior knowledge come from only two classes. Moreover, we no longer have semantic classes, *e.g.,* swimming pools, which are absent in some of the Zurich test images.

Table IV reports the evaluation results of building extraction on the Vaihingen dataset. On this dataset, the highest F1 score among the APs were achieved with the Area-FPWS-AP, which outperformed one of the best methods in the literature, namely AP, by 0.51%, 1.90% and 2.59% in terms of OA, F1, and mIOU, respectively. Regarding the use of prior knowledge on watershed-APs, both Area-CPWS-AP and Area-FPWS-AP (resp. Vol-CPWS-AP and Vol-FPWS-AP) outperformed Area-WS-AP (resp. Vol-WS-AP) with respect to all metrics, except for precision, which validates the interests of semantic prior knowledge in the context of building extraction. The largest improvement was observed for the watershed-AP filtered with area, with the Area-FPWS-AP being 1.24%, 7.70%, 4.08% and 5.47% better than Area-WS-AP with respect to OA, recall, F1 and mIOU scores, respectively. All three FPWS-AP presented very similar results and outperformed their respective WS-AP counterparts by more than 6%, 3%, and 4%, respectively, in terms of recall, F1, and mIOU scores.

Classification results on one image of the Vaihingen dataset are illustrated in Fig. 10. The true positives, false negatives and false positives are represented in white, red, and cyan, respectively. On this image, we can observe that the area and volume watershed-APs produce less noisy classification results when compared to the other methods. In particular, the Vol-CPWS-AP presents a higher precision than all other methods.

Table XI presents the evaluation of building extraction on the Zurich dataset. With respect to the baseline NIR+RGB, all APs (except for AP-minT) provided better scores for most metrics. For each metric, the highest score was achieved by one of the watershed-APs, with Area-FPWS-AP giving the highest OA, F1, and mIOU scores among all tested methods. In general,

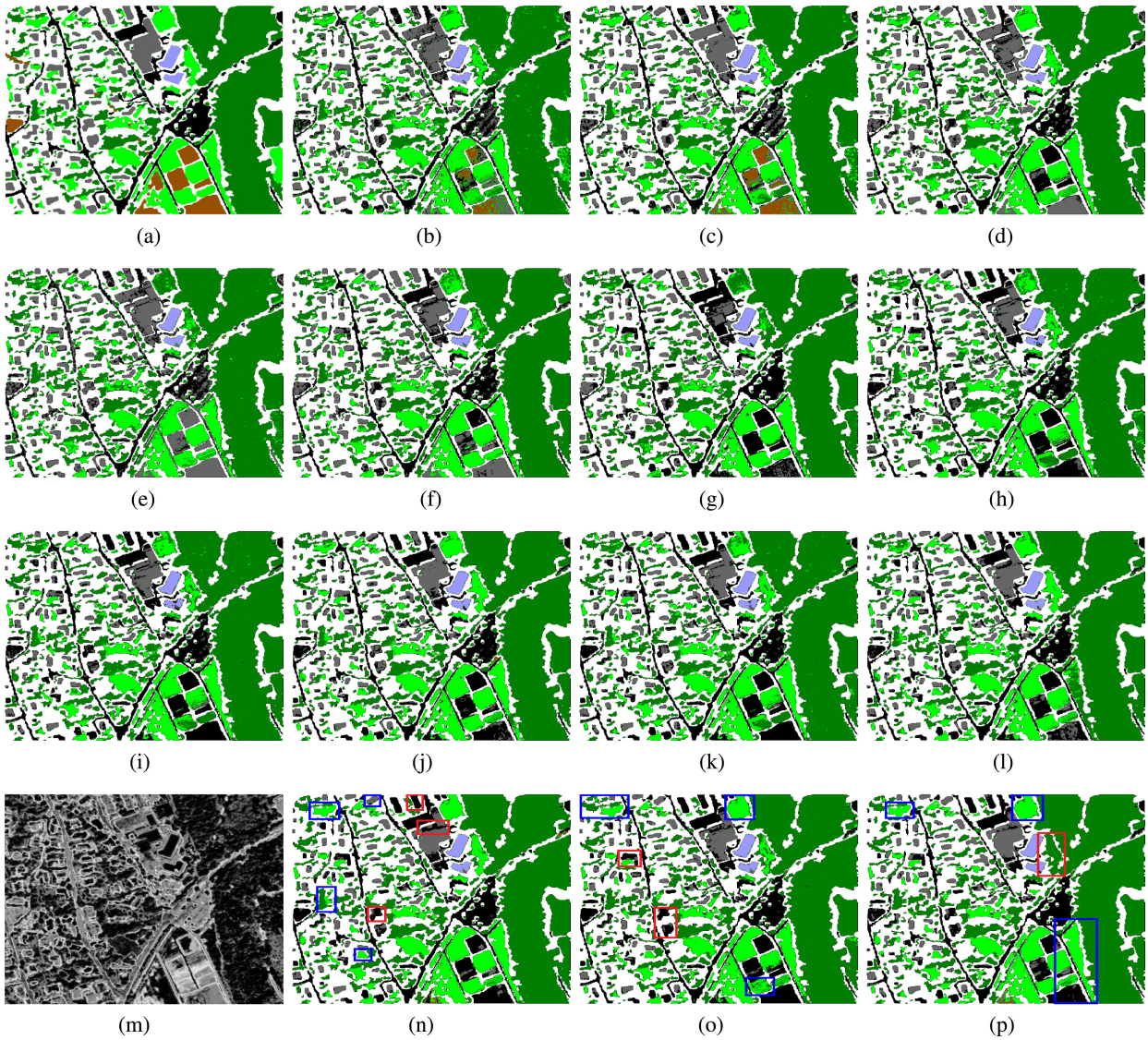

Fig. 9. Ground-truth, classification results and classification probability map μ (used to computed the CPWS-APs) of the image *zh18* of the Zurich dataset: roads, buildings, trees, grass, bare soil, water, railways, swiming pools, background. (a) Ground-truth. (b) NIR+RGB. (c) NIR+RGB$_{5x5}$. (d) AP-maxT. (e) AP-minT. (f) AP. (g) SDAP. (h) α-AP. (i) ω-AP. (j) Area-WS-AP. (k) Dyn-WS-AP. (l) Vol-WS-AP. (m) Probability map (μ). (n) Area-CPWS-AP. (o) Dyn-CPWS-AP. (p) Vol-CPWS-AP.

FPWS-APs led to more pixels being classified as belonging to the *building* class, increasing the number of true and false positives, as attested by the lower precision and higher recall scores of FPWS-APs when compared to CPWS-APs. Still, Area-FPWS-APs (resp. Dyn-FPWS-APs) achieved F1 scores at least 4% higher than Area-WS-AP and Area-CPWS-AP (resp. Dyn-WS-AP and Dyn-CPWS-AP).

In conclusion, watershed-APs computed with the help of semantic prior knowledge showed its usefulness in both land-cover classification and building extraction on multispectral data. On both tested datasets, the highest scores were achieved by one of the watershed-APs, whether it was with CPWS-APs or FPWS-APs. Based on our experiments, the effectiveness of considering semantic prior knowledge either during the construction phase of hierarchical watersheds or during the filtering step of watershed-APs depends on the task and on the dataset under study. For land-cover pixel classification, CPWS-APs seem to be more adapted than FPWS-APs when there is a high imbalance between the number of semantic classes, which are present in each test image, such as in the Zurich dataset. On the other hand, FPWS-APs performed better than CPWS-APs on the Zurich dataset when the data was divided in only two classes: building and background.

VII. Conclusion

This article proposed the watershed-AP as an extension of AP to hierarchical watersheds computed from (edge) weighted graphs. Besides, the relevance of using semantic prior knowledge was considered in the construction and filtering of such

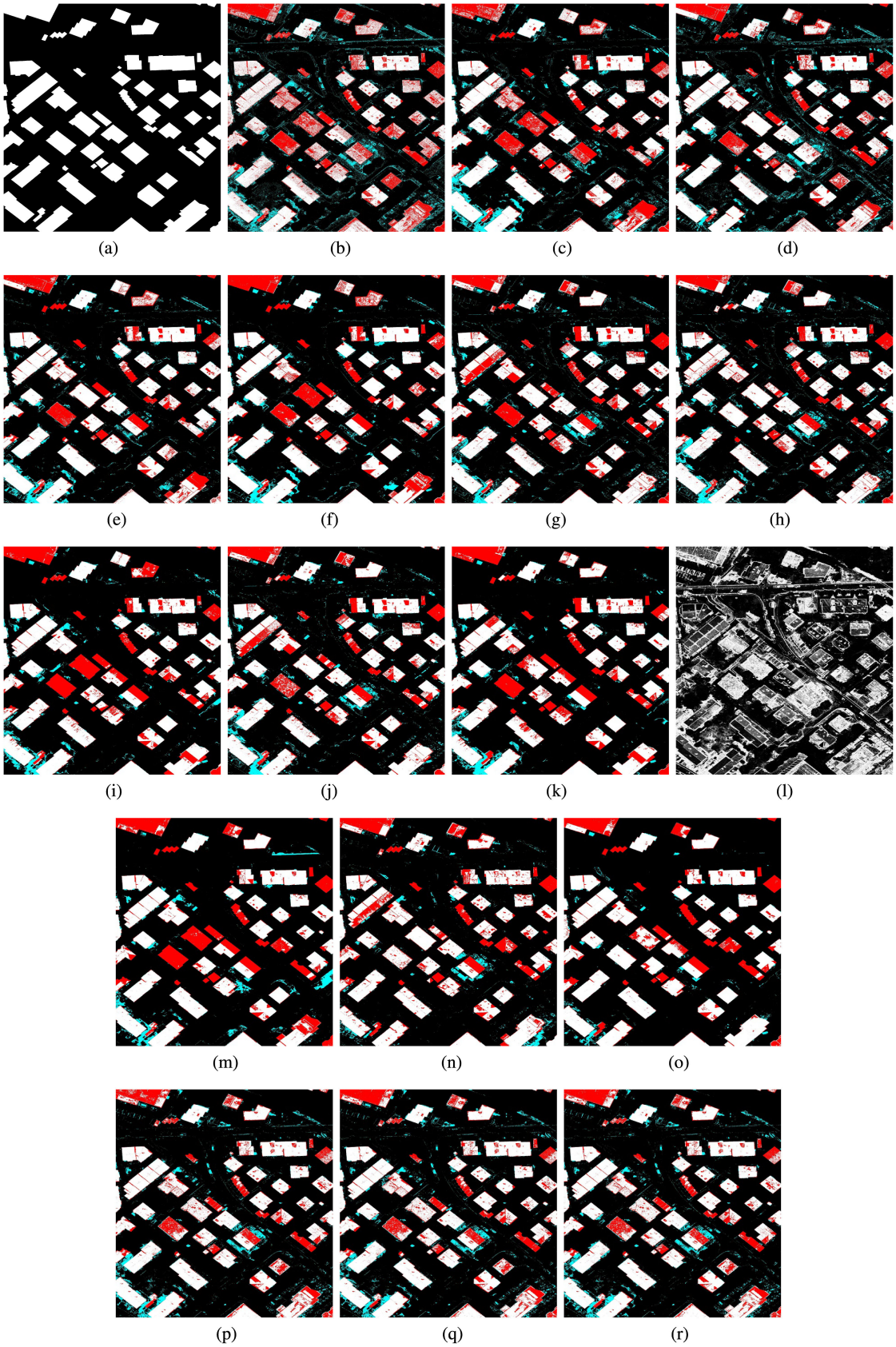

Fig. 10. Ground-truth, classification results and classification probability map μ (used to computed the CPWS-APs) of the image number 30 of the Vaihingen dataset: ☐ true positive, ■ false negative, ■ false positive. (a) Ground-truth. (b) NIR+RG. (c) AP-maxT. (d) AP-minT. (e) AP. (f) SDAP. (g) α-AP. (h) ω-AP. (i) Area-WS-AP. (j) Dyn-WS-AP. (k) Vol-WS-AP. (l) Probability map (μ). (m) Area-CPWS-AP. (n) Dyn-CPWS-AP. (o) Vol-CPWS-AP. (p) Area-FPWS-AP. (q) Dyn-FPWS-AP. (r) Vol-FPWS-AP.

TABLE X
AVERAGE CLASSIFICATION RESULTS OF BUILDING EXTRACTION ON THE VAIHINGEN DATASET

Method	**Dim.**	OA	Precision	Recall	F1	mIOU
NIR+RG	3	86.97 ± 0.03	77.77 ± 0.15	66.63 ± 0.17	71.77 ± 0.08	55.97 ± 0.10
AP-maxT [3]	48	88.12 ± 0.16	82.30 ± 0.37	66.55 ± 0.39	73.59 ± 0.37	58.22 ± 0.46
AP-minT [3]	48	88.98 ± 0.07	85.02 ± 0.19	67.61 ± 0.39	75.32 ± 0.22	60.42 ± 0.29
AP [3]	90	90.29 ± 0.07	89.24 ± 0.27	69.29 ± 0.21	78.01 ± 0.16	63.95 ± 0.21
SDAP [32]	48	89.53 ± 0.08	88.90 ± 0.18	66.15 ± 0.28	75.86 ± 0.20	61.11 ± 0.26
α-AP [25]	48	88.89 ± 0.12	85.55 ± 0.32	66.56 ± 0.36	74.87 ± 0.30	59.83 ± 0.38
ω-AP [25]	48	89.21 ± 0.12	85.10 ± 0.28	68.62 ± 0.46	75.98 ± 0.31	61.26 ± 0.40
Area-WS-AP	48	89.56 ± 0.07	89.40 ± 0.20	65.84 ± 0.24	75.83 ± 0.17	61.07 ± 0.23
Dyn-WS-AP	48	89.08 ± 0.07	85.88 ± 0.18	67.14 ± 0.32	75.36 ± 0.21	60.47 ± 0.27
Vol-WS-AP	48	89.69 ± 0.12	89.12 ± 0.26	66.69 ± 0.33	76.29 ± 0.29	61.66 ± 0.38
Watershed-APs computed with semantic prior knowledge:						
Area-CPWS-AP	48	89.90 ± 0.15	90.13 ± 0.49	66.71 ± 0.62	76.66 ± 0.40	62.16 ± 0.53
Dyn-CPWS-AP	48	89.20 ± 0.14	**92.34 ± 0.40**	61.67 ± 0.78	73.95 ± 0.48	58.67 ± 0.61
Vol-CPWS-AP	48	**90.86 ± 0.20**	91.96 ± 0.61	69.32 ± 0.61	79.05 ± 0.49	65.36 ± 0.67
Watershed-APs filtered with semantic prior knowledge:						
Area-FPWS-AP	45	90.80 ± 0.03	87.48 ± 0.17	**73.54 ± 0.24**	**79.91 ± 0.09**	**66.54 ± 0.13**
Dyn-FPWS-AP	45	89.92 ± 0.14	83.70 ± 0.53	73.86 ± 0.35	78.47 ± 0.26	64.57 ± 0.36
Vol-FPWS-AP	45	90.79 ± 0.04	87.43 ± 0.17	**73.54 ± 0.29**	79.89 ± 0.12	66.51 ± 0.16

The training samples are extracted from the training images (Id 1,3,5,7,13,17,21,23,26,32,37]). The samples are divided in only two classes: background and buildings.

TABLE XI
AVERAGE CLASSIFICATION RESULTS OF BUILDING EXTRACTION ON THE ZURICH DATASET FOR THE IMAGES 16-20 OF ZURICH

Method	**Dim.**	OA	Precision	Recall	F1	mIOU
NIR+RGB	4	87.65 ± 0.06	57.88 ± 0.31	47.26 ± 0.22	52.03 ± 0.20	35.16 ± 0.18
AP-maxT [3]	64	89.32 ± 0.06	69.65 ± 0.51	43.64 ± 0.88	53.65 ± 0.59	36.66 ± 0.55
AP-minT [3]	64	88.14 ± 0.13	62.56 ± 0.93	40.63 ± 0.57	49.26 ± 0.49	32.68 ± 0.43
AP [3]	120	89.67 ± 0.15	74.90 ± 0.54	40.76 ± 1.19	52.78 ± 1.07	35.86 ± 0.99
SDAP [32]	64	89.90 ± 0.09	71.53 ± 0.33	47.75 ± 0.97	57.26 ± 0.69	40.12 ± 0.68
α-AP [25]	64	90.27 ± 0.06	74.26 ± 0.59	47.94 ± 0.39	58.26 ± 0.24	41.11 ± 0.24
ω-AP [25]	64	90.27 ± 0.08	73.86 ± 0.81	48.45 ± 0.57	58.51 ± 0.33	41.36 ± 0.33
Area-WS-AP	64	90.79 ± 0.05	76.91 ± 0.33	49.99 ± 0.39	60.59 ± 0.28	43.47 ± 0.29
Dyn-WS-AP	64	90.40 ± 0.03	74.80 ± 0.22	48.67 ± 0.36	58.97 ± 0.25	41.81 ± 0.25
Vol-WS-AP	64	91.06 ± 0.06	**77.63 ± 0.28**	51.82 ± 0.51	62.15 ± 0.38	45.09 ± 0.40
Watershed-APs computed with semantic prior knowledge:						
Area-CPWS-AP	64	90.67 ± 0.12	76.80 ± 0.72	48.96 ± 1.12	59.79 ± 0.80	42.64 ± 0.81
Dyn-CPWS-AP	64	89.79 ± 0.08	68.10 ± 0.81	52.53 ± 1.04	59.30 ± 0.45	42.15 ± 0.45
Vol-CPWS-AP	64	90.89 ± 0.14	75.98 ± 1.04	52.25 ± 0.88	61.91 ± 0.65	44.84 ± 0.68
Watershed-APs filtered with semantic prior knowledge:						
Area-FPWS-AP	60	**91.08 ± 0.11**	73.03 ± 0.71	58.75 ± 0.25	**65.11 ± 0.33**	**48.27 ± 0.36**
Dyn-FPWS-AP	60	90.46 ± 0.08	68.76 ± 0.40	**59.87 ± 0.43**	64.01 ± 0.31	47.07 ± 0.33
Vol-FPWS-AP	60	90.85 ± 0.23	71.60 ± 1.30	58.70 ± 0.42	64.50 ± 0.66	47.61 ± 0.73

The samples are divided in only two classes: background and buildings.

hierarchies. We evaluated and validated our approach on the pixel classification and building extraction of two multispectral remote sensing images, which showed the potential of hierarchical watersheds in this field. On both datasets, the best results were achieved by watershed-APs computed or filtered with prior knowledge from training pixels.

As future work, we will to explore the versatility of hierarchical watersheds by considering other methods to obtain the gradient of remote sensing images, as well as different ways of obtaining prior knowledge from training pixels and of incorporate such knowledge into the watershed-AP construction pipeline. Also, the proposed approach has a great potential to be adapted and applied to other remote sensing data such as synthetic aperture radar images or LiDAR point clouds.

REFERENCES

[1] I. Destival, "Mathematical morphology applied to remote sensing," *Acta Astronautica*, vol. 13, no. 6–7, pp. 371–385, 1986.
[2] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2042–2055, Sep. 2002.
[3] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[4] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
[5] D. S. Maia, M.-T. Pham, E. Aptoula, F. Guiotte, and S. Lefèvre, "Classification of remote sensing data with morphological attributes profiles: A decade of advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 43–71, Sep. 2021.

[6] J. Cousty, G. Bertrand, L. Najman, and M. Couprie, "Watershed cuts: Minimum spanning forests and the drop of water principle," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 8, pp. 1362–1374, Aug. 2009.

[7] J. Cousty, L. Najman, and B. Perret, "Constructive links between some morphological hierarchies on edge-weighted graphs," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Berlin, Heidelberg: Springer, 2013, pp. 86–97.

[8] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation," *Math. Morphol. Image Process.*, vol. 34, pp. 433–481, 1993.

[9] D. S. Maia, M.-T. Pham, and S. Lefèvre, "Watershed-based attribute profiles for pixel classification of remote sensing data," in *Proc. Int. Conf. Discrete Geometry Math. Morphol.*, Cham: Springer, 2021, pp. 120–133.

[10] S. Lefèvre, "Knowledge from markers in watershed segmentation," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Berlin, Heidelberg: Springer, 2007, pp. 579–586.

[11] P. A. De Miranda, A. X. Falcão, and J. K. Udupa, "Synergistic ARC-weight estimation for interactive image segmentation using graphs," *Comput. Vis. Image Understanding*, vol. 114, no. 1, pp. 85–99, 2010.

[12] S. Derivaux, G. Forestier, C. Wemmert, and S. Lefèvre, "Supervised image segmentation using watershed transform, fuzzy classification and evolutionary computation," *Pattern Recognit. Lett.*, vol. 31, no. 15, pp. 2364–2374, 2010.

[13] S. Derivaux, S. Lefevre, C. Wemmert, and J. Korczak, "Watershed segmentation of remotely sensed images based on a supervised fuzzy pixel classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2006, pp. 3712–3715.

[14] S. Aksoy *et al.*, "Performance evaluation of building detection and digital surface model extraction algorithms: Outcomes of the PRRS 2008 algorithm performance contest," in *Proc. IAPR Workshop Pattern Recognit. Remote Sens.*, 2008, pp. 1–12.

[15] S. Lefevre, A. Puissant, and F. Levoy, "Weakly supervised image segmentation: Application to mapping and monitoring of salt marsh vegetation in the mont-saint-michel bay from high resolution imagery," in *Proc. ESA-EUSC-JRC 2011-Image Inf. Mining: Geospatial Intell. Earth Observ.*, 2011, pp. 4.

[16] N. Courty, E. Aptoula, and S. Lefèvre, "A classwise supervised ordering approach for morphology based hyperspectral image classification," in *Proc. 21st Int. Conf. Pattern Recognit.*, 2012, pp. 1997–2000.

[17] S. Velasco-Forero and J. Angulo, "Supervised ordering in $\mathbb{R}^p$: Application to morphological processing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3301–3308, Nov. 2011.

[18] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "Marker-based hierarchical segmentation and classification approach for hyperspectral imagery," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 1089–1092.

[19] A. Fehri, S. Velasco-Forero, and F. Meyer, "Prior-based hierarchical segmentation highlighting structures of interest," *Math. Morphol.-Theory Appl.*, vol. 3, no. 1, pp. 29–44, 2019.

[20] S. Lefèvre, L. Chapel, and F. Merciol, "Hyperspectral image classification from multiscale description with constrained connectivity and metric learning," in *Proc. 6th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2014, pp. 1–4.

[21] C. He, S. Li, D. Xiong, P. Fang, and M. Liao, "Remote sensing image semantic segmentation based on edge information guidance," *Remote Sens.*, vol. 12, no. 9, Art. no. 1501, 2020.

[22] L. E. Cué La Rosa, R. Queiroz Feitosa, P. Nigri Happ, I. Del'Arco Sanches, and G. A. Ostwald Pedro daCosta, "Combining deep learning and prior knowledge for crop mapping in tropical regions from multitemporal SAR image sequences," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2029.

[23] M.-T. Pham, E. Aptoula, and S. Lefèvre, "Classification of remote sensing images using attribute profiles and feature profiles from different trees: a comparative study," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4511–4514.

[24] P. Bosilj, E. Kijak, and S. Lefèvre, "Partition and inclusion hierarchies of images: A comprehensive survey," *J. Imag.*, vol. 4, no. 22018, Art. no. 33.

[25] P. Bosilj, B. B. Damodaran, E. Aptoula, M. Dalla Mura, and S. Lefèvre, "Attribute profiles from partitioning trees," in *Proc. Int. Symp. Math. Morphol. Appl. Signal Image Process.*, 2017, pp. 381–392.

[26] M.-T. Pham, S. Lefèvre, E. Aptoula, and L. Bruzzone, "Recent developments from attribute profiles for remote sensing image classification," in *Proc. Int. Conf. Pattern Recognit. Artif. Intell.*, 2018, pp. 102–107.

[27] M. Grimaud, "New measure of contrast: The dynamics," in *Image Algebra and Morphological Image Processing III*, vol. 1769. Int. Soc. Optics Photon., 1992, pp. 292–305.

[28] C. Vachier and F. Meyer, "Extinction value: A new measurement of persistence," in *Proc. IEEE Workshop Nonlinear Signal Image Process.*, 1995, vol. 1, pp. 254–257.

[29] L. Najman, J. Cousty, and B. Perret, "Playing with kruskal: Algorithms for morphological trees in edge-weighted graphs," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Berlin, Heidelberg: Springer, 2013, pp. 135–146.

[30] M. Volpi and V. Ferrari, "Semantic segmentation of urban scenes by learning local class interactions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 1–9.

[31] Accessed: Aug. 16, 2021. [Online]. Available: https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-vaihingen/

[32] M. Dalla Mura, J. Benediktsson, and L. Bruzzone, "Self-dual attribute profiles for the analysis of remote sensing images," in *Proc. Int. Symp. Math. Morphol. Appl. Signal Image Process.*, 2011, pp. 320–330.

[33] B. Perret, G. Chierchia, J. Cousty, S. Guimarães, Y. Kenmochi, and L. Najman, "Higra: Hierarchical graph analysis," *SoftwareX*, vol. 10, 2019, Art. no. 100335.

[34] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmentation of remote sensing images with sparse annotations," *IEEE Geoscience Remote Sensing Lett.*, vol. 19, pp. 1–5, 2021.

[35] Y. Cui, L. Chapel, and S. Lefèvre, "Scalable bag of subpaths kernel for learning on hierarchical image representations and multi-source remote sensing data classification," *Remote Sens.*, vol. 9, no. 3,2017, Art. no. 196.