



Classification of Remote Sensing Data With Morphological Attribute Profiles

©SHUTTERSTOCK.COM/LAVIZZARA

A decade of advances

**DEISE SANTANA MAIA, MINH-TAN PHAM,
ERCHAN APTOULA, FLORENT GUIOTTE,
AND SÉBASTIEN LEFÈVRE**

Morphological attribute profiles (APs) are among the most prominent methods for spatial-spectral pixel analysis of remote sensing images. Since their introduction a decade ago to tackle land cover classification, many studies have been contributed to the state of the art, focusing not only on their application to a wider range of tasks but also on their performance improvement and extension to more complex Earth observation data.

Despite the overwhelming proliferation of deep learning-based methods in the past five years, APs are far from obsolete, due mainly to their high flexibility, low computational cost, lower training data requirement, and rigorous mathematical foundation. In this survey, an entire decade's worth of more than 100 AP-related contributions to the field of remote sensing have been compiled, providing an extensive panorama of this robust and effective tool. Moreover, a collective experimental comparison of the reviewed

AP variations is provided as well, not only in terms of classification performance but, for the first time, in terms of their generalization capacity, too.

BACKGROUND

Classification constitutes one of the paramount tasks of remote sensing image analysis for Earth observation. Its performance is critical for the success of land use and cover mapping and monitoring. The rapid development of high-resolution (HR) and very-HR (VHR) image acquisition technologies has led to increasingly complex images with higher levels of detail. Consequently, the importance of the joint exploitation of spatially contextual information along with spectral pixel characteristics became clear early on [1].

Morphological profiles (MPs) [2] were introduced almost two decades ago to address this exact issue. In essence, they produce multiscale descriptions of their input through the application of a sequence of morphological reconstruction-based filters using structuring elements of various sizes (and shapes) [2]. Their rigorous mathematical foundation and inherent ability to capture spatial-spectral information have led to the development of several variants [3]–[6].

However, as spatial resolutions and image sizes progressively increased, their relatively high computational cost has been drastically accentuated. Moreover, their initially celebrated capacity of capturing size and shape variations eventually became extremely limited when confronted with thematic classes differing in terms of alternative properties, such as contrast, homogeneity, and so on. In an effort to overcome these shortcomings, MPs have been

Digital Object Identifier 10.1109/MGRS.2021.3051859
Date of current version: 1 March 2021

generalized into morphological APs through the seminal work of Dalla Mura et al. [7].

Unlike MPs, APs employ attribute filters (AFs) [8], a powerful class of connected morphological filters capable of removing entire connected components (CCs) with respect to arbitrarily defined attributes (e.g., geometric, statistical, and so on), thus eliminating the size/shape limitation of MPs. Furthermore, APs can be generated efficiently through quasilinear [9] and parallel [10] algorithms as well as tree-based hierarchical representation of their inputs [11], thus equipping them with a high level of scalability—an invaluable property in remote sensing, where gigapixel images are becoming the norm. In fact, [12] presents an application of AP at a terapixel scale!

It is thus not surprising that, since their introduction 10 years ago, a great number of AP-related publications have appeared (4,540 on Google Scholar as of June 2020), tackling various aspects of remote sensing image analysis. Many among them have been dedicated to further improving APs through a rich variety of extensions, focusing on every single stage of their calculation.

Even though recent years have witnessed the overwhelming proliferation of deep learning [13] (overshadowing, performancewise, most nondeep feature extraction methods), APs continue to withstand the test of time. In addition to their aforementioned invaluable properties, this is also due to their ability to perform even with limited amounts of training data as well as their capacity to accommodate the arbitrary modalities and challenging types of images so often encountered in the remote sensing domain.

This article (a short version of which appeared at the 2018 International Conference on Pattern Recognition and Artificial Intelligence) presents a survey of APs and contributes to the state of the art in the following ways:

- We provide a comprehensive review of 10 years of advances on APs by decomposing their calculation into four stages and grouping the reviewed studies accordingly. Evidently, this is not the first survey on APs [14], [15]. Contrary to [14] from five years ago, this survey is not limited by the vanilla definition of APs and their focus on hyperspectral data, and, compared to [15], it additionally addresses APs on partition tree structures, threshold-selection techniques, and AP postprocessing.
- We present the results of an extensive series of classification experiments with multiple real data sets intended to measure the performance of the various reviewed AP variants. However, the experiments have been conducted for the first time in terms of connectivity type as well as spectral quantization level.
- More importantly, we present the results of pixel-classification experiments intended to measure not only the performance of the various reviewed AP variants but their generalization capacity as well. In the big data era, generalization is a core and sought-after property of any content description tool, one for which APs have been criticized, as an image's tree representation contains

both training and testing elements [16]. Mixing training/testing sets is, unfortunately, not uncommon with deep learning either [17]. We underline this validation malpractice, often encountered in the state of the art, and propose a solution through an image's spatial subdivision and independent tree construction.

- To promote reproducible research and ease generalization, we also provide the first available open source library for APs (Simple Attribute Profiles [SAP]), available at <https://gitlab.inria.fr/fguillette/sap>.

THE PRINCIPLE OF APs

APs are multilevel image-description tools obtained by successively applying a set of morphological AFs [7]. Unlike usual image filtering operators, which are directly performed on the pixel level, AFs work on the CC level based on the concept of image connectivity. In particular, AFs are applied on CCs with regard to a predicate based on an arbitrary statistical or geometric property thereof. Consequently, they exhibit a higher level of flexibility with respect to operators by reconstruction, which are severely limited by characterizing only the size and shape of their input. This advantage naturally extends to APs versus MPs as well [2], [7].

The generation of APs [7] from an input image can be summarized as a four-step process (see Figure 1):

- 1) construction of the image's hierarchical tree representation, where CCs are denoted as nodes
- 2) computation of one or more relevant attributes describing the geometrical and statistical features from each tree node
- 3) filtering the tree by preserving/removing nodes according to their attribute values compared against predefined thresholds
- 4) reconstruction of the image from the filtered tree.

Step one can be performed using different pixel connectivity rules. For 2D images, four- and eight-connectivity are the most common. Steps three and four can be implemented for different attributes (with varying threshold values) to finally produce a set of filtered images (by stacking them), forming the APs.

More formally, according to the seminal work of [7], given a grayscale image $X: E \rightarrow \mathbb{Z}$, $E \subseteq \mathbb{Z}^2$, the calculation of APs on X is achieved by applying a sequence of AFs based on a min-tree (i.e., attribute-thickening operators $\{\phi_k\}_{k=1}^K$) and max-tree (i.e., attribute-thinning operators $\{\gamma_k\}_{k=1}^K$) as follows:

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

where $\phi_k(X)$ denotes the filtered image obtained by applying the attribute thickening ϕ with regard to the threshold k . A similar explanation is made for $\gamma_k(X)$. As observed, the resulting $AP(X)$ is a stack of $(2K + 1)$ images including the original image, K filtered images from the thickening profiles, and the other K from the thinning profiles.

A toy example of AFs is presented in Figure 2. Given the grayscale image $X: E \rightarrow [0, 1, 2]$ of Figure 2(a), we first obtain the max-tree T of X using four-connectivity. Then, we compute the area (number of pixels) of the nodes of T . Subsequently, we prune the nodes of T with an area less than a given parameter k . In our case, k is equal to eight, and the nodes composed of fewer than eight pixels are pruned from the tree. Finally, we reconstruct the image from the pruned tree, resulting in X' . Hence, X' is the area thinning of X for $k=8$.

RECENT ADVANCES FROM APs

Each of the four AP construction stages has received various forms of extensions and contributions from the scientific community. Moreover, AP-based image analysis pipelines often include some form of preprocessing, usually to adapt multiband input [since (1) expects single-band data] as well as postprocessing steps to increase description capability.

We now revisit the recently proposed developments that have provided significant contributions to the AP framework for remote sensing image classification. Here, our study focuses on the following key concepts:

- ▶ the adaptation of APs to various modalities besides single-band images and, in particular, multiband data (the “Input Data” section)
- ▶ the construction of APs using various hierarchical image representations (the “Tree Construction” section)
- ▶ the determination of attributes and thresholds (the “Attribute and Threshold Selection” section)
- ▶ tree-filtering rules used in the construction of APs (the “Tree Filtering” section)
- ▶ postprocessing techniques intended for feature enhancement (the “Postprocessing of Output Profiles” section).

For other AP-related notions, including profiles computed with different morphological filters and image-reconstruction techniques, we refer readers to the related references for additional details (the “Extensions and Generalization of APs” section).

INPUT DATA

Since APs were originally proposed to deal with only single-band images [7], their adaptation to other kinds of remote sensing data became necessary early on. In particular, their adaptation to multichannel images (multispectral and

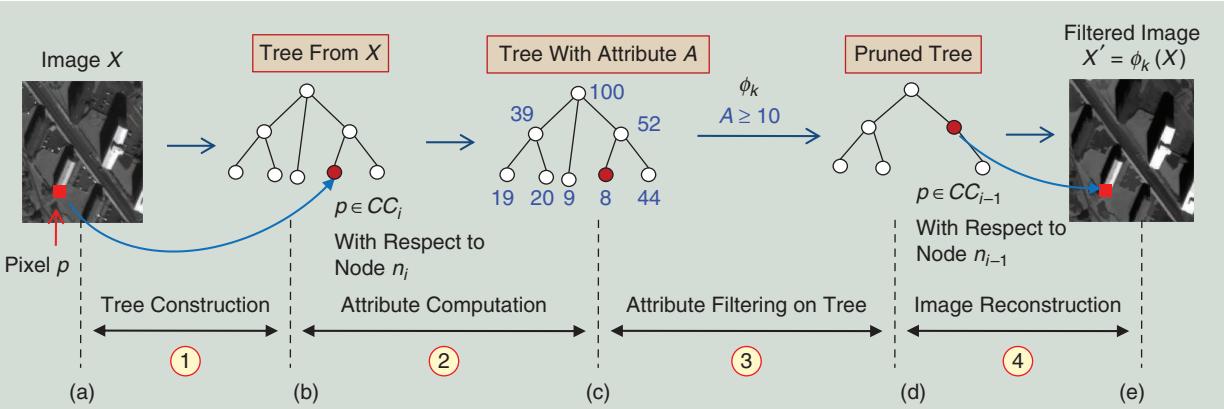


FIGURE 1. The AP generation framework, which involves four main stages: (1) tree construction, (2) attribute computation, (3) tree-based attribute filtering (pruning), and (4) image reconstruction from the filtered (pruned) tree. The (a) image, (b) max-tree, (c) area attribute, (d) area filtering, and (e) reconstruction are shown. (Source: [18]; used with permission.)

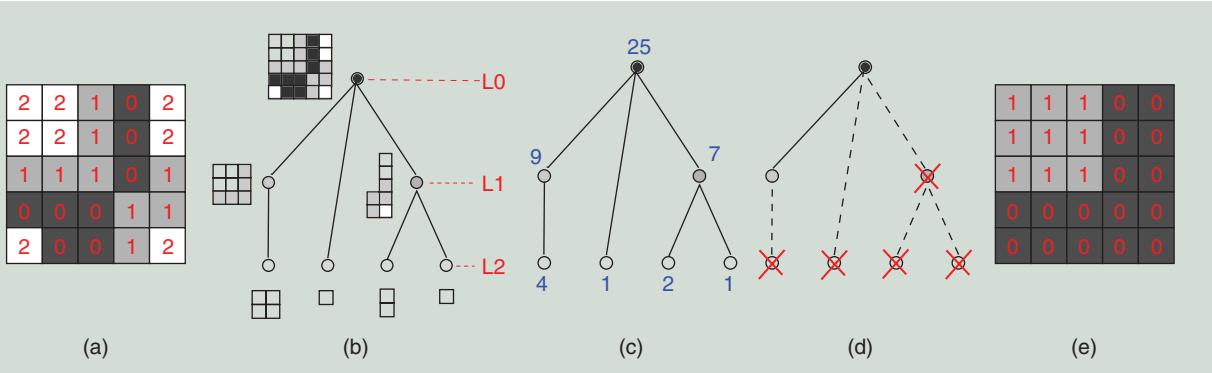


FIGURE 2. (a) The original grayscale image $X: E \rightarrow [0, 1, 2]$. (b) The max-tree T of X computed using four-connectivity. (c) The area (number of pixels) of the nodes of T . (d) The area filtering of T with regard to the threshold $k = 8$. (e) The image X' reconstructed from the pruned tree.

hyperspectral) became an intensive research topic. The main idea in this regard has been to reduce the various and often correlated image bands into fewer components through some dimension-reduction tool and extract APs independently from each of them, followed by their subsequent merging.

The initial extension of APs to hyperspectral images was proposed in [19], namely, extended AP (EAP), which consists of employing principal component analysis (PCA) to this end—an unsupervised yet often suboptimal tool due to its linear nature. Alternatives to PCA that have been additionally studied include independent component analysis (ICA) [20], [21], kernel PCA (KPCA) [22], feature space discriminant analysis (FSDA) [23], and other supervised methods, such as discriminant analysis feature extraction (DAFE) [24], nonparametric weighted feature extraction (NWFE) [25], and the sparse Hilbert–Schmidt independence criterion and surrogate kernel [26], among others.

Given that an independent AP calculation from each band (or each image component, assuming dimension reduction has been applied to the image bands) ignores any and all correlational information among them, [27] proposed a vector alternative calculating APs collectively and simultaneously from all available bands/components. The core idea relies on representing all bands through a single hierarchical representation, leading to vector APs.

It would be also possible to compute APs on derived features of optical remote sensing data, such as edge/contour information obtained by Sobel gradient filtering of panchromatic images [10] or the normalized difference vegetation index from multispectral images [28] for urban and crop field classification. Then, to deal with VHR optical images, where highly textural information becomes significant, another adaptation of APs was proposed in [29]. The raw input data are replaced by the textural features, thus considerably improving the classification performance on the tested textured images.

Furthermore, although APs are usually explored for the supervised classification of optical images (either panchromatic, multispectral, or hyperspectral), some studies have investigated them within unsupervised scenarios. For example, in [30], the authors exploited the differential APs (i.e., which compute the difference between successive APs to form differential profiles) for unsupervised anomaly detection in hyperspectral images, where they observed that the anomalies and background of an image are enhanced in the thinning and thickening profiles, respectively. Some other examples are the retrieval of building height using panchromatic angular images [31] and change detection in temporal panchromatic images [32].

While there has been strong focus on the application of APs to optical remote sensing data alternative remote sensing image types have received far less attention. One may witness some tentative works on synthetic aperture radar (SAR) and polarimetric SAR images for segmentation [33], building detection [34], crop field and land cover

classification [35], [36], and change detection [37]–[39] using the original APs and differential APs; on passive microwave remote sensing image analysis [40]; on lidar data for building detection [41] and land cover classification [16], [28], [42]–[44]; on satellite image time series classification using *Sentinel-2* data [45], [46]; and on the fusion of APs and extinction profiles (EPs) (a variant of APs discussed in the “Extensions and Generalization of APs” section) of hyperspectral and lidar data using composite kernel support vector machine (SVM) [47], [48], and deep learning approaches [49], [50] for land cover classification. This is still an open topic for ongoing and future research.

TREE CONSTRUCTION

Although AFs were introduced more than 20 years ago [8], the relatively late popularity of APs is mainly due to computational issues revolving around the efficient calculation of CCs, which were resolved, to a significant degree, through the tree representation of images [11]. They are of paramount importance for the computation of APs since the trees need to be computed only once and then multiple filtering outputs can be derived easily from them.

Even though the seminal work of Dalla Mura et al. [7] relies on component trees for the implementation of APs, the type of tree is independent from the rest of the procedure; thus, component trees can be replaced by alternative tree types. Consequently and not surprisingly, a number of reported works explore such options; e.g., tree of shapes (ToS) as well as alpha and omega trees, each with its own set of properties. Despite a plethora of tree representations for modeling CCs in mathematical morphology (not all of which have been implemented with the purpose of AP construction), they can all be categorized into inclusion and partitioning hierarchies. We invite interested readers to refer to the recent comprehensive survey of partition and inclusion hierarchies of images conducted by Bosilj et al. [51].

Inclusion hierarchies constitute partial partitions of a given image with nested supports, and their components are formed by creating, inflating, and merging image blocks [52]. Moreover, inclusion trees require the presence of a total ordering relation imposed on the set of image pixel values, which evidently renders their extension to multivariate images challenging since ordering vectors is not straightforward.

Partitioning hierarchies, on the other hand, constitute full partitions of a given image, where the leaves of the hierarchy form the finest partition and are iteratively merged until a single root node is formed. In addition, partitioning hierarchies, or simply partitioning trees, most often require only a similarity metric for determining the merging or not between neighboring components, hence making them particularly suitable for processing multivariate images.

The vast majority of the reported work on APs relies on max- and min-trees, which belong both to the category of inclusion trees. More formally, given a grayscale image

$X: E \rightarrow \mathbb{Z}$, $E \subseteq \mathbb{Z}^2$, its upper-level sets are defined as $\{X \geq t\}$ with $t \in \mathbb{Z}$ (respectively, lower-level sets as $\{X \leq t\}$), i.e., the set of images obtained by thresholding an image at all possible values of its pixels. The CCs ($CC \subseteq E$) composing the upper- or lower-level sets are referred to as *peak components*. These two tree types (which are dual with respect to complementation) model the inclusion relations between these peak components; thus, max-trees are excellent for modeling regions that are brighter than their surroundings and min-trees, for regions that are darker. That is why, in addition, we employ both attribute thinnings (i.e., max-tree filtering) and thickenings (i.e., min-tree filtering) during the construction of APs [see (1)].

As the construction of two trees per image is both memorywise and computationally inefficient and results in longer feature vectors per image, thus also affecting classification performance and complexity, self-dual APs (SDAPs) were introduced to target these issues [53]. More specifically, they rely on the use of the ToS [54], which has been designed to provide a unified representation for both bright and dark image structures.

The ToS is constructed by filling the holes of the aforementioned peak components, and the shapes represented by the nodes do not intersect and are either disjoint or nested. Consequently, the ToS is an inclusion hierarchy that unlike the component tree, is also contrast invariant and self-dual. SDAPs have been empirically shown to outperform APs consistently in terms of classification performance while also producing shorter feature vectors per pixel [25], [53], [55].

Motivated by the success of the ToS and by the useful qualities of partitioning trees in the aforementioned context, α - and ω -trees were recently applied to AP implementation [56]. In particular, the α -tree is constructed based on the local range, where every tree node corresponds to an α -CC [57]. For instance, for $\alpha > 0$, an α -CC is defined as the CC of the maximal size such that only the neighboring pixels with gray-level difference less than or equal to α are considered connected.

Although α -trees can lead efficiently to a complete and self-dual image representation, due to the locality of the metric used, gray-level variations within α -CCs can be much higher than α (i.e., the chaining effect [57]). This undesirable effect can be mitigated through the use of ω -trees [58], which constitute a subset of α -trees, constraining every α -CC with their global range (i.e., the maximal dissimilarity between any two pixels belonging to that component). The hierarchy remains self-dual, complete, and capable of capturing regions of low, intermediate, and high gray levels, but global range provides better grouping per level than just a local measure.

Lefèvre et al. [59] have focused on the similarity metric requirement of partitioning trees for extension to multiband images and used metric learning to adapt them to hyperspectral images. Bosilj et al. [56] and Koç et al. [60], on the other hand, have empirically tested APs, SDAPs, and α - and ω -APs against each other and established the superiority of SDAPs in noise-free conditions, while partitioning

tree-based profiles outperformed their inclusion tree counterparts in terms of noise robustness.

ATTRIBUTE AND THRESHOLD SELECTION

Undoubtedly, APs possess many desirable practical and theoretical properties that render them particularly suitable for the task of spatial-spectral description in our context. However, they are not without flaws, and the main source of criticism so far has been in terms of their sensitivity to parameter selection [61]–[63]; by *parameters*, we mean particularly the attributes employed to characterize every tree node and, most importantly, the set of associated threshold values. The attribute- and threshold-selection parameters are discussed in the “Attributes” and “Thresholds” sections, respectively.

ATTRIBUTES

From a theoretical point of view, any function $a: \mathcal{P}(E) \rightarrow \mathbb{R}$, where $\mathcal{P}(E)$ denotes the power set of E , computable on an arbitrary collection of pixels, can, in fact, be employed as an attribute for AP construction. In practice, it is used during filtering by comparing a given CC’s ($CC \subseteq E$) attribute value against a predetermined threshold in the form of a binary predicate (e.g., in case of area, “is the CC’s area greater than 300 pixels?”). It thus provides a great degree of freedom as far as the object-based analysis of an image is concerned. The pioneering article of APs [7] introduced four such attributes: area, moment of inertia, diagonal length of the bounding box, and standard deviation; the first three describe a geometric property related to the shape of the tree node under study, and the last provides its statistical pixel intensity distribution.

Although these four attributes are, by far, the most widely encountered in the state of the art, APs can accommodate (from a theoretical point of view) a vast pool of attributes. Examples include entropy and homogeneity [7] as well as the diameter of the equivalent circle and area of the convex hull for automatic threshold selection [64]; complexity (perimeter over area) [65]; perimeter and area of the bounding box used to evaluate threshold-free (TF) APs [66]; solidity (area over area of the convex hull) and orientation (between the major axis of the convex hull and x-axis) [67]; coefficient of variation (CoV) and normalized radar cross section (NRCS) tailored for SAR images [33], [39], where CoV is the ratio of the standard deviation divided by the mean value of pixel intensities and NRCS, expressed in decibels, is the RCS per unit area of the surface. Furthermore, in [68], it has been observed that, when dealing with multiband input, one can extend the pool of attribute measures to include multidimensional functions exploiting all available bands simultaneously, and two new attributes have been proposed: higher-dimensional spread and dispersion.

As far as the selection of attributes is concerned, there is no straightforward rule or limitation. Formally, one should use the attributes that are “meaningful” for the data under consideration or, in other words, the attributes with respect

to which the objects of interest differentiate themselves from the rest. In cases where this cannot be determined or known a priori, it is customary to most often combine attributes that are expected to provide complementary information, e.g., some geometrical (such as area or moment of inertia) and some statistical (such as standard deviation), known as *extended multiattribute profiles* [7]. Nevertheless, to tackle the potentially long feature vectors resulting from the combination of various attributes, the application of dimension-reduction methods is not uncommon [62].

THRESHOLDS

For the optimal performance of any given attribute, the set of corresponding thresholds is supposed to span the range of values between which lie the attribute values of the tree nodes representing the objects of interest. In the presence of objects of interest of varying scales, however, as is often the case in practice, the size of the threshold set (or the subdivision level of said range) also becomes of paramount importance so as to capture the objects' response through AFs.

An example for illustrating this principle is the case of distinguishing minerals of various sizes with sieves; evidently, if you possess only two sieves (one large and one very small), you cannot expect to detect minerals of in-between sizes. Consequently, it is no surprise that the selection of the threshold set has a profound effect on the description performance of APs [27]. Although using a wide threshold range with a very fine subdivision level might, at first attempt, seem to be an intuitive countermeasure against this issue, it is to be avoided since it not only triggers the Hughes phenomenon but also provides no guarantee of capturing the differences among the various scales of the objects of interest.

During the early years of APs, threshold sets were determined exclusively manually based on expert knowledge, usually with four thresholds per attribute [7], [20], [53]. Even though some attributes, such as moment of inertia, possess many desirable invariance properties (against scale and rotation), rendering them and their manual threshold sets robust against content variations, most attributes are, unfortunately, heavily affected by the spatial and spectral resolutions as well as size of the objects and regions of interest.

Consequently, it was not long before attempts started being made to solve this issue. One line of research focused on developing analytical approaches based on expert knowledge in the form of equations for producing thresholds. For instance, [61] and [63] have proposed such equations for area depending on the input's spatial resolution as well as for standard deviation based on the image's mean pixel intensity, where the number of thresholds, however, is still user-defined.

Further approaches relying on (semi)supervised learning were put forward in [69], which proposed clustering the attributes of a given tree presentation into a user-defined number of clusters, while [70] investigated the evolutionary optimization of a very large set of thresholds. Then, [71] and

[72] presented a framework based on granulometric characteristic functions, where the authors employ morphological filters to first assess the input image's content and then determine the corresponding thresholds adaptively.

More recently, unsupervised threshold-selection methods [64], [73] and TF APs [66], [74], [75] have been proposed. In [73], from the AP computed from all possible threshold values of an attribute, a genetic algorithm selects the subset of reconstructed images that convey the highest amount of information. A disadvantage of this technique is the expensive computation of APs for all possible threshold values. A more efficient threshold-selection method is proposed in [64]: first, the set of attribute values of all nodes of a component tree is sorted in nondecreasing order; then, an increasing curve is built from the sorted attribute values; and finally, the points on the curve with the largest gradient values are the selected threshold values.

In [74] and [75], a TF approach was introduced, where every tree node is described based on simple statistical properties of the sequence of attribute values belonging to the nodes in the path connecting the node under study to the root of the tree. Properties such as the highest change of attribute value have shown promise. This strategy removes the need for thresholds. Moreover, by no longer employing the same global thresholds for every node, it instead uses a node-adaptive description strategy.

TREE FILTERING

In terms of tree filtering, relatively few developments have taken place since the inception of APs. Nevertheless, this section provides an overview of filtering rules for the sake of completeness.

Tree filtering is the stage where AF is performed to a given image using its tree representation. More specifically, given an attribute and a corresponding threshold value, one removes certain nodes (either single nodes or entire branches) of the tree that do not satisfy the threshold according to some predefined strategy, such as max, min, direct, Viterbi, and subtractive [7].

The effect of the selected filtering strategy depends on whether or not the attribute under study is increasing. Formally, an attribute $a(\cdot)$ assessed on region CC_i is said to be increasing if the following property holds [76]:

$$\forall CC_1 \subseteq CC_2, \Rightarrow a(CC_1) \leq a(CC_2). \quad (2)$$

Common increasing attributes include the area and diagonal length of the bounding box. When the attribute is increasing, filtering is straightforward. More specifically, if a node does not satisfy the underlying predicate, it is removed along with all its descendants since increasingness guarantees that they do not satisfy the predicate either. After removal, a node's pixel values become those of its highest ancestor node that satisfies the predicate. In this case, all of the aforementioned filtering strategies lead to the same outcome.

On the other hand, if the attribute is not increasing (e.g., moment of inertia, standard deviation, and so on), then

filtering is no longer straightforward, as whether a node's descendants need to be removed or not can no longer be determined by the node under study alone. At this point, one of the following is used [11]:

- ▶ *Max*: This rule prunes the nodes along a branch starting from the leaves up until reaching the first node that satisfies the predicate and needs to be preserved.
- ▶ *Min*: This strategy prunes the nodes along a branch starting from the leaves up until reaching the last node that does not satisfy the predicate and needs to be removed.
- ▶ *Viterbi*: This approach relies on dynamic programming through the Viterbi algorithm. It formulates filtering as an optimization problem in terms of node removal and preservation costs, which solves for minimal cost.
- ▶ *Direct*: This strategy simply consists of removing the nodes that do not satisfy the underlying predicate. Its eventual descendants are transferred to the first ancestor node that satisfies the predicate and thus needs to be preserved. Although simple, the direct filtering strategy is notorious [77] for its difficulty in dealing with shape-based object analysis. This is mostly due to contrast loss-related issues, which also constitute the main motivation behind the design of the subtractive strategy.
- ▶ *Subtractive*: This behaves almost identically to the direct strategy, with the only difference being that, after removing a node not satisfying the underlying predicate, an additional propagation step is performed on the descendant nodes. In particular, the pixel intensity associated with the descendant nodes is lowered in the case of the max-tree (and increased in the case of the min-tree) so that their contrast with respect to the local background remains consistent once removal takes effect. In the case of the ToS, in addition to lowering the pixel intensity of the remaining descendant nodes, the subtractive rule can introduce new intensity values that were not present in the original image [55], [78].

As far as partitioning trees are concerned, the reader is referred to [56]. The subtractive strategy has been shown to outperform its alternatives when dealing with nonincreasing attributes, especially with moment of inertia [77].

The reader is referred to [55] and [78] for the results of an empirical comparison among filtering strategies.

POSTPROCESSING OF OUTPUT PROFILES

APs, i.e., the sequence of filtered images in (1), can be directly fed into supervised classifiers, such as an SVM or random forest (RF), for classification on a pixel basis. Such direct application has provided better performance compared to MPs [7] in terms of classification accuracy as well as computational cost. However, since APs often lead to feature vectors with a relatively high redundancy level, depending on the number of employed thresholds [24], the postprocessing of these features has been addressed in several studies. Many among them have proposed applying different feature-selection techniques to extract more informative features and reduce their dimension. In [24] and [61], both

linear (PCA and ICA) and nonlinear methods (ICA, KPCA, DAFE, DBFE, NWFE, and so on) are investigated. A general framework and systematic survey of spatial-spectral approaches combining APs with these feature-selection techniques is presented in [14].

Other works have focused on extraspatial processing of APs for better characterization of structural and textural information from the image content. Recent studies [79], [80] claim that, when dealing with VHR remote sensing images where regions and objects appear more heterogeneous, APs may not provide a complete spatial characterization of pixels. Therefore, some efforts have been realized to improve APs through the histogram or some first-order statistical features of the local patch around the pixel under study. As a result, local histogram-based APs (HAPs) [79], [81] and local feature-based APs (LFAPs) [80], [82] have been proposed and proven to be more efficient for better dealing with local textures. The extensions of these extraspatial processing methods to self-dual profiles and hyperspectral images have been studied in [80].

Some further notable extensions to APs include their sparse representation in an attempt to increase their description capacity [83]. In detail, through the collection of representative samples of low-dimensional class-dependent structures, any sample can then be sparsely and more effectively represented and classified. Moreover, the combination of APs with classifier ensembles has also been investigated intensively in [84] and [85].

Last but not least, we have observed an increasing tendency to combine APs with convolutional neural networks (CNNs) in the classification of satellite images. While CNNs require large training sets to provide optimal features from the raw data, APs can produce effective features from scratch, thanks to their inherent expert knowledge. CNNs can exploit APs to produce even stronger features from them without the need for large training sets. Hence, APs simplify the learning process of CNNs by reducing the number of training samples required for a satisfactory classification result, which, consequently, reduces training time.

On the downside, combining CNNs with APs increases the design and computational complexity of the classification task. For instance, we refer readers to deep learning approaches on APs [86], [87], EPs (discussed in the "Extensions and Generalization of APs" section) [88], and SDAPs [89]. In those works, spatial features are extracted in two phases: first using APs (and their variants) and then with CNNs. More precisely, pixel features obtained from APs are fed into CNNs, leading to better classification results when compared to the AP and CNN methods individually while considerably increasing the test time [86], [90].

EXTENSIONS AND GENERALIZATION OF APs

As mentioned previously, the main advantages of APs in comparison to MPs are the efficient computation of APs through hierarchical image representations and the possibility of extracting information other than that constrained

by the size and shape of structuring elements. However, using APs in remote sensing images also has a few limitations.

First, apart from TF APs [66], [74], [75], the quality of an AP depends on the selected set of thresholds. A bad selection can lead to redundant information in the AP [61], [62]. An alternative solution to alleviate this redundancy problem is to replace the AFs used in the computation of APs by extinction filters, resulting in EPs [91].

An extinction filter acts on the regional extrema (minima or maxima) of an image: each extremum is either completely preserved or pruned. Let X be a grayscale image. The extinction value (with respect to a given attribute) of any maximum M of X is the maximum attribute value k such that M is still included in a maximum of $\gamma_k(X)$, where γ_k is the thinning operator with parameter k . Similarly, the extinction value of any minimum M of X is the maximum attribute value k such that M is still included in a minimum of $\phi_k(X)$, where ϕ_k is the thickening operator with parameter k . This way, to compute an EP, the filtering parameter is the number of minima or maxima to be preserved instead of a threshold value, which makes EPs less sensitive to image resolution [91].

Since [91], EPs and their extension to hyperspectral images (extended EPs) have been successfully applied to the land cover classification of hyperspectral data [49], [92] and fusion of hyperspectral and lidar data [30]. To further reduce redundancy, composite kernels are used to fuse the spatial information of EPs with hyperspectral data in [93] and with lidar data in [47].

Another limitation of APs is that, very often, clusters of pixels associated with distinct semantic objects, such as roads and buildings, are connected by narrow paths of similar intensity value. This leads to pixels of different semantic classes being connected throughout several levels of a component tree. Consequently, the attribute values of several CCS describe the union of objects of different classes instead of an object of a single class. In [94], the authors address those problems in the context of hyperspectral image classification. To overcome those issues, attribute connected filters are replaced by partial reconstruction filters, which allows the disconnection of regions connected by narrow paths and improves the overall classification accuracy. This idea is explored in [43], which shows the interest of using partial reconstruction in the classification of hyperspectral and lidar images in comparison to attribute connected filters.

Another related approach, called *invariant APs* (IAPs), was recently proposed in [95] to overcome other limitations of APs, including the sensibility of APs to geometric transformations, like rotation, and to the surrounding of pixels of a same material. Different from APs, IAPs are not computed from a hierarchical representation of the input data. Instead, hierarchical information is indirectly extracted from the original image by performing convolutions of different sizes and computing the Fourier transform for different values of the Fourier order. Then, IAPs are obtained by stacking spatially invariant profiles, acquired from the

segmentations of the convoluted images, and frequency invariant profiles, acquired from the histogram of oriented gradients of the Fourier transforms.

Finally, Pham et al. [96] propose a generalization of APs called *feature profiles* (FPs). They generalize step four of the generation of APs that consists of reconstructing an image from a filtered tree. To build an AP, this reconstruction is originally performed by projecting the gray values of the nodes of the filtered tree onto the image pixels. In [96], this reconstruction step is extended by taking into consideration not only the gray values of the nodes but also other statistical and geometrical features. The resulting images compose the so-called FP. The experiments with remote sensing images of [96] and [97] show the interest of projecting attributes like area and moment of inertia in the context of image classification. Hence, we also consider FPs later in the "Experimental Study" section.

EXPERIMENTAL STUDY

This section describes our experimental study to evaluate the performance of standard APs as well as some of their recent variants. The contributions of this section are twofold: the evaluation of 1) newer variants of APs and 2) the impact of connectivity and quantization parameters on the performance of APs. This is the first study addressing the effect of connectivity and quantization parameters in this context.

Experiments were mostly performed in Python using publicly available libraries. (Source codes are available at <https://gitlab.inria.fr/dsantana/attributes-profiles-survey-source-codes>.) First, APs and some of their variants were computed with the SAP package. (The documentation and source codes are provided at <https://gitlab.inria.fr/fguiotte/sap>.) To the best of our knowledge, this is the first available open source library for computing APs and some of their extensions. The SAP package relies on the Higra [98] library (the documentation and source codes are given at <https://github.com/higra>), which provides efficient implementation and postprocessing of morphological trees in C++. Then, classification was performed with the scikit-learn Python library.

Supervised classification has been conducted on both grayscale and hyperspectral images for the sake of comprehensiveness. In this section, we introduce the data sets, experimental setup commonly encountered in the state of the art [7], [14], and classification results that have been obtained.

In more detail, the standard setup involves calculating the hierarchical tree representation (as explained in detail in the "Principle of APs" section) from the entire input image, computing the APs (or their variants) from this representation, and then subdividing the resulting features based on the locations of training and validation/testing pixels. The next section elaborates on the reasons that this approach constitutes a validation malpractice and proposes an alternative strategy for better assessing the generalization capacity of APs.

DATA DESCRIPTION

The experiments were conducted with two publicly available data sets for the sake of reproducibility. To show performance variations depending on the number of spectral bands, one hyperspectral and one panchromatic data set have been selected.

PAVIA UNIVERSITY DATA SET

The Pavia data set is a hyperspectral image acquired by the Reflective Optics System Imaging Spectrometer airborne sensor with 1.3-m spatial resolution over the region of Pavia University, Italy. (The Pavia data set and its ground truth can be downloaded from http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.) The image consists of 610×340 pixels with 103 spectral bands (from 0.43 to 0.86 μm) after the noisy bands are removed. The ground truth covers nine thematic classes: trees, asphalt, bitumen, gravel, metal sheets, shadows, meadows, self-blocking bricks, and bare soil.

For this image, the standard training set (provided by the IEEE Geoscience and Remote Sensing Society Data and Algorithm Standard Evaluation website: <http://dase.grss-ieee.org/>), composed of 3,921 pixels (see Table 1), was adopted for the classification task. The test set was composed of all of the remaining 40,002 pixels in the ground truth that are not in the training set. In the remainder of this article, this partition into the training/test set of Pavia will be denoted as *Pavia*₁.

The reader may note that some works in the literature consider all ground-truth pixels as testing pixels, which is not done here to provide fair classification results with a nonvoid intersection between training and testing pixels. The false-color image (made by combining bands 31, 56, and 102), ground-truth map, and training set are shown in Figure 3. Following the standard approach of handling hyperspectral images [7], we first performed the PCA on this data set, and the first four PCs (involving more than 99% of the total variance) were preserved for our experiments. The APs and their extensions were computed independently on each selected PC and concatenated, leading to EAPs. This way, we assessed the family of EAPs and their variants on the Pavia data set.

GRAY–POTSDAM DATA SET

The Potsdam data set is composed of 38 aerial HR images of $6,000 \times 6,000$ pixels with 5-cm spatial resolution over the city of Potsdam, Germany. (The Potsdam data set and its ground truth can be downloaded from <http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html>.) For each image, the red, green, blue, and infrared bands are available. The ground-truth annotations of this data set cover six thematic classes including impervious surface, building, low vegetation, tree, car, and clutter/background (water bodies, tennis courts, swimming pools, and so on). Experiments were performed on one image of this data set, namely, *top_potsdam_7_7.tif*, whose ground truth is composed of several CCs of each of the six classes.

To highlight the strength of the spatial information extracted from APs in the context of image classification, we considered a grayscale version of the original red, green, blue (RGB) image. The original RGB image was converted into a grayscale image using the formula $0.3R + 0.59G + 0.11B$, which gives an approximation of the luminance in the National TV Standards Committee color space [99]. The input grayscale image and its thematic ground-truth map are shown in Figure 4(a) and 4(b), respectively.

To prevent biased results toward the majority class, 10 training sets were obtained by random sampling of the same number of pixels from each of the six thematic classes (see Table 2). In total, for each random split, 360,000 pixels (1% of the ground-truth samples) were selected for training, and the remaining 35,640,000 pixels were used for testing.

TABLE 1. THE NUMBER OF TRAINING AND TEST SAMPLES OF THE STANDARD (PAVIA₁) AND NEW (PAVIA₂) PARTITIONS OF THE PAVIA DATA SET.

CLASS	NUMBER OF SAMPLES		
	TRAINING SET	TEST SET	TOTAL
Asphalt	548	6,304	6,852
Meadows	540	18,146	18,686
Gravel	392	1,815	2,207
Trees	524	2,912	3,436
Metal sheets	265	1,113	1,378
Bare soil	532	4,572	5,104
Bitumen	375	981	1,356
Bricks	514	3,364	3,878
Shadows	231	795	1,026

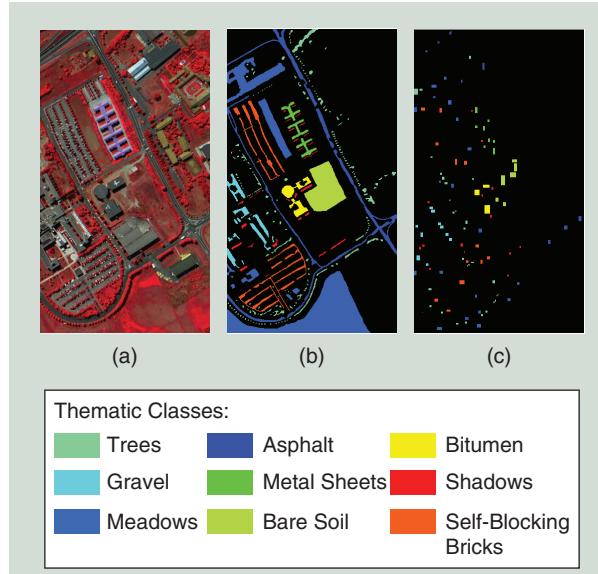


FIGURE 3. The 610×340 Pavia University data: the (a) false-color image made by bands 31, 56, and 102; (b) ground truth including nine thematic classes; and (c) training set.

TABLE 2. THE NUMBER OF TRAINING AND TEST SAMPLES PER CLASS OF THE PARTITIONS GRAYPOTSDAM₁ AND GRAYPOTSDAM₂ OF THE GRAY-POTSDAM DATA SET.

CLASS	NUMBER OF SAMPLES		
	TRAINING SET	TEST SET	TOTAL
Background	60,000	1,565,250	1,625,250
Trees	60,000	5,785,203	5,845,203
Cars	60,000	631,810	691,810
Buildings	60,000	12,362,473	12,422,473
Low vegetation	60,000	8,639,455	8,699,455
Impervious surfaces	60,000	6,655,809	6,715,809

In the remainder of this article, those random partitions into the training/test set of Gray–Potsdam are denoted as *GrayPotsdam*₁. Due to the dimensions of Gray–Potsdam, the random training pixels of *GrayPotsdam*₁ cannot be visualized when the image is downsized. The reader can refer to the source codes of our experiments (available at <https://gitlab.inria.fr/dsantana/attributes-profiles-survey-source-codes>) to visualize the training/test splits of the Gray–Potsdam data set.

EXPERIMENTAL SETUP

The experiments were conducted using a wide variety of parameters, shown in Table 3, some of which have been employed for the first time in the state of the art. To ease the readability of the results presented in this and the following section, we present a summary of our experimental settings in Tables 4 and 5. Table 4 shows the parameter settings used for each experiment, and Table 5 contains the description of each data set split employed in our experiments.

The popular choices of classifiers for APs are SVM and RF [14]. As suggested in [14], RFs usually perform better than SVMs in this context. Moreover, RFs require lower training and prediction time [80]. Hence, our supervised classification was conducted on the two data sets using the RF classifier [100] with 100 trees, as employed in other studies [19], [96]. In the scikit-learn implementation of RFs, a random shuffling of the data is applied before training. Hence, the classification output for each run is not reproducible unless a fixed value is assigned to the *random_state* parameter of the *RandomForestClassifier* class.

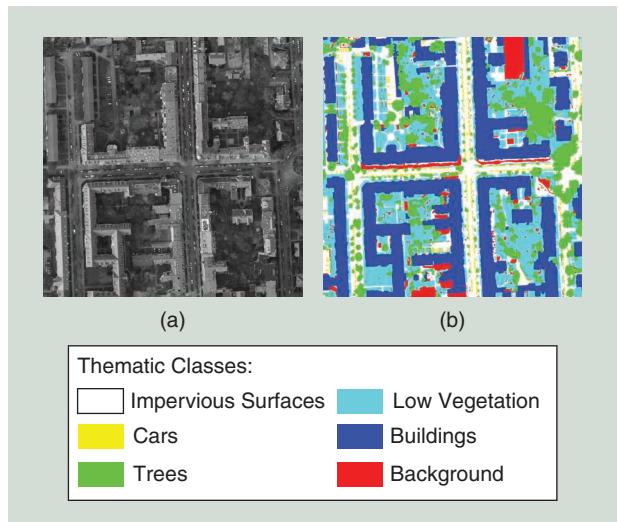


FIGURE 4. (a) The grayscale version of the image *top_potsdam_7_7.tif* of the Potsdam data set and (b) its thematic ground truth.

To take this into consideration, the classification results reported in this and the following section are the average and standard deviation of the scores obtained for 10 runs on each data set. The number of variables involved in the training was set to the square root of the feature vector length. To evaluate and compare the classification accuracy of different approaches, overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) have been taken into account [7]. As far as AF is concerned, for the sake of design simplicity and computational cost, we have limited our experiments to two attributes: area and the moment of inertia. The former is increasing, and the latter is not. For the Pavia data set, 14 area thresholds were computed automatically using the formula proposed in [61],

$$\lambda_a, X = \frac{1,000}{\nu} \{ a_{\min}, a_{\min} + \delta_a, a_{\min} + 2\delta_a, \dots, a_{\max} \}, \quad (3)$$

which has received wide acclaim in related works [14], [27], [96], [101]. In (3), a_{\min} and a_{\max} are initialized by 1 and 14, respectively, with a step increase δ_a equal to one, and ν represents the spatial resolution of the input data. The resulting thresholds for the Pavia data set follow:

$$\lambda_{a,\text{Pav}} = \{ 770; 1,538; 2,307; 3,076; 3,846; 4,615; 5,384; 6,153; 6,923; 7,692; 8,461; 9,230; 10,000; 10,769 \}.$$

TABLE 3. THE CHOICE OF METHODS AND PARAMETERS EMPLOYED IN THE REPORTED TESTS.

DATA SETS	ATTRIBUTES	TREE TYPES	QUANTIZATION (B)	CONNECTIVITY	POSTPROCESSING	THRESHOLDS
Pavia University	Area	Max-min	8	4	HAPs	Manual
Gray–Potsdam	Moment of inertia	ToS α -tree ω -tree Max-tree Min-tree	16 64	8	FPs LFAPs	Automatic Without

For the Gray-Potsdam data set, the thresholds obtained with the previous formula were not sufficient to cover the size variations of the targeted classes. To obtain a set of

thresholds spanning a larger range of values without increasing the number of thresholds, we considered the 14th first values of the geometric sequence whose n th term is given

TABLE 4. THE EXPERIMENTAL SETTINGS EMPLOYED IN THE REPORTED TESTS.

TABLES WITH EVALUATION RESULTS	DATA SET SPLIT	ATTRIBUTES	TREE TYPES	QUANTIZATION (B)	CONNECTIVITY	POSTPROCESSING	THRESHOLDS
Table 6	<i>Pavia</i> ₁	All	All	8	4	All	All
Table 8	<i>GrayPotsdam</i> ₁	All	All	8	4	FPs	Manual
Table 10	<i>GrayPotsdam</i> ₁	All	All	8	4	LFAPs	Without
Table 11	<i>Pavia</i> ₁	All	All	8	8	FPs	Manual
					4	LFAPs	Without
Table 12	<i>Pavia</i> ₁	All	All	8	4	All	All
				16			
				32			
Table 13	<i>Pavia</i> ₁	All	All	8	4	FPs	All
	<i>Pavia</i> ₂						
Table 14	<i>Pavia</i> ₂	All	All	8	4	FPs	All
Table 15	<i>GrayPotsdam</i> ₁	All	All	8	4	FPs	Manual
	<i>GrayPotsdam</i> ₂						Without
Table 16	<i>Potsdam</i> ₂	All	All	8	4	FPs	Manual
							Without
Table 17	<i>Potsdam</i> ₃	All	All	8	4	FPs	Manual
Table 18	<i>Potsdam</i> ₃	All	All	8	4	FPs	Manual

TABLE 5. THE PAVIA AND GRAY–POTSDAM DATA SET SPLITS EMPLOYED IN THE REPORTED TESTS.

DATA SET SPLIT	DESCRIPTION	OBJECTIVE
<i>Pavia</i> ₁ (Figure 3)	<ul style="list-style-type: none"> ▶ Standard split obtained from http://dase.grss-ieee.org/ ▶ Commonly used in the literature 	<ul style="list-style-type: none"> ▶ Provide evaluation scores of AP extensions, which are comparable with the results in the literature (in which all the whole data are preprocessed with PCA, and training/testing features are extracted from the same tree, leading to the leakage of training features)
<i>Pavia</i> ₂ [Figure 8(b)]	<ul style="list-style-type: none"> ▶ 10 sets of training/test samples randomly extracted from a restricted region of the <i>Pavia</i> data ▶ Lower variability of training pixels when compared to <i>Pavia</i>₁ ▶ Better separation among training/testing pixels than <i>Pavia</i>₁ ▶ The same number of training samples per class as <i>Pavia</i>₁ 	<ul style="list-style-type: none"> ▶ Evaluate the impact of having large ground-truth regions not contributing to the training set ▶ Generalize APs to data sets with lower levels of leakage of training features, which still remain due to all the whole data being preprocessed with PCA, as well as to training and testing features being extracted from the same tree ▶ Provide a more realistic partition of the <i>Pavia University</i> data set when compared to <i>Pavia</i>₁
<i>GrayPotsdam</i> ₁	<ul style="list-style-type: none"> ▶ 10 sets of training/test samples randomly extracted from all of the ground-truth CCs of Gray–Potsdam ▶ The same number of training samples per semantic class 	<ul style="list-style-type: none"> ▶ Evaluate APs and their extensions on Gray–Potsdam using the training/test splitting method commonly used in the literature
<i>GrayPotsdam</i> ₂ (Figure 9)	<ul style="list-style-type: none"> ▶ 10 sets of training/test samples randomly extracted from a restricted set of ground-truth CCs of Gray–Potsdam ▶ The same number of training and test samples as <i>GrayPotsdam</i>₁ 	<ul style="list-style-type: none"> ▶ Similar to <i>Pavia</i>₂: introduce CCs in the test set that do not include any training pixels. ▶ Test if the features extracted from APs really reflect the geometrical characteristics of the objects belonging to a certain class or if the success of APs is mainly due to the leakage of training/testing features
<i>GrayPotsdam</i> ₃ (Figure 11)	<ul style="list-style-type: none"> ▶ Data set divided in half ▶ The same number of training pixels per class as <i>GrayPotsdam</i>₁ and <i>GrayPotsdam</i>₂, but a different number of test pixels ▶ Test set composed of all pixels from the lower half ▶ 10 sets of training samples randomly extracted from the upper half 	<ul style="list-style-type: none"> ▶ Generalize APs to multiple-image data sets ▶ Compute training and testing features from different trees obtained from the two halves of the image, with no leakage of training/testing features

by 200×2^n . The resulting thresholds for the Gray-Potsdam data set follow:

$$\lambda_{a,\text{Pot}} = \{200; 400; 800; 1,600; 3,200; 6,400; 12,800; 25,600; 51,200; 102,400; 204,800; 409,600; 819,200; 1,638,400\}.$$

As far as the moment of inertia is concerned, the manually set thresholds used in several studies [19], [25], [27] were adopted here as well:

$$\lambda_{i,\text{Pot}} = \lambda_{i,\text{Pav}} = \{0.2, 0.3, 0.4, 0.5\}.$$

Moreover, as mentioned in Table 3, we consider two additional experimentation parameters, namely, the input image's connectivity and quantization level, the effect of which, to the best of our knowledge, has not been previously studied in the state of the art. For instance, in [49], [91], and [102], the authors use four-connectivity, but, in most published studies, these two parameters are seldom mentioned.

For the sake of simplicity and to avoid introducing new acronyms, EAP will be referred to as AP in the remainder of this article. Hence, whenever we discuss results on the Pavia data set, it should be understood that the APs and their variants correspond to EAPs in this data set.

In more detail, we compare APs and EAPs generated from different kinds of trees, including the max-tree (AP-maxT); the min-tree (AP-minT); a max-tree along with a min-tree (as in standard APs) [7]; SDAPs [53] from the ToS; and α -APs and ω -APs from α - and ω -trees, respectively [56]. We also provide the results of some effective postprocessing techniques, including the HAPs/histogram self-dual attribute profiles (HSDAPs) [79], [80], LFAPs/local feature self-dual

attribute profiles (LFSDAPs) [80], and of some extensions of APs, including TF APs [74] and FPs [96].

A far more limited comparative study of these parameters is reported in [97]. We obtained HAPs and HSDAPs using histograms of seven bins and a window size of 7×7 pixels, which are the optimal parameter settings according to the experiments of [96]. LFAPs and LFSDAPs were computed using the mean and standard deviation of 7×7 -sized windows as the local features. As shown in [96], feature and histogram profiles are fairly robust to the choice of window size, but the experiments of [96] with 7×7 windows provided superior results. FPs were obtained by projecting the average gray levels (FP_μ), area (FP_a), and both average gray levels and area ($FP_{\mu+a}$) of tree nodes during the reconstruction step, as done in [96]. Among all tested methods, TF AP [74] is the only one not implemented in the SAP library.

RESULTS

This section starts with the presentation and discussion of results obtained from experiments employing the commonly encountered 8-b quantization and four-connectivity [49], [91], [102] in the state of the art. Then, it continues with an evaluation of the effect of the aforementioned two parameters on classification performance.

CLASSIFICATION RESULTS FOR THE PAVIA DATA SET

The overall classification results, classification results per class, and classification maps for the Pavia data set obtained using the partition *Pavia*₁ are presented in Tables 6 and 7 and Figure 5, respectively.

For this data set, the tree type underlying the APs appears to have an important influence on performance. More specifically, the α -APs and ω -APs outperformed both APs and SDAPs. In particular, using the ω -tree, one can achieve an average OA of 96.33%, i.e., 5.76% and 2.23% better than standard APs and SDAPs, respectively. However, this improvement is not observed for all classes. While the accuracy of ω -AP for the gravel class increases by 20.58% with respect to APs, the accuracy for the asphalt class decreases by 5.98%. Then, by postprocessing the AP and SDAP with histogram and local FPs, we improved AP (respectively, SDAP) by more than 4% (respectively, 1%) in terms of OA.

Regarding the FPs, FP_μ and $FP_{\mu+a}$ provided much better results than FP_a , as already shown in the original article on FPs [96]. Finally, the TF profile, which has fewer dimensions than all other tested profiles, outperformed AP by 3.41%, 3.13%, and 4.51% in terms of OA, AA, and κ , respectively. Among all tested methods, the best classification accuracy was achieved by FP_μ , with OA = 96.76 and $\kappa \times 100 = 95.65$. Compared to standard APs, enhancements of 6.19% in OA and 8.12% in κ were adopted.

Furthermore, our classification results on the Pavia data set are competitive with respect to some recent deep learning approaches discussed in [17], in which the authors compare several CNN architectures from the literature for hyperspectral image classification. Among those tested, the best results for

TABLE 6. THE CLASSIFICATION RESULT OF THE PAVIA DATA SET OBTAINED BY DIFFERENT METHODS USING THE DEFAULT FOUR-CONNECTIVITY AND 1-B QUANTIZATION.

CLASSIFICATION RESULT				
METHOD	DIMENSION	OA (%)	AA (%)	$\kappa \times 100$
Four PCs	4	65.27 ± 0.25	74.88 ± 0.2	56.93 ± 0.27
AP-maxT	80	89.21 ± 0.63	87.37 ± 0.19	85.42 ± 0.8
AP-minT	80	87.11 ± 2	92.62 ± 0.61	83.23 ± 2.41
AP	152	90.57 ± 2.6	93.13 ± 0.67	87.53 ± 3.32
SDAP	80	94.1 ± 0.21	93.85 ± 0.37	92.16 ± 0.27
α -AP	80	95.46 ± 0.59	95.25 ± 1.23	93.91 ± 0.79
ω -AP	80	96.33 ± 0.4	97.17 ± 0.89	95.08 ± 0.54
HAP	1,064	94.73 ± 0.3	92.96 ± 0.31	92.84 ± 0.42
HSDAP	340	95.35 ± 0.29	94 ± 0.49	93.79 ± 0.39
LFAP	304	94.75 ± 0.29	94.12 ± 0.41	92.97 ± 0.39
LFSDAP	160	96.34 ± 0.16	92.81 ± 0.2	95.06 ± 0.22
FP_μ	152	96.76 ± 0.15	97.05 ± 0.19	95.65 ± 0.21
FP_a	152	85.48 ± 0.98	93.94 ± 0.55	81.36 ± 1.16
$FP_{\mu+a}$	304	96.15 ± 0.14	96.74 ± 0.32	94.82 ± 0.19
TF-AP	72	93.98 ± 0.53	96.26 ± 0.29	92.04 ± 0.68

The values in bold represent the best OA, AA and kappa scores among all tested methods.

Pavia ($OA = 84.32 \pm 0.72$ and $\chi = 0.799 \pm 0.009$), obtained with the same data set split ($Pavia_1$) employed here, achieved with a 3D CNN [103], are still inferior to most of our results given in Table 6.

The reader may note that our classification results differ from those presented in the original articles of (extended) APs [19], SDAPs [25], LFAPS [80], and α - and ω -APs [56]. More precisely, the differences can be explained by the various sets of attributes and thresholds employed in some of the articles [19], [25], [80], the use of pre- and postprocessing techniques other than PCA [25], and number of RF trees [25]. Moreover, even using the same set of thresholds and same number of RF trees, as was done in [56], the quantization and connectivity parameters, which are not explicitly given in those articles, may play a role in the final results, as discussed in the “Assessment of Connectivity and Quantization” section. Nevertheless, in terms of conclusions, there is no incoherence between our article and the articles cited at the beginning of this paragraph.

CLASSIFICATION RESULTS FOR THE GRAY-POTSDAM DATA SET

The overall and per-class classification results for the Gray-Potsdam data set using $GrayPotsdam_1$ are presented in Tables 8 and 9, respectively. As already mentioned, the reported results are the average scores over 10 runs on the different random training test splits of $GrayPotsdam_1$. Since LFAPs perform better than HAPs in general, as attested by [80], and due to the expensive computation of HAPs, we consider only LFAPs in our experiments with the Gray-Potsdam data set.

In the case of the Gray-Potsdam data set, it can be observed that AP variants boost accuracy consistently at various degrees. In particular, the α -APs and ω -APs could outperform APs on each single max-tree or min-tree but still fall below the standard APs. On the other hand, SDAP performed better than AP, α -AP, and ω -AP. Then, by postprocessing the output profiles, LFAP and LFSdap outperformed AP and SDAP by 3.52% and 4.61%, respectively, in terms of OA. Among the FPs, the best result was achieved by $FP_{\mu+a}$, which outperforms the APs by more than 2% in terms of OA, AA, and χ . Finally, the TF AP presented lower scores than APs, but it outperformed AP-maxT, AP-minT, α -AP, and ω -AP, despite having the smallest number of dimensions among all methods. In conclusion, the best classification result was obtained by LFSdap, with 80.80% of OA, which represents an improvement of 5.70% with respect to the standard AP.

Figure 6(a) and (b) presents a crop of the Gray-Potsdam data set and its ground truth, respectively, composed of the first 500 lines (from top to bottom) and 1,200 columns (from left to right) of the original data. Figure 6(c)–(o) illustrates the classification maps obtained on the crop of Figure 6(a) using the aforementioned methods. We can see that the classification based solely on pixel gray values is very noisy in most regions of the image. By incorporating spatial information from APs, we see a more

TABLE 7. THE CLASSIFICATION RESULTS PER CLASS OF THE PAVIA DATA SET OBTAINED BY DIFFERENT METHODS USING THE DEFAULT FOUR-CONNECTIVITY AND 1-B QUANTIZATION.

METHOD	DIMENSION	ACCURACY PER CLASS (%)								
		MEADOW	ASPHALT	GRAVEL	TREE	SOIL	BITUMEN			
Four PCs	4	53.2 ± 0.45	71.25 ± 0.44	38.39 ± 0.53	98.28 ± 0.1	98.79 ± 0.17	67.34 ± 0.58	66.74 ± 0.96	83.28 ± 0.44	96.64 ± 0.34
AP-maxT	80	92.33 ± 0.14	93.66 ± 1.41	43.32 ± 0.67	95.61 ± 0.77	99.63 ± 0.08	71.07 ± 0.24	97.83 ± 0.14	95.28 ± 0.13	97.61 ± 0.46
AP-minT	80	92.81 ± 0.08	80.28 ± 4.42	86.02 ± 1.65	98.97 ± 0.08	99.9 ± 0.08	84.19 ± 0.08	99.93 ± 0.12	98.02 ± 0.13	93.47 ± 3.01
AP	152	95.71 ± 0.21	87.38 ± 5.81	73.17 ± 3	99.09 ± 0.29	99.65 ± 0.05	85.64 ± 0.14	100 ± 0	99.24 ± 0.18	98.25 ± 2.03
SDAP	80	97.15 ± 0.28	92.51 ± 0.47	77.15 ± 0.29	93.21 ± 0.7	99.83 ± 0.06	99.08 ± 0	98.94 ± 0.25	98.15 ± 0.5	88.62 ± 3.07
α -AP	80	89.39 ± 0.21	96.63 ± 0.43	76.22 ± 11.05	99.73 ± 0.07	99.61 ± 0.07	98.93 ± 0.06	99.49 ± 0	99.35 ± 0.1	97.82 ± 0.44
ω -AP	80	89.73 ± 0.56	96.70 ± 0.42	93.75 ± 7.84	99.75 ± 0.09	99.64 ± 0.04	98.95 ± 0.06	99.49 ± 0	99.36 ± 0.11	97.12 ± 0.4
HAP	1,064	99.86 ± 0.16	97.68 ± 0.42	63.83 ± 1.71	97.74 ± 0.11	99.96 ± 0.06	80.73 ± 2.22	99.99 ± 0.03	98.13 ± 0.07	98.73 ± 0.33
HSDAP	340	99.7 ± 0.18	94.49 ± 0.46	68.59 ± 2.44	95.25 ± 0.19	99.99 ± 0.03	99.68 ± 0.39	99.55 ± 0.1	99.08 ± 0.06	89.70 ± 2.48
LFAP	304	90.33 ± 0.9	96.23 ± 0.36	72.68 ± 3.79	96.79 ± 0.1	99.2 ± 0.24	98.08 ± 0.09	99.36 ± 0.17	97.28 ± 0.1	97.12 ± 0.79
LFSdap	160	99.42 ± 0.18	98.42 ± 0.12	65.82 ± 0.51	92.88 ± 0.8	99.52 ± 0.18	97.06 ± 0.56	99.16 ± 0.14	98.75 ± 0.13	84.26 ± 1.25
FP_μ	152	94.97 ± 0.13	97.1 ± 0.21	88.65 ± 0.91	97.07 ± 0.19	99.89 ± 0.04	98.43 ± 1.57	100 ± 0	97.43 ± 0.53	99.92 ± 0.06
FP_α	152	95.51 ± 0.25	73.87 ± 2.2	92.31 ± 4.14	99.31 ± 0.2	99.93 ± 0.04	86.98 ± 1.18	99.98 ± 0.04	99.72 ± 0.1	97.8 ± 1.82
$FP_{\mu+\alpha}$	304	96.03 ± 0.23	96.61 ± 0.21	92.43 ± 1.28	97.93 ± 0.43	99.89 ± 0.04	91 ± 0.96	99.97 ± 0.05	98.57 ± 0.38	98.23 ± 1.78
TF-AP	72	95.79 ± 0.26	90.31 ± 1.14	82.42 ± 2.41	99.64 ± 0.08	99.75 ± 0.1	100 ± 0	99.23 ± 0.19	99.56 ± 0.06	

The values in bold represent the highest scores in each column, i.e., the highest scores achieved for each of the semantic classes of Pavia among all tested methods.

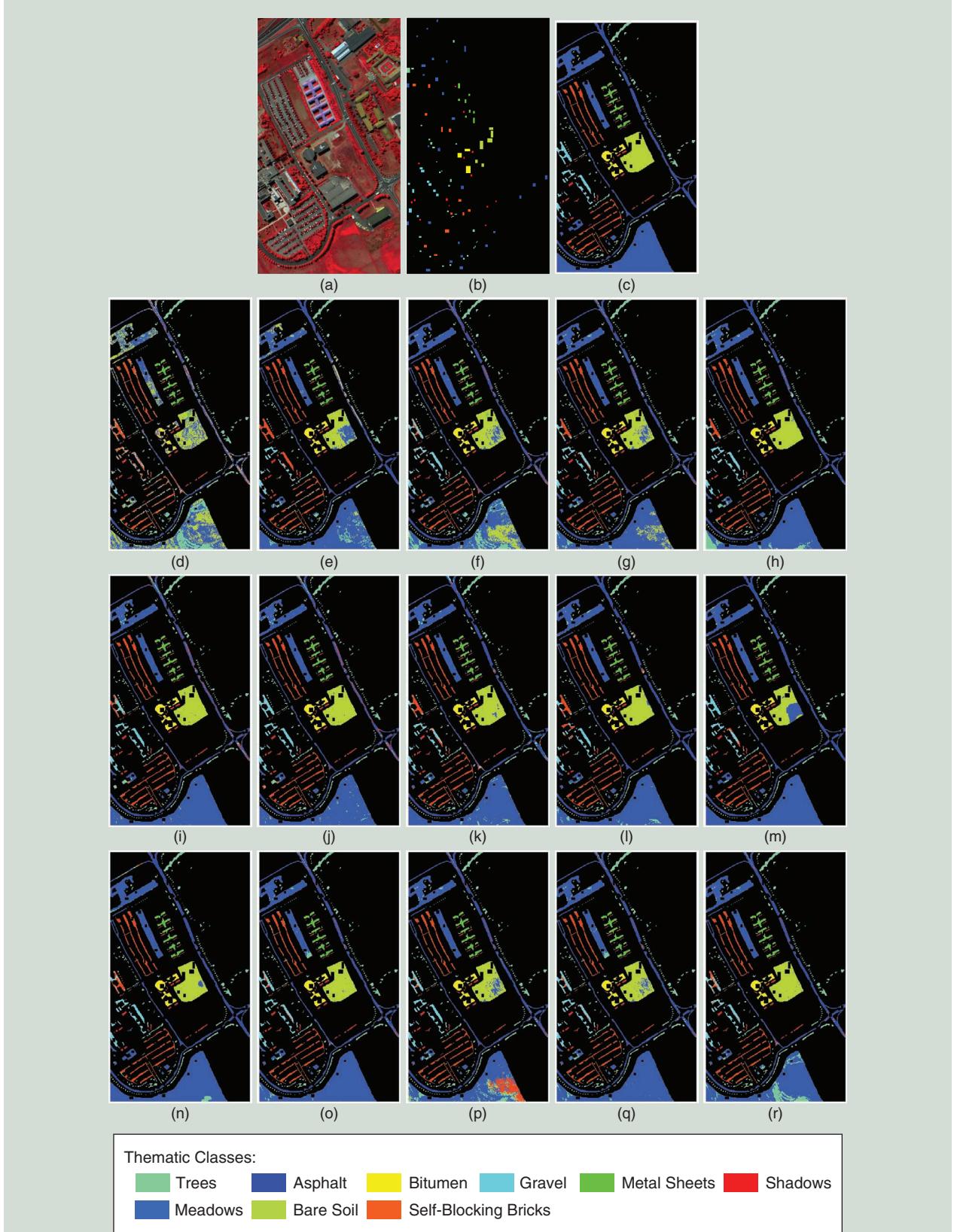


FIGURE 5. The classification results of the Pavia data corresponding to the results of Table 6: the (a) image, (b) training set, (c) test set, (d) four PCs, (e) AP-maxT, (f) AP-minT, (g) AP, (h) SDAP, (i) α -AP, (j) ω -AP, (k) LFAP, (l) LFSADP, (m) HAP, (n) HSDAP, (o) FP_{μ_t} , (p) FP_{μ_r} , (q) $FP_{\mu+a_t}$, and (r) TF-AP. Inline image: trees; Inline image: gravel; Inline image: meadows; Inline image: asphalt; Inline image: metal; Inline image: bare soil; Inline image: bitumen; Inline image: shadows; Inline image: bricks.

structured result, with a clearer separation between the regions of different classes.

Furthermore, the postprocessing of APs and SDAPs with local features successfully reduces the noise in all classes, especially the regions containing trees (in green) and buildings (in blue). In terms of future research directions, it would be interesting to investigate the combination potential of the AP variants and extensions in an effort to discover whether they provide complementary information and, eventually, higher performances.

ASSESSMENT OF CONNECTIVITY AND QUANTIZATION

In this section, we discuss the influence of the connectivity parameters (four versus eight) and of the quantization parameters (64, 16, and 8 b) in the classification results obtained with different methods. As previously stated, the connectivity parameter is rarely mentioned in published works, though it can have a nonnegligible impact in the construction of tree representations.

In general, trees computed with four-connectivity are “finer” than those obtained with eight-connectivity. In other words, given the trees T_4 and T_8 obtained from the same image using four- and eight-connectivity, respectively, every node of T_4 is a subset of a node of T_8 .

For instance, in Figure 7(b) and (c), we show the max-trees of the image of Figure 7(a) computed with four- and eight-connectivity, respectively. It can be verified that every node of max- T_4 is a subset of a node of max- T_8 . As another example, the max-tree of the four-connected Gray–Potsdam image is composed of approximately 31% more nodes than the max-tree of the eight-connected image (4,725, and 207 versus 3,606, and 550 nodes). Concerning the max-trees computed

on the first PC of Pavia, the relative difference in the number of nodes is even higher: using four-connectivity leads to approximately 43.7% more nodes than using eight-connectivity (25,030 versus 17,413 nodes). Those observations raise the question of whether the connectivity has as much of an impact on the APs as it has on the number of tree nodes.

Similarly to the connectivity, the quantization parameter heavily affects the trees’ depth and number of nodes. For the Pavia data set (downloaded from http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes),

TABLE 8. THE CLASSIFICATION RESULTS OF THE GRAY–POTSDAM DATA SET OBTAINED BY DIFFERENT METHODS USING THE DEFAULT FOUR-CONNECTIVITY AND 1-B QUANTIZATION.

METHOD	DIMENSION	CLASSIFICATION RESULT		
		OA (%)	AA (%)	$\kappa \times 100$
Gray values	1	46.56 ± 0.42	34.3 ± 0.02	30.84 ± 0.32
AP-maxT	20	63.65 ± 0.14	59.68 ± 0.02	52.65 ± 0.15
AP-minT	20	58 ± 0.23	56.55 ± 0.02	47.13 ± 0.21
AP	38	75.1 ± 0.05	77.56 ± 0.02	67.83 ± 0.05
SDAP	20	76.19 ± 0.08	78.25 ± 0.02	69.2 ± 0.09
α -AP	20	68.32 ± 0.07	67.94 ± 0.04	59.47 ± 0.08
ω -AP	20	68.1 ± 0.06	67.78 ± 0.03	59.23 ± 0.06
LFAP	76	78.62 ± 0.04	81.21 ± 0.04	72.25 ± 0.05
LFSDAP	40	80.8 ± 0.03	83.44 ± 0.02	75.05 ± 0.04
FP_μ	38	77.14 ± 0.05	79.41 ± 0.02	70.42 ± 0.06
FP_a	38	77.19 ± 0.05	79.46 ± 0.02	70.47 ± 0.05
$FP_{\mu+a}$	76	77.94 ± 0.04	80.17 ± 0.03	71.43 ± 0.05
TF-AP	18	72.34 ± 0.05	74.93 ± 0.04	64.33 ± 0.06

The values in bold represent the highest scores in the columns OA, AA, and κ .

TABLE 9. THE CLASSIFICATION RESULTS PER CLASS OF THE GRAY–POTSDAM DATA SET OBTAINED BY DIFFERENT METHODS USING THE DEFAULT FOUR-CONNECTIVITY AND 1-B QUANTIZATION.

METHOD	DIMENSION	ACCURACY PER CLASS (%)					
		BACKGROUND	TREES	CARS	BUILDINGS	LOW VEGETATION	IMPERVIOUS SURFACES
Gray values	1	10.24 ± 1.34	16.23 ± 1.23	11.45 ± 0.22	49.82 ± 1.51	70.86 ± 1.51	47.21 ± 1.12
AP-maxT	20	53.82 ± 0.19	31.8 ± 0.63	59.4 ± 0.54	68.53 ± 0.29	77.42 ± 0.58	67.08 ± 0.62
AP-minT	20	54.74 ± 0.8	51.14 ± 1.22	58.37 ± 0.19	61.34 ± 0.83	67.38 ± 0.3	46.32 ± 1.28
AP	38	85.43 ± 0.21	63.49 ± 0.17	92.02 ± 0.08	84.15 ± 0.16	72.42 ± 0.21	67.85 ± 0.22
SDAP	20	82.57 ± 0.48	64.89 ± 0.22	92.82 ± 0.09	83.93 ± 0.14	74.43 ± 0.23	70.87 ± 0.48
α -AP	20	71.7 ± 0.14	50.39 ± 0.3	75.38 ± 0.21	80.68 ± 0.18	62.65 ± 0.13	66.86 ± 0.24
ω -AP	20	71.07 ± 0.14	50.26 ± 0.42	75.73 ± 0.24	80.24 ± 0.21	62.62 ± 0.17	66.76 ± 0.22
LFAP	76	85.18 ± 0.11	70.54 ± 0.17	96 ± 0.05	84.38 ± 0.09	75.87 ± 0.09	75.3 ± 0.07
LFSDAP	40	87.98 ± 0.09	74.25 ± 0.12	97.3 ± 0.03	86.15 ± 0.1	78 ± 0.08	76.94 ± 0.08
FP_μ	38	86.55 ± 0.22	66.47 ± 0.17	93.06 ± 0.05	85.73 ± 0.15	74.66 ± 0.17	69.97 ± 0.18
FP_a	38	86.96 ± 0.22	66.42 ± 0.18	92.98 ± 0.06	85.9 ± 0.13	74.7 ± 0.19	69.81 ± 0.21
$FP_{\mu+a}$	76	87.39 ± 0.16	67.6 ± 0.14	93.4 ± 0.05	86.41 ± 0.13	75.47 ± 0.19	70.74 ± 0.18
TF-AP	18	82.76 ± 0.18	59.09 ± 0.16	90.47 ± 0.1	81.7 ± 0.14	69.24 ± 0.14	66.32 ± 0.24

Values in bold represent the highest accuracy scores in each column.

the value of a pixel at every band is represented as a 16-b unsigned integer. However, as mentioned, the APs are computed on the four PCs of the Pavia data set. Those components are obtained with the *PCA.fit()* method of the scikit-learn Python library, which returns real-valued (64-b float) components. In general, rounding those real values to 16- or 8-b integers reduces the number of distinct values in the components and, consequently, the time and space complexity to compute their respective trees. Though the computation time and space complexity are not critical for the small-sized Pavia data set, they are of great importance for larger data sets. Hence, we study the effect of

approximating the 64-b values to 16- and 8-b values in the classification of Pavia.

Since the postprocessing techniques LFAP and HAP do not depend directly on the connectivity and quantization parameters, they are not included in this set of experiments. Moreover, as the ToS is not yet implemented with eight-connectivity in the Higras package, it is not considered in this section. In Tables 10 and 11, we present the classification results of Gray-Potsdam and Pavia data, respectively, with four- and eight-connectivity.

For the Potsdam data, the connectivity parameter had little but consistent influence on the classification results with

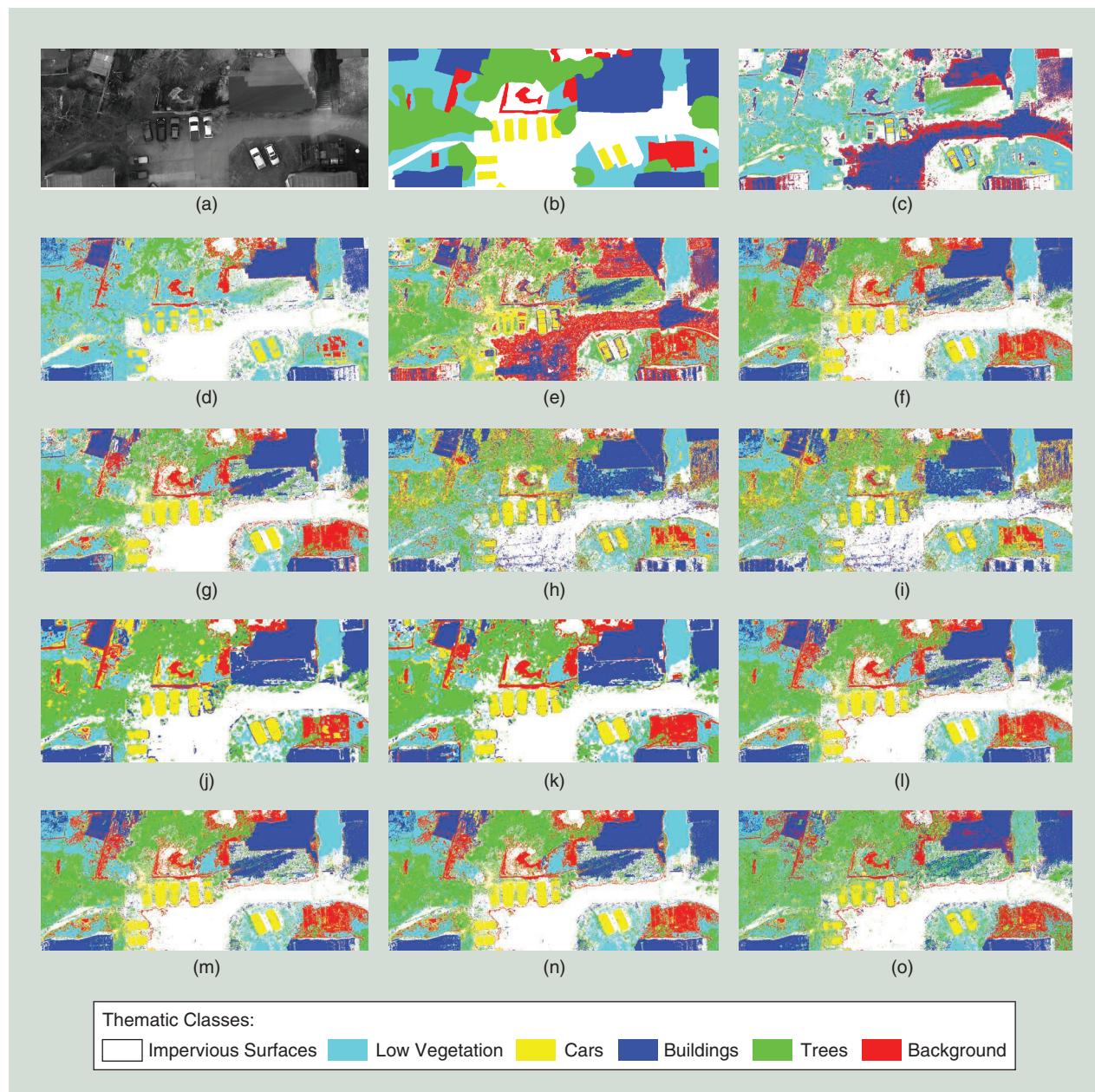


FIGURE 6. The classification results of a crop of Gray-Potsdam corresponding to the results of Table 8: the (a) crop of Gray-Potsdam, (b) crop of ground truth, (c) grayscale, (d) AP-maxT, (e) AP-minT, (f) AP, (g) SDAP, (h) α -AP, (i) ω -AP, (j) LFAP, (k) LSFDAP, (l) FP_μ , (m) FP_α , (n) $FP_{\mu+a}$, and (o) TF-AP.

different methods. The experiments with four-connectivity provided better scores in general. We attribute this result to the larger number of tree nodes obtained using four- instead of eight-connectivity. Hence, we conclude that the finer regions obtained with four-connectivity provide valuable attributes for performing classification on this data set.

Regarding the Pavia data, changing the connectivity parameter had a different impact on each method. All methods perform better on the four-connected Pavia data, except for FP_a . Notably, changing the connectivity parameter of the AP-minT from eight to four led to improvements of 8.89% and 11.02% in terms of OA and χ , respectively. On the other hand, changing the connectivity parameter of the FP_a from four to eight led to improvements of 4.09% and 4.72% in terms of OA and χ , respectively.

Overall, though, four-connectivity can be observed to almost consistently outperform eight-connectivity. In terms of future research directions, advanced connectivity concepts, such as hyperconnectivity [104] and mask-based connectivity [105], appear as promising options.

We now focus on the assessment of the quantization parameter in the classification of the Pavia data set. Since the

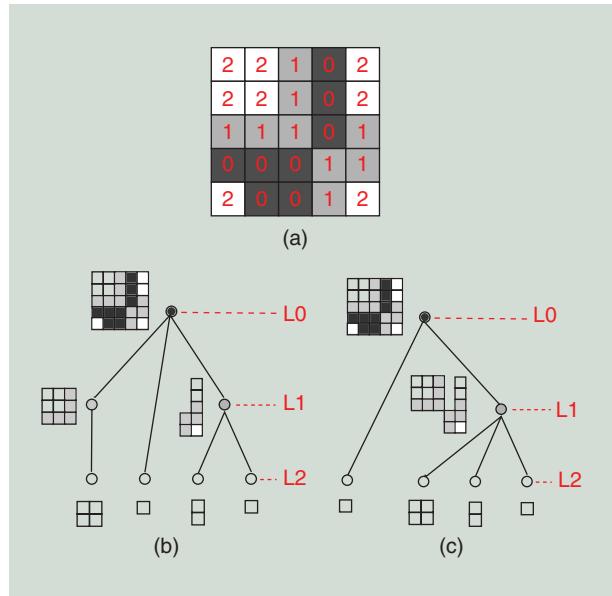


FIGURE 7. (a) The original grayscale image $X: E \rightarrow [0, 1, 2]$. (b) The max-tree of X computed using four-connectivity. (c) The max-tree of X computed using eight-connectivity.

TABLE 10. A COMPARISON OF THE CLASSIFICATION RESULTS OF GRAY-POTSDAM FOR DIFFERENT CONNECTIVITY PARAMETERS.

METHOD	DIMENSIONS	CLASSIFICATION RESULT (FOUR-CONNECTIVITY)			CLASSIFICATION RESULT (EIGHT-CONNECTIVITY)		
		OA (%)	AA (%)	$\chi \times 100$	OA (%)	AA (%)	$\chi \times 100$
AP-maxT	20	63.65 ± 0.14	59.68 ± 0.02	52.65 ± 0.15	63.03 ± 0.16	58.9 ± 0.02	51.84 ± 0.17
AP-minT	20	58 ± 0.23	56.55 ± 0.02	47.13 ± 0.21	57.36 ± 0.23	55.75 ± 0.03	46.3 ± 0.21
AP	38	75.1 ± 0.05	77.56 ± 0.02	67.83 ± 0.05	73.91 ± 0.07	76.42 ± 0.03	66.36 ± 0.07
α -AP	20	68.32 ± 0.07	67.94 ± 0.04	59.47 ± 0.08	67.67 ± 0.08	66.72 ± 0.04	58.64 ± 0.08
ω -AP	20	68.1 ± 0.06	67.78 ± 0.03	59.23 ± 0.06	67.4 ± 0.08	66.46 ± 0.03	58.34 ± 0.08
FP_μ	38	77.14 ± 0.05	79.41 ± 0.02	70.42 ± 0.06	75.95 ± 0.07	78.23 ± 0.02	68.91 ± 0.08
FP_a	38	77.19 ± 0.05	79.46 ± 0.02	70.47 ± 0.05	75.98 ± 0.09	78.28 ± 0.03	68.96 ± 0.1
$FP_{\mu+a}$	76	77.94 ± 0.04	80.17 ± 0.03	71.43 ± 0.05	76.7 ± 0.06	78.96 ± 0.03	69.86 ± 0.07
TF-AP	18	72.34 ± 0.05	74.93 ± 0.04	64.33 ± 0.06	71.9 ± 0.08	74.49 ± 0.03	63.81 ± 0.09

The values in bold represent the highest scores when comparing the scores on the left (four connectivity) with the scores on the right (eight-connectivity).

TABLE 11. A COMPARISON OF THE CLASSIFICATION RESULTS OF PAVIA FOR DIFFERENT CONNECTIVITY PARAMETERS.

METHOD	DIMENSIONS	CLASSIFICATION RESULT (FOUR-CONNECTIVITY)			CLASSIFICATION RESULT (8-CONNECTIVITY)		
		OA (%)	AA (%)	$\chi \times 100$	OA (%)	AA (%)	$\chi \times 100$
AP-maxT	80	89.21 ± 0.63	87.37 ± 0.19	85.42 ± 0.8	87.88 ± 1.16	89.56 ± 0.3	84.17 ± 1.43
AP-minT	80	87.11 ± 2	92.62 ± 0.61	83.23 ± 2.41	78.22 ± 0.5	88.26 ± 0.14	72.21 ± 0.57
AP	152	90.57 ± 2.6	93.13 ± 0.67	87.53 ± 3.32	90.54 ± 1.55	93.8 ± 0.4	87.31 ± 1.99
α -AP	80	95.46 ± 0.59	95.25 ± 1.23	93.91 ± 0.79	94.73 ± 0.4	95.66 ± 0.27	92.97 ± 0.52
ω -AP	80	96.33 ± 0.4	97.17 ± 0.89	95.08 ± 0.54	94.71 ± 0.34	95.69 ± 0.17	92.94 ± 0.44
FP_μ	152	96.76 ± 0.15	97.05 ± 0.19	95.65 ± 0.21	93.35 ± 0.15	93.55 ± 0.16	90.9 ± 0.2
FP_a	152	85.48 ± 0.98	93.94 ± 0.55	81.36 ± 1.16	89.57 ± 1.18	94.03 ± 0.32	86.08 ± 1.5
$FP_{\mu+a}$	304	96.15 ± 0.14	96.74 ± 0.32	94.82 ± 0.19	94.21 ± 0.32	95 ± 0.28	92.09 ± 0.44
TF-AP	72	93.98 ± 0.53	96.26 ± 0.29	92.04 ± 0.68	92.65 ± 0.24	95.26 ± 0.17	90.32 ± 0.31

The values in bold represent the highest scores when comparing the columns on the left (four-connectivity) with their corresponding columns on the right (four-connectivity).

TABLE 12. A COMPARISON OF THE CLASSIFICATION RESULTS OF PAVIA FOR DIFFERENT QUANTIZATION PARAMETERS.

METHOD	DIMENSION	CLASSIFICATION RESULT (8 B)			CLASSIFICATION RESULT (16 B)			CLASSIFICATION RESULT (64 B)		
		OA (%)	AA (%)	$\chi \times 100$	OA (%)	AA (%)	$\chi \times 100$	OA (%)	AA (%)	$\chi \times 100$
Four PCs	4	65.27 ± 0.25	74.88 ± 0.2	56.93 ± 0.27	68.6 ± 0.1	78.08 ± 0.09	60.93 ± 0.11	68.51 ± 0.24	78.08 ± 0.22	60.81 ± 0.29
AP-maxT	80	89.21 ± 0.63	87.37 ± 0.19	85.42 ± 0.8	88.97 ± 0.2	87.39 ± 0.15	84.99 ± 0.27	89.07 ± 0.37	87.42 ± 0.17	85.13 ± 0.49
AP-minT	80	87.11 ± 2	92.62 ± 0.61	83.23 ± 2.41	91.58 ± 1.64	94.55 ± 0.65	88.89 ± 2.06	92.42 ± 1.07	94.77 ± 0.83	89.97 ± 1.35
AP	152	90.57 ± 2.6	93.13 ± 0.67	87.53 ± 3.32	95.43 ± 0.21	94.01 ± 0.47	93.83 ± 0.28	95.5 ± 0.19	94.25 ± 0.52	93.93 ± 0.26
SDAP	80	94.1 ± 0.21	93.85 ± 0.37	92.16 ± 0.27	92.29 ± 0.32	91.36 ± 0.42	89.66 ± 0.42	92.49 ± 0.47	91.34 ± 0.61	89.92 ± 0.61
α -AP	80	95.46 ± 0.59	95.25 ± 1.23	93.91 ± 0.79	93.87 ± 0.72	93.85 ± 0.22	91.82 ± 0.93	94.08 ± 0.17	93.96 ± 0.09	92.1 ± 0.22
ω -AP	80	96.33 ± 0.4	97.17 ± 0.89	95.08 ± 0.54	93.75 ± 0.54	92.97 ± 0.76	91.64 ± 0.71	93.75 ± 0.44	92.78 ± 0.98	91.64 ± 0.59
FP_μ	152	96.76 ± 0.15	97.05 ± 0.19	95.65 ± 0.21	97.09 ± 0.7	97.43 ± 0.47	96.1 ± 0.93	97.10 ± 1	97.64 ± 0.33	96.11 ± 1.31
FP_a	152	85.48 ± 0.98	93.94 ± 0.55	81.36 ± 1.16	82.72 ± 0.74	92.64 ± 0.63	77.99 ± 0.95	83.09 ± 0.42	93.05 ± 0.29	78.5 ± 0.55
$FP_{\mu+a}$	304	96.15 ± 0.14	96.74 ± 0.32	94.82 ± 0.19	97.48 ± 0.31	98.3 ± 0.19	96.62 ± 0.41	97.55 ± 0.29	98.23 ± 0.26	96.71 ± 0.39

The values in bold represent the highest OA among the scores achieved using the three different quantization settings.

RGB values of the Potsdam data set are already provided as 8-b values, the quantization parameter is not assessed on Gray–Potsdam.

In Table 12, we present the classification results of the Pavia data set with 8- (default), 16-, and 64-b quantization. Regarding the classification based solely on the four PCs, approximating the 64-b float values to 16-b values had little influence on the results. However, approximating the 64-/16-b to 8-b values had a larger impact on the classification results, with a decrease of more than 3% in terms of accuracy and χ scores.

A more significant loss is observed in the classification results with AP-minT and AP, with a decrease of more than 5% in terms of accuracy and χ scores when the 64-b float values are approximated to 8-b values. On the other hand, we have opposite results for SDAP, α -AP, and ω -AP: using 8-b for quantization provides the best scores. Consequently, we could draw different conclusions regarding the performance of AP compared to SDAP, α -AP, and ω -AP depending on the quantization settings: if 8 b, then SDAP, α -AP, and ω -AP outperform AP; otherwise, AP outperforms those three methods.

Regarding FPs, the best FP_μ and $FP_{\mu+a}$ results are achieved with 64-b quantization, whereas the best FP_a is achieved with 8-b quantization. It is noteworthy that, as discussed previously, FP_μ and FP_a also presented opposite results with respect to connectivity parameters. This may suggest that having finer (respectively, coarser) components trees obtained with four-connectivity and 64-b quantization (respectively, eight-connectivity and 8-b quantization), for example, leads to better FP_μ (respectively, FP_a) scores.

Our experiments show that, in addition to the choice of attributes and threshold values, the connectivity and quantization parameters have a great impact as well on the performance of APs with different tree representations. More importantly, this effect varies greatly depending on the underlying tree type. In summary, these two parameters, which are almost always silently set to default values (unmentioned in published studies), appear to merit the same level of attention and care that threshold/attribute selection enjoys.

DISCUSSION OF THE GENERALIZATION OF APs

The experiments described in the previous section follow the same approach of other experiments in the literature: the training and testing features are obtained from the same tree computed on the whole data (or on the PCs of the data). This approach is reasonable when the aim is to completely classify the pixels of an image whose annotated pixels are evenly spread across this image [12]. That was the case of the training sets considered previously, which allowed us to obtain an improvement of more than 30% in terms of classification accuracy using APs with respect to spectral pixel values. However, this technique raises doubts in situations where one encounters distinct images for training and testing or, alternatively, when the training pixels are not evenly spread across the data. This

issue has already been discussed in [17], where the authors show that having training samples evenly spread across all ground-truth CCs is not a realistic scenario for evaluating classification methods in the remote sensing context.

In the following, we discuss the generalization of APs in those two scenarios. First, in the “Generalization of APs to Other Partially Annotated Images” section, we address the problem of having data composed of a single

image but with a better separation between training and testing pixels. To do so, we perform experiments using a new split of the Gray–Potsdam and Pavia data sets. Then, in the “Generalization of APs to Multiple-Image Data Sets” section, we approach the case where the training and testing sets belong to different images. We split the Gray–Potsdam image so that APs can be computed separately for training and testing pixels.

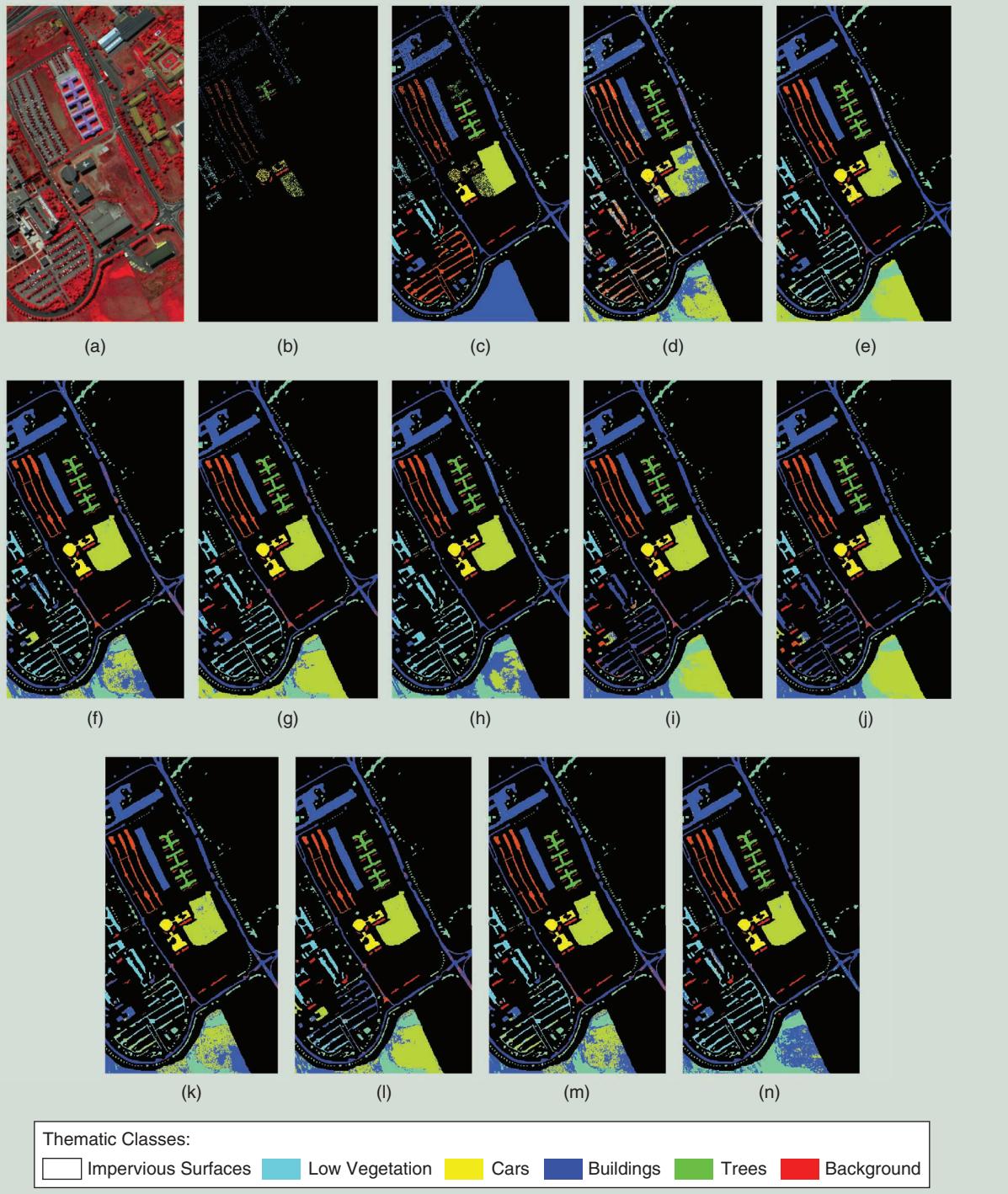


FIGURE 8. The classification results of the Pavia data set (using the partition *Pavia₂*) based on the proposed training and testing sets: the (a) image, (b) training set, (c) test set, (d) four PCs, (e) AP-maxT, (f) AP-minT, (g) AP, (h) SDAP, (i) α -AP, (j) ω -AP, (k) FP_μ , (l) FP_α , (m) $FP_{\mu+\alpha}$, and (n) TF-AP.

GENERALIZATION OF APs TO OTHER PARTIALLY ANNOTATED IMAGES

The “standard” training set of the Pavia University data set, i.e., the training set of *Pavia₁*, is composed of pixels belonging to most of the ground-truth (and testing) CCs. Hence, the training set of each semantic class accounts for most of the variability in terms of spectral signatures and geometric properties of the CCs of said class. However, the availability of training samples from every CC of each semantic class in a real-world scenario, dealing commonly with remote sensing data sets representing geographically large areas, is evidently unrealistic. The data-splitting procedure used in this section aims to simulate, to a limited degree, the aforementioned real-world conditions.

We compare the classification results presented in the previous section with those based on a new partition of the Pavia and Gray–Potsdam data sets. (The proposed split of the Pavia and Gray–Potsdam data sets is available at <https://gitlab.inria.fr/dsantana/attributes-profiles-survey-source-codes>.) In the new partition of Pavia [see Figure 8(b) and (c)], 10 sets of training samples were randomly extracted from a restricted region (composed of 85 CCs) of the ground

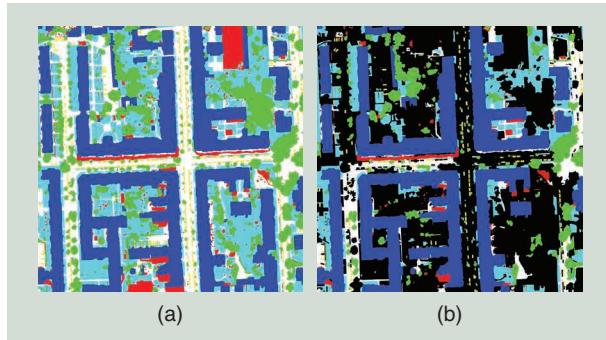


FIGURE 9. (a) The original Gray–Potsdam ground truth. (b) The CCs that contributed to the training set of the partition *GrayPotsdam₂*.

truth, which is composed of 229 CCs in total. To provide comparable results with the standard partition of this data set, we selected the same number of training samples per class, resulting in 3,921 training samples (see Table 1) and 40,002 testing samples for each of the 10 random splits.

Similarly, 10 sets of training samples of Gray–Potsdam were extracted from nearly half of the ground-truth CCs. In total, 726 ground-truth CCs (see Figure 9) contributed to the new training sets of Gray–Potsdam. The same numbers of training and test samples given in Table 2 were obtained. These new partitions of the Pavia and Gray–Potsdam data sets are denoted as *Pavia₂* and *GrayPotsdam₂*, respectively. Both the training and testing sets include pixels belonging to all classes of those data sets. The experiments were performed using the default connectivity and quantization settings given in the “Experimental Study” section which are four-connectivity and 8-b quantization. We applied the same set of thresholds for the area and moment of inertia attributes used in the previous experiments on Gray–Potsdam and Pavia.

The main challenge of using those new data set partitions is that a sample of a given semantic class is not always representative of other samples of the same class in different regions of the image. This is particularly true for the bare soil and meadows classes of Pavia. In the classification results based only on pixel intensities, several meadow samples are classified as bare soil and vice versa, as shown in Figure 5(d). A similar observation can be drawn for the building and impervious surface classes of Gray–Potsdam [see Figure 6(c)].

Tables 13 and 14 present the classification results computed on the partitions *Pavia₁* (Figure 3) and *Pavia₂* (Figure 8). First, in Table 13, we see a degradation in the baseline classification results (four PCs) using the new partition, which is due to the training set not including diversified samples of every class. Then, we can observe an even larger degradation in terms of AP results. With *Pavia₁*, all APs improve the

TABLE 13. THE CLASSIFICATION RESULTS OF THE PAVIA DATA SET PERFORMED ON THE TRAINING/TESTING SETS FROM STANDARD PARTITION PAVIA₁ (FIGURE 3) AND NEW PARTITION PAVIA₂ (FIGURE 8).

METHOD	DIMENSIONS	CLASSIFICATION RESULT (PAVIA₁)			CLASSIFICATION RESULT (PAVIA₂)		
		OA (%)	AA (%)	$\times 100$	OA (%)	AA (%)	$\times 100$
Four PCs	4	65.27±0.25	74.88±0.2	56.93±0.27	59.58±0.74	73.44±0.45	50.76±0.76
AP-maxT	80	89.21±0.63	87.37±0.19	85.42±0.8	61.23±0.68	82.27±0.36	54.42±0.69
AP-minT	80	87.11±2	92.62±0.61	83.23±2.41	68.72±2.72	79.89±1.06	62.03±2.88
AP	152	90.57±2.6	93.13±0.67	87.53±3.32	64.49±1.35	84±0.74	57.9±1.38
SDAP	80	94.1±0.21	93.85±0.37	92.16±0.27	76.63±3.7	82.05±1.08	70.85 ± 4.09
α -AP	80	95.46±0.59	95.25±1.23	93.91±0.79	64.72±0.89	76.02±1.19	57.32±1.11
ω -AP	80	96.33±0.4	97.17 ± 0.89	95.08±0.54	63.62±1.5	74.59±0.78	55.96±1.55
FP _{μ}	152	96.76 ± 0.15	97.05±0.19	95.65 ± 0.21	75.57 ± 1.43	85.36 ± 2.14	69.61±1.63
FP _{a}	152	85.48±0.98	93.94±0.55	81.36±1.16	66.4±1.17	81.48±0.39	59.8±1.17
FP _{$\mu+a$}	304	96.15±0.14	96.74±0.32	94.82±0.19	72.06±2.91	82.83±0.83	65.81±3.15
TF-AP	72	93.98±0.53	96.26±0.29	92.04±0.68	69.25±5.6	84.46±1.8	63.22±5.94

The values in bold represent the highest score in each column.

baseline results by 20.21–31.49% in terms of OA, while, with *Pavia*₂, the improvement ranges from 1.65% to 15.99%.

Moreover, with *Pavia*₂, as shown in Table 14, none of the AP-based methods outperforms the baseline with respect to the brick class, and we observe a less remarkable improvement for the meadow class. On the other hand, the best scores achieved with *Pavia*₂ for the other seven classes are comparable with the best scores obtained for *Pavia*₁.

For the Pavia data set, we can conclude that the classification results with AP-based methods presented in the “Experimental Study” section cannot be generalized to more realistic scenarios (with better separation between training and test pixels). In fact, it appears that APs are much less useful when the training and test samples belong to regions with dissimilar spatial and geometric properties. For instance, the training and test pixels of the asphalt class all belong to thin and elongated regions, which may be why the results for this class are improved by all AP-based methods with respect to the baseline. On the other hand, the pixels in the training and test sets of the meadow class belong to components with very different shapes, which may partially explain the worse results on this class.

Figure 8(d)–(n) illustrates the classification results of Table 13. We can observe that all AP-based methods considerably improve the classification results of the bare soil class (in light green) with respect to the baseline results [Figure 8(d)]. This can be due to the training and test pixels sharing nodes/features at higher levels of the trees. On the other hand, the classification results of the shadow and meadow classes are worsened by at least half of those methods.

We now compare the results between the partitions *GrayPotsdam*₁ and *GrayPotsdam*₂ of the Gray–Potsdam data set. Table 15 presents the classification results obtained with the partitions *GrayPotsdam*₁ (already given in Table 8) and *GrayPotsdam*₂. Similarly to Pavia, we can observe a significant drop in classification accuracy for all AP-based methods when the new partition *GrayPotsdam*₂ is used. While APs improve the baseline scores by up to 31.38% on the *GrayPotsdam*₁ partition, the classification improvements do not exceed 10.55% on the new partition *GrayPotsdam*₂. Moreover, the classification results of two of the six classes (low vegetation and impervious surfaces) are worsened by most of the AP-based methods, as shown in Table 16.

Figure 10(d)–(n) illustrates the classification results on the crop of Gray–Potsdam given in Figure 10(b) [the same as Figure 6(b)] obtained with the partition *GrayPotsdam*₂. The regions of the cropped image, which contributed to the training set of *GrayPotsdam*₂, are shown in Figure 10(c). Our main observation is that the classification results of the regions that do not contribute to the training sets are much poorer if compared to other regions of the same class. For instance, among the pixels belonging to the largest CC of the impervious class (which does not contribute to the training sets), virtually no pixels are correctly classified by the tested methods. In contrast, the pixels of this

TABLE 14. THE CLASSIFICATION RESULTS PER CLASS OF THE PAVIA DATA SET PERFORMED ON THE PARTITION PAVIA₂ [FIGURE 8(B) AND (C)].

METHOD	DIMENSION	ACCURACY PER CLASS (%)					
		METAL	SOIL	BITUMEN	BRICK	SHADOW	MEADOW
Four PCs	4	98.09 ± 0.29	98.43 ± 0.47	75.59 ± 3.21	60.84 ± 3.01	99.28 ± 0.16	
AP-maxT	80	95.06 ± 1.34	98.25 ± 0.19	94.68 ± 0.75	97.43 ± 0.42	99.9 ± 0.15	
AP-minT	80	92.49 ± 0.8	98.64 ± 0.21	97.61 ± 0.46	99.9 ± 0.09	57.41 ± 1.14	
AP	152	97.5 ± 1.34	98.85 ± 0.17	98.34 ± 0.87	100 ± 0	96 ± 6.05	
SDAP	80	98.34 ± 0.44	64.6 ± 8.24	99.44 ± 0.2	99.34 ± 0.19	37.45 ± 0.45	
α -AP	80	95.71 ± 0.55	32.89 ± 0.04	99.1 ± 0.18	99.1 ± 0.25	37.16 ± 0.03	54.16 ± 0.13
ω -AP	80	95.65 ± 0.41	41.58 ± 3.25	99.06 ± 0.19	99.15 ± 0.28	41.13 ± 10.79	75.18 ± 1.71
FP _{μ}	152	96.86 ± 0.93	60.86 ± 2.52	98.62 ± 1.14	99.7 ± 0.13	38.09 ± 1.46	67.01 ± 7.05
FP _{σ}	152	95.33 ± 1.21	40.76 ± 2.82	99.53 ± 0.49	98.85 ± 0.2	96.41 ± 0.88	38.17 ± 2.37
FP _{$\mu-\sigma$}	304	96.32 ± 1.34	53.13 ± 6.58	99.84 ± 0.09	98.88 ± 0.16	100 ± 0	78.65 ± 17
TF-AP	72	95.72 ± 0.59	45.59 ± 11.73	62.79 ± 6.58	98.99 ± 0.15	99.69 ± 0.12	59.01 ± 1.31
						58.70 ± 9.1	58.63 ± 1.76
							99.95 ± 0.08

The values in bold represent the highest accuracy scores in each column.

region are fairly well classified when using the partition *GrayPotsdam*₁, as shown in Figure 6.

Overall, using the new partitions *Pavia*₂ and *GrayPotsdam*₂, we observed a very significant drop of performance across all approaches. Moreover, the drop in performance occurred at different degrees for each thematic class of the data sets. This may be linked to the fact that some classes are composed of regions with similar geometric properties (e.g., the asphalt class of *Pavia* and the car class of *Gray-Potsdam*), while this is not true for the other classes (e.g., the meadow class of *Pavia* and the low vegetation class of *Gray-Potsdam*). However, a deeper investigation is necessary to confirm this assumption.

GENERALIZATION OF APs TO MULTIPLE-IMAGE DATA SETS

Though the partition *Pavia*₂ better separates training and test samples when compared to *Pavia*₁, the training and test

pixels of *Pavia*₂ may still share features obtained from nodes at higher levels of the tree. The same holds for the partition *GrayPotsdam*₂. In this section, we go one step further to completely separate the computation of training and test features.

As mentioned previously, the Potsdam data set is composed of several patches covering a large urban scene. The original International Society for Photogrammetry and Remote Sensing (ISPRS) labeling contest consisted of providing the classification labels for pixels on the testing patches based on the information given by the training patches. Hence, extracting training and testing features from the same tree is possible only if all patches are connected in the real scene. Otherwise, we would need to compute tree representations separately on the training and testing patches.

To approach the generalization of APs to this kind of problem, we split the *Gray-Potsdam* image in such a way that two independent component trees can be computed

TABLE 15. THE CLASSIFICATION RESULTS OBTAINED WITH THE PARTITIONS GRAYPOTSDAM₁ AND GRAYPOTSDAM₂ OF THE GRAY-POTSDAM DATA SET.

METHOD	DIMENSIONS	CLASSIFICATION RESULT (GRAYPOTSDAM ₁)			CLASSIFICATION RESULT (GRAYPOTSDAM ₂)		
		OA (%)	AA (%)	$\times 100$	OA (%)	AA (%)	$\times 100$
Four PCs	1	46.56 ± 0.42	34.3 ± 0.02	30.84 ± 0.32	47.76 ± 0.29	34.34 ± 0.04	32.15 ± 0.21
AP-maxT	20	63.65 ± 0.14	59.68 ± 0.02	52.65 ± 0.15	52.8 ± 0.19	45.39 ± 0.09	39.13 ± 0.16
AP-minT	20	58 ± 0.23	56.55 ± 0.02	47.13 ± 0.21	46.58 ± 0.16	44.32 ± 0.09	32.97 ± 0.14
AP	38	75.1 ± 0.05	77.56 ± 0.02	67.83 ± 0.05	57.11 ± 0.19	55.54 ± 0.12	44.5 ± 0.23
SDAP	20	76.19 ± 0.08	78.25 ± 0.02	69.2 ± 0.09	55.77 ± 0.14	54.14 ± 0.09	42.92 ± 0.15
α -AP	20	68.32 ± 0.07	67.94 ± 0.04	59.47 ± 0.08	58.31 ± 0.1	53.23 ± 0.21	46.38 ± 0.12
ω -AP	20	68.1 ± 0.06	67.78 ± 0.03	59.23 ± 0.06	58.17 ± 0.13	53.36 ± 0.2	46.26 ± 0.18
FP _{μ}	38	77.14 ± 0.05	79.41 ± 0.02	70.42 ± 0.06	57.35 ± 0.29	56.16 ± 0.14	44.87 ± 0.34
FP _{a}	38	77.19 ± 0.05	79.46 ± 0.02	70.47 ± 0.05	55.53 ± 0.29	54.81 ± 0.25	42.63 ± 0.36
FP _{$\mu+a$}	76	77.94 ± 0.04	80.17 ± 0.03	71.43 ± 0.05	56.62 ± 0.4	55.68 ± 0.19	43.9 ± 0.46
TF-AP	18	72.34 ± 0.05	74.93 ± 0.04	64.33 ± 0.06	55.52 ± 0.09	54.09 ± 0.07	42.75 ± 0.09

The values in bold represent the highest accuracy scores in each column.

TABLE 16. THE CLASSIFICATION RESULTS PER CLASS OF THE GRAY-POTSDAM DATA PERFORMED ON THE PARTITION GRAYPOTSDAM₂.

METHOD	DIMENSION	BACKGROUND	TREES	CARS	BUILDINGS	ACCURACY PER CLASS (%)	
						LOW VEGETATION	IMPERVIOUS SURFACES
Gray values	4	11.42 ± 0.7	15.64 ± 1.46	8.37 ± 0.26	48.01 ± 0.89	73.23 ± 1.48	49.38 ± 0.94
AP-maxT	80	33.17 ± 0.58	23.1 ± 1.17	47.9 ± 0.42	62.83 ± 0.6	73.86 ± 1.11	31.5 ± 0.5
AP-minT	80	47.83 ± 1.01	28.56 ± 0.68	46.29 ± 0.29	61.01 ± 0.35	33.92 ± 0.47	48.29 ± 0.88
AP	152	45.98 ± 0.32	45.31 ± 0.25	77.78 ± 0.25	80.02 ± 0.34	37.7 ± 0.38	46.44 ± 0.25
SDAP	80	46.42 ± 0.41	41.47 ± 0.28	74.75 ± 0.38	77.93 ± 0.21	37.66 ± 0.35	46.59 ± 0.17
α -AP	80	35.68 ± 1.42	43.5 ± 0.29	63.38 ± 0.35	79.05 ± 0.17	37.86 ± 0.23	59.91 ± 0.34
ω -AP	80	35.34 ± 1.29	43.44 ± 0.26	65.02 ± 0.42	78.67 ± 0.15	37.82 ± 0.43	59.88 ± 0.36
FP _{μ}	152	46.48 ± 0.21	47.49 ± 0.41	79.3 ± 0.43	80.08 ± 0.73	36.93 ± 0.33	46.71 ± 0.31
FP _{a}	152	48.67 ± 1.36	44.95 ± 0.3	76.3 ± 0.38	77.22 ± 0.67	35.63 ± 0.3	46.08 ± 0.33
FP _{$\mu+a$}	304	47.53 ± 0.62	46.52 ± 0.36	78.34 ± 0.44	79.18 ± 0.99	35.85 ± 0.41	46.67 ± 0.47
TF-AP	72	40.13 ± 0.26	43.3 ± 0.2	78.68 ± 0.18	76.15 ± 0.27	50.54 ± 0.28	35.73 ± 0.21

The values in bold represent the highest accuracy scores in each column.

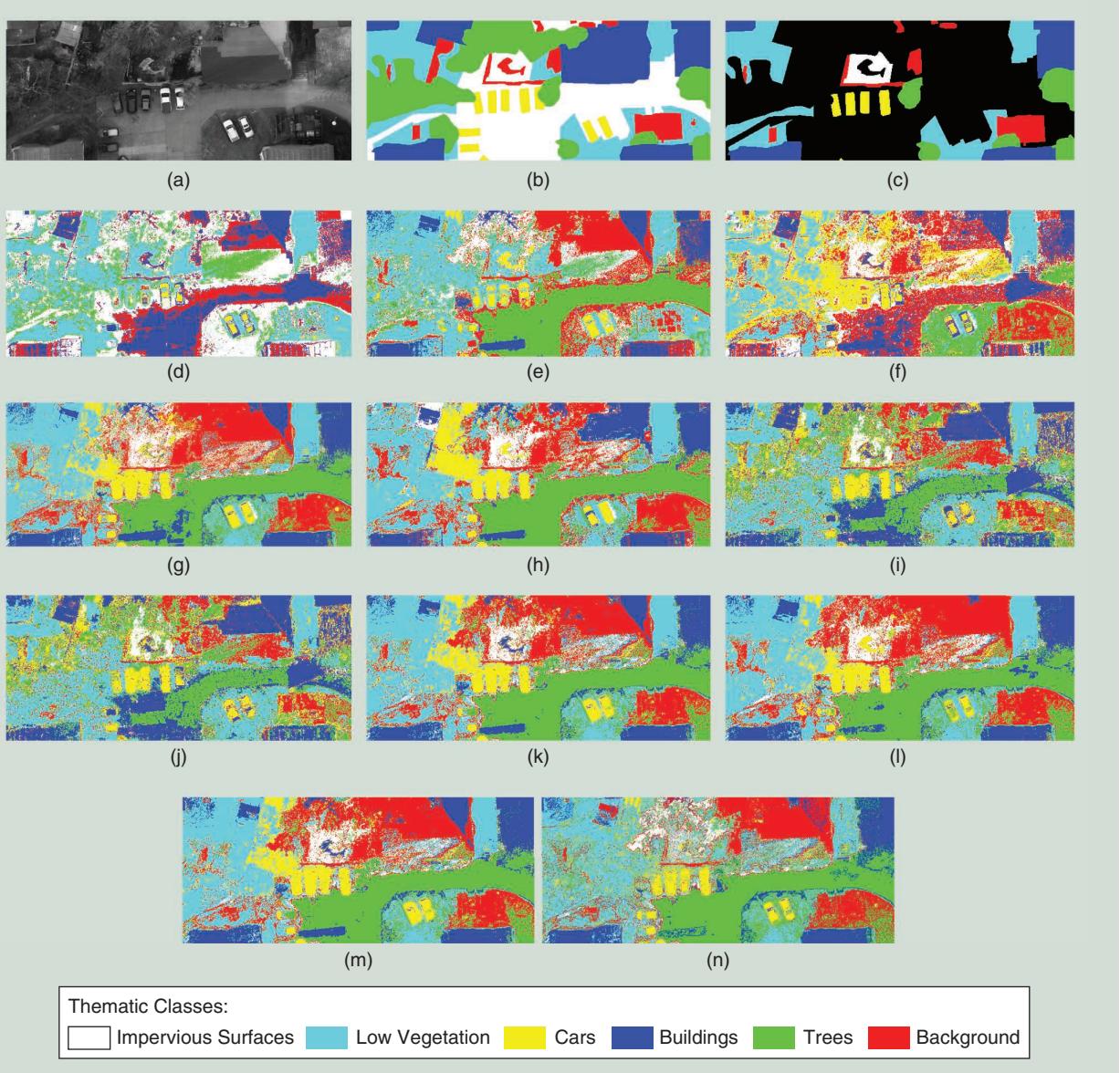


FIGURE 10. The classification results of a crop of Gray-Potsdam corresponding to the results of Table 15: the (a) crop of Gray-Potsdam, (b) crop of ground truth, (c) regions that include training pixels of *GrayPotsdam*₂, (d) grayscale, (e) AP-maxT, (f) AP-minT, (g) AP, (h) SDAP, (i) α -AP, (j) ω -AP, (k) $FP_{\mu t}$, (l) $FP_{\mu+a}$, and (n) TF-AP.

for the training and testing samples. In this way, we ensure that the training and testing pixels do not share any nodes along the trees. Alternatively, we could have chosen two neighboring image patches from the Potsdam data set, but this would not contribute more to the generalization of the method than splitting Gray-Potsdam in two.

The Gray-Potsdam data set was divided in two halves (see Figure 11) such that 10 sets of training and test samples were extracted from the upper and lower halves, respectively. From the upper half, we randomly selected 60,000 pixels of each class, as was done for the other partitions of this data set. The test set is composed of all 18,000,000 pixels in the lower half of the image. This partition of Gray-Potsdam is denoted as *GrayPotsdam*₃ in the remainder of this section.

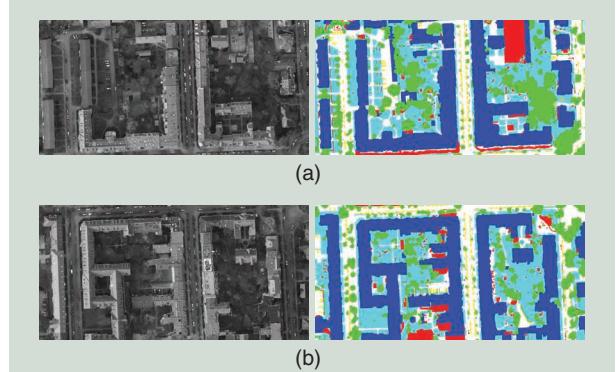


FIGURE 11. The training and testing sets from the Gray-Potsdam data set, made on the (a) upper and (b) lower halves. Both grayscale images and their ground truth are provided.

To evaluate the generalization of APs in this scenario, we performed experiments following two approaches: the standard one, where a single hierarchical representation is computed on the whole image and then used to extract training and test features, and a new approach, where two independent trees are computed on the training and test images and, hence, the training and test features are obtained from distinct trees. Tables 17 and 18 present the overall and per-class classification results,

respectively, with the partition $GrayPotsdam_3$ following those two approaches.

Our first observation is that the baseline results on the partition $GrayPotsdam_3$ present slightly higher scores when compared to the baseline results on $GrayPotsdam_1$ and $GrayPotsdam_2$. On the other hand, we observe an even larger drop in performance for all AP-based approaches with respect to the previous experiments. Following the usual approach, with a single tree computed from the whole data,

TABLE 17. THE CLASSIFICATION RESULTS OF THE GRAY-POTSDAM DATA SET WITH THE PARTITION GRAYPOTSDAM₃ FOLLOWING TWO APPROACHES: TRAINING AND TEST FEATURES EXTRACTED FROM A SINGLE TREE COMPUTED ON THE WHOLE DATA AND FROM INDEPENDENT TREES COMPUTED ON THE TRAINING AND TEST IMAGES.

METHOD	DIMENSIONS	CLASSIFICATION RESULT (POTSDAM ₃ , SINGLE TREE)			CLASSIFICATION RESULT (POTSDAM ₃ , TWO TREES)		
		OA (%)	AA (%)	$\times 100$	OA (%)	AA (%)	$\times 100$
Gray level	1	48.04 ± 0.67	33.22 ± 0.14	31.68 ± 0.58	48.04 ± 0.67	33.22 ± 0.14	31.68 ± 0.58
AP-maxT	20	50.89 ± 0.42	39.41 ± 0.19	35.11 ± 0.45	41.07 ± 1.09	33.45 ± 0.6	25.22 ± 0.97
AP-minT	20	36.34 ± 0.31	32.82 ± 0.23	21.87 ± 0.25	37.6 ± 0.64	31.24 ± 0.37	22.22 ± 0.56
AP	38	47.99 ± 0.26	40.22 ± 0.17	32.48 ± 0.3	48.02 ± 0.3	37.49 ± 0.28	31.71 ± 0.37
SDAP	20	48.47 ± 0.31	37.98 ± 0.28	32.35 ± 0.39	48.28 ± 0.32	36.97 ± 0.38	31.68 ± 0.37
α -AP	20	53.8 ± 0.26	43.32 ± 0.19	40.2 ± 0.29	46.33 ± 0.22	31.43 ± 0.31	26.66 ± 0.43
ω -AP	20	53.89 ± 0.22	43.8 ± 0.18	40.31 ± 0.25	46.21 ± 0.38	31.05 ± 0.29	26 ± 0.5
FP _{μ}	38	49.66 ± 0.22	39.76 ± 0.17	34.42 ± 0.27	49.77 ± 0.64	37.31 ± 0.26	33.30 ± 0.67
FP _{a}	38	45.03 ± 0.51	37.99 ± 0.27	29.61 ± 0.47	46.58 ± 0.38	36.33 ± 0.3	29.57 ± 0.45
FP _{$\mu+a$}	76	49.82 ± 0.32	40.7 ± 0.37	34.89 ± 0.38	51.92 ± 0.39	38.68 ± 0.33	35.64 ± 0.53

The values in bold represent the highest scores in each column.

TABLE 18. THE CLASSIFICATION RESULTS PER CLASS OBTAINED WITH THE PARTITION GRAYPOTSDAM₃ FOLLOWING TWO APPROACHES: TRAINING AND TEST FEATURES EXTRACTED FROM A SINGLE TREE COMPUTED ON THE WHOLE DATA AND FROM INDEPENDENT TREES COMPUTED ON THE TRAINING AND TEST IMAGES.

METHOD	TREES	BACKGROUND	ACCURACY PER CLASS (%)				
			TREES	CARS	BUILDINGS	LOW VEGETATION	IMPERVIOUS SURFACES
Gray values	—	1.59 ± 0.94	15.46 ± 1.25	10.99 ± 0.4	49.76 ± 1.9	76.20 ± 1.05	45.28 ± 1.35
AP-maxT	1	4.14 ± 0.35	19.17 ± 0.78	43.18 ± 0.52	60.55 ± 0.86	73.2 ± 0.78	36.2 ± 0.92
	2	7.14 ± 1.16	29.22 ± 3.02	44.43 ± 1.09	62.89 ± 0.93	33.01 ± 4.28	24.01 ± 2.68
AP-minT	1	29.45 ± 1.08	38.56 ± 1.08	25.41 ± 1.13	40.48 ± 0.84	42.99 ± 0.89	20 ± 1.28
	2	20.89 ± 0.79	38.66 ± 2.29	17.87 ± 0.64	41.74 ± 1.31	47.67 ± 1.75	20.58 ± 2.49
AP	1	10.84 ± 0.26	40.69 ± 0.74	50.81 ± 0.71	65.48 ± 0.63	49.32 ± 1.4	24.17 ± 0.77
	2	5.48 ± 0.37	38.66 ± 0.97	38.95 ± 0.51	67.15 ± 0.46	48.29 ± 1.29	26.43 ± 1.41
SDAP	1	8.71 ± 0.24	34.81 ± 0.95	40.25 ± 1.04	67.77 ± 0.33	52.98 ± 1.35	23.36 ± 0.58
	2	4.98 ± 0.49	30.96 ± 0.53	40.03 ± 1.59	69.6 ± 0.35	50.58 ± 0.79	25.64 ± 0.76
α -AP	1	9.57 ± 0.89	34.49 ± 0.88	43.74 ± 0.38	68.1 ± 0.13	54.04 ± 1.24	49.99 ± 1.6
	2	3.94 ± 0.56	11.46 ± 2.21	25.43 ± 1.39	71.73 ± 0.34	56.21 ± 1.45	19.84 ± 0.92
ω -AP	1	9.43 ± 0.88	34.39 ± 0.67	46.55 ± 0.37	67.97 ± 0.33	54.37 ± 0.98	50.07 ± 1.33
	2	2.39 ± 0.68	12.3 ± 1.9	24.72 ± 0.74	72.07 ± 0.87	55.21 ± 1.76	19.63 ± 0.78
FP _{μ}	1	10.23 ± 0.24	40.04 ± 0.84	42.02 ± 0.62	68.84 ± 0.45	49.44 ± 1.06	27.97 ± 0.9
	2	4.13 ± 0.31	36.16 ± 0.87	35.64 ± 0.67	71.73 ± 1.43	52.23 ± 1.42	23.97 ± 1.54
FP _{a}	1	10.27 ± 0.6	42.05 ± 1.49	48.75 ± 1.19	63.71 ± 1.09	41.17 ± 1.7	21.99 ± 1.04
	2	5.35 ± 1.1	32.49 ± 1.2	40.71 ± 1.33	66.74 ± 0.93	46.99 ± 1.45	25.67 ± 1.66
FP _{$\mu+a$}	1	10 ± 0.22	41.32 ± 1.5	48.38 ± 1.46	69.9 ± 0.52	49.26 ± 1.26	25.32 ± 0.88
	2	2.42 ± 0.39	37.94 ± 1.18	36.36 ± 0.81	75.43 ± 0.43	50.85 ± 1.69	29.07 ± 0.83

six over nine methods outperform the baseline. The best method (ω -AP) improves the baseline by 5.85, 10.58, and 8.63% in terms of OA, AA, and χ scores, respectively. On the other hand, with two trees computed independently on the training and test images, only three over nine methods outperform the baseline to a smaller degree. The best approach ($FP_{\mu+a}$) improves the baseline by 3.88, 5.46 and 3.96% in terms of OA, AA, and χ scores, respectively.

Comparing the results obtained with these two approaches, we conclude that the results obtained with a single tree are better in general. We attribute this to the fact that training and test pixels share nodes at higher levels of the trees computed on the whole data, leading to more similar attribute values.

In terms of accuracy per class (see Table 18), we see that, for the classes background, trees, cars, and buildings, most of the AP-based methods, computed with either one or two independent trees, improve the baseline results. In particular, the largest improvements were observed for the class cars, which happens to be composed of the most homogeneous regions in terms of shape and size. However, this was not the case for the low vegetation and impervious surface classes: the performance on those two classes is degraded when APs are used for classification.

From these results, we conclude that the features extracted from APs can be useful in the classification of data sets composed of several images extracted from a larger mosaic of images from the same scene. If a single tree representation can be computed from the whole scene, in a semisupervised scenario, the results can be more promising than if independent trees are computed on each image separately. Moreover, the accuracy scores per class raise the question of whether other geometric and topological attributes could boost the performance for the classes with less-homogeneous shapes (like low vegetation and impervious surfaces).

The experiments described in this section highlight some of the challenges we can encounter when using APs in different contexts of image classification, as in, for example, the selection of suitable attributes. Even though the improvements are less significant when compared to the standard data partitions employed in published works (as explained in the “Experimental Study” section), we are still able to benefit from the spatial features extracted from APs. This opens a path for further investigations on the extension of APs for image classification using multiple images.

FUTURE DIRECTIONS

Despite a decade of firm advances and wide acclaim by the scientific community, many AP-related research questions still remain unanswered, while new ones are added, often due to constant technological advances in terms of image acquisition. The proliferation of satellites and active and passive sensor types raises the important issue of how to represent and analyze, through AP, heterogeneous as well as multiresolution data, e.g., optical and SAR information. Likewise, the ever-increasing temporal resolutions

also present the challenge of handling multitemporal data in the context of hierarchical image representation and processing. Also, an additional significant research direction is the AP-based analysis of multivariate data (either multispectral or hyperspectral) since there is no widely accepted multivariate morphology framework as of yet.

Moreover, given the regular availability of large-scale data through missions, such as *Sentinel-1* and *Sentinel-2*, more general research directions include the computation of domain-invariant features and boosting the already-high scalability of AP. Furthermore, given the capacity of AP for effective content description through relatively few training samples and the groundbreaking content description performance of deep networks, the combination of their potentials constitutes a powerful concept worthy of pursuing. Last but not least, regarding the generalization of APs to real-world scenarios, future directions include experiments with multiple images of a data set (e.g., training and testing on multiple image patches of the Potsdam data set) and a study of the generalization capacity of APs across data sets (with similar image resolutions).

CONCLUSIONS

APs have replaced MPs as an effective spatial–spectral pixel description tool and reinforced them in terms of both computational efficiency and flexibility, thus rendering APs one of the paramount approaches of the field during the last decade, prior to the advent of deep learning. They have been employed extensively by a plethora of researchers and referenced in hundreds of publications while having been extended in a wide variety of ways.

In contrast to past works, this survey has provided an extensive review of an entire decade of AP-related developments, organized according to each of the AP calculation steps as well as in terms of adaptation strategies to multivariate data, underlying tree representations, attribute selection, and very recent postprocessing strategies. Furthermore, a comprehensive series of experiments has been conducted with multiple data sets to quantify the relative performances of major AP variants using standard and additional parameters, which have been investigated for the first time in the state of the art. It has turned out that image-quantization level and data connectivity have a significant effect on classification performance despite being often overlooked.

Our results confirm that, despite their age, the recent AP variants are powerful enough to compete in the case of some data sets even against deep learning. In addition, for the sake of reproducibility, all experiments were conducted with a publicly accessible software library.

Moreover, one more significant contribution of this survey is raising the issue of data division in the context of the underlying evaluation protocol. We have underlined the pitfalls of using a single tree structure for the entire image, as is commonly conducted in the state of the art, and propose a solution through a spatial subdivision of the image with multiple resulting trees to

simulate a real deployment scenario. Our findings show very important performance gaps that can otherwise lead to false generalization conclusions.

ACKNOWLEDGMENTS

This work was supported by the Agence Nationale de la Recherche (ANR) Multiscale project under reference ANR-18-CE23-0022. Erchan Aptoula was supported by the Tubitak project 118E258. The authors would like to thank Prof. Paolo Gamba and the ISPRS for making available the Pavia University and Potsdam data sets, respectively.

AUTHOR INFORMATION

Deise Santana Maia (deise.santana-maia@irisa.fr) is with the Université Bretagne Sud, Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Vannes, 56000, France.

Minh-Tan Pham (minh-tan.pham@irisa.fr) is with the Université Bretagne Sud, Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Vannes, 56000, France.

Erchan Aptoula (eaptoula@gtu.edu.tr) is with the Gebze Technical University, Institute of Information Technologies, Gebze, 41400, Kocaeli, Turkey.

Florent Guiotte (florent.guiotte@irisa.fr) is with the Université Bretagne Sud, Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Vannes, 56000, France.

Sébastien Lefèvre (sebastien.lefeuvre@irisa.fr) is with the Université Bretagne Sud, Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Vannes, 56000, France.

REFERENCES

- [1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York: Wiley, 2003.
- [2] 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, 2001. doi: 10.1109/36.905239.
- [3] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles," *Pattern Recogn.*, vol. 37, no. 6, pp. 1097–1116, 2004. doi: 10.1016/j.patcog.2004.01.006.
- [4] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2005, vol. 1, pp. 176–179.
- [5] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005. doi: 10.1109/TGRS.2004.842478.
- [6] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008. doi: 10.1109/TGRS.2008.922034.
- [7] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010. doi: 10.1109/TGRS.2010.2048116.
- [8] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vision Image Understanding*, vol. 64, no. 3, pp. 377–389, 1996. doi: 10.1006/cviu.1996.0066.
- [9] L. Najman and M. Couprise, "Building the component tree in quasi-linear time," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3531–3539, Nov. 2006. doi: 10.1109/TIP.2006.877518.
- [10] F. Merciol, T. Balem, and S. Lefèvre, "Efficient and large-scale land cover classification using multiscale image analysis," in *Proc. ESA Conf. Big Data from Space (BiDS)*, 2017.
- [11] P. Salembier, A. Oliveras, and L. Garrido, "Antietensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998. doi: 10.1109/83.663500.
- [12] F. Merciol et al., "Geobia at the terapixel scale: Toward efficient mapping of small woody features from heterogeneous VHR scenes," *ISPRS Int. J. Geo-Inform.*, vol. 8, no. 1, p. 46, 2019. doi: 10.3390/ijgi8010046.
- [13] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogram. Remote Sens.*, vol. 152, pp. 166–177, 2019. doi: 10.1016/j.isprsjprs.2019.04.015.
- [14] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral-spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, 2015. doi: 10.1109/TGRS.2014.2358934.
- [15] P. Ghamisi et al., "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 6, no. 3, pp. 10–43, 2018. doi: 10.1109/MGRS.2018.2854840.
- [16] F. Guiotte, S. Lefèvre, and T. Corpetti, "Rasterization strategies for airborne lidar classification using attribute profiles," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, 2019, pp. 1–4. doi: 10.1109/JURSE.2019.8808945.
- [17] N. Audebert, B. Le Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, 2019. doi: 10.1109/MGRS.2019.2912563.
- [18] 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 Recognition and Artificial Intelligence (ICPRAI)*, 2018.
- [19] M. Dalla Mura, J. Atli Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, 2010. doi: 10.1080/01431161.2010.512425.
- [20] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, 2011. doi: 10.1109/LGRS.2010.2091253.
- [21] Y. Wu, L. Zheng, D. Xie, R. Zhong, and Q. Chen, "Classification of high-resolution multispectral satellite remote sensing images using

- extended morphological attribute profiles and independent component analysis," in *Proc. 9th Int. Conf. Digital Image Process. (ICDIP 2017)*, 2017, vol. 10420, p. 104203I, doi: 10.1117/12.2281770.
- [22] S. Bernabe, P. R. Marpu, A. Plaza, M. Dalla Mura, and J. A. Benediktsson, "Spectral-spatial classification of multispectral images using kernel feature space representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 288–292, 2014. doi: 10.1109/LGRS.2013.2256336.
- [23] M. Imani and H. Ghassemian, "Attribute profile based feature space discriminant analysis for spectral-spatial classification of hyperspectral images," *Comput. Electr. Eng.*, vol. 62, pp. 555–569, 2017. doi: 10.1016/j.compeleceng.2016.09.029.
- [24] P. R. Marpu, M. Pedernana, M. Dalla Mura, S. Peeters, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *Int. J. Image Data Fusion*, vol. 3, no. 3, pp. 269–298, 2012. doi: 10.1080/19479832.2012.702687.
- [25] G. Cavallaro, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Extended self-dual attribute profiles for the classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1690–1694, 2015. doi: 10.1109/LGRS.2015.2419629.
- [26] B. B. Damodaran, N. Courty, and S. Lefèvre, "Sparse Hilbert Schmidt independence criterion and surrogate-kernel-based feature selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2385–2398, 2017. doi: 10.1109/TGRS.2016.2642479.
- [27] E. Aptoula, M. Dalla Mura, and S. Lefèvre, "Vector attribute profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3208–3220, 2016. doi: 10.1109/TGRS.2015.2513424.
- [28] B. B. Damodaran, J. Höhle, and S. Lefèvre, "Attribute profiles on derived features for urban land cover classification," *Photogrammetric Eng. Remote Sens.*, vol. 83, no. 3, pp. 183–193, 2017. doi: 10.14358/PERS.83.3.183.
- [29] M.-T. Pham, S. Lefèvre, and F. Merciol, "Attribute profiles on derived textural features for highly textured optical image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1125–1129, 2018. doi: 10.1109/LGRS.2018.2820817.
- [30] A. Taghipour and H. Ghassemian, "Hyperspectral anomaly detection using attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1136–1140, 2017. doi: 10.1109/LGRS.2017.2700329.
- [31] G. A. Licciardi, A. Villa, M. Dalla Mura, L. Bruzzone, J. Chanussot, and J. A. Benediktsson, "Retrieval of the height of buildings from worldview-2 multi-angular imagery using attribute filters and geometric invariant moments," *IEEE J. Select. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 1, pp. 71–79, 2012. doi: 10.1109/JSTARS.2012.2184269.
- [32] N. Falco, M. Dalla Mura, F. Bovolo, J. A. Benediktsson, and L. Bruzzone, "Change detection in VHR images based on morphological attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, 2012. doi: 10.1109/LGRS.2012.2222340.
- [33] M. Boldt, A. Thiele, K. Schulz, and S. Hinz, "SAR image segmentation using morphological attribute profiles," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XL-3, no. 3, p. 39, 2014. doi: 10.5194/isprsarchives-XL-3-39-2014.
- [34] L. Xue, X. Yang, and Z. Cao, "Building extraction of SAR images using morphological attribute profiles," in *Communications, Signal Processing, and Systems*. Berlin: Springer-Verlag, 2012, pp. 13–21.
- [35] P. R. Marpu, K.-S. Chen, C.-Y. Chu, and J. A. Benediktsson, "Spectral-spatial classification of polarimetric SAR data using morphological profiles," in *Proc. 3rd Int. Asia-Pacific Conf. Synthetic Aperture Radar (APSAR)*, 2011, pp. 1–3.
- [36] A. Tombak, E. Aptoula, and K. Kayabol, "Pixel-based classification of SAR images using features," in *Proc. 26th Signal Process. Commun. Appl. Conf.*, Cesme, Turkey, 2018.
- [37] M. Boldt, K. Schulz, A. Thiele, and S. Hinz, "Using morphological differential attribute profiles for change categorization in high resolution SAR images," *Inter. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XL-1/W1, no. 1, pp. 29–34, 2013. doi: 10.5194/isprsarchives-XL-1-W1-29-2013.
- [38] M. Boldt, A. Thiele, K. Schulz, and S. Hinz, "Feature extraction for change analysis in SAR time series," in *Earth Resources and Environmental Remote Sensing/GIS Applications VI*, vol. 9644. International Society for Optics and Photonics, 2015, p. 964410.
- [39] M. Boldt, A. Thiele, K. Schulz, F. J. Meyer, and S. Hinz, "Practical approach for synthetic aperture radar change analysis in urban environments," *J. Appl. Remote Sens.*, vol. 13, no. 3, p. 034528, 2019. doi: 10.1117/1.JRS.13.034528.
- [40] S. Lefèvre and E. Aptoula, "Morphological tools for spatial and multiscale analysis of passive microwave remote sensing data," in *Proc. 14th Specialist Meeting Microw. Radiometry and Remote Sens. Environ. (MicroRad)*, 2016, pp. 145–150. doi: 10.1109/MICRORAD.2016.7530523.
- [41] D. Mongus, N. Lukac, D. Obrul, and B. Žalik, "Detection of planar points for building extraction from lidar data based on differential morphological and attribute profiles," in *Proc. VCM 2013—The ISPRS Workshop on 3D Virtual City Modeling*, 2013, vol. 2, pp. 21–26.
- [42] M. Pedernana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and lidar data using extended attribute profiles," *IEEE J. Sel. Topics Sig. Proc.*, vol. 6, no. 7, pp. 856–865, 2012. doi: 10.1109/JSTSP.2012.2208177.
- [43] W. Liao et al., "Taking optimal advantage of fine spatial resolution: Promoting partial image reconstruction for the morphological analysis of very-high-resolution images," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 8–28, 2017. doi: 10.1109/MGRS.2017.2663666.
- [44] F. Guiotte, S. Lefèvre, and T. Corpetti, "Voxel-based attribute profiles on lidar data for land cover mapping," in *Proc. IGARSS 2019—IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2491–2494.
- [45] C. Tuna, F. Merciol, and S. Lefèvre, "Attribute profiles for satellite image time series," in *Proc. IGARSS 2019—IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 126–129.
- [46] C. Tuna, B. Mirmahboub, F. Merciol, and S. Lefèvre, "Component trees for image sequences and streams," *Pattern Recog. Lett.*, vol. 129, pp. 255–262, 2020. doi: 10.1016/j.patrec.2019.11.038.
- [47] M. Zhang, P. Ghamsari, and W. Li, "Classification of hyperspectral and lidar data using extinction profiles with feature fusion," *Remote Sens. Lett.*, vol. 8, no. 10, pp. 957–966, 2017. doi: 10.1080/2150704X.2017.1335902.

- [48] C. Kwan, D. Gribben, B. Ayhan, S. Bernabe, A. Plaza, and M. Selva, "Improving land cover classification using extended multi-attribute profiles (EMAP) enhanced color, near infrared, and LiDAR data," *Remote Sens.*, vol. 12, no. 9, p. 1392, 2020. doi: 10.3390/rs12091392.
- [49] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, 2016. doi: 10.1109/LGRS.2016.2600244.
- [50] D. Hong, L. Gao, R. Hang, B. Zhang, and J. Chanussot, "Deep encoder-decoder networks for classification of hyperspectral and LiDAR data," *IEEE Geosci. Remote Sens. Lett.*, 2020.
- [51] P. Bosilj, E. Kijak, and S. Lefèvre, "Partition and inclusion hierarchies of images: A comprehensive survey," *J. Imaging*, vol. 4, no. 2, p. 33, 2018. doi: 10.3390/jimaging4020033.
- [52] C. Ronse, "Ordering partial partitions for image segmentation and filtering: Merging, creating and inflating blocks," *J. Math. Imaging Vis.*, vol. 49, no. 1, pp. 202–233, 2014. doi: 10.1007/s10851-013-0455-2.
- [53] M. Dalla Mura, J. Benediktsson, and L. Bruzzone, "Self-dual attribute profiles for the analysis of remote sensing images," in *Proc. Int. Symp. Math. Morpho. Its Appl. Signal Image Proc.*, 2011, pp. 320–330.
- [54] P. Monasse and F. Guichard, "Scale-space from a level lines tree," *J. Vis. Commun. Image Represent.*, vol. 11, no. 2, pp. 224–236, 2000. doi: 10.1006/jvci.1999.0441.
- [55] G. Cavallaro, M. Dalla Mura, J. A. Benediktsson, and A. Plaza, "Remote sensing image classification using attribute filters defined over the tree of shapes," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3899–3911, 2016. doi: 10.1109/TGRS.2016.2530690.
- [56] P. Bosilj, B. B. Damodaran, E. Aptoula, M. Dalla Mura, and S. Lefèvre, "Attribute profiles from partitioning trees," in *Int. Symp. Math. Morpho. Its Appl. Signal Image Proc.*, 2017, pp. 381–392.
- [57] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1132–1145, 2008. doi: 10.1109/TPAMI.2007.70817.
- [58] P. Soille, "On genuine connectivity relations based on logical predicates," in *Proc. Int. Conf. Image Anal. Process.*, 2007, pp. 487–492.
- [59] 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 Signal Proc.: Evolution Remote Sens. (WHISPERS)*, 2014, pp. 1–4.
- [60] S. G. Koç, E. Aptoula, P. Bosilj, B. B. Damodaran, M. D Mura, and S. Lefevre, "A comparative noise robustness study of tree representations for attribute profile construction," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, 2017, pp. 1–4. doi: 10.1109/SIU.2017.7960159.
- [61] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, 2014. doi: 10.1109/TGRS.2013.2292544.
- [62] P. Ghamisi, J. A. Benediktsson, G. Cavallaro, and A. Plaza, "Automatic framework for spectral-spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE J. Select. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 2147–2160, 2014. doi: 10.1109/JSTARS.2014.2298876.
- [63] P. R. Marpu, M. Pedergnana, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral-spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, 2013. doi: 10.1109/LGRS.2012.2203784.
- [64] A. Das, K. Bhardwaj, S. Patra, and L. Bruzzone, "A novel threshold detection technique for the automatic construction of attribute profiles in hyperspectral images," *IEEE J. Select. Topics Appl. Earth Observat. Remote Sens.*, vol. 13, pp. 1374–1384, 2020. doi: 10.1109/JSTARS.2020.2981164.
- [65] A. Das, K. Bhardwaj, and S. Patra, "Morphological complexity profile for the analysis of hyperspectral images," in *Proc. 4th Int. Conf. Recent Adv. Inf. Technol. (RAIT)*, 2018, pp. 1–6.
- [66] K. Bhardwaj, S. Patra, and L. Bruzzone, "Threshold-free attribute profile for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7731–7742, 2019. doi: 10.1109/TGRS.2019.2916169.
- [67] U. Bhangale, S. S. Durbha, R. L. King, N. H. Younan, and R. Vatsavai, "High performance GPU computing based approaches for oil spill detection from multi-temporal remote sensing data," *Remote Sens. Environ.*, vol. 202, pp. 28–44, 2017. doi: 10.1016/j.rse.2017.03.024.
- [68] E. Aptoula, "Hyperspectral image classification with multidimensional attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2031–2035, 2015. doi: 10.1109/LGRS.2015.2443860.
- [69] Z. Mahmood, G. Thoonen, and P. Scheunders, "Automatic threshold selection for morphological attribute profiles," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2012, pp. 4946–4949.
- [70] M. Pedergnana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, 2013. doi: 10.1109/TGRS.2012.2224874.
- [71] G. Cavallaro, N. Falco, M. Dalla Mura, L. Bruzzone, and J. A. Benediktsson, "Automatic threshold selection for profiles of attribute filters based on granulometric characteristic functions," in *Int. Symp. Math. Morpho. Its Appl. Sig. Image Proc.*, 2015, pp. 169–181.
- [72] G. Cavallaro, N. Falco, M. Dalla Mura, and J. A. Benediktsson, "Automatic attribute profiles," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1859–1872, 2017. doi: 10.1109/TIP.2017.2664667.
- [73] K. Bhardwaj and S. Patra, "An unsupervised technique for optimal feature selection in attribute profiles for spectral-spatial classification of hyperspectral images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 138, pp. 139–150, 2018. doi: 10.1016/j.isprsjprs.2018.02.005.
- [74] E. Aptoula and S. Guner Koc, "Attribute profiles without thresholds," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2018.
- [75] U. Derbashi and E. Aptoula, "Attribute profiles without thresholds through histogram based tree path description," in *Proc. IEEE Mediterranean and Middle-East Geosci. Remote Sens. Symp. (M2GRSS)*, 2020.
- [76] P. Salembier and M. H. F. Wilkinson, "Connected operators," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 136–157, Nov. 2009. doi: 10.1109/MSP.2009.934154.
- [77] E. R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson, "Connected shape-size pattern spectra for rotation and

- scale-invariant classification of gray-scale images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 29, no. 2, pp. 272–285, Feb. 2007. doi: 10.1109/TPAMI.2007.28.
- [78] G. Cavallaro, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "A comparison of self-dual attribute profiles based on different filter rules for classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2014, pp. 1265–1268.
- [79] B. Demir and L. Bruzzone, "Histogram-based attribute profiles for classification of very high resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2096–2107, 2016. doi: 10.1109/TGRS.2015.2496167.
- [80] M.-T. Pham, S. Lefèvre, and E. Aptoula, "Local feature-based attribute profiles for optical remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1199–1212, 2018. doi: 10.1109/TGRS.2017.2761402.
- [81] R. Battiti, B. Demir, and L. Bruzzone, "Compressed histogram attribute profiles for the classification of VHR remote sensing images," in *Proc. SPIE Remote Sens.*, 2015, p. 96,430R.
- [82] M.-T. Pham, S. Lefèvre, E. Aptoula, and B. B. Damodaran, "Classification of VHR remote sensing images using local feature-based attribute profiles," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2017, pp. 1083–1086.
- [83] B. Song et al., "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, 2014.
- [84] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, 2015. doi: 10.1109/TGRS.2015.2409195.
- [85] R. Bao, J. Xia, M. Dalla Mura, P. Du, J. Chanussot, and J. Ren, "Combining morphological attribute profiles via an ensemble method for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 359–363, 2016.
- [86] E. Aptoula, M. C. Ozdemir, and B. Yanikoglu, "Deep learning with attribute profiles for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1970–1974, 2016. doi: 10.1109/LGRS.2016.2619354.
- [87] T. Tian, L. Gao, W. Song, K.-K. R. Choo, and J. He, "Feature extraction and classification of VHR images with attribute profiles and convolutional neural networks," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 18,637–18,656, 2018. doi: 10.1007/s11042-017-5331-4.
- [88] H. Teffahi, H. Yao, S. Chaib, and N. Belabid, "A novel spectral-spatial classification technique for multispectral images using extended multi-attribute profiles and sparse autoencoder," *Remote Sens. Lett.*, vol. 10, no. 1, pp. 30–38, 2019. doi: 10.1080/2150704X.2018.1523581.
- [89] S. G. Koç and E. Aptoula, "Hyperspectral image classification with convolutional networks trained with self-dual attribute profiles," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, 2017, pp. 1–4.
- [90] K. Zhu, Y. Chen, P. Ghamisi, X. Jia, and J. A. Benediktsson, "Deep convolutional capsule network for hyperspectral image spectral and spectral-spatial classification," *Remote Sens.*, vol. 11, no. 3, p. 223, 2019. doi: 10.3390/rs11030223.
- [91] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016. doi: 10.1109/TGRS.2016.2561842.
- [92] J. Xia, P. Ghamisi, N. Yokoya, and A. Iwasaki, "Random forest ensembles and extended multiextinction profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 202–216, 2018. doi: 10.1109/TGRS.2017.2744662.
- [93] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, 2017. doi: 10.1109/TGRS.2017.2768479.
- [94] W. Liao, M. Dalla Mura, J. Chanussot, R. Bellens, and W. Philips, "Morphological attribute profiles with partial reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1738–1756, 2016. doi: 10.1109/TGRS.2015.2488280.
- [95] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 2020.
- [96] M.-T. Pham, E. Aptoula, and S. Lefèvre, "Feature profiles from attribute filtering for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 11, no. 1, pp. 249–256, 2018. doi: 10.1109/JSTARS.2017.2773367.
- [97] 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 2018—IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4511–4514.
- [98] B. Perret, G. Chierchia, J. Cousty, S. Guimarães, Y. Kenmochi, and L. Najman, "Higra: Hierarchical graph analysis," *SoftwareX*, vol. 10, p. 100,335, 2019. doi: 10.1016/j.softx.2019.100335.
- [99] D. Pascale, "A review of RGB color spaces," *Tech. Rep.*, 2003. [Online]. Available: http://www.babelcolor.com/index_htm_files/A%20review%20of%20RGB%20color%20spaces.pdf
- [100] A. Liaw and M. Wiener, "Classification and regression by RandomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [101] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Select. Topics Appl. Earth Observat. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, 2016. doi: 10.1109/JSTARS.2016.2634863.
- [102] P. G. Bascoy, P. Quesada-Barriuso, D. B. Heras, F. Argüello, B. Demir, and L. Bruzzone, "Extended attribute profiles on GPU applied to hyperspectral image classification," *J. Supercomput.*, vol. 75, no. 3, pp. 1565–1579, 2019. doi: 10.1007/s11227-018-2690-1.
- [103] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, 2017. doi: 10.3390/rs9010067.
- [104] M. H. Wilkinson, "Hyperconnectivity, attribute-space connectivity and path openings: Theoretical relationships," in *Proc. Int. Symp. Mathemat. Morphol. and Its Appl. Signal Image Process.*, 2009, pp. 47–58.
- [105] G. K. Ouzounis and M. H. Wilkinson, "Mask-based second-generation connectivity and attribute filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 990–1004, 2007. doi: 10.1109/TPAMI.2007.1045.