



**HAL**  
open science

## Ultimate levelings with strategy for filtering undesirable residues based on machine learning

Wonder Alves, Charles F Gobber, Dennis da Silva, Alexandre Morimitsu, Ronaldo F Hashimoto, Beatriz Marcotegui

### ► To cite this version:

Wonder Alves, Charles F Gobber, Dennis da Silva, Alexandre Morimitsu, Ronaldo F Hashimoto, et al.. Ultimate levelings with strategy for filtering undesirable residues based on machine learning. 14th International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Jul 2019, Saarbrücken, Germany. <10.1007/978-3-030-20867-7\_23>. <hal-02430523>

**HAL Id: hal-02430523**

**<https://minesparis-psl.hal.science/hal-02430523v1>**

Submitted on 7 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Ultimate levelings with strategy for filtering undesirable residues based on machine learning

Wonder A. L. Alves<sup>1</sup>, Charles F. Gobber<sup>1</sup>, Dennis da Silva<sup>2</sup>, Alexandre Morimitsu<sup>2</sup>, Ronaldo F. Hashimoto<sup>2</sup>, and Beatriz Marcotegui<sup>3</sup>

<sup>1</sup> Informatics and Knowledge Management Graduate Program,  
Universidade Nove de Julho, São Paulo, Brazil

<sup>2</sup> Department of Computer Science, Institute of Mathematics and Statistics,  
Universidade de São Paulo, São Paulo, Brazil

<sup>3</sup> Centre for Mathematical Morphology, Mines-ParisTech, Fontainebleau, France

**Abstract.** Ultimate levelings are operators that extract important image contrast information from a scale-space based on levelings. During the residual extraction process, it is very common that some residues are selected from undesirable regions, but they should be filtered out. In order to avoid this problem some strategies can be used to filter residues extracted by ultimate levelings. In this paper, we introduce a novel strategy to filter undesirable residues from ultimate levelings based on a regression model that predicts the correspondence between objects of interest and residual regions. In order to evaluate our new approach, some experiments were carried out with a plant dataset and the results show the robustness of our method.

## 1 Introduction

Residual operators are transformations that involve combinations of morphological operators with differences. Morphological gradient, top-hat transforms, skeleton by maximal balls and ultimate opening are some examples of residual operators widely used in image processing applications.

There is a class of important residual operators called *ultimate levelings* [5]. They are residual operators that analyze the evolution of the residual values between two consecutive operators on a scale-space of levelings and keep the maximum residues for each pixel. The residual value of these operators can reveal important contrast information in images. This class of operators includes maximum difference of openings (resp., closings) by reconstruction [15], differential morphological profiles [21], ultimate attribute openings (resp., closings) [22], differential attribute profiles [9], shape ultimate attribute openings (resp., closings) [12], differential area profiles [19] and ultimate grain filters [3]. They have successfully been used as a preprocessing step in various applications such as texture features extraction [15], segmentation of high-resolution satellite imagery [20,9], text location [23,1] and segmentation of building façades [12].

Due to the design of the ultimate levelings, some residues extracted by them can be from undesirable regions. In this sense, several researches have been pro-

posed in recent years, introducing strategies to filter undesirable residues during the residual extraction process [5,12,3,2]. They are based on attributes extracted from residual regions. However, when the number of attributes increases the strategy construction becomes difficult, because you need to find thresholds to apply the strategy. In order to deal with this problem we can made use of machine learning. In addition, the combination of machine learning and morphological operators is successfully applied in a large range of problems, such that: morphological operators learning [16], convolutional nets and watershed cuts [8], morphological profiles [9], and others.

Given the above considerations, the main contribution of this paper is a novel strategy to filter undesirable residues based on machine learning. Since ultimate levelings can be efficiently computed from morphological trees, information can be extracted from branches of the trees to obtain features for a regression or a classification model. This idea is inspired by maximally stable extremal regions [17] and morphological profiles [9], but we made it considering contrast information extracted by the residues. In order to apply and evaluate our new approach, we chose the plant bounding box detection problem [18].

The remainder of this paper is structured as follows. Sections 2 and 3 briefly recall some definitions and properties about the morphological trees and the ultimate levelings. The original contribution of this paper is given in Section 4, where we introduce a novel strategy to filter undesirable residues from the ultimate levelings based on a regression model that predicts the correspondence between objects of interest and residual regions. Experimental results are shown in Section 5. Finally, Section 6 concludes this work and presents some future research directions.

## 2 Theoretical Background

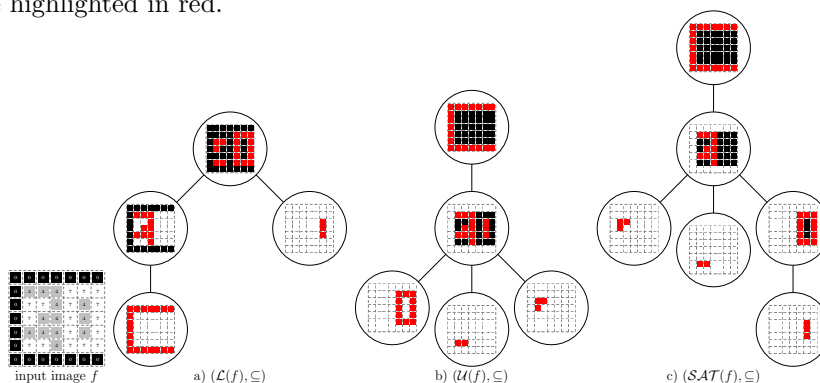
For decades, image representations through trees have been proposed to carry out tasks of image processing and analysis, such as: filtering, segmentation, pattern recognition, contrast extraction, compression and others. In this scenario, the image is represented by means of a tree, then all tasks are performed through information extraction or modifications in the tree itself, and finally an image is reconstructed from the modified tree.

In order to build the trees considered in this paper, we need the following definitions. First, we consider images as mappings from a Cartesian grid  $\mathcal{D} \subset \mathbb{Z}^2$  to a discrete set of  $k \geq 1$  integers  $\mathbb{K} = \{0, 1, \dots, k - 1\}$ . These mappings can be decomposed into *lower (strict)* and *upper (large)* level sets, i.e., for any  $\lambda \in \mathbb{K}$ ,  $\mathcal{X}_\downarrow^\lambda(f) = \{p \in \mathcal{D} : f(p) < \lambda\}$  and  $\mathcal{X}_\uparrow^\lambda(f) = \{p \in \mathcal{D} : f(p) \geq \lambda\}$ . From these sets, we define two other sets  $\mathcal{L}(f)$  and  $\mathcal{U}(f)$  composed by the connected components (CCs) of the lower and upper level sets of  $f$ , i.e.,  $\mathcal{L}(f) = \{\tau \in \mathcal{CC}(\mathcal{X}_\downarrow^\lambda(f)) : \lambda \in \mathbb{K}\}$  and  $\mathcal{U}(f) = \{\tau \in \mathcal{CC}(\mathcal{X}_\uparrow^\lambda(f)) : \lambda \in \mathbb{K}\}$ , where  $\mathcal{CC}(\mathcal{X})$  denotes the sets of either 4 or 8-CCs of  $\mathcal{X}$ , respectively.

The ordered pairs consisting of the CCs of the lower and upper level sets and the usual inclusion set relation, i.e.,  $(\mathcal{L}(f), \subseteq)$  and  $(\mathcal{U}(f), \subseteq)$ , induce two

dual trees [24,13,7] called *component trees*. It is possible to combine them into a single tree in order to obtain the so-called *tree of shapes*. Then, let  $\mathcal{P}(\mathcal{D})$  denote the *powerset* of  $\mathcal{D}$  and let  $sat : \mathcal{P}(\mathcal{D}) \rightarrow \mathcal{P}(\mathcal{D})$  be the operator of saturation [7] (or filling holes). Then,  $\mathcal{SAT}(f) = \{sat(\tau) : \tau \in \mathcal{L}(f) \cup \mathcal{U}(f)\}$  be the family of CCs of the upper and lower level sets with holes filled. The elements of  $\mathcal{SAT}(f)$ , called *shapes*, are nested by the inclusion relation and thus the pair  $(\mathcal{SAT}(f), \subseteq)$  induces a tree which is called *tree of shapes* [7].

In tree of shapes, and also component trees (max-tree and min-tree), each pixel  $p \in \mathcal{D}$  is associated only to the smallest Connected Component (CC) of the tree containing it; and through parenthood relationship, it is also associated to all its ancestor nodes. Then, we denote by  $\mathcal{SC}(\mathcal{T}, p)$  the smallest CC containing  $p$  in tree  $\mathcal{T}$ . Similarly, we say  $p \in \mathcal{D}$  is a *compact node pixel* (CNP) of a given CC  $\tau \in \mathcal{T}$  if and only if  $\tau$  is the smallest CC containing  $p$ , i.e.,  $\tau = \mathcal{SC}(\mathcal{T}, p)$ . Fig. 1 shows examples of min-tree, max-tree, and tree of shapes, where CNPs are highlighted in red.



**Fig. 1.** Min-tree (a), max-tree (b) and tree of shapes (c) as compact representations of  $(\mathcal{L}(f), \subseteq)$  and  $(\mathcal{U}(f), \subseteq)$ , and  $(\mathcal{SAT}(f), \subseteq)$ , respectively, of the image  $f$ . Only Compact Node Pixels (CNPs) are stored and they are highlighted in red.

### 3 Ultimate levelings

Ultimate levelings constitute a wider class of residual operators defined from a scale-space of levelings  $\{\psi_i : i \in \mathcal{I}\}$  [5,3,2]. An ultimate leveling analyzes the evolution of residual values from a family of consecutive primitives, i.e.  $r_i^+(f) = [\psi_i(f) - \psi_{i+1}(f) \vee 0]$  and  $r_i^-(f) = [\psi_{i+1}(f) - \psi_i(f) \vee 0]$ , keeping the maximum positive and negative residues for each pixel. Thus, contrasted objects can be detected if a relevant residue is generated when they are filtered out by one of these levelings.

More precisely, the ultimate leveling  $\mathcal{R}$  is defined for any image  $f$  as follows:

$$\mathcal{R}(f) = \mathcal{R}^+(f) \vee \mathcal{R}^-(f), \quad (1)$$

where  $\mathcal{R}^+(f) = \sup_{i \in \mathcal{I}} \{r_i^+(f)\}$  and  $\mathcal{R}^-(f) = \sup_{i \in \mathcal{I}} \{r_i^-(f)\}$ .

Residual values of these operators can reveal important contrasted structures in the image. In addition to these residues, other associated information can be obtained such as properties of the operators that produced the maximum residual value. For example, W. Li [15] introduced a function  $q_{\mathcal{I}_{\max}} : \mathcal{D} \rightarrow \mathcal{I}$  that associates to each pixel the major index that produces the maximum non-null residue, i.e.,

$$p \in \mathcal{D}, q_{\mathcal{I}_{\max}}(p) = \begin{cases} q_{\mathcal{I}_{\max}}^+(p), & \text{if } [\mathcal{R}(f)](p) > [\mathcal{R}^+(f)](p), \\ q_{\mathcal{I}_{\max}}^-(p), & \text{otherwise.} \end{cases} \quad (2)$$

where  $q_{\mathcal{I}_{\max}}^+(p) = \max\{i + 1 : [r_i^+(f)](p) = [\mathcal{R}^+(f)](p) > 0\}$  and  $q_{\mathcal{I}_{\max}}^-(p) = \max\{i + 1 : [r_i^-(f)](p) = [\mathcal{R}^-(f)](p) > 0\}$ .

The ultimate levelings can be efficiently implemented thanks to the theorem proposed in W. A. L. Alves et. al [4], which shows that an increasing family of levelings  $\{\psi_i : i \in \mathcal{I}\}$  can be obtained through a sequence of pruned trees  $(\mathcal{T}_f^0, \mathcal{T}_f^1, \dots, \mathcal{T}_f^{\mathcal{I}_{\max}})$  from the structure of the max-tree, min-tree or tree of shapes  $\mathcal{T}_f$  constructed from the image  $f$ . Then, the  $i$ -th positive (resp. negative) residue  $r_i^+(f)$  can be obtained from the set of nodes  $\mathcal{N}r(i) = \mathcal{T}_f^i \setminus \mathcal{T}_f^{i+1}$ , i.e.,  $\forall \tau \in \mathcal{N}r(i)$ ,

$$r_{\mathcal{T}_f^i}^+(\tau) = \begin{cases} \text{level}(\tau) - \text{level}(\text{Parent}(\tau)), & \text{if } \text{Parent}(\tau) \notin \mathcal{N}r(i), \\ \text{level}(\tau) - \text{level}(\text{Parent}(\tau)) \\ + r_{\mathcal{T}_f^i}^+(\text{Parent}(\tau)), & \text{otherwise,} \end{cases} \quad (3)$$

where  $\text{level}(\tau)$  and  $\text{Parent}(\tau)$  are functions that represent the gray level and the parent node of  $\tau$  in  $\mathcal{T}_f$ , respectively. Thus, the  $i$ -th positive (resp. negative) residue  $r_i^+(f)$  is given as follows:

$$\forall p \in \mathcal{D}, [r_i^+(f)](p) = \begin{cases} r_{\mathcal{T}_f^i}^+(\mathcal{SC}(\mathcal{T}_f^i, p)), & \text{if } \mathcal{SC}(\mathcal{T}_f^i, p) \in \mathcal{N}r(i), \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Those facts lead to efficient algorithms for computing ultimate levelings and its variations [3,10].

Ultimate levelings are operators that extract residual information from primitive families. During the residual extraction process, it is very common that undesirable regions of the input image contain residual information that should be filtered out. These undesirable residual regions often include desirable residual regions due to the design of the ultimate levelings which consider maximum residues. Thus, residual information can be improved by filtering of residues extracted from undesirable regions [5,12,3,2]. We can decide whether a residue  $r_i^+(f)$  (resp.,  $r_i^-(f)$ ) is filtered out or not, just checking nodes  $\tau \in \mathcal{N}r(i)$  that satisfy a given filtering criterion  $\Omega : \mathcal{P}(\mathcal{D}) \rightarrow \{\text{desirable}, \text{undesirable}\}$ . Thus, we just calculate the ultimate leveling  $\mathcal{R}$  for residues  $r_i^+$  (resp.,  $r_i^-$ ) such that satisfy the criterion  $\Omega$ . So, positive (resp. negative) residues are redefined as follows:

$$\forall \tau \in \mathcal{N}r(i), r_{\mathcal{T}_f^i}^{\Omega+}(\tau) = \begin{cases} r_{\mathcal{T}_f^i}^+(\tau), & \text{if } \exists C \in \mathcal{N}r(i) \\ & \text{such that } \Omega(C) \text{ is desirable;} \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

and thus redefined the ultimate levelings with strategy for filtering from undesirable residues as follows:  $\mathcal{R}_\Omega(f) = \mathcal{R}_\Omega^+(f) \vee \mathcal{R}_\Omega^-(f)$ , where,  $\mathcal{R}_\Omega^+(f) = \sup\{r_i^{\Omega^+}(f) : i \in \mathcal{I}\}$  and  $\mathcal{R}_\Omega^-(f) = \sup\{r_i^{\Omega^-}(f) : i \in \mathcal{I}\}$ .

Given the above considerations, in this paper we are interested in the maximum desirable residues provided by Ultimate Levelings. In this sense, is presented in the Section 4 a new approach to construct a strategy based on machine learning techniques.

#### 4 Strategy to filter undesirable residues based on machine learning

In this section, we present the construction of a strategy to filter undesirable residues from the ultimate levelings based on a regression that predicts the correspondence between objects of interest and residual regions.

For this, we recall that an increasing family of levelings  $\{\psi_i : i \in \mathcal{I}\}$  can be obtained through a sequence of pruned trees  $(\mathcal{T}_f^0, \mathcal{T}_f^1, \dots, \mathcal{T}_f^{\mathcal{I}^{MAX}})$  from the structure of the max-tree, min-tree or tree of shapes  $\mathcal{T}_f$  constructed from the image  $f$ . In addition, thanks to Eq.4, a sequence of residues  $(r_0(f), r_1(f), \dots, r_{|\mathcal{I}-1|}(f)) : r_i(f) = \psi_i(f) - \psi_{i+1}(f)$  can be obtained by the sequence of residual nodes  $(\mathcal{N}r(0), \mathcal{N}r(1), \dots, \mathcal{N}r(|\mathcal{I}| - 1)) : \mathcal{N}r(i) = \mathcal{T}_f^i \setminus \mathcal{T}_f^{i+1}$ . Thus, from a sequence of residual nodes, we can use the tree structure to define strategies for filtering undesirable residues of the ultimate levelings.

Given this consideration, our idea is to represent each residue through an attribute vector extracted from overlapping residual regions with low contrast. Therefore, we define a partial order relation  $\preceq$  on the collection of residual nodes  $\mathcal{B}r = \{\mathcal{N}r(i) : i = 0, 1, \dots, |\mathcal{I}| - 1\}$  such that for any  $\mathcal{N}r(i), \mathcal{N}r(j) \in \mathcal{B}r$ , we write  $\mathcal{N}r(i) \preceq \mathcal{N}r(j)$  if and only if the nodes belonging to  $\mathcal{N}r(i)$  are descendants of some node belonging to  $\mathcal{N}r(j)$  in  $\mathcal{T}_f$ .

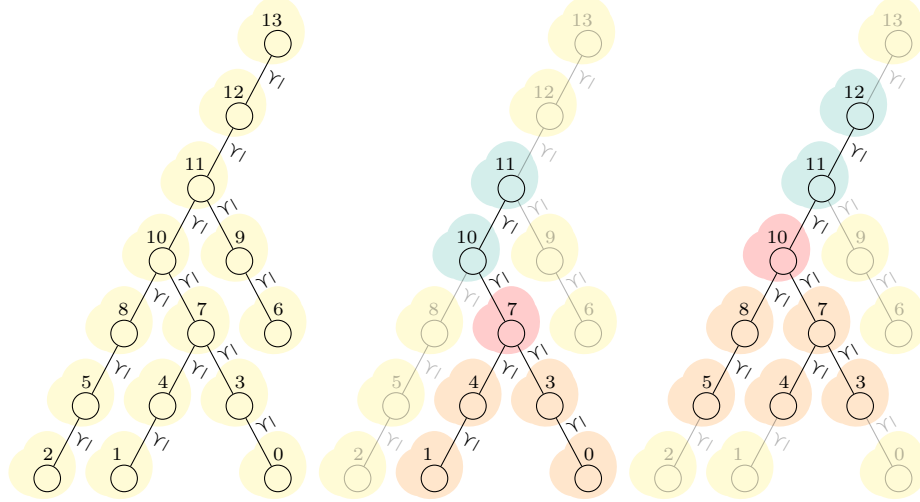
This is equivalent to  $\bigcup\{p \in \tau : \tau \in \mathcal{N}r(i)\} \subseteq \bigcup\{p \in \tau : \tau \in \mathcal{N}r(j)\}$ . Now, we use this poset  $(\mathcal{B}r, \preceq)$  in order to extract an attribute vector on a sequence of residual nodes, that is defined as follows:

$$\mathcal{B}_\Delta(i) = (\mathcal{N}r(i - k) \in \mathcal{B}r : -\Delta \leq k \leq \Delta), \quad (6)$$

where  $\mathcal{N}r(i - \Delta) \preceq \mathcal{N}r(i - \Delta + 1) \preceq \dots \preceq \mathcal{N}r(i) \preceq \mathcal{N}r(i + 1) \preceq \dots \preceq \mathcal{N}r(i + \Delta)$  and  $\Delta \in \mathbb{N}$  is a parameter that defines the amount of residual regions. In addition, if there are bifurcations in the path between the nodes  $\mathcal{N}r(i - \Delta)$  to  $\mathcal{N}r(i + \Delta)$ , we choose the path with biggest accumulated area to define the sequence. This is based on the fact that all descendent nodes of a given node are subset of it, then the path of descendant with biggest accumulated area contains more region in common with the node, it means that the nodes of the path are more stable [14]. Note that the length of  $\mathcal{B}_\Delta(i)$  is  $2\Delta + 1$ . An example of poset  $(\mathcal{B}r, \preceq)$  and the sequence  $\mathcal{B}_\Delta(i)$  is showed in Fig. 2. Then, we define an attribute vector  $A(i) \in \mathbb{R}^{n(2\Delta+1)}$  on a sequence  $\mathcal{B}_\Delta(i)$  for any  $i \in \mathcal{I}$ , as follows:

$$A(i) = (\kappa_j(\mathcal{N}r) : \mathcal{N}r \in \mathcal{B}_\Delta(i) \text{ and } j = 1, 2, \dots, n), \quad (7)$$

where  $\kappa_j : \mathbb{P}(\mathcal{D}) \rightarrow \mathbb{R}$  is an attribute (feature) extracted from the largest residual region belonging to  $\mathcal{N}r \in \mathcal{B}_\Delta(i)$  and  $n$  is the number of attributes.



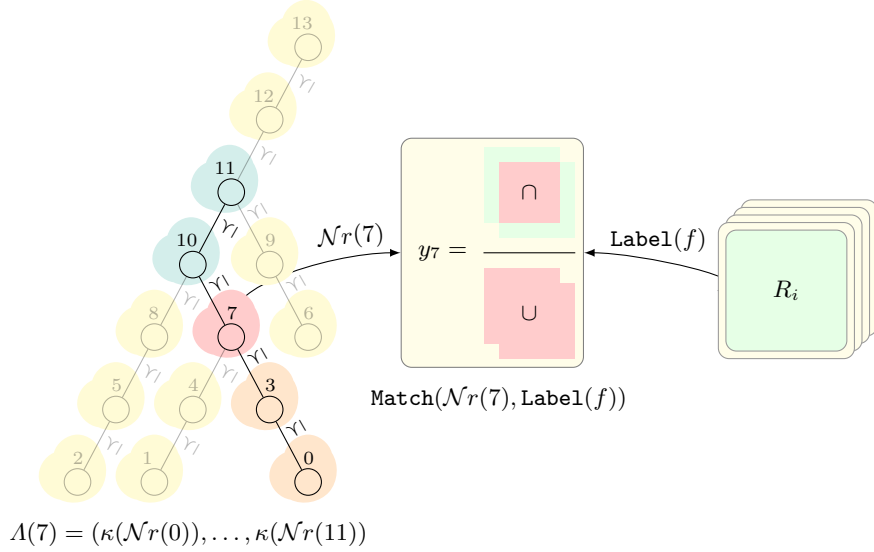
**a)** The labels  $0, 1, \dots, 13$  are related to the sets  $\mathcal{N}r(0), \mathcal{N}r(1), \dots, \mathcal{N}r(13)$ . **b)** Ascendant (in green) and descendant paths (in orange) of  $\mathcal{N}r(7)$  (in red). **c)** Ascendant (in green) and descendant paths (in orange) of  $\mathcal{N}r(10)$  (in red).

**Fig. 2.** An example of feature extraction using our approach for  $\Delta = 2$ . In **b)** the path formed by  $\mathcal{N}r(10) \preceq \mathcal{N}r(11)$  is the ascendant path of  $\mathcal{N}r(7)$  and there are two descendant paths of  $\mathcal{N}r(7)$  formed by  $\mathcal{N}r(0) \preceq \mathcal{N}r(3)$  and by  $\mathcal{N}r(1) \preceq \mathcal{N}r(4)$ . In **c)** there is a similar situation.

To construct the regression, we define a training set  $(X, Y)$  with  $m \in \mathbb{N}$  labeled samples such that  $X \subseteq \mathbb{R}^m \times \mathbb{R}^{n(2\Delta+1)}$  is a set of attribute vectors of the residual regions and  $Y \subseteq \mathbb{R}^m$  is a set of measure of similarities. Thus, we have that  $A(i) \in X$  is an attribute vector that represents the residual region of  $r_i$  and  $\text{Match}(\mathcal{N}r(i), \text{Label}(f)) \in Y$  is the value of measure of similarity between the residual region of  $r_i$  and a labeled region  $R \in \text{Label}(f)$  of an image  $f$ . This measure of similarity is defined as follows:

$$\text{Match}(\mathcal{N}r(i), \text{Label}(f)) = \max_{R \in \text{Label}(f)} \frac{|R \cap \tau_{\text{region}}|}{|R \cup \tau_{\text{region}}|}, \quad (8)$$

where  $\tau_{\text{region}} = \bigcup\{p \in \tau : \tau \in \mathcal{N}r(i)\}$  is residual region of  $r_i$ . An example of training set extraction is shown in Fig. 3. We remember that, a machine learning model usually tries to find a function  $h : \mathbb{R}^{n(2\Delta+1)} \rightarrow \mathbb{R}$ , also called the hypothesis, that best fits the training set  $(X, Y)$ .



**Fig. 3.** An example of training set computation. For each residual node set  $\mathcal{N}r(i)$  we extract an attribute vector  $\Lambda(i)$  and the matching value  $y_i$ . Thus, the pair  $(\Lambda(i), y_i)$  is a sample.

Once  $h$  has been trained, we can decide when a residue  $r_i^+(f)$  (resp.,  $r_i^-(f)$ ) is desirable, if the prediction of  $\Lambda(i)$  is greater than a threshold  $\varepsilon$ , that is:

$$\forall \tau \in \mathcal{N}r(i), r_{\mathcal{T}_f^i}^{h+}(\tau) = \begin{cases} r_{\mathcal{T}_f^i}^+(\tau), & \text{if } h(\Lambda(i)) > \varepsilon, \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

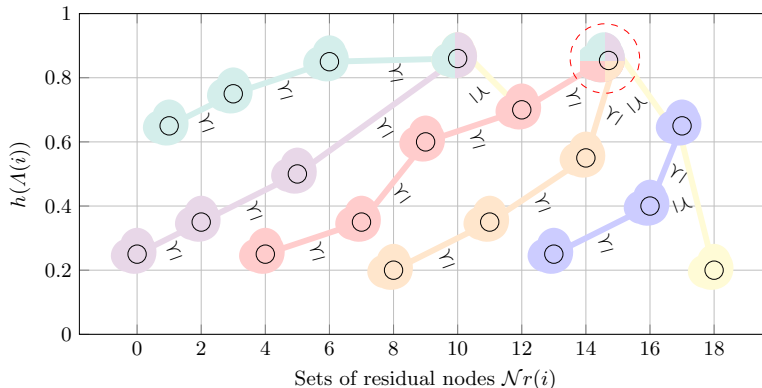
and thus redefined the ultimate levelings with strategy for filtering from undesirable residues as follows:  $\mathcal{R}_h(f) = \mathcal{R}_h^+(f) \vee \mathcal{R}_h^-(f)$ , where,  $\mathcal{R}_h^+(f) = \sup\{r_i^{h+}(f) : i \in \mathcal{I}\}$  and  $\mathcal{R}_h^-(f) = \sup\{r_i^{h-}(f) : i \in \mathcal{I}\}$ .

#### 4.1 Selecting disjoint residual regions

As mentioned in [5], the ultimate levelings can produce nesting of residual regions. Selecting disjoint residual regions (DRR) is a good way to deal with this. Thus, let  $\pi_F$  be the set of  $\mathcal{N}r(i)$  that contains nodes in the path started from a leaf  $F$  to the root of  $\mathcal{T}_f$ . During this path, we say that the best residual region is the one with greatest prediction, that is:

$$\text{DRR} = \left\{ S : S = \arg \max_{\mathcal{N}r(i) \in \pi_F} \{h(\Lambda(i)) : \nexists \mathcal{N}r(j), \mathcal{N}r(j) \preceq \mathcal{N}r(i) \text{ with } h(\Lambda(j)) > h(\Lambda(i))\} \right\} \quad (10)$$

Thus, we select all best residual node sets applying Eq. 10 for each leaf  $F$  in  $\mathcal{T}_f$  and then we redefined Eq. 9 only with the residues  $\mathcal{N}r(i)$  belonging to the set DRR. An illustration of the disjoint residual node sets selection that apply the Eq. 10 is shown in Fig. 4.



**Fig. 4.** An example of DRR selection. We start exploring paths from leaves to root of the tree according to Eq. 10. Residual node sets of more than one color represent residual node sets that are the best for two or more paths. For example the residual node set  $\mathcal{N}r(15)$  (marked by a dashed red circle) is the best of the paths  $\mathcal{N}r(1) \preceq \dots \preceq \mathcal{N}r(15) \preceq \dots \preceq \mathcal{N}r(18)$ ,  $\mathcal{N}r(0) \preceq \dots \preceq \mathcal{N}r(15) \preceq \dots \preceq \mathcal{N}r(18)$ ,  $\mathcal{N}r(4) \preceq \dots \preceq \mathcal{N}r(15) \preceq \dots \preceq \mathcal{N}r(18)$  and  $\mathcal{N}r(8) \preceq \dots \preceq \mathcal{N}r(15) \preceq \dots \preceq \mathcal{N}r(18)$ .

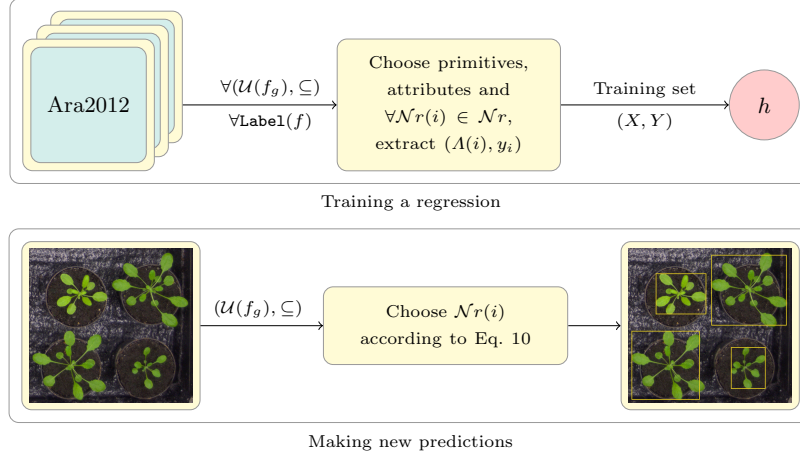
## 5 Results and experiments

In this section we present the results of experiments conducted with our new approach. To this purpose, we choose an important vision task problem called plant bounding box detection. Plants have a very complex morphology, so the detection of their bounding box is considered a difficult task. A good plant dataset in literature is provided by Minervine et al. [18]. In respect of plant bounding box detection task, the dataset is composed by three subsets: Ara2012, Ara2013-Canon and Ara2013-Rpi, totalizing 70 images with size between  $3108 \times 2324$  and  $2592 \times 1944$  pixels.

An overview of our approach applied to plant bounding box detection is shown in Fig. 5. First, we constructed regression model using features based on contrasts, colors and shapes. After, this regression model is used to select the desirable residues (see Eq. 9). Finally we define the bounding boxes using the DRR extracted of the desirable residues (see Eq. 10). The main parameters are: (1) the primitives obtained by the area attribute with min area of 50 and max area of 95,000 and the bounding box area attribute with min area of 150 and max area of 400,000; (2) the attributes chose to extract the features of the residual regions: 3 contrast attributes (residue, maximum residue and altitude), 3 color attributes (green level, distance of RGB foreground and distance of RGB background) and 3 shape attributes (TBMR, compactness, eccentricity). Totalizing 9 type of feautres, i.e.,  $k = 9$ ; (3) the minimal value  $\varepsilon$  to avoid false bounding boxes in DRR (see Eq. 10).

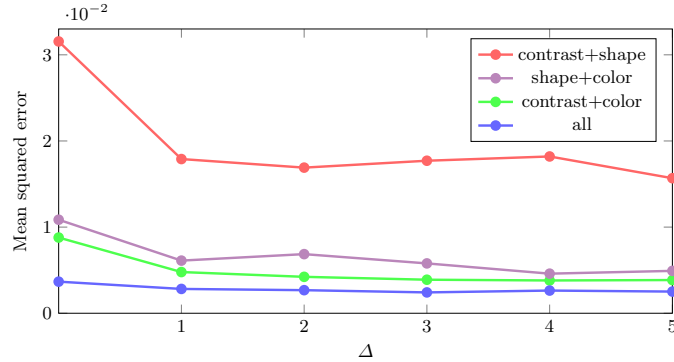
Thus we trained  $h$  with a Multilayer Neural Network using 3 hidden layers with (50, 30, 10) units and a learning rate of 0.005 found by a grid search. The model was trained on the 1.59 million of training samples and evaluated on

the 32 thousand of validation and test samples. We notice overfitting when the number of neurons and hidden layers were set to high values.



**Fig. 5.** Overview of our approach. First we train a regression model  $h$ . Then, we can make new predictions to detect bounding boxes according to Eq. 10.

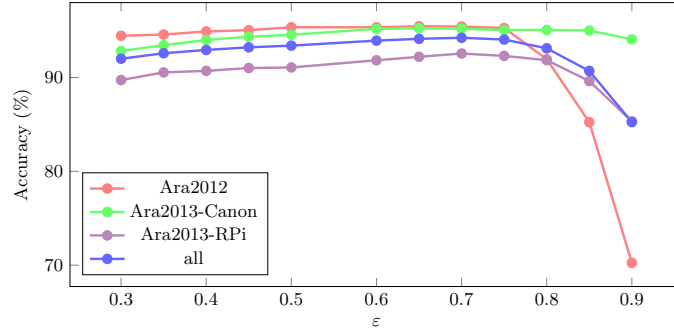
In order to study the influence of  $\Delta$  in plant bounding box detection, we conducted some experiments varying this parameter. We remember that the choice of the type of features and the number of them are usually challenges in machine learning. Thus, to select the best combination of both, we performed a grid search tuned by cross-validation as shown in Fig. 6.



**Fig. 6.** Evolution of  $\Delta$  during model validation. Note that, the number of features is  $k(2\Delta + 1)$ , where  $k$  is the number of attributes. For example, consider  $\Delta = 3$ , the number of features for all attributes is  $9 \times (2 \times 3 + 1) = 63$ .

The results are summarized in Fig. 6. They reveal that  $\Delta$  parameter and the number of features have a big influence in the model, because when we increase  $\Delta$  and the number of features the MSE of all models tends to decrease. The best  $h$  was obtained with  $\Delta = 3$  and 63 features.

Another parameter that influences the result of plant bounding box detection is  $\varepsilon$ . This parameter determines the minimum prediction value for a residual node set to be considered desirable. In Fig. 7 is shown the evolution of  $\varepsilon$  during tests. The best accuracy was obtained using  $\varepsilon = 0.7$ .



**Fig. 7.** Evolution of  $\varepsilon$  during tests.

Previously, we presented results in plant bounding box detection using non supervised strategies [11,6]. In order to compare them with our new approach, we chose the metrics provided by Minervine et al [18]. Those measures are: *SBD* that is a kind of accuracy metric; *DiC* that is the difference between predicted bounding box and the annotated bounding box, and  $|DiC|$  that is the absolute value of *DiC*. In a general way, the results obtained by our new approach are the best results of all our works in plant bounding box detection. Those results are summarized in Table 1.

**Table 1.** Results obtained with our new approach compared to our previously works. In a general way the results of our new approach is the best of our works.

Dataset	Approach	<i>SBD</i> [%]	<i>DiC</i>	$ DiC $
Ara2012	Mumford Shah Energy [11]	92.00	0.20	0.20
	MSER [6]	91.90	0.20	0.20
	TBMR [6]	93.60	0.00	0.00
	Our	<b>95.43</b>	-0.18	0.43
Ara2013-Canon	Mumford Shah Energy [11]	87.50	0.10	0.30
	MSER [6]	89.40	0.10	0.10
	TBMR [6]	92.00	0.20	0.20
	Our	<b>95.20</b>	0.14	0.37
Ara2013-RPi	Mumford Shah Energy [11]	80.30	0.10	0.40
	MSER [6]	83.20	0.20	0.40
	TBMR [6]	84.00	0.30	0.30
	Our	<b>92.55</b>	-0.33	0.48
Mean (All)	Mumford Shah Energy [11]	85.80	0.00	0.30
	MSER [6]	87.60	0.10	0.20
	TBMR [6]	89.30	0.10	0.10
	Our	<b>94.23</b>	-0.11	0.42

## 6 Conclusion

This paper presented a novel approach to construct a strategy to filter undesirable residues from the ultimate levelings based on a regression that predicts the correspondence between objects of interest and residual regions. As it is known, during the residual extraction process, it is very common that undesirable regions of the input image contain residual information that should be filtered out. In this sense, we designed a filter based on the regression to filter out residues extracted from undesirable regions. The results obtained applying the filter in plant bounding box detection reveal the robustness of our approach. Further studies should investigate the following areas: *(i)* apply regression prediction as a weight of Ultimate Levelings residues, *(ii)* explore classifiers models instead of regressions and *(iii)* methods to determine  $\Delta$  parameter.

## 7 Acknowledgements

This study was financed in part by the CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Finance Code 001); FAPESP - Fundação de Amparo a Pesquisa do Estado de São Paulo (Proc. 2018/15652-7); CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (Proc. 428720/2018-8).

## References

1. Alves, W., Hashimoto, R.: Text regions extracted from scene images by ultimate attribute opening and decision tree classification. In: Proceedings of the 23rd SIBGRAPI Conference on Graphics, Patterns and Images. pp. 360–367 (2010)
2. Alves, W., Morimitsu, A., Castro, J., Hashimoto, R.: Extraction of numerical residues in families of levelings. In: Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI - Conference on. pp. 349–356 (2013)
3. Alves, W.A.L., Hashimoto, R.F.: Ultimate grain filter. In: IEEE International Conference on Image Processing. pp. 2953–2957 (2014)
4. Alves, W.A.L., Morimitsu, A., Hashimoto, R.F.: Scale-space representation based on levelings through hierarchies of level sets. In: Proc. of the 12th International Symposium on Mathematical Morphology and its Applications to Image and Signal Processing. vol. 9082, pp. 265–276 (2015)
5. Alves, W.A., Hashimoto, R.F., Marcotegui, B.: Ultimate levelings. *Computer Vision and Image Understanding* **165**, 60 – 74 (2017)
6. Alves, W.A.L., Gobber, C.F., Hashimoto, R.F.: Plant bounding box detection from desirable residues of the ultimate levelings. In: Image Analysis and Recognition - 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27-29, 2018, Proceedings. pp. 474–481 (2018)
7. Caselles, V., Monasse, P.: Geometric Description of Images as Topographic Maps. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (2010)
8. Couprie, C., Farabet, C., Najman, L., LeCun, Y.: Convolutional nets and watershed cuts for real-time semantic labeling of rgb-d videos. *J. Mach. Learn. Res.* **15**, 3489–3511 (2014)

9. Dalla Mura, M., Benediktsson, J.A., Waske, B., Bruzzone, L.: Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing* **48**, 3747–3762 (2010)
10. Fabrizio, J., Marcotegui, B.: Fast implementation of the ultimate opening. In: *Proceedings of the 9th International Symposium on Mathematical Morphology*. pp. 272–281 (2009)
11. Gobber, C.F., Alves, W.A., Hashimoto, R.F.: Ultimate leveling based on mumford-shah energy functional applied to plant detection. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. pp. 220–228 (2017)
12. Hernández, J., Marcotegui, B.: Shape ultimate attribute opening. *Image and Vision Computing* **29**, 533 – 545 (2011)
13. Jones, R.: Connected filtering and segmentation using component trees. *Computer Vision and Image Understanding* **75**(3), 215 – 228 (1999). <https://doi.org/https://doi.org/10.1006/cviu.1999.0777>
14. Kimmel, R., Zhang, C., Bronstein, A., Bronstein, M.: Are MSER features really interesting? *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **33**(11), 2316–2320 (Nov 2011)
15. Li, W., Haese-Coat, V., Ronsin, J.: Residues of morphological filtering by reconstruction for texture classification. *Pattern Recognition* **30**, 1081 – 1093 (1997)
16. Masci, J., Angulo, J., Schmidhuber, J.: A learning framework for morphological operators using counter-harmonic mean. In: Hendriks, C.L.L., Borgefors, G., Strand, R. (eds.) *Mathematical Morphology and Its Applications to Signal and Image Processing*. pp. 329–340 (2013)
17. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing* **22**, 761–767 (2004)
18. Minervini, M., Fischbach, A., Schar, H., Tsafaris, S.A.: Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters* **81**, 80–89 (2016)
19. Ouzounis, G.K., Pesaresi, M., Soille, P.: Differential area profiles: Decomposition properties and efficient computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(8), 1533–1548 (Aug 2012)
20. Pesaresi, M., Benediktsson, J.: A new approach for the morphological segmentation of high-resolution satellite imagery. *Geoscience and Remote Sensing, IEEE Transactions on* **39**, 309–320 (2001)
21. Pesaresi, M., Benediktsson, J.: Image segmentation based on the derivative of the morphological profile. In: Goutsias, J., Vincent, L., Bloomberg, D. (eds.) *Mathematical Morphology and its Applications to Image and Signal Processing, Computational Imaging and Vision*, vol. 18, pp. 179–188. Springer US (2000)
22. Retornaz, T., Marcotegui, B.: Scene text localization based on the ultimate opening. In: *Proceedings of the 8th International Symposium on Mathematical Morphology and its Applications to Image and Signal Processing*. pp. 177–188 (2007)
23. Retornaz, T., Marcotegui, B.: Ultimate opening implementation based on a flooding process. *The 12th International Congress for Stereology* (2007)
24. Salembier, P., Oliveras, A., Garrido, L.: Anti-extensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing* **7**(4), 555–570 (1998)