Fuzzy ARTMAP Driven Features Selection: Towards a System to Design Information Fusion for the Geographic Discovery

P. F. Prado[1][*] and I. C. S. Duarte[2]

[1]*Department of Earth Physics and Thermodynamics, University of Valencia, Valencia, Valencian Community 50 46100, Spain*
[2]*Department of Biology, Federal University of São Carlos, São Carlos, São Paulo 18052-780, Brazil*

ABSTRACT. The availability of ground-level and overhead images is increasing fast. However, the potentialities of joint usage of these images on information fusion systems for in-deep geographic discovery, by answering the typical binomial quest: what and where are still to be completely evaluated. It is, therefore, necessary to establish a variety of procedures to study these potentialities including on how to assess domain-invariance between this imagery related to ground-level and overhead capture domains for example. We addressed in this paper a novel framework for one prospect for an optimal set of handcrafted domain-invariant features for ground the optimized usage of multiple sources of information on a feature level based fusion system. All of this was done framed by the usage of small datasets and, also, of interpretable classifiers (human-centered computing) and it paves the way towards the proposition of application software to be used potentially as a didactic tool with emphasis on the botanic domain. The results also endorse a domain-invariance that was already implicit on the results of previously published papers. Broadly evaluating, these results connecting image datasets on contrasting scales can launch new investigations in the scope of scale-free phenomena.

Keywords: methodological framework, dimensionality reduction, feature selection, training, interpretability, automatic, tree-bark, land use

1. Introduction

Traditionally in remote sensing, land-use classification is done using overhead imagery (Leung and Newsam, 2015). However, detailing overlapping classes, such as between shopping centers or warehouses, is challenging. These authors proposed an information fusion driven approach by using both digital images, captured at ground-level and overhead, because they complement each other, thus providing a more in-depth exploration of the geography of a place under study (Leung and Newsam, 2015). Tokarczyk et al. (2015) stated that, comparatively, ground-level imagery potentially is endowed with fine texture details that provide high intra-class variability and tend to minimize inter-class variability regarding an automatic classification task.

Wegner et al. (2016) evaluated information fusion on classifying trees in the urban environment of Pasadena USA. The task was to classify trees among 18 species, using convolutional neural networks fusing information of Red, Green, and Blue (RGB) aerial imagery with that registered on street view pictures in multiple zooms. The authors considered the results promising and extensible to any publicly available objects in the context of the called World-Wide Light Field (WWLF) delimited by all outdoors pictures available.

Luo et al. (2008) studied the fusion of overhead imagery and ground-level geotagged imagery to allow events recognition. Ordinary shots, taken at ground-level using normal cameras, are complementary to aerial images in terms of allowing the automatic recognition of the event (Luo et al., 2008). The authors used scale-invariant feature transform (SIFT) to describe the texture and feature descriptors locally, based on mean pixel intensity of the channel value, concerning the Hue, Saturation, and Value (HSV) color space. On the pairwise comparison results suggested, overhead imagery has more relevant information than the ground-level imagery for event recognition. The approach that proposed jointly usage of the two images had a better performance than the one in which these datasets image were considered separately, in 9 out of the 11 evaluated classes (Luo et al., 2008). Regarding classifying images samples of golf or forest classes, of the Flickr dataset, the ground-level option performed better than overhead imagery. The classifier used in the overhead images was a multiclass ADABoost and for ground images a support-vector machine (SVM) a typical shallow architecture for machine learning.

The main difference between shallow learning and deep learning architecture relates to the number of layers of artificial neurons used: shallow uses few ones comparatively to the deep learning. Shallow learning can obtain elevated classification accuracy with a small dataset with few layers if supported with

---

[*] Corresponding author. Tel.: +55 15 32297543; fax: +55 15 32297543
E-mail address: piefedo@alumni.uv.es (P. F. Prado).

relevant features descriptors extracted by experts (Pasupa and Sunhem, 2016). Deep learning notoriously uses more layers and perform automatically (it does not rely on expert knowledge) the extraction of feature descriptors but tends to present overfitting issues (i.e., over considering samples of the training phase) when classifying on small datasets. Some efforts on augmenting deep learning can potentially generate comparable performance on classifying on small datasets (Pasupa and Sunhem, 2016), but the loss in classifier's interpretability and the possibility of verification of the results remains because the features descriptors are automatically generated and do not rely on the more interpretable expert knowledge and in workable equations.

In the context of data impossibility (due to economic, physical and temporal constraints for example) like frequently verified on the botanic domain, when it's not realistic any attempt to obtain the large dataset, the usage of deep learning in a project is not recommendable due to overfitting issues. In this case, the traditional and consolidated option is the shallow learning and its interpretability qualities. The feature-level fusion driven framework described here uses a shallow learning architecture called Fuzzy ARTMAP classifier (Carpenter et al., 1992).

The objective of this framework proposition is to offer a tool to one prospect the domain-invariant features descriptors that would allow dimensionality reduction. All of this was done framed by the usage of small datasets, of interpretable classifiers (human-centered computing) and it paves the way towards the proposition of application software to be used potentially as a didactic tool with emphasis on the botanic domain. The results also endorse a domain-invariance that was already implicit on results of previously published papers. Broadly evaluating, these results connecting image datasets on contrasting scales can launch new investigations in the scope of scale-free phenomena.

2. Material and Methods

2.1. The Subset of BARKTEX–SB

The BARKTEX benchmark consists of 408 (RGB) treebark (trunk) images on the visible electromagnetic spectrum (Lakmann and Priese, 1998). A distinctive feature is that each one of the 408 images exclusively relates to one specific tree without repetitions and this fact tends to minimize bias due to features of a single tree. The color and texture feature descriptors required to achieve high accuracy on classifying these images have been under investigation (Porebski et al., 2018).

The subset of the BARKTEX (SB) was based on 48 selected images (128×192 pixels), selected to avoid external interference such as lichens and shadows. Afterward, these images were cropped to keep only the bark patches. These 48 images equally represent 6 species of trees named: Betula, Fagus, Quercus, Robinia, Pinus, and Picea. The identification codes of the images of SB can be found in the supplementary material.

2.2. The UCMERCED Benchmark–UC

This benchmark dataset is formed by overhead (aerial) or-

tho-imagery distributed, equally, among 21 landuse classes on a pixel spatial resolution of 0.30 m (1 foot). This dataset was made of 100 images (for each class) manually selected from the National Map of the United States Geological Survey (USGS) of the following regions: Birmingham, Boston, Buffalo, Columbus, Dallas, Harrisburg, Houston, Jacksonville, Las Vegas, Los Angeles, Miami, Napa, New York, Reno, San Diego, Santa Barbara, Seattle, Tampa, Tucson and Ventura. The images have 256×256 pixels (Yang and Newsam, 2010). All the 2,100 images of the UCMERCED were used as the dataset for experiments called UC.

2.3. The Features Available to Selection

The feature selection occurred among a pre-defined set of traditional and global features descriptors, which was the same for all experiments carried out. Some feature descriptors were based on channels of a collection of color spaces consisting of RGB, HSV, Lab (CIELa *b*), XYZ, and YC_bC_r called features of the "group colors" (GC). The feature descriptors calculated using gray-level images were called the "group no colors" (GNC). The gray-level image calculation was performed according to the routines of the rgb2gray function of the MATLAB software.

The values of each feature descriptor resulted in the averaged values delimited by sliding windows of 50×50 pixels regarding SB and 150×150 pixels concerning UC. These sizes were adopted as an educated guess to consider the diversity of images encompassed at each benchmark.

Given a 2D image $p(i,j,k)$ which pixels intensities, ranging from 0 to 255, can be subdivided or grouped (bins) for a histogram construction purpose in L distinct intervals. Each one of these bins is indexed by a number in the following interval $1 \leq l \leq L$ of integers, where i, j refers to horizontal and vertical coordination respectively and $k = 1, 2, 3$ stands to each one of the color channels and $k = 4$ relates to the grayscale image. The histogram h_k of each channel (color and levels of gray) is calculated by:

$$h_k(l) = \frac{n_l}{n_k} \tag{1}$$

where n_l refers to the number of pixels on the channel k related to the bin l of the total L and n_k the total number of pixels of the image in the channel k.

The set of feature descriptors named PEPTOC composed by the called first-order statistics (Gonzalez and Woods, 2006) was formed by: mean (m_k), standard deviation (μ_k), smoothness (S_k), third moment (T_k) and uniformity (U_k) and entropy (ET_k), formulated as follows:

$$m_k = \sum_{l=1}^{L} z_l h_k(z_l) \tag{2}$$

$$\mu_k = \sqrt{\sum_{l=1}^{L} z_l^2 h_k(z_l)} \tag{3}$$

$$S_k = 1 - \frac{1}{1 + \mu_k^{\,2}} \tag{4}$$

$$T_k = \sum_{l=1}^{L} z_l^{\,3} h_k(z_l) \tag{5}$$

$$U_k = \sum_{l=1}^{L} h_k(z_l)^2 \tag{6}$$

$$ET_k = -\sum_{l=0}^{L} h_k(z_l) \cdot \log_2\big(h_k(z_l) + \delta\big) \tag{7}$$

where z_l is one of the possible intensity values of each channel and δ is a small number used to avoid the logarithm of zero.

The Haralick texture features descriptors (Haralick et al., 1973) contrast ($H_{1,\theta}$), correlation ($H_{2,\theta}$), energy ($H_{3,\theta}$) and homogeneity ($H_{4,\theta}$), all of them parameterized using $d = 1$ and $\theta = 0, 45°, 90°, 135°$ for the co-occurrence matrix calculation, as preconized elsewhere (Gonzalez and Woods, 2009) and below:

$$H_{1,\theta} = \sum_{r=0}^{L-1}\sum_{s=0}^{L-1} |r-s|^2 \, C(r,s,\theta) \tag{8}$$

$$H_{2,\theta} = \sum_{r=0}^{L-1}\sum_{s=0}^{L-1} \frac{(r-\mu_r).(s-\mu_s)}{\sigma_r.\sigma_s} C(r,s,\theta) \tag{9}$$

$$H_{3,\theta} = \sum_{r=0}^{L-1}\sum_{s=0}^{L-1} C(r,s,\theta)^2 \tag{10}$$

$$H_{4,\theta} = \sum_{r=0}^{L-1}\sum_{s=0}^{L-1} \frac{C(r,s,\theta)}{1+|r-s|} \tag{11}$$

The Fourier feature descriptors used are based on Fourier spectrum (Martins et al., 2007), covering $0 \sim 180°$ with a $20°$ step. This step is an educated guess and calculated as follows:

$$F_a = \sum_{\theta=\theta_a}^{\theta_a + 20°} \left| \Im\big(p(i,j,4)\big) \right| \tag{12}$$

By calculating these formulations, using channels of distinct color spaces and for the gray-level image. As the next step, it was constructed the set of 76 features descriptors (Table 1) which described the initial dimensional space to be potentially automatically reduced using the sequential forward selection.

SB and UC datasets were used on a train test ratio of 0.5: 0.5, randomly selected onto each class of images. This procedure was done for each experimental run (E). The total amount of E was 24 equally divided between experimenting with SB and UC. This number 12 was based on the adopted 10 reported elsewhere (Xia et al., 2017).

The feature descriptor selection was driven by the "Se-quential Forward Selection" (SFS) algorithm (Fukunaga, 1990; Pudil et al, 1994), a traditional dimensionality reduction algorithm already used for automatically classifying on tree-bark images elsewhere (Porebski et al, 2010a). The pseudo-code of SFS is described in the supplementary material (Porebski et al., 2010a).

The usage of the Fuzzy ARTMAP artificial neural network's architecture (Carpenter et al., 1992) for classifying on SB images is a novelty according to our best knowledge. The same observation does not hold concerning UCMERCED imagery. This classifier was implemented as MATLAB scripts (.m type). The parametrization used was $\alpha = 0.001$, $\beta = 1$, $\rho = 1$, based on a work on medical diagnosis supported by digital images (Sáez, 2013).

2.4. The Steps Used for Results Analysis

In order to evaluate if the machine learning solution captured properly the complexity of each benchmark used than we defined the image classification accuracy $IBC_{n,\,E}$. It is the ratio between the quantities of correctly grouped (or classified) images (i.e., the same classification described by the benchmark information) divided by the quantity of images grouping/classifications attempted in each experimental run n of experiment E, as follows:

$$IBC_{n,E} = (A/B) \tag{13}$$

where A is the number of correctly, classified images after the nth selection of the image feature descriptors by the SFS algorithm in the experiment E. B is the total number of images classified i.e., 48 for SB and 2,100 for UC. Our focus was to define the minimum set of features descriptors, which allowed the maximum accuracy of each classification experimental run.

After all the available features descriptors were tested, the $IBC_{n,\,E}$ was utilized as a reference for the calculation of the normalized accuracy $IBCN_{n,\,E}$ which value equals to one indicates the conditions related to the maximum $IBC_{n,\,E}$ was reached in the execution E. The same normalization rationale was extended to the nth feature descriptor selected (n) and to that related to the maximum $IBC_{n,\,E}$ (i.e., the n_{max}) was used as a reference to calculate the called $DNM_{n,\,E}$ as follows:

$$DNM_{n,E} = \left(\frac{n}{n_{\max}}\right) \tag{14}$$

We used a unitary increment (IU) with a signal: the positive increment to indicate the selection of a feature descriptor among those of the GNC subgroup. The negative increment was used to signal the selection of features descriptors among those of the GC subgroup. The AIU index (initialized at value 0) proposition relied on the summation of these signaled increments (IU). It was used to evaluate the dynamic of the choices by the SFS algorithm. This selection history was represented graphically up to DNM equal to 100% for each execution E, in SB and UC experiments.

Table 1. Dark Hachure Indicates the Available Feature Descriptors for Image Classification

	Mean (m_k)	Standard deviation (μ_k)	Smoothness (S_k)	Third moment (T_k)	Uniformity (U_k)	Entropy (ET_k)	Haralick (All)	Fourier (All)
GNC	▓	▓	▓	▓	▓	▓	▓	
R	▓	▓	▓	▓	▓	▓		
G	▓	▓	▓	▓	▓	▓		
B	▓	▓	▓	▓	▓	▓		
A	▓	▓	▓	▓	▓	▓		
B (LAB)	▓	▓	▓	▓	▓	▓		
H	▓							
S	▓							
V	▓							
X	▓					▓		
Y	▓	▓	▓	▓	▓	▓		
Z	▓	▓	▓	▓	▓	▓		
Y	▓							
C$_b$	▓							
C$_r$	▓							

2.5. An Inter-Dataset Approach for Comparison Concerning Features Selection

A heuristic driven by the concept of the center of "mass" of data clusters was proposed and math modeled by the product and ratio of equations fitted to the observed trends. Adjusted equations (Teq) were fitted to the dispersion in the AIU × DNM plot, and other equations were fitted to the IBCN × DNM pattern for both: SB and UC. A quotient $R_{SB\text{-}UC}$ was then calculated according to this equation:

$$R_{SB-UC} = \left(\frac{IBC_{SB}}{IBC_{UC}}\right) \cdot \left(\frac{TeqAIU_{SB}}{TeqAIU_{UC}}\right) \cdot \left(\frac{TeqIBCN_{SB}}{TeqIBCN_{UC}}\right) \qquad (15)$$

where (IBC_{SB}/IBC_{UC}) is the ratio between the mean ($n = 12$) of maximum IBC of among all E of SB and UC. The objective of this heuristic was to compare, on an inter-dataset approach, the dynamics of feature selection among GNC or GC subgroup of feature descriptors considering also the IBC and IBCN values.

2.6. The Analysis of the Subgroup of Features by the Lens of the Increasing IBC

Histograms presenting the "Contributing" fraction among the "Total" selected features that increased the IBC concerning SB and UC experiments were created. We used the earth mover's distance (Rubner et al., 1998) to endow this framework with a numeric measure for the comparison between histograms.

3. Results and Discussion

The IBC verified in SB varied from 87 to 100% using 10 to 25 feature descriptors, whereas in UC we verified IBC in the range of 70 to 74% with the selection of 20 to 40 feature descriptors. The mean IBC on UC classification was 72.67% using on average 34 feature descriptors. The inferred selection

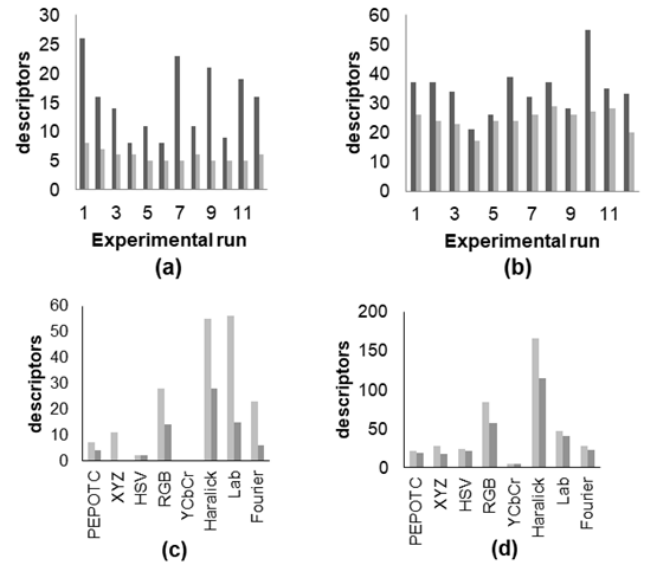patterns of feature descriptors in SB and UC were graphically summarized (Figure 1).

Figure 1. Column 1 refers to SB and Column 2 refers to UC experiments. Row 1 stands to the pairwise comparison of contributing feature descriptors along all the experimental run. For row 1, dark grey relates to the total of descriptors used for the achieve IBC maximum and light grey relates to the set of contributing descriptors i.e. those that caused positive IBC increment. Row 2 focuses on detail these contributing feature descriptor framed by subgroups clustering. The values showed stand to the summation along with all 12 experiments for SB and UC. For row 2, the total amount of each type of selected feature descriptors is shown in light gray and the number of feature descriptors, among the total, that provided a positive increment of the IBC, are shown in dark gray.

The GNC subgroup was the most important feature de-

scriptor on both: experimenting with SB and UC. This is evidenced by the initial phase of the monitoring of dynamics of the feature descriptor selection when the most significant part of the IBC increase occurred (Figure 2).

The ratio R_{SB-UC} shows evidence that SB, accumulatively, relied more on GNC than UC did. This holds until approximately DNM 6%. The $R_{SB-UC} < 1$ shows evidence of an inversion on this tendency, however, both still, relying more on GNC cumulatively. Approximately, in a 55% DNM value, the reversion of the signal of the SB/UC ratio showed more selection of GC feature descriptors than GNC in SB cumulatively, contrasting with the pattern verified in UC.
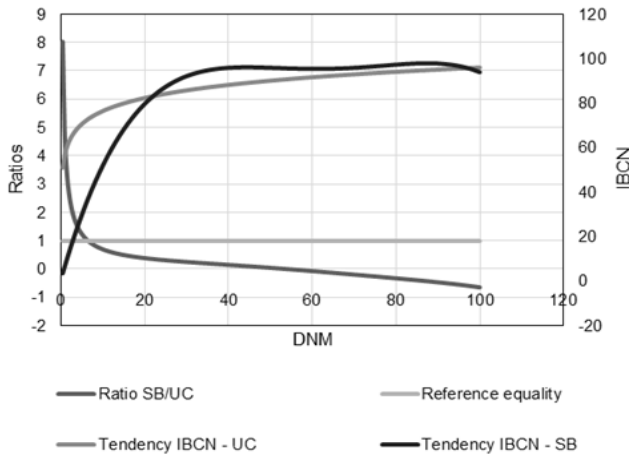

Figure 2. Outputted values by the heuristic used to evaluate the comparison of the selection among GNC and GC feature descriptors on SB and UC. On the secondary axis the corresponding IBCN.

Regarding the global ratio Contributing/Total feature descriptors, 37.91% was inferred in SB compared to 73.13% in UC. The selections in SB or UC were different in each run *E* quantitatively and qualitatively. This is an expected consequence of the methodological step related to randomly changing training and testing datasets. In the general framework of the ratio of contributors of each subgroup about the general total sum of feature descriptors selected, the results highlight some similarities indeed (Figure 3).

Intending to indicate quantitatively the distance between the histograms of Figure 3, we calculated the called "Earth mover's Distance", resulting in a value of 0.54. It can be visually inferred in Figure 3 that the prevalence of similarities occurred concentrated on the frame of the GNC feature descriptors.

Focusing on the SB, the mean ($n = 12$) maximum accuracy on the classification task on SB was 90.97% using a mean value of 17 feature descriptors. We verified the importance of the Haralick features descriptors for the classification, as well as the relevance of RGB color space. Both observations are consistent with that described by Porebski et al. (2010b) using a multiple color space (total of 28), obtained an IBC of 70.8% using 6 feature descriptors in BARKTEX images and the first

and most important feature descriptors was a Haralick feature descriptor, related to the channel B of the RGB (Porebski et al., 2010b).
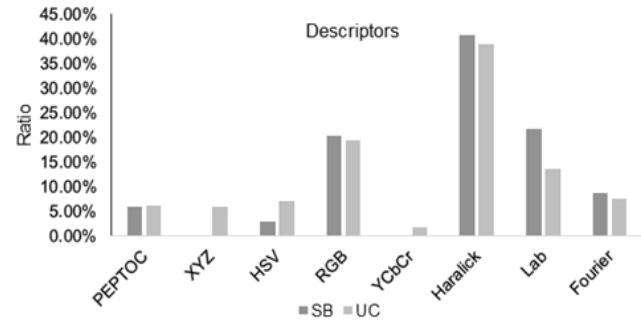

Figure 3. Pairwise comparison of ratio: contributing feature descriptors/total feature descriptors selected on the subgroup approach.

Wan et al. (2004) investigated several methods of texture analysis based on gray and color levels, with a variable amount of feature descriptors. These authors used 160 images of tree barks, divided into 9 classes and obtained IBC between 62 and 77% with feature descriptors extracted only from gray levels images. The insertion of feature descriptors extracted from colored images resulted in IBC of 73 and 89%. Noteworthy that, using the same 1-NN classifier, the maximum increment due to use color images reported by them was 15%.

Porebski et al. (2014) subdivided images of BARKTEX, composing a new benchmark called NEW BARKTEX. They adopted a 0.5:0.5 test-training ratio and obtained 75.9% of average IBC. The classifier used was the 1-NN jointly with Haralick feature descriptors based on HSV and I1 I2 I3 color spaces. These authors inferred a minimum of 48.3% of IBC with the same feature descriptors but using another classifier called Rank Sum.

Regarding the IBC on UC (72.67% using on average 34 feature descriptors), Xia et al. (2017), using global image feature descriptors and the same UC images, obtained IBC ranging from 28 to 44%. Castelluccio et al. (2015) presented reports of classification on UCMERCED with IBC ranging from 73 to 97%.

Yang and Newsam (2010) extended the concept applied on the calculation of co-occurrence matrices in gray-levels images and proposed the method called Spatial Pyramidal co-occurrence using Kernel (SPCK) for the analysis of the spatial occurrence of visual words. These authors did obtain IBC ranging from 73.8 to 76% on the UC dataset.

Notoriously, the lower (mean) maximum IBC verified and the higher amount of feature descriptors selected on UC experiments in this work shows evidence that to classify on UC is a more complex task comparatively. This possibly relates to the diversity of classes of this benchmark containing textures ranging along the not ordinated to highly ordinated. In addition, the variety of color patterns among UC dataset is worth mentioning. Kannan and Cottrell (2012) studied the effects related to the

selection of distinct methods to convert color images into grey-levels concerning the image classification task. Moreover, the Luminance method, embedded on the rgb2gray function used in this work, was preconized as the most suitable for texture discrimination and this probably favored classifying on SB than on the more complex UC dataset.

Bertrand et al. (2018) studied the automatic classification of tree-bark images using handcrafted feature descriptors, inspired on visual clues used by botanic researchers on tree identification. According to this visual approach, there are six different classes of bark called smooth lenticels, furrows, ridges, cracks, scales and strips. Bertrand et al. (2018) also used a train-test ratio of 0.5:0.5 to divide the images of the dataset. These authors used the Canny filters to map edges, which subsidized the proposition of horizontal and vertical visual words. Furthermore, the Gabor wavelet was also used with four different sinusoidal scales to provide independence about some specific distance of shooting for image capture. The H-channel of the HSV color space was used to consider the so-called typical yellow aspect of the tree-bark (Bertrand et al., 2018). As a possible substitution for H, the same authors also considered the concatenation of channels and b of the Lab color space. This option, indeed, was deprecated based on the justification of the minimization of the size of the vector of feature descriptors.

The promising results of Bertrand et al. (2018) obtained using Gabor filter-based feature descriptors resonate with the relevance of the Haralick feature descriptors in the SB classifications (Figure 3). Gabor's filter-based feature descriptors also performed well when used to classify UCMERCED images (Yang and Newsam, 2010) and this inference corroborates the similarities observed in the SB and UC results regarding the Haralick feature descriptors. The possible existence and relevance, in UC, of subtler differences between textures on the comparison with SB, possibly explains the worse performance of the Haralick feature descriptors in UC as one can infer in Figure 3. This resonates with the statement that Gabor filters performed better to treat subtle textural overlays of satellite imagery on the comparison with Haralick features (Newsam et al., 2004). This same relation of Gabor filters with Haralick also corroborates the relatively inferior performance of the Fourier descriptors observed in our results. This would occur due to the lack of the joint consideration of frequency and space that was better captured by the Haralick descriptors possibly.

The performance of feature descriptors related to the RGB color space correlates with the statement by Porebski et al. (2010a) that the RGB color space is strongly suited for classifying on BARTEX images. Histograms calculated from HSL color space channels presented a high capacity to classify images of some UCMERCED classes (Yang and Newsam, 2010). This was indeed evidenced by the performance of feature descriptors calculated from HSV channels (Figure 3).

The UC and SB results appear closely related by the lens of the GNC feature descriptor results (Figure 3), but not when evaluating using GC, except for those related to RGB. The remarks that RGB, a (Lab), and b (Lab) related feature descriptors are calculated with shared formulations and considering

the results on Lab, these suggest that the classification on the UC possibly requires a more diverse gamut than SB. This rationale, indeed, correlates with the observation that satellite images have an intense variability due to seasonal factors and location (Luo et al., 2008) what probably does not apply as intensely to SB. This additional variability was possibly tackled by the feature descriptors based on HSV and YC_bC_r color spaces standing with the discriminative power related to HSL on classifying on UC (Yang and Newsam, 2010). Luo et al. (2008) also pointed out that HSV color channels are important for the classification of satellite images.

In short, an approximately single pattern in the choice of feature descriptors was verified in the SB and UC experiments concerning the most relevant feature descriptors: Haralick and RGB (Figure 3). Secondarily, the Fourier and PEPTOC feature descriptors had fairly similar contribution patterns. The relevance of the Lab channels in the SB results occurred as expected, considering the previous results obtained by Bertrand et al. (2018).

Remarkably, Mathew et al. (2018) used deep learning networks trained on IMAGENET images, a dataset of images of ordinary objects commonly seen in daily life, to classify tree-bark images, focusing on species rather than individuals. These authors also opted for the Lab channels: *a* and *b* to address the grey/red tendency of tree barks under analysis (Mathew et al., 2018).

Penatti et al. (2015) evaluated the feature descriptors acquired by training a deep learning network to classify ground-level imagery of ordinary objects from the IMAGENET dataset. These authors classified on UCMERCED images using traditional feature descriptors and those extracted by the deep learning network after classifying IMAGENET images. The latter ones performed better than the former (traditional) approach supported by "handcrafted" feature descriptors defined in a nonautomatic way by domain experts (Penatti et al., 2015). Jointly considered the results by Mathew et al. (2018) and Penatti et al. (2015) relates the UCMERCED and BARKTEX benchmarks by the lens of shared set of automatically extracted features descriptors as represented by Figure 4.

3.1. Related Works Concerning this Framework for Dimensionality Reduction

Targeting to fuse the information tree-bark and leaf information on a system Bertrand et al. (2018) adopted the called feature-level fusion option by simply concatenating fea-tures, without previous step for dimension reduction, to be used on the tree identification task. Luo et al. (2008) did a fusion of a more expressive amount of data of satellite imagery and ground-level images, also bypassed a dimension reduction step, but, to avoid the curse-of-dimensionality risky related the concatenation of an expressive amount of features on a single vector of feature descriptors, they did choose a decision level fusion strategy.

The knowledge about the advantages of dimensionality reduction has been extensively studied especially regarding the
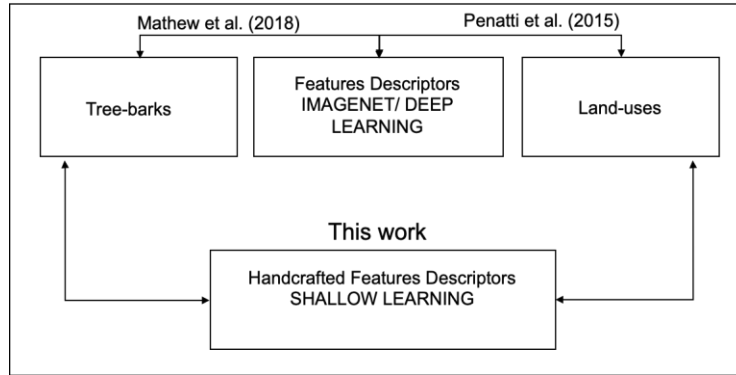
Figure 4. Schematics of indirect endorsement a domain-invariance that was already implicit on results previously published.

data input of the Fuzzy ARTMAP classifier which has an issue concerning the tendency to the proliferation of categories. Thomas and Wilscy (2011) used principal component analysis and linear discriminant analysis coupled to a simplified fuzzy ARTMAP classify on faces images for person recognition purposes. Sonawale and Ade (2015) reported that the feature selection grounded by the called Constraint selection algorithm resulted in increased accuracy of an ARTMAP classifier from 51.60 to 93.50%. The same authors pointed out that dimensionality reduction allowed to increase the accuracy of the fuzzy ARTMAP classification on approximately 19% on other paper reporting about an intrusion detection application.

Sonawale and Ade (2015) also stated that dimensionality reduction strategies related to subclass selection, like this used on this framework, allow the selection of a small group of original features. Because some original features are kept, the learning model is endowed with interpretability, comparatively, with the alternative option of the automatic feature extraction and related latent space usage.

To evoke some analogy of the approach of this work with domain adaptation methods seems to offer some valuable insights. Domain adaptation is to use features learned using training data of a source domain to solve classifications tasks on another one related target shifted-domain, commonly, without any sample labeled. Pan et al. (2011) proposed the transfer component analysis (TCA) as a new learning method. This author state that this method presents "the integration of unsupervised feature extraction and distribution matching in a latent space".

Comparatively, this heuristic framework presented by us is supported by the selection of handcrafted (and interpretable) features instead relying on automatic feature extraction. Regarding the dimensionality reduction it was approached by the SFS implementation allowing one to select the set of cross-domain relevant features in analogy to the "distribution matching in a latent space". Another important difference is that our method indeed relies on labeled sample on both of domains explored.

The Fuzzy ARTMAP classifier used in this paper endows the framework with desirable characteristics, like the convergence guarantee but also negative ones like the sensitivity to

the order of examples presented during a training phase.

Cavallo et al. (2018) proposed a framework to allow users to verify the effects of dynamically changing the input and the output on a task of dimensionality reduction. This framework is based on feature extraction driven by principal component analysis and an auto encoder usage (deep learning), contrasting with our shallow learning driven feature selection. Differently from our proposition, the application of this tool is not specific to data encoded on images but to dimensionality reduction drive exploratory analysis. Nevertheless, both share the experimental running ground, one by allowing interactive visual manipulation and our intrinsically related to the automatized exhaustive search of the SFS. Some interactivity (not visual) and manipulation are also possible in our framework by changing the pre-defined (a priori) set of feature descriptors for selection.

Lample et al. (2017) presented an application software that also explores domain-invariance based on the called Fader Networks that relied on encoder-decoder (deep learning). This software targets preserving an invariant representation of a personality image on a latent space to ground attributes modification on it for generating synthetic images with an eyeglass, for example among others possible attributes. The Sound table and sliding buttons inspired these solutions described in this paper. The preservation of the core or background information was valorized including the interactive possibility of totally reversion of the attribute modification. The interpretability is not relevant in this application but to keep track of the core information that allows the unique identification of a person is indeed a grounding base. Then, comparatively, to vary attribute has the contrasting effect of increasing the dimensionality without keeping the track of the minimal dimensionality that allows the personal identification with a resemblance to geographic discovery i.e., the most accurate whatwhere definite binomial.

Efforts to minimize the "black-box" character of the deep learning methods are still a work in progress indeed (Shrikumar et al., 2017). These authors proposed the method DeepLIFT. The rationale of this method is to offer in-sights on deep learning by allowing one to perform a kind of perturbation study. This study would allow one to gain deep insights about the learning model under evaluation.

The majority of the works related to the shallow domain adaption methods use a dimensionality reduction step based on automatic feature extraction and construction of latent spaces (Csurka et al., 2017) diverging essentially of this work regarding classifiers interpretability promotion. The same observation applies to some prospective analogies with called multi-view representation learning (Li et al., 2019) assuming that contrasting domains could stand to different views.

This interpretability issue indeed seems to be the fingerprint of an old, relevant and ongoing scientific debate concerning the delicate balance between applying mathematics and keep sensing physically the problem under analysis.

4. Conclusion

This work achieved the initial objectives by offering a methodological framework that allowed us to evaluate similarities between two disparate image domains. In addition to quantitative operational aspects, the proposed methodology presents didactic advantages and favors the understanding of its operation by people not relying on deep learning networks, and consequently not demanding large quantities of images reinforcing the aptitude for use in the botanical domain, in which the availability of images of several distinct individuals of the same species of tree is generally scarce (Bertrand et al., 2018). Another advantage is the use of feature descriptors defined by an explicit mathematical formula. Thus, avoiding the black-box approach, which is, intrinsically, unfriendly to human interpretation. These results suggest the potential to the proposition of an application software based in this framework described in this paper.

The obtained results confirm previously reported connections between the benchmarks of IMAGENET, UCMERCED, and tree-bark such as BARKTEX, notoriously, through the feature descriptors of the GNC subgroup and the respective representations in grey levels. Knowledge of similarities, even if partial, can support minimal and optimized fusion systems handling captured images at different levels to detail the geographic discovery in a region under analysis.

A possible extrapolation of this results to other fields of studies, considering the distance used for capturing images as an interferer ("background"), is to consider that the proposed methodology may be interpreted as a tool to use digital images to investigate scale-free phenomena frequently reported in nature like for example as done by Papale (2018) and many others.

References

Bertrand, S., Ameur, R.B., Cerutti, G., Coquin, D., Valet, L., and Tougne, L. (2018). Bark and Leaf Fusion Systems to Improve Automatic Tree Species Recognition. *Ecological Informatics*. https://doi.org/10.1016/j.ecoinf.2018.05.007

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.D. (1992). A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks.* 3 (5), 698-713. https://doi.o rg/10.1 109/72.159059

Cavallo, M. and Demiralp, C. (2018). A Visual Interaction Framework for Dimensionality Reduction Based Data Exploration, Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, April 21-26, Montreal QC, Canada. https://doi.org/ 10.1145/3170427.3186508

Castelluccio, M., Poggi, G., Sansone, C., and Verdoliva, L. (2015). Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv:1508.00092.

Csurka, G. (eds) (2017). Domain Adaptation in Computer Vision Applications. Advances in Computer Vision and Pattern Recognition. Springer, Cham. https://doi.org/10.1007/978-3-3195834711

Fukunaga, K. (1990). Statistical pattern recognition (second edition). San Diego, CA: Academic Press.

Gonzalez, R.C. and Woods, R.E. (2006). Digital Image Processing. 3rd edition. Prentice-Hall.

Gonzalez, R.C., Woods, R.E. and Eddins, S.L. (2009). Digital Image Processing using Matlab. 2nd edition. Gatesmark Publishing. https://doi.org/10.1117/1.3115362

Haralick, K., Robert M., Shanmugam, K., and Dinstein, I. (1973). Textural Descriptores for Image Classification. *IEEE Transactions on Systems*. Man and Cybernetics. 3(6), 610-621. https://doi.org/10.1109/TSMC.1973.4309314

Kannan, C. and Cottrell, G.W, (2012). Color-to-Grayscale: Does the Method Matter in Image Recognition? *PLoS One*. 7(1), e29740. https://doi.org/10.1371/journal.pone.0029740.

Lakmann, R. and Priese, L. (1998). A reduced covariance color texture model for micro-textures. Proc. 10thScandinavian Conf. on Image Analysis, Lappeenranta, Finland, 947-953.

Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L. and Ranzato, M.A. (2017). Fader networks: manipulating images by sliding attributes. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Ulrike von Luxburg, Isabelle Guyon, Samy Bengio, Hanna Wallach, and Rob Fergus (Eds.). Curran Associates Inc., USA, 5969-5978.

Leung, D. and Newsam, S. (2015). Land cover classification using geo-referenced photos. *Multimedia Tools and Applications*. 74, 11741. https://doi.org/10.1007/s11042-014-2261-2

Li, Y., Yang, M. and Zhang, Z. (2019). A Survey of Multi-View Representation Learning. *IEEE Transactions on Knowledge and Data Engineering,* 31(10), 1863-1883. https://doi.org/10.1109/TKDE.2018.2872063

Luo, J., Yu, J., Joshi, D. and Hao, W. (2008). Event recognition: viewing the world with a third eye. In Proceedings of the 16th ACM international conference on Multimedia. ACM, New York, NY, USA, 1071-1080. https://doi.org/10.1145/14593 59.14 59574

Martins, A.C.G., Simões, A.S. and Prado, G.I. (2007). Classificação de texturas usando redes neurais artificiais e o modelo de Spot Noise. In: VI ENIA - Encontro Nacional de Inteligência Artificial, Rio de Janeiro. Anais do VI ENIA-Encontro Nacional de Inteligência Artificial. CD-ROM.

Mathew C, Philippe G, and Jonathan G. (2018). Tree species identification from bark images using convolutional neural networks. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1075-1081). IEEE.

Newsam, S., Wang L., Bhagavathy, S. and Manjunath, B.S. (2004). Using texture to analyze and manage large collections of remote sensed image and video data. *Applied optics*. https://doi.org/10.1364/AO.43.000210

Pan, S.J., Tsang, I.W., Kwok, J.T., and Yang, Q. (2011). Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks.* 22(2), 199-210. https://doi.org/10.1109/TNN.2010.2091281

Papale, P. (2018). Global time-size distribution of volcanic eruptions

on Earth. *Scientific Reports*. 8, 6838. https://doi.org/10.1038/s4 1598-018-25286-y

Pasupa, K. and Sunhem, W. (2016) A comparison between shallow and deep architecture classifiers on small dataset. 8th International Conference on Information Technology and Electrical Engineering (ICI TEE), Yogyakarta, 1-6. https://doi.org/10.1109/ICITEED.201 6.78 63293

Penatti, O.A.B., Nogueira, K. and Dos Santos, J.A. (2015) Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, pp. 44-51. https://doi.org/10.1109/CVPRW.2015.7301382

Porebski, A., Vandenbroucke, N. and Macaire, L. (2010a). Comparison of feature selection schemes for color texture classification. 2nd International Conference on Image Processing Theory, Tools and Applications, 2010. Paris, 32-37 https://doi.org/10.1109/IPTA.201 0.5586760

Porebski, A., Vandenbroucke, N. and Macaire, L. (2010b). A multi-color space approach for texture classification: experiments with outex, vistex and barktex image databases. 5th European Conf. on Colour in Graphics, Image and Vision (CGIV'10), Joensuu (Finland), 314-319, Society for Imaging Science and Technology (IS&T), (ISBN: 978-0-89208-291-9).

Porebski, A., Vandenbroucke, N., Macaire, L. and Hamad, D. (2014). A new benchmark image test suite for evaluating colour texture classification schemes. *Multimedia Tools and Applications*. 70 (1), 543-556. https://doi.org/10.1007/s11042-013-1418-8

Porebski, A., Truong Hoang, V., Vandenbroucke, N. and Hamada, D. (2018). Multi-color space local binary pattern-based feature selection for texture classification. *Journal of Electronic Imaging*. 27 (1). https://doi.org/10.1117/1.JEI.27.3.039801

Pudil, P., Novovicova, J. and Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*. 15, 1119-1125. https://doi.org/10.1016/0167-8655(94)90127-9

Rubner, Y, Tomasi, C. and Guibas, L.J. (1998). A metric for distributions with applications to image databases. Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 59-66. https://doi.org/10.1109/ ICCV.1998.710701

Sáez, A., Rivas, E., Montero-Sánchez, A., Paradas, C., Acha, B., Pascual, A. and Escudero, L.M. (2013). Quantifiable diagnosis of muscular dystrophies and neurogenic atrophies through network analysis. *BMC Medicine*. 11(1), 77, 2013. https://doi.org/10.1186/17 41-7015-11-77

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning Important Features Through Propagating Activation Differences. In Proceedings of the 34th International Conference on Machine Learning. Volume 70 (3145-3153). JMLR.

Sonawale, S., and Ade, R. (2015). Dimensionality Reduction: An Effective Technique for Feature Selection. *International Journal of Computer Applications* (0975-8887), 117(3). https://doi.org/10. 5120/20535-2893

Thomas, A. and Wilscy, M. (2011). Face Recognition Using Simplified Fuzzy Artmap. *Signal & Image Processing*. 1.10.5121/sipi j.2010.1212. https://doi.org/10.5121/sipij.2010.1212

Tokarczyk, P. Wegner, J.D., Walk, S. and Schindler, K. (2015). Features, Color Spaces, and Boosting: New Insights on Semantic Classification of Remote Sensing Images, *IEEE Transactions on Geoscience and Remote Sensing*, 53(1), 280-295. https://doi.org /10.1109/TGRS.2014.2321423

Wan, Y.Y., Du, J.X., Huang, D.S., Chi, Z., Cheung, Y.M., Wang, X.F., and Zhang, G.J. (2004). Bark texture feature extraction based on statistical texture analysis. International Symposium on Intelligent Multimedia, Video and Speech, Hong Kong Anais da 2004 International Symposium on Intelligent Multimedia, Video and Speech. p. 482-485

Xia, G.S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y. and Lu, X (2017). AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. I*EEE Transactions on Geoscience and Remote Sensing*. 1-17. https://doi.org/10.1109/TGRS.2017.268 5945

Yang, Y.; Newsam, S. (2010) Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010, San Jose, USA https://doi.org/10.1145/186 9790.1869829

Wegner, J.D., Branson, S., Hall, D., Schindler, K., and Perona, P. (2016). Cataloging Public Objects Using Aerial and Street-Level Images - Urban Trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6014-6023). https://doi.org/10.1109/CVPR.2016.647