

Locally Adaptive Hierarchical Cluster Termination With Application To Individual Tree Delineation

Ashlin Richardson, *Graduate Student Member, IEEE*, and Donald Leckie

Abstract—A clustering termination procedure which is locally adaptive (with respect to the hierarchical tree of sets representative of the agglomerative merging) is proposed, for agglomerative hierarchical clustering on a set equipped with a distance function. It represents a multi-scale alternative to conventional scale dependent threshold based termination criteria.

We trim the tree at specific locations by studying cumulative extreme values of rates of change of parameters along paths of the agglomeration hierarchy, each path representing the "history" of successive merges with respect to an initial set. Thus the method considers the smallest localities. Moreover, a cumulative extreme value of the rate of a parameter indicates the parameter is changing more rapidly than it has yet changed, which, in the context of geometric parameters we interpret to mean: the geometry at that merging step is changing more drastically than at any preceding step. We refer to this qualitative phenomenon as geometric "paradigm shift". The method is sensitive to extreme changes in geometry that may happen at any scale.

The proposed termination is presented in the context of a motivating example in forest mapping. Automated approaches for delineating Individual Tree Crowns (ITCs) from high resolution imagery offer the possibility of improving mapping consistency, accuracy, and effectiveness. The Individual Tree Crown (ITC) methods of Gougeon et al. [1], [2] distinguish individual tree crowns by following the darker "valley" material between them. Such is a highly effective *modus operandi* for conically shaped trees (softwoods). Hardwood trees, on the other hand, are of greater structural complexity resulting in less distinction between internal and external shadow, hence oversegmentation may result.

As a possible avenue for remediating this exceptional situation, the proposed termination method offers assistance in detecting instances of oversegmentation by proposing geometrically meaningful candidate groups of segments, from hierarchical clustering using an appropriate "distance". After developing the approach in this motivating context, we rank the resulting clusters of segments by a scale-independent dispersion measure.

Index Terms—Machine Learning, Pattern Recognition, Data Mining, Hierarchical Agglomeration, Clustering, Termination, Cluster Termination, Locally Adaptive, Multi-scale, Automatic Termination, Forestry, Individual Tree Crown, ITC.

I. MOTIVATING EXAMPLE: OVERSEGMENTATION

THE motivating context for the approach is taken to be an Individual Tree Crown (ITC) delineation, which specifically refers to a quantity of geospatial regions, or segments, with each segment ideally corresponding to the crown area of an individual tree. Each such segment is known as an isolation (ISOL). The ITC method [1], based on valley following, is most successful for coniferous trees, which are largely conical, so that the tree crown material is exhibited as a distinguishable

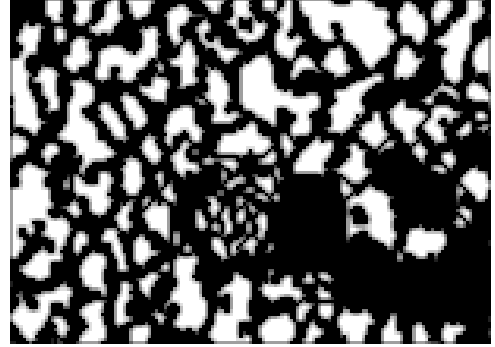


Fig. 1: ISOLS (Individual Tree Crown Segments)

peak, accompanied by the surrounding basal area in the image brightness. Hardwood trees are more like broccoli than carrots, showing complicated structure of extrema in the brightness within the tree-top area, making problematic the distinction between basal areas and the darker valley material. This extreme situation may result in oversegmentation. In this application, the purpose of the proposed approach is to help identify such a situation by specifying candidate groups of segments whose geometries are consistent with instances of oversegmentation (the "split-case").

A. Algorithmic Overview

The possible hierarchical relationships between groups of local extrema in the image are, in general, unclear. How would one distinguish a small group of broccoli plants, from a group of small branches of a large broccoli plant? In general it may not be possible to resolve this situation from imagery alone. The approach considered here is intended to attack the simpler problem: the identification of candidate ISOL groups which potentially represent an oversegmented tree. We base the approach upon the following principles: that potential candidate groups of ISOLS should be spatially compact, though not necessarily circular, and that some restraints should be placed on their proportions, according to elementary parameters reflecting their geometry.

The approach is summarized in three steps: Hierarchical Agglomerative Clustering (HAC) using a geometric distance function, a locally adaptive termination procedure for the resulting hierarchy incorporating geometric information, and finally a ranking of the resulting candidate groups of segments, according to a scale-independent dispersion measure.

II. HIERARCHICAL AGGLOMERATIVE CLUSTERING (HAC)

For illustrating the distance function used in the Hierarchical Agglomerative Clustering (HAC), we assume an unreasonably

A. Richardson is with the Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada, e-mail: ashy@uvic.ca

D. Leckie is with the Pacific Forestry Centre, National Resources Canada, Victoria, BC, Canada, e-mail: dleckie@nrcan.gc.ca



Fig. 2: At left: (counterclockwise from top right) ISOL areas 1,2,3,4. At right: the ISOL edge pixels.

simple input segmentation, consisting of only four ISOLS (Fig. 2). After this illustration, we exhibit case study data featuring a cluster corresponding to the oversegmented case, demonstrating the mechanics of the termination.

The first principle (spatial compactness of candidate groups) we express through Hierarchical Agglomerative Clustering (HAC), as applied to groups of ISOLS. The rich history of HAC includes Wishart's proposal of Hierarchical Mode Analysis [3] and recent generalizations thereof, including those of Stuetzle [4], [5] and Carlsson [6]. Standard HAC assumes a choice of distance function between sets. In this example, we choose a geometrically motivated distance, the **connective distance**, which we define summararily. This distance is intended to represent the darker valley area corresponding to the interstitial matter between the segments.

A. Connective Links Using Compass Directions

For each ISOL in the image, we locate the boundary (edge) pixels. For a given ISOL, we find **connective links** with adjacent ISOLS according to the following heuristic: from a given edge pixel, we consider a path emanating in each of the eight compass directions. The path continues in a straight line across interstitial pixels not associated with any ISOL. Supposing the path reaches a pixel corresponding to an ISOL other than the originating ISOL, the path is deemed a **connective link** between the two ISOLS. We denote the set of **connective links** between A and B as $L(A, B)$ so that $l \in L(A, B)$ represents a given link, and $p \in l$ represents a pixel representative of the connective link. Also, for a given ISOL A we will have occasion to use the pixels representative of A, which we will denote by $P(A)$. The edge pixels belonging to A we will refer to by $E(A)$.

B. Connective Distance Between ISOLS

We compute the connective links between any applicable pairs of ISOLS, making possible the definition of the **connective distance** between any two ISOLS, using $L(A, B)$. Since an individual connective link $l \in L(A, B)$ is comprised of the representative pixels, we define the pixel count $|L(A, B)|$ corresponding to this collection of links as the count of the pixels within the union $|\cup L(A, B)|$ of the individual links from A to B (or from B to A). The **connective distance** $d(A, B)$ is taken to be this pixel count, formally:

$$d(A, B) = |L(A, B)| = |\cup L(A, B)|.$$

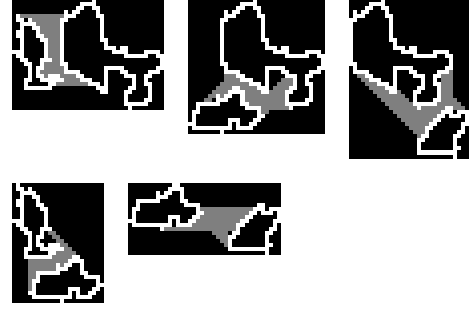


Fig. 3: Connective area between ISOL pairs: (1-2), (1-3), (1-4), (2-3), (3-4) respectively.

In Fig. 3 we show the **connective area** for the example, in grey (the ISOL edges are in white). Next it is necessary to define the connective area between two groups of ISOLS.

C. Connective Distance Between Groups Of ISOLS

Given two groups of isols \mathbb{A} and \mathbb{B} , we define the **connective area** between them as the number of pixels in the union of the set $L(A, B)$ of **connective links** between any pair of ISOLS A, B with A in group \mathbb{A} and B in the other group \mathbb{B} :

$$d(\mathbb{A}, \mathbb{B}) = \left| \bigcup_{A \in \mathbb{A}, B \in \mathbb{B}} L(A, B) \right|. \quad (1)$$

We should also define the set of **connective links** between two groups, say group \mathbb{A} and group \mathbb{B} , as the set of all connective links between any pair of ISOLS A and B, with ISOL A in group \mathbb{A} and ISOL B in group \mathbb{B} :

$$L(\mathbb{A}, \mathbb{B}) = \bigcup_{A \in \mathbb{A}, B \in \mathbb{B}} L(A, B).$$

D. Hierarchical Agglomerative Clustering

Each iteration of the Hierarchical Agglomerative Clustering (HAC) operates upon a list (S) of groups of ISOLS. The list is initialized with groups consisting of the individual ISOLS:

$$S = \{\{ISOL_1\}, \{ISOL_2\}, \dots\}.$$

Each iteration of the HAC simply merges together the two closest groups in the list S (closest in the sense of (1), the **connective distance** between groups of isols).

1) The two closest groups are:

$$(\mathbb{A}, \mathbb{B}) = \underset{\mathbb{A} \in S, \mathbb{B} \in S, \mathbb{A} \neq \mathbb{B}}{\operatorname{argmin}} \{d(\mathbb{A}, \mathbb{B})\}.$$

2) Upon determining the closest \mathbb{A}, \mathbb{B} in the list S , the groups are united (where the = sign means assignment):

$$\mathbb{A} = \mathbb{B} = \mathbb{A} \cup \mathbb{B}.$$

Then we return to 1), repeating until all groups are united. In this application, the end result is the group of all ISOLS. Supposing there are N iterations indexed by $i = 1, \dots, N$ we denote, while developing in the next section the **agglomeration hierarchy**, the merged result $\mathbb{A} \cup \mathbb{B}$ at iteration i as:

$$h_i = \mathbb{A} \cup \mathbb{B}.$$

The merging for the trivial example is shown in Fig. (4-6).

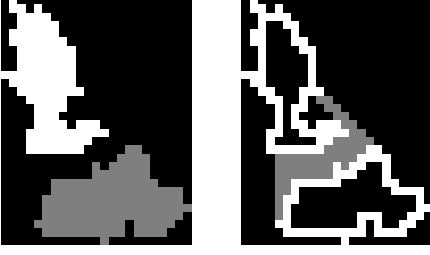


Fig. 4: Left: the sets joined at the first iteration. Right: the associated connective area between groups (grey).

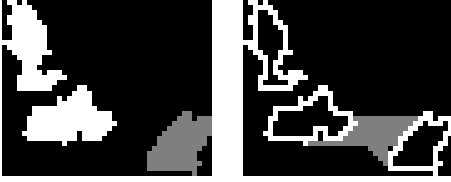


Fig. 5: Left: the sets joined at the second iteration. Right: the associated connective area between groups (grey).

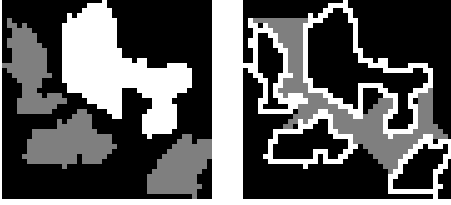


Fig. 6: Left: the sets joined at the third iteration. Right: the associated connective area between groups (grey).

E. The Agglomeration Hierarchy

The **agglomeration hierarchy** is the list of all groups that have ever been part of the list S , that is, any groups of ISOLS that existed in any step of the merging procedure; the associations between such groups indicate the ancestral history of each, detailing exactly which groups of ISOLS are merged. Considering the ISOL groups in the history to correspond directly to the nodes of the graph (the **vertex set**) we view the agglomeration hierarchy as a graph. Ancestral relationships between the groups merged are taken to correspond with the links between nodes on the graph - ordered pairs (parent, child) of ISOL groups. In graph terminology, these links comprise the **edge set**. For the trivial example, Fig. (7) shows the simple hierarchy which results from agglomeration.

Because the merging started with the list of singleton groups corresponding to individual ISOLS, each such group $\{ISOL_j\}$ where $j = 1, \dots, M$ (M is the number of ISOLS) is a member of the vertex set of the hierarchy. As any group in the hierarchy corresponds to a vertex (node) of the graph, each h_i , i.e. the merging result at each iteration i , also corresponds to a vertex (node). The last merging result is the group consisting of all of the ISOLS in the image, i.e.,

$$h_N = \{ISOL_1, ISOL_2, \dots, ISOL_M\}.$$

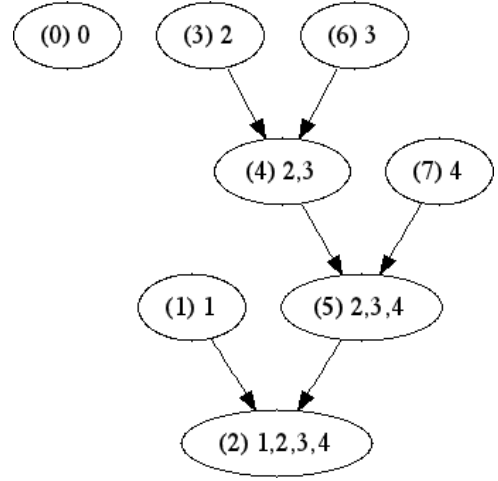


Fig. 7: An example **agglomeration hierarchy**. Arrows represent the **edge set**; ellipsoids represent the **vertex set**.

Again, the links between nodes of the graph may be represented as ordered pairs. Supposing at some stage the groups \mathbb{A} and \mathbb{B} are merged to form $h_i = \mathbb{A} \cup \mathbb{B}$, then we add to the edge set of the hierarchy the two edges (\mathbb{A}, h_i) and (\mathbb{B}, h_i) . That is, graphically, lines or arrows drawn from both \mathbb{A} and \mathbb{B} to h_i indicate that \mathbb{A} and \mathbb{B} were merged to form h_i .

F. Linear Paths On The Agglomeration Hierarchy

1) *Immediate Successors* : Examining the history of an individual ISOL, $\{ISOL_j\}$, in terms of its membership in successively larger groups that form, we start at $\mathbb{A} = \{ISOL_j\}$ simply following the edges of the graph to the next merged group involving \mathbb{A} , (the edge set has a unique edge (\mathbb{A}, h_i) starting at \mathbb{A} and ending at h_i). Because any node \mathbb{A} in the hierarchy has one successor, that is, the group \mathbb{A} is merged with the nearest other group (\mathbb{B}) in the list S to form h_i , in general for a group of ISOLS \mathbb{A} it seems handy to reserve the notation $h(\mathbb{A})$ to represent the "successor" of \mathbb{A} . So a linear path from $\{ISOL_j\}$ can be written as:

$$\{\{ISOL_j\}, h(\{ISOL_j\}), h(h(\{ISOL_j\})), \dots, h(h(\dots h(\{ISOL_j\}) \dots)) = h_N\}.$$

That is, starting with the ISOL, we hop down node by node (by repeated applications of $h()$) until we reach the bottom:

$$h(h(\dots h(\{ISOL_j\}) \dots)) = h_N.$$

Thus, the terminal node h_N has no successors:

$$h(h_N) = \{\}.$$

2) *Immediate Ancestors*: In later sections we find it useful to denote the set of **immediate ancestors** of a group by h^{-1} . That is, $h^{-1}(\mathbb{A})$ represents the two (or zero) groups merged to form the group \mathbb{A} . So, $h^{-1}(\mathbb{A})$ either has two elements since it resulted from a merge, or $h^{-1}(\mathbb{A}) = \{\} = \emptyset$, i.e., having no ancestors, the group consists of a single ISOL. Supposing further that H is a list of different nodes in the hierarchy, we extend the definition of ancestor set to apply to H , so that we

can find the **immediate ancestors** of all elements of H , rather than just those of a single element:

$$h^{-1}(H) = \{h^{-1}(h_\lambda) \mid h_\lambda \in H\}.$$

The analogous definition for the **immediate successors** of a list of groups is:

$$h(H) = \{h(h_\lambda) \mid h_\lambda \in H\}.$$

3) *All Successors And Ancestors For A Given Hierarchical Node*: The above definitions allow us to refer to **successors at depth k** as:

$$h^k(H) = \bigcup_{h_i \in H} \{h(\dots h(h_i))\}$$

(k applications of h) and the **ancestors at height k** by:

$$h^{-k}(H) = \bigcup_{h_i \in H} \{h^{-1}(\dots h^{-1}(h_i))\}$$

(k applications of h^{-1}) to define the **set of all successors**:

$$h^\infty(H) = \{H\} \cup \bigcup_{k=1}^{\infty} \{h(H)\}$$

and the **set of all ancestors**:

$$h^{-\infty}(H) = \{H\} \cup \bigcup_{k=1}^{\infty} \{h^{-k}(H)\}.$$

Supposing H is a node, the set of all successors of H is the linear path beginning at H and ending at the terminal node representing all the ISOLS in the image. For a given element h_i of the hierarchy, shortly we will have occasion to refer to the pixels representative of that group:

$$P(h_i) = \bigcup_{A \in \cup h^\infty(h_i)} P(A). \quad (2)$$

This is possible from combining the pixels $P(A)$ from ISOLS belonging to any of the ancestor groups.

III. TERMINATION VIA GEOMETRIC PARAMETER RATES

A. Possible Geometric Parameters Of Interest

1) A_{merge} : By A_{merge} we refer to the connective area $d(\mathbb{A}, \mathbb{B})$ when ISOL groups \mathbb{A} and \mathbb{B} are merged to form $h_i = \mathbb{A} \cup \mathbb{B}$. We can also think of this as $P(h_i)$ as above (2).

2) LW_{ratio} : The parameter LW_{ratio} represents the ratio between "length" and "width" of a hypothetical pixel region corresponding to the connective area $d(\mathbb{A}, \mathbb{B})$ when ISOL groups \mathbb{A} and \mathbb{B} are merged to form $h_i = \mathbb{A} \cup \mathbb{B}$, assuming that such a hypothetical pixel region corresponding to the connective area is in fact rectangular - of course, this is not the case. Yet the parameter necessarily captures the interplay between "length" and "thickness" of the connective area in an analogous sense. In particular, we calculate the estimated "length" of the connective area associated with the merge as:

$$\hat{l} = \frac{1}{|L(\mathbb{A}, \mathbb{B})|} \sum_{\mathbb{L} \in L(\mathbb{A}, \mathbb{B})} |\mathbb{L}|.$$

That is, taking the total length of connective links between groups: \mathbb{A} and \mathbb{B} , that is, $\sum_{\mathbb{L} \in L(\mathbb{A}, \mathbb{B})} |\mathbb{L}|$, dividing this by the total number of links between \mathbb{A} and \mathbb{B} , that is, $|L(\mathbb{A}, \mathbb{B})|$, resulting in \hat{l} , we arrive at the average length of a link between the

ISOL groups \mathbb{A} and \mathbb{B} . Assuming such a region is rectangular, we estimate the "width" of the hypothetical region as:

$$\hat{w} = \frac{A_{merge}}{\hat{l}}.$$

Then the hypothetical ratio between "length" and "width" for the region is:

$$LW_{ratio} = \frac{\hat{l}}{\hat{w}} = \hat{l} / \frac{A_{merge}}{\hat{l}} = \frac{(\hat{l})^2}{A_{merge}}.$$

3) N_{pix} : The parameter N_{pix} represents the total area of ISOLS belonging to the group formed by the current merge. Supposing the current hierarchical node (a group of ISOLS) is H , $\cup H$ refers to the ISOLS within the group H . Then the total ISOL area corresponding to H is the number of pixels from component ISOLS:

$$\sum_{A \in H} P(A).$$

4) N_{edge} : The parameter N_{edge} represents the perimeter, as the count of all edge pixels, of ISOLS belonging to the group formed by the current merge. Then the total perimeter of ISOLS corresponding to H is the number of edge pixels corresponding to all component ISOLS:

$$\sum_{A \in H} E(A).$$

5) $A_{cumulative}$: The parameter $A_{cumulative}$ represents the interstitial area corresponding to the group of ISOLS formed by the current merge: $A_{cumulative}$ is the pixel count corresponding to the combined connective areas pertaining to merges ancestral to (and including) the current merge. We define the **connective links** of the hierarchical element h_i by:

$$L(H) = \bigcup_{\mathbb{A} \in h^{-1}(H), \mathbb{B} \in h^{-1}(H), \mathbb{A} \neq \mathbb{B}} L(\mathbb{A}, \mathbb{B}).$$

This allows us to define the cumulative set of connective links, by following the hierarchy "upwards" by listing all ancestors $h^{-\infty}(H)$ of the current node H , combining together the connective links associated with all such ancestors:

$$\bigcup_{h_i \in h^{-\infty}(H)} L(h_i).$$

Combining and counting the pixels from all such links, we arrive at the combined area from the connective links of all merges in the history of the node H :

$$A_{cumulative} = \left| \bigcup \left[\bigcup_{h_i \in h^{-\infty}(H)} L(h_i) \right] \right|.$$

B. Geometric Parameters On Linear Paths

Along the linear path originating from $\{ISOL_j\}$:

$$\{\{ISOL_j\}, h(\{ISOL_j\}), h(h(\{ISOL_j\})), \dots, h(h(\dots h(\{ISOL_j\}) \dots)) = h_N\}$$

we can plot any of the above parameters, or ratios thereof. This is a locally adaptive method; in particular, the locality considered is the starting ISOL itself. The locality expands (with successive merges) along the linear path, until the

terminal node h_N is reached. Therefore the linear path allows a multi-scale study of the clustering behaviour, from local to completely global. We watch for transitions between progressively greater scales by performing calculus of the parameters upon linear paths on the hierarchy. The length of a path starting at node H is the size of the set of successors:

$$|h^\infty(H)|.$$

C. Rates Of Geometric Parameters Along Such Linear Paths

Given one of the parameters above (or ratios thereof) we examine the values along the above linear path. In particular, finite differences of such parameters along the linear path provide valuable information. We suppose a parameter takes a sequence of values along the linear path as follows:

$$f_0, f_1, f_2, \dots, f_{|h^\infty(H)|} = \{f_j\}_{j=0, \dots, |h^\infty(H)|},$$

where the nodes along the path from the node h_0 are:

$$h_0, h(h_0), h(h(h_0)), \dots, h_{|h^\infty(H)|}(h_0) = \{h_j\}_{j=0, \dots, |h^\infty(H)|}.$$

For a given node H we denote by $i(H)$ the merge iteration at which H was formed (not defining $i(H)$ for H without ancestors), and if in addition, the starting element h_0 is understood, we denote by $j(H)$ the j -index of $H = h_j$ in the linear sequence starting at h_0 . We can define j^{th} differences in two ways, as follows:

$$D_j = \frac{f_{j+1} - f_j}{i(h_{j+1}) - i(h_j)} \quad (3)$$

$$D_j = \frac{f_{j+1} - f_j}{j+1 - j} = f_{j+1} - f_j. \quad (4)$$

The first is interesting, as it expresses the rate of change of the parameter f with respect to the iteration number i ; this could be the subject of further research. In this implementation we use (4), expressing the rate of change of the parameter f with respect to the index of the location within the linear path.

D. Extrema Of Rates Of Geometric Parameters

We apply the following view: when considering a location on a path being followed, an extreme value of greater magnitude than encountered thus far (according to the rate of change of the parameter f) should indicate that a qualitative change has occurred at the merge corresponding to the path location considered. We use the first order difference (4) noting the use of higher order differences as a possible subject of further inquiry. The current implementation scales the parameter in the range $[0, 1]$, calculating (4) for $j = 0$ to $|h^\infty(H)| - 1$. Then we take the **cumulative maximum** of (4) as:

$$C_{max}(J) = \max\{D_j \mid j \leq J\}. \quad (5)$$

A point j where $C_{max}(j) \neq C_{max}(j-1)$ is deemed a **break point**, representing a candidate location indicative of a substantive qualitative change in the merging behaviour. We denote the set of break points for (5) by:

$$B_{h_0} = \{j \mid C_{max}(j) \neq C_{max}(j-1)\},$$

the indices along the linear sequence where the "upward steps" in the transformed sequence (5) are observed.

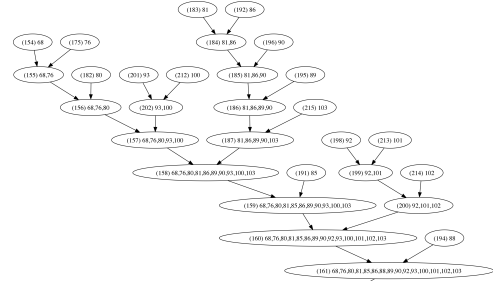


Fig. 8: A small component of the hierarchy for the case study.

E. Histogram Thresholding Of Break Point Counts

Supposing H represents all nodes within the hierarchy, and H_0 represents the singleton nodes within the hierarchy: given a starting element h_0 , we denote by $\chi_H(B_{h_0})$ the indicator function for the breakpoints of the transformed sequence (5) over the path starting at h_0 . That is, $\chi_H(B_{h_0})|_h = 1$ for a given hierarchical node h if $h \in B_{h_0}$, and is 0 otherwise. That is, for a given node, $\chi_H(B_{h_0})$ indicates whether or not the transformed sequence has a break point at the given node. Then, we count (at each node) all break points recommended by all the individual paths beginning at singleton nodes:

$$\sum_{h_0 \in H_0} \chi_H(B_{h_0}).$$

Evaluating this function at a node h gives a count of break points experienced at h , from all paths considered:

$$F(h) = \sum_{h_0 \in H_0} \chi_H(B_{h_0})|_h.$$

We evaluate the histogram of values $\{F(h) \mid h \in H\}$ (shown in Fig. 12 for the case study) obtained by $F(h)$ to perform a significance analysis, setting an arbitrary threshold $p = 25\%$ to obtain the results shown. The significance value is taken to be $F_{significance}$ for which $p = 25\%$ of the values of the histogram of $F(h)$ are to the right of $F_{significance}$.

We consider hierarchical nodes with a break point count above $F_{significance}$ to be indicative of the desired "paradigm shift", or change in qualitative merging behaviour (with respect to the parameters of interest). Accordingly, we trim the hierarchy by nullifying any nodes (and their successors) whose break point count is above $F_{significance}$. The terminal nodes $h^\infty(H_0)$ of the trimmed hierarchy H' are the computational output of this approach. Since the termination trims the hierarchy giving attention to the location of trimming according to the collective behaviour of the individual linear paths, the approach is locally adaptive to the hierarchical structure resulting from hierarchical agglomeration.

IV. CASE STUDY

The case study is the area of 131 ISOLS shown in Fig. 1. The agglomeration hierarchy for this example is complicated, so a small sample is shown in Fig. 8. For the termination procedure, we plot the parameters and results corresponding to both A_{merge} and $\frac{LW_{ratio}}{A_{cumulative}}$. Termination using either parameter successfully isolated the desired feature: the circular ISOL

TABLE I: Scores for A_{merge} based termination candidates

$i(h_i)$	MAD	Max.A.D.	$S(h_i)$
161	1887	15.42	122.3
150	4766	28.00	170.1
226	4396	20.60	213.3
82	17374	48.11	361.0
72	27809	46.21	601.7
27	27022	42.35	638.0

arrangement near the image centre. In Fig. (9) the parameters A_{merge} and $\frac{LW_{ratio}}{A_{cumulative}}$ are shown along a path beginning at ISOL 103, which is inside the desired image feature. Next, transformations of A_{merge} and $\frac{LW_{ratio}}{A_{cumulative}}$ are shown: the first difference (4) in Fig. (10) and the cumulative maximum of the first difference (5) in Fig. (11). Again, break points are determined according to the "steps" exhibited by the graph of (5). This procedure calculates the transformed parameter along all trajectories starting at the initial hierarchical nodes H_0 - in each case, the local indices along the linear path are converted into global (nonlinear) indices representing the hierarchical elements themselves, in order that a count of break points may be accumulated over the entire hierarchy H .

Once the terminal nodes of the hierarchy are obtained from the adaptive termination procedure, we discard those representing groups of ISOLS numbering below a threshold: for this study, we chose 7. Such a choice limits the display of groups that do not represent substantial clusters.

V. CANDIDATE RANKING PROCEDURE

Supposing the hierarchical element h_i is a (terminal) candidate cluster, we rank candidate clusters by the following score, based on the coordinates of the representative pixels $P(h_i)$:

$$S(h_i) = \frac{\frac{1}{|P(h_i)|} \sum_{x \in P(h_i)} |x_i - m(P(h_i))|}{\max_{x \in P(h_i)} |x_i - m(P(h_i))|}$$

where we take the measure of centrality m to be:

$$m(X) = \frac{1}{|X|} \sum_{(x,y) \in X} (x,y).$$

The numerator is the Mean Absolute Deviation (MAD); the denominator is the Maximum Absolute Deviation. Then $S(h_i)$ represents a dispersion measure that is normalized by the object's scale (according to the Maximum Absolute Deviation). Fig. 13 and Fig. 14 show the results ordered in increasing values of $S(h_i)$, for the termination according to A_{merge} and $\frac{LW_{ratio}}{A_{cumulative}}$ respectively. Table I and II record the values of the dispersion $S(h_i)$ in both cases, for each terminal node h_i .

VI. CONCLUSION

To comprehensively analyze the hierarchy of agglomeration, a framework for calculus along paths descending the agglomeration hierarchy was developed. The framework accomplishes

TABLE II: Scores for $\frac{LW_{ratio}}{A_{cumulative}}$ based candidates

$i(h_i)$	MAD	Max.A.D.	$S(h_i)$
106	1958.4	16.09	121.6
161	1887.7	15.42	122.3
225	2983.9	20.37	146.4
23	5886.5	28.27	208.1
138	6922.4	31.00	223.2
100	9353.1	31.28	299.0
14	30478	51.43	592.5

a locally adaptive termination procedure for hierarchical agglomerative clustering; this represents an alternative to conventional threshold termination rules based on fixing either the threshold distance, the number of clusters, or even the number of initial clusters, for example [7] in the case of clustering for land-cover type discrimination using polarimetric SAR imagery. Finally, a case study demonstrated the potential application of the new methodology to results of Gougeon's Individual Tree Crown (ITC) delineation [1]. We suggest further that the cumulative extreme points associated with a hierarchy offer the possibility of geometrically meaningful simplifications of hierarchies, for visualization purposes.

Advancement of the framework in the context of multi-spectral and hyperspectral imagery is a direction for further inquiry pertaining to ITC delineation, and to image analysis in general. The new termination used in conjunction with hierarchical clustering for multispectral imagery along with three dimensional visualization of the spatial and spectral extent of clusters in the spirit of [8] could potentially lead to valuable interactive and visually intuitive insights into the spatial and spectral separability of land-cover types.

REFERENCES

- [1] F. A. Gougeon, "A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images," *Canadian journal of remote sensing*, vol. 21, no. 3, pp. 274 – 284, 1995.
- [2] D. G. Leckie, F. A. Gougeon, N. Walsworth, and D. Paradine, "Stand delineation and composition estimation using semi-automated individual tree crown analysis," *Remote Sensing of Environment*, vol. 85, no. 3, pp. 355 – 369, 2003.
- [3] D. Wishart, *Mode analysis: a generalization of nearest neighbor which reduces chaining effects*. AP London, 1969.
- [4] W. Stuetzle and R. Nugent, "A generalized single linkage method for estimating the cluster tree of a density," *Journal of Computational and Graphical Statistics*, vol. 19, no. 2, pp. 397 – 418, 2010.
- [5] W. Stuetzle, "Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample," *Journal of Classification*, vol. 20, no. 1, pp. 25 – 47, 2003.
- [6] G. Carlsson and F. Mmoli, *Multiparameter Hierarchical Clustering Methods*, ser. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg, 2010.
- [7] J.S. Lee and M.R. Grunes and E. Pottier and L. Ferro-Famil, "unsupervised terrain classification preserving polarimetric scattering characteristics", *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42,number. 4, pp. 722 – 731, 2004.
- [8] A. Richardson, D. Goodenough, H. Chen, B. Moa, G. Hobart, and W. Myrvold, "Unsupervised nonparametric classification of polarimetric sar data using the k-nearest neighbor graph," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, 2010, pp. 1867 –1870.

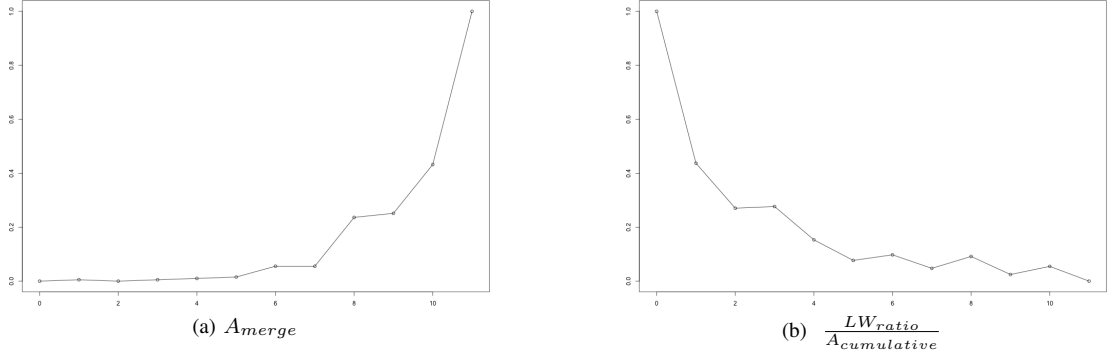


Fig. 9: Along the linear path starting at ISOL 103: the parameter without transformations.

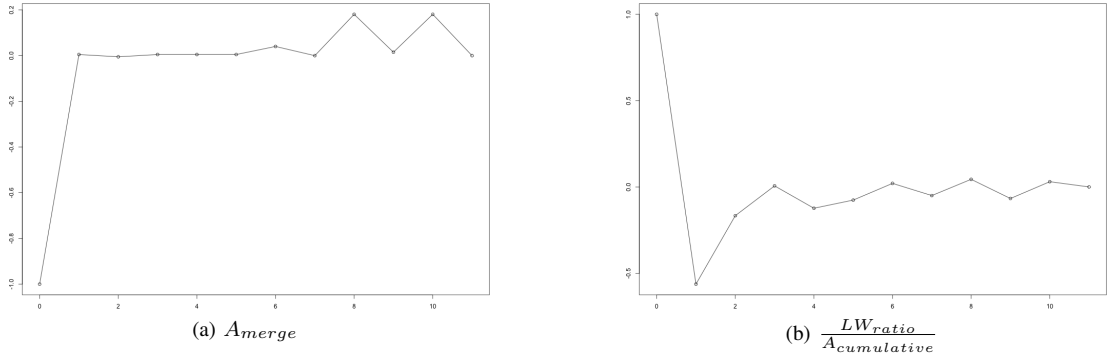


Fig. 10: Along the linear path starting at ISOL 103: the first difference of the parameter.

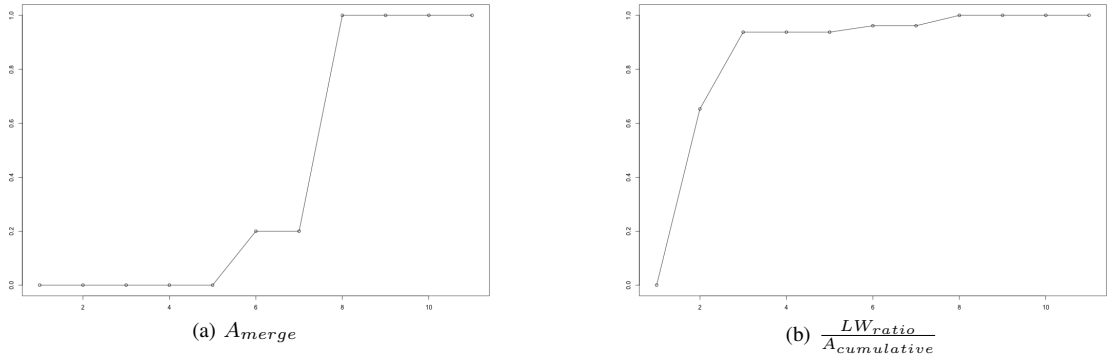


Fig. 11: Along the linear path starting at ISOL 103: the cumulative maximum of the first difference.

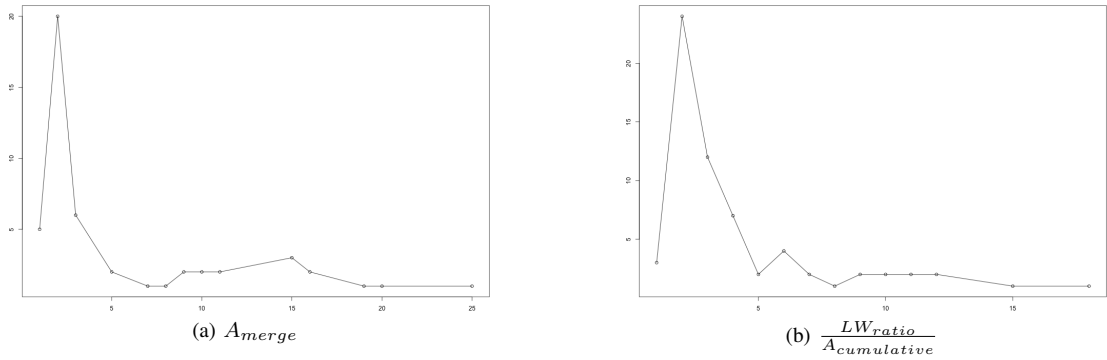


Fig. 12: The histogram of break-point counts (the number of hierarchical elements with a particular count).

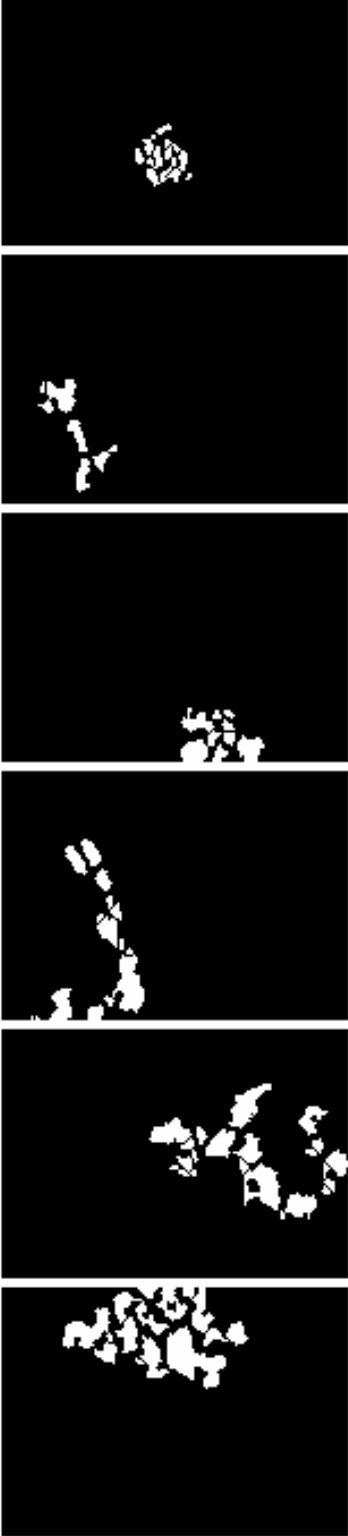


Fig. 13: A_{merge} based candidates ranked by $S(h_i)$



Fig. 14: $\frac{LW_{ratio}}{A_{cumulative}}$ based candidates ranked by $S(h_i)$