IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 March 2024, accepted 15 May 2024, date of publication 20 May 2024, date of current version 29 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3403309

RESEARCH ARTICLE

Multi-Spectral Source-Segmentation Using Semantically-Informed Max-Trees

MOHAMMAD HASHEM FAEZI[1], REYNIER PELETIER[2], AND MICHAEL H. F. WILKINSON[1], (Senior Member, IEEE)

[1]Bernoulli Institute, University of Groningen, 9700 AB Groningen, The Netherlands
[2]Kapteyn Institute, University of Groningen, 9700 AB Groningen, The Netherlands

Corresponding author: Mohammad Hashem Faezi (m.h.faezi@rug.nl)

This work was supported by the Centre for Data Science and Systems Complexity (DSSC), University of Groningen.

ABSTRACT In this paper, we propose an innovative approach to multi-band source-segmentation that addresses the constraints of single-band max-tree-based methods and effectively manages component-graph complexity. Our method extends multiple max-trees by integrating semantically meaningful nodes, derived from statistical tests, into a structured graph. This integration enables the exploration of correlations among cross-band emissions, enhancing segmentation accuracy. Evaluation with artificial multi-band astronomical images shows our method's superior accuracy in detecting and segmenting multi-spectral imagery. We achieve 98% accuracy in identifying correlated cross-band sources. Compared to state-of-the-art methods, our approach improves detection precision from 0.92 to 0.95 without sacrificing recall. Furthermore, quantitative analysis demonstrates significant speed enhancements, particularly on 3-channel images sized at 1,000 pixels squared, our method achieves up to an approximately 31× acceleration when compared to a component-graph-based approach. The versatility and effectiveness of the proposed method suggest applications in remote sensing and multi-spectral large-scale image data analysis.

INDEX TERMS Hierarchical structures, connected components, multi-band source-segmentation.

I. INTRODUCTION

Multi-spectral image data is essential for scientific research in fields such as remote sensing [1], exploration [2], and medicine [3]. Fluorescence microscopy images [4] use multiple color channels to visualize cellular components and molecular interactions, providing insights into biological processes [5]. Multi-band imaging is particularly useful in astronomy, where it is used to study a wide range of objects, from stars and galaxies to planetary nebulae and supernovae [6].

The most commonly used astronomical source detection and segmentation method SExtractor [7] and its alternatives [8], [9], [10], [11] are limited to single-band image processing. As a result, researchers have had to turn to alternative methods to extract information from multi-channel images, such as component graphs [12] where the dense graph structure results in a more complex construction, storage, and processing procedure, thus becoming the major obstacle in handling large-scale multi-band image data.

Several studies have aimed to extend the max-tree structure to multivariate images [13], [14], [15]. The primary challenge in this endeavor is the absence of a natural order relationship between pixel values in multi-valued images, which possess a vectorial nature. To address this issue, Carlinet and Géraud [16] employed a graph structure to extend the tree of shapes to multivariate image data, based on the level of inclusion within the graph, rather than on attributes.

In this study, we introduce a new algorithm called MMTO (multi-band max-tree objects) designed to accurately find correlated cross-frame sources in multi-spectral image data. This method extends precise grey-level max-tree node attributes, into a topologically ordered graph in which the ordering of initial tree structures are preserved, thereby enabling the efficient processing of hierarchical node partitions in a semantic manner.

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

A. RELATED WORK

Robust algorithms are necessary for the automatic detection of objects in astronomical images. These images exhibit a large dynamic range of intensities, a high quantity of noise, and numerous objects with uncertain boundaries, posing challenges to the detection process. Several methods have been proposed for faint object detection in astronomical image data. In this section, we will describe some of these methods briefly.

SExtractor [7] is widely used in astronomy for source extraction. It identifies objects as groups of pixels where the flux from a source dominates. Segmentation is done through thresholding, identifying connected pixels above a certain threshold as detections. Each pixel's value is a combination of background signal and light from objects. *SExtractor* constructs a background map by computing local background estimators in a grid covering the frame.

After extraction, objects pass through a filter to split them into final components, especially in crowded fields or with low detection thresholds. *SExtractor* uses a multi-thresholding deblending method, which struggles with closely spaced components.

NoiseChisel [9] is a faint object detection method designed to operate effectively amidst noise while requiring minimal signal inputs. The approach involves convolving a small *FWHM*[1] kernel with the image to enhance the signal-to-noise ratio by smoothing the image and reducing high spatial frequencies.

A low threshold, determined from the cumulative distribution of pixel values in the convolved image, is then applied to the smoothed image. The threshold is set to less than the sky value, to detect the faintest pixels. After applying the threshold, the image results in a binary image with pixels either having a count above or below the threshold.

Foreground pixels neighboring background pixels are iteratively converted to background pixels using erosion and dilation operations, collectively known as opening, to separate objects. To address potential false detection due to proximity, a classification scheme is employed to preserve valuable faint pixels while removing spurious detection.

ProFound [10] is another detection method that uses a single threshold stage to discriminate pixels containing sources, and processes these pixels in descending order of brightness, with a watershed process being used to allocate less bright neighboring pixels (within some tolerance) to the object of the brightest pixel in a region until all pixels bordering the object are either allocated to other objects, are marked as background, or have higher flux than neighboring pixels within the object.

Using the saddle-point segmentation approach, *ProFound* lacks the flexibility to create segments deeply embedded within segments and does suffer memory limitation issues for very large images.

DeepScan [8] is an automatic detection method that has been designed specifically to detect extended low surface brightness objects without fragmenting the source structure, using the *DBSCAN*[2] algorithm, that gradually identifies neighboring pixels, grouping them based on the density of pixels above a signal-to-noise threshold within a certain distance (ϵ radius). One notable issue with this method is that there is no source deblending other than that of the source mask (*SExtractor* is used to create masks).

In *DeepScan*, like many detection methods, there is a trade-off between completeness and output purity, determined by input parameters: clustering radius (ϵ), confidence parameter (k), and detection threshold. Higher ϵ values detect fainter objects but may cause source confusion. k should be sufficient to avoid spurious detection yet not too high to sacrifice completeness.

MTObjects uses morphological operators for source detection. Its hierarchical approach handles nested objects efficiently by constructing a max-tree from a smoothed, background-subtracted image. Statistical tests are then employed to identify nodes with flux levels statistically differing from their backgrounds.

The algorithm starts by estimating the background through statistical tests on image tiles. Selected background tiles have their mean subtracted from the image. Then, MTObjects identifies significant branches in a max-tree, evaluating their power attributes. Finally, the hierarchical structure is labeled for object representation.

A comparison[3] [17] of SExtractor, ProFound, NoiseChisel, and MTObjects, determined that MTObjects produces the highest scores on all quality measures. All the mentioned methods have been effective in single-band processing but cannot fully utilize multi-band information.

CGO (component-graph objects) [18] is a framework that has been proposed for multi-band image segmentation. This method extends MTObjects to multi-band data, on a component-graph structure, to enhance object detection and deblending capacity. The framework extends the principles of component-graphs to address the limitations posed by single-band processing, where the lack of combined information often leads to reduced sensitivity in detecting celestial objects and difficulties in deblending overlapping sources. As a result of intricate data structure, CGO fails to efficiently process high-resolution large-scale images within a reasonable time frame.

In this study, we introduce a method that offers significant advantages over the component-graph methodology, exhibiting superior efficiency in construction, storage, and processing while achieving heightened accuracy by leveraging partial grey-level information.

The preliminary sections of this article are structured as follows. Section II offers a concise overview of grey-scale astronomical image representation using the max-tree data

[1] Full width at half maximum.

[2] Density-based spatial clustering of applications with noise.

[3] DeepScan was excluded, due to its time complexity issue.

structure. Within this section, we define sources as partitions of max-tree nodes. Section III introduces the expansion of these sources into a graph structure, serving as a framework for cross-frame source comparison. This section encompasses precise definitions, feature selection process based on max-tree attributes and time complexity analysis.

Section IV presents our experimental results. Initially, we describe the multi-band complex simulations with ground truth, facilitating a robust evaluation. Following this, we illustrate the accuracy of MMTO in detecting correlated cross-band sources and highlight its advantages over the state-of-the-art methods. This demonstration showcases the method's superior efficiency in time performance while achieving heightened accuracy.

II. BACKGROUND

A. ASTRONOMICAL IMAGE DATA

Astronomical image data is a representation of the electromagnetic radiation received from celestial objects and captured by telescopes. This data is collected using detectors that are sensitive to different ranges of the electromagnetic spectrum, such as visible light, infrared, or ultraviolet. The detectors convert the incoming light into an electrical signal, which is then quantified and stored as 16-bit floating-point numerical data.

The resulting data is processed to remove various types of noise, correct for instrumental effects (such as flat fielding, bad pixel correction), stack various frames to get a higher signal-to-noise ratio and get rid of artifacts. The final step in the processing is to convert the data into an image format, where the brightness of each pixel represents the intensity of the light received from a particular part of the sky. Multiple channels are used to represent the intensity across various electromagnetic spectrum.

Astronomical image data provide crucial information about the physical properties of celestial objects, such as their size, temperature, mass, and chemical composition. This information are used to study the formation and evolution of stars, galaxies, and other astronomical objects, as well as to understand the properties and dynamics of the universe as a whole. The analysis of astronomical image data requires specialized software and expertise in areas such as image processing and statistical analysis.

B. SOURCE REPRESENTATION USING MAX-TREES

In mathematical morphology, digital images can be represented in various ways, instead of just considering the grid of pixels and their values. Alternative representations often organize the data from a structural perspective using region- and connectivity-based methods [19]. Such representations rely on tree or graph hierarchies of connected components.

The max-tree [20] is a commonly used data structure to represent and analyze the hierarchical structure of digital images, constructed based on the concept of regional maxima, where each node in the tree corresponds to a connected region of the image.

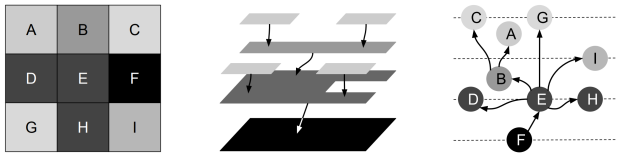


FIGURE 1. Representation of a grey-scale image and its corresponding max-tree with the parent-child relationships. (a) Original image, (b) it's schematic inclusion of peak components, and (c) max-tree representation.

This tree representation allows efficient and effective analysis of image properties such as shape, size, and topology. The max-tree has proven to be a valuable tool for image analysis and is widely used in various applications, such as morphological filtering [21], text recognition [22], feature extraction [23], and image segmentation [24] in a hierarchical manner. Figure 1 illustrates an example of a grey-scale image and its corresponding max-tree.

We consider the emissions as groups of max-tree nodes that reconstruct a specific partition of connected pixels with different intensities in a neighborhood on the 2D grid (typically the 4- or 8- connectivity). To discriminate these significant nodes from irrelevant max-tree nodes and extrema, *i.e.* noise and background, and assign them into source segments, we used the simplification and tree partitioning method that MTObjects uses.

MTObjects statistical tests aim at identifying significant nodes within a tree structure, particularly focusing on nodes representing emissions in an image. These tests use attributes derived from pixel intensities to distinguish between nodes representing objects and those due to noise. The significance test, employs the power attribute, which measure the intensity difference within a node and its closest significant ancestor. This test evaluates whether the observed intensity within a node is significantly higher than the background level, assuming that the distribution of the power attribute scaled by the variance σ^2 for noise nodes follows a χ^2 distribution [11].

Algorithm 1 gets the initial max-tree of a grey-level image and returns the simplified tree including only the significant nodes. Algorithm 2 partitions the simplified tree into sub-trees, each one representing a source. Figure 2 illustrates a part of an astronomical image simulation, significant nodes, and its corresponding segment map. A more detailed description of the algorithms used to find significant nodes can be found in [11].

To extract the attributes of the sources, we represent their corresponding groups of pixels as sub-trees of a simplified max-tree.

Definition 1: Let T be a max-tree. A *sub-tree* $\llcorner T$ of T is a non-empty set of nodes satisfying the following condition: for every node $n_i \in \llcorner T$, there exists another node $n_j \in \llcorner T$ such that $n_i = \text{parent}(n_j)$ or $n_j = \text{parent}(n_i)$, if and only if $|\llcorner T| \neq 1$.

According to the definition, every sub-tree $\llcorner T$ has a node with the largest area (local root), which we call a *semantic*

Algorithm 1 Get Significant Nodes of a Max-Tree

Require: Initial max-tree T
Ensure: Significant node labeled max-tree T
for each node n in T **do**
  if significantTest(T, n) is True **then**
    mark n as significant
  end if
end for
return T

Algorithm 2 Max-Tree Node Partitioning

Require: Initial max-tree T
Ensure: Tree partition $\mathcal{P}T$ of T
for each node n in T (root to leaf iteration) **do**
  if n has no significant ancestor **then**
    mark n as semantic node $\mathcal{S}_n$
  else if checkNewSource(n) is True **then**
    mark n as semantic node $\mathcal{S}_n$
  else
    assign n to it's immediate semantic node ancestor
  end if
end for
return tree partitions as $\mathcal{P}T$

node. We denote this node by $\mathcal{S}_n$, where n is the semantic node index in T. $\mathcal{S}_n$ determines the segment boundary of the corresponding sub-tree.

Definition 2: A *tree partition* $\mathcal{P}T$ is a division of a max-tree T into a set of non-empty sub-trees $\llcorner T_0, \ldots, \llcorner T_\ell$ such that each tree node belongs to exactly one sub-tree. This can be expressed mathematically as

$$\forall i, j \in 0, \ldots, \ell, \quad \llcorner T_i \cap \llcorner T_j = \varnothing \quad \text{if} \quad i \neq j. \tag{1}$$

If N is the set of nodes in the max-tree T, then

$$N = \bigcup_{i=0}^{\ell} \llcorner T_i. \tag{2}$$

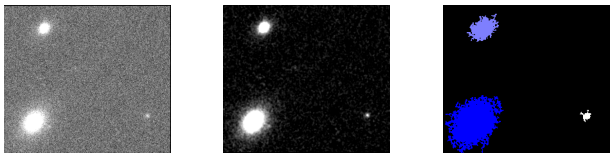
The partitioning of the max-tree into sub-trees provides a framework for grouping connected pixels and extracting their grey-level information based on the image partition they reconstruct.

III. MMTO ALGORITHM

A. GRAPH REPRESENTATION OF MULTI-BAND SOURCES

We use a directed acyclic graph $G = (N, E)$ to represent the multi-band sources, where $N = n_1, n_2, \ldots, n_\ell$ is the set of nodes and E is the set of edges. The graph is topologically ordered by a relation $\prec$ on N such that for every edge $(n_i, n_j) \in E$, $n_j \prec n_i$.

Proposition 1: Let $T_1, T_2, \ldots, T_\ell$ be separate max-trees with roots $r_1, r_2, \ldots, r_\ell$ respectively. If a new node r_0 is added such that $r_1, r_2, \cdots, r_\ell \prec r_0$ and the edges $(r_1, r_0), (r_2, r_0), \ldots, (r_\ell, r_0)$ are created, then the resulting graph G is also topologically ordered.



(a) (b) (c)

FIGURE 2. (a) A small piece of the astronomical simulated image, (b) image reconstruction of max-tree significant nodes, and (c) segmentation map of groups of pixels with different intensities as a single source.

Proof: Let G be the graph resulting from adding r_0 node and $(r_1, r_0), (r_2, r_0), \ldots, (r_\ell, r_0)$ edges.

In a topologically ordered graph, if (n_i, n_j) is a directed edge, then n_j precedes n_i. Since r_0 precede $r_1, r_2, \ldots, r_\ell$, adding these edges doesn't introduce any directed cycles in G, preserving its topological ordering, thus, the proposition holds.

Definition 3: A semantic tree $\mathcal{T}$ is a partitioned max-tree with semantic nodes (local roots) holding grey-level information of corresponding partition nodes.

Figure 3 illustrates how a semantic tree is derived from a max-tree partition of significant nodes.

We integrate a set of semantic trees $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_\ell$, to a graph $\mathcal{G}$ by assigning their roots to a new root node while preserving the node area orders. Figure 6-a illustrates the initial graph, constructed from two semantic trees.

Proposition 2: If new edges are added to a topologically ordered graph without violating the partial order relationship between the vertices, the resulting graph will also be topologically ordered.

Proof: Let $\mathcal{G}$ be a topologically ordered graph and N be it's set of nodes with a partial order relationship denoted by $\prec$. For every directed edge (n_i, n_j), n_j precedes n_i.

We need to show that the resulting graph $\mathcal{G}'$, obtained by adding new edges to $\mathcal{G}$, remains topologically ordered. $\mathcal{G}$ was initially a directed acyclic graph (DAG), as no cycles were present. The addition of new edges with directions consistent with $\prec$ does not introduce cycles into $\mathcal{G}'$, because the new edges maintain the acyclic property of $\mathcal{G}$. Thus, $\mathcal{G}'$ is also a DAG. Moreover, $\mathcal{G}'$ preserves the partial order relationship $\prec$ among its nodes since the new edges adhere to this relationship. Therefore, $\mathcal{G}'$ is a topologically ordered graph.

MMTO, utilizes partial grey-level information of max-trees as vector elements and identifies correlated cross-tree semantic nodes using the hierarchical order of inclusion, center-to-center distance, and cosine similarity of the vector elements

B. SOURCE COMPARISON THROUGH SEMANTIC SIMILARITY

In multi-band source representation, similar sources should be identified as the same entities with intensity variations, as a result of different image acquisition parameters of different bands. To address this, we utilize the cosine

similarity measure to quantify the similarity of sources that are located in close proximity to one another in individual frames, thereby merging them into the same source in the multi-band segmentation map. To achieve this, we first map the semantic nodes of two (or more) semantic trees into a vector space. Then, we compute the cosine similarity of cross-tree semantic nodes, which represent sub-trees precise grey-level information, using

$$similarity(\vec{V}_n, \vec{V}_m) = \frac{\vec{V}_n \cdot \vec{V}_m}{\|\vec{V}_n\| \|\vec{V}_m\|} \tag{3}$$

where $\vec{V}_i$ is the vector corresponding to $\mathcal{S}_i$ semantic node. Using the cosine similarity measure, we can find the correlated sources that display a high degree of similarity, while maintaining the order of sources that are dissimilar providing a means of enhancing the precise identification of multi-band emissions.

To ensure the accuracy of cosine similarity measurements, we employed preprocessing on our dataset. The goal is to identify a subset of features that are most relevant to the target variable, for a robust subsequent analyses such as similarity computation. First, we applied winsorizing to limit the effect of outliers by replacing extreme values with less extreme ones. Let's assume we have an array X. We can represent winsorizing mathematically as

$$x_i' = \begin{cases} p_\alpha(X), & \text{if } x_i < p_\alpha(X) \\ p_{1-\alpha}(X), & \text{if } x_i > p_{1-\alpha}(X) \\ x_i, & \text{otherwise} \end{cases} \tag{4}$$

where x_i is the i-th element of the original array, x_i' is the winsorized i-th element, $p_\alpha(X)$ is the $\alpha - th$ percentile, and $p_{1-\alpha}(X)$ is the $(1-\alpha) - th$ percentile of the array. The parameter α represents the proportion of extreme values to be winsorized. We set $\alpha = 0.05$, and the values above the 95th percentile and below the 5th percentile will be winsorized. Next, we normalized the features using min-max normalization to ensure that each feature has a similar range of values, preventing features with larger values dominating the cosine similarity calculation. By winsorizing and normalization techniques, we can fortify our feature selection process against outliers and ensure that the selected features are highly informative for the task of similarity measurement.

We selected five potential features, namely mean-normalized area, mean-normalized flux, flux/area, first Hu moment, and circularity of the max-tree attributes, to construct vectors that capture information about the intensity, size, and shape of each individual semantic node. The area of a sub-tree is defined as the number of pixels covered by its corresponding semantic node

$$a_n = \sum_{(x,y)\in\mathcal{S}_n} 1, \tag{5}$$

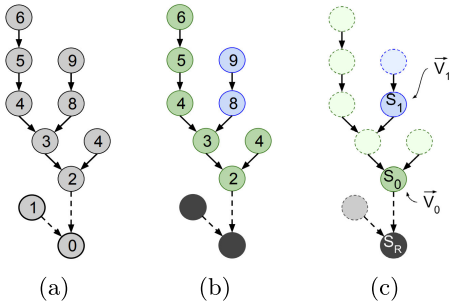


FIGURE 3. (a) Simplified max-tree (solid and dashed edges represent the hierarchical inclusion between significant and non-significant nodes respectively). (b) A potential tree partition with three sub-trees (significant nodes are partitioned based on their attributes; non-significants are assigned to the root). (c) Semantic tree $\mathcal{T}$ and the vectors assigned to semantic nodes constructed from their corresponding sub-tree nodes attributes.

while the flux is the sum of the intensity of these pixels, defined as

$$f_n = \sum_{(x,y)\in\mathcal{S}_n} i(x,y) \tag{6}$$

where x and y are 2D image data coordinates.

In order to account for differences in size and intensity of emissions across individual bands, these elements are mean-normalized by dividing them by the mean values of all semantic nodes on individual semantic trees as

$$\bar{a} = \frac{\sum_{(x,y)\in\mathcal{S}_n\in\mathcal{T}} 1}{\sum_{\mathcal{S}_n\in\mathcal{T}} 1} \tag{7}$$

and

$$\bar{f} = \frac{\sum_{(x,y)\in\mathcal{S}_n\in\mathcal{T}} i(x,y)}{\sum_{\mathcal{S}_n\in\mathcal{T}} 1} \tag{8}$$

Given a semantic node $\mathcal{S}_n$ the raw moments M_{kl} are defined as

$$\mathrm{M}_{kl} = \sum_x \sum_y x^k y^l \tag{9}$$

where x and y are the coordinates of every pixel in $\mathcal{S}_n$. The centroid $(\bar{x},\bar{y})$ of $\mathcal{S}_n$ is given by

$$\bar{x} = \frac{\mathrm{M}_{10}}{\mathrm{M}_{00}} \quad \text{and} \quad \bar{y} = \frac{\mathrm{M}_{01}}{\mathrm{M}_{00}}. \tag{10}$$

Central moments of $\mathcal{S}_n$ used in the following are

$$\mu_{00} = \mathrm{M}_{00}, \tag{11}$$

$$\mu_{20} = \mathrm{M}_{20} - \bar{x} * \mathrm{M}_{10} \tag{12}$$

and

$$\mu_{20} = \mathrm{M}_{02} - \bar{y} * \mathrm{M}_{01}. \tag{13}$$

Normalised moment of inertia (first Hu moment invariant [25]) of $\mathcal{S}_n$ is given by

$$fhm = \eta_{20} + \eta_{02} \tag{14}$$

where

$$\eta_{\text{kl}} = \frac{\mu_{\text{kl}}}{\mu_{00}^{1+\frac{k+l}{2}}}. \tag{15}$$

The circularity c_n of the source can be computed as

$$c_n = \frac{a_n}{p_n^2}. \tag{16}$$

Here, p_n is the perimeter length of $\mathcal{S}_n$ defined as

$$p_n = \sum_{(x,y)\in\partial\mathcal{S}_n} 1 \tag{17}$$

where $\partial\mathcal{S}_n$ denotes the boundary of the semantic node $\mathcal{S}_n$. This equation computes the perimeter length of a connected component in terms of the sum of the lengths of its boundary edges.

To assess the performance of the similarity measure, we used the

$$\text{F-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{18}$$

a widely used evaluation metric that combines precision, defined as

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \tag{19}$$

and recall, defined as

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \tag{20}$$

into a single score. Specifically, we computed precision and recall by comparing the cosine similarity between artificial sources in an image data simulation with ground truth to a predefined similarity threshold.

True positives were defined as the number of source pairs that had a cosine similarity greater than the threshold and were marked as similar sources in the ground truth, while false positives were the number of pairs that had a cosine similarity greater than the threshold but were marked as dissimilar sources. Similarly, false negatives were the number of source pairs that had a cosine similarity less than the threshold but were marked as similar sources, and true negatives were the number of source pairs that had a similarity less than the threshold and were marked as dissimilar sources in the ground truth. The F1-score was then computed using these values, where a perfect score is 1.0 and a score of 0 indicates poor performance.

According to the data presented in figure 4, the elements of

$$\vec{V}_n = \left\langle \frac{a_n}{\bar{a}}, \frac{f_n}{\bar{f}}, \frac{f_n}{a_n}, fhm_n \right\rangle \tag{21}$$

achieve the highest F-score when the optimal similarity threshold of 0.94 is used. The procedure of mapping sub-trees into a vector space is summarised in algorithm 3. Each semantic node in the semantic tree, and its corresponding vector represents a source emission in the image data.

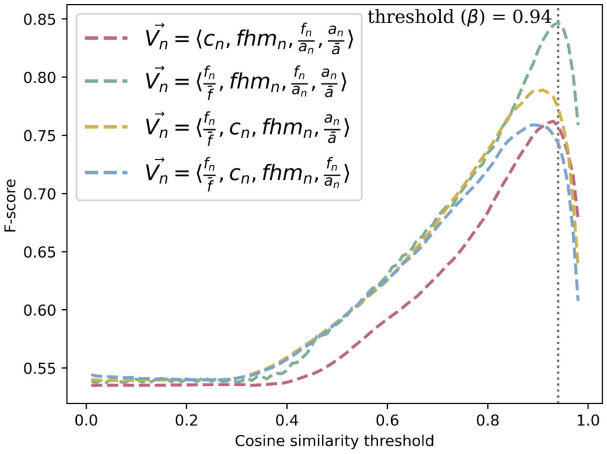


FIGURE 4. Line plot showing the F-score of cosine similarity as a function of cosine similarity threshold. Each line represents a different feature set, with color indicators. The plot highlights the optimal threshold value for each feature set, with corresponding F-score values. Results suggest that the feature set $\frac{a_n}{\bar{a}}$, $\frac{f_n}{\bar{f}}$, $\frac{f_n}{a_n}$ and fhm_n provides the best performance across all threshold values, with the highest F-score of .846 at a similarity threshold of 0.94.

Algorithm 3 Semantic Tree Construction

Require: Tree partition $\mathcal{PT}$
Ensure: Semantic tree $\mathcal{T}$

$V \leftarrow \varnothing$
for each sub-tree $\llcorner\mathcal{T}$ in $\mathcal{PT}$ **do**
  $a_n \leftarrow \text{getArea}(\llcorner\mathcal{T})$
  $f_n \leftarrow \text{getFlux}(\llcorner\mathcal{T})$
  $fhm_n \leftarrow \text{getFirstHuMoment}(\llcorner\mathcal{T})$
  $\vec{V}_n \leftarrow \left\langle a_n, f_n, \frac{f_n}{a_n}, fhm_n \right\rangle$
  $V \leftarrow V \cup \{\vec{V}_n\}$
end for
$V \leftarrow \text{winsorize}(V)$
$V \leftarrow \text{normalize}(V)$
$V \leftarrow \text{meanNormalizeAreas}(V)$
$V \leftarrow \text{meanNormalizeFluxes}(V)$
return $\mathcal{T}$ and V

C. CROSS-TREE CORRELATED SOURCE MERGING

Let $\mathcal{G}$ be the topologically ordered graph derived by integration of $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_\ell$ semantic trees. Let $\mathcal{S}_n \in \mathcal{T}_i$ be a semantic node that we want to find its correlated source $\mathcal{S}_m \in \mathcal{T}_j$. Using the hierarchical order of nodes in $\mathcal{G}$, we only compare the cosine similarity and center of mass to center of mass distance of siblings of $\mathcal{S}_n$ in $\mathcal{T}_j$, to prevent irrelevant comparisons. Let $\mathcal{S}_m$ be a sibling of $\mathcal{S}_n$ (both $\mathcal{S}_n$ and $\mathcal{S}_m$ have the same parent in $\mathcal{T}_j$); $\mathcal{S}_m$ and $\mathcal{S}_n$ are the same sources in different bands if their center of mass distance is less and their corresponding vectors $\vec{V}_n$ and $\vec{V}_m$ have a cosine similarity greater than certain thresholds; we add an edge from $parent(\mathcal{S}_m, \mathcal{T}_j)$ to $\mathcal{S}_n$ to merge their segment border in the final segmentation map (adding this edge keeps $\mathcal{G}$ a topologically ordered graph as the area of $parent(\mathcal{S}_m, \mathcal{T}_j)$ is larger than the area of $\mathcal{S}_n$).

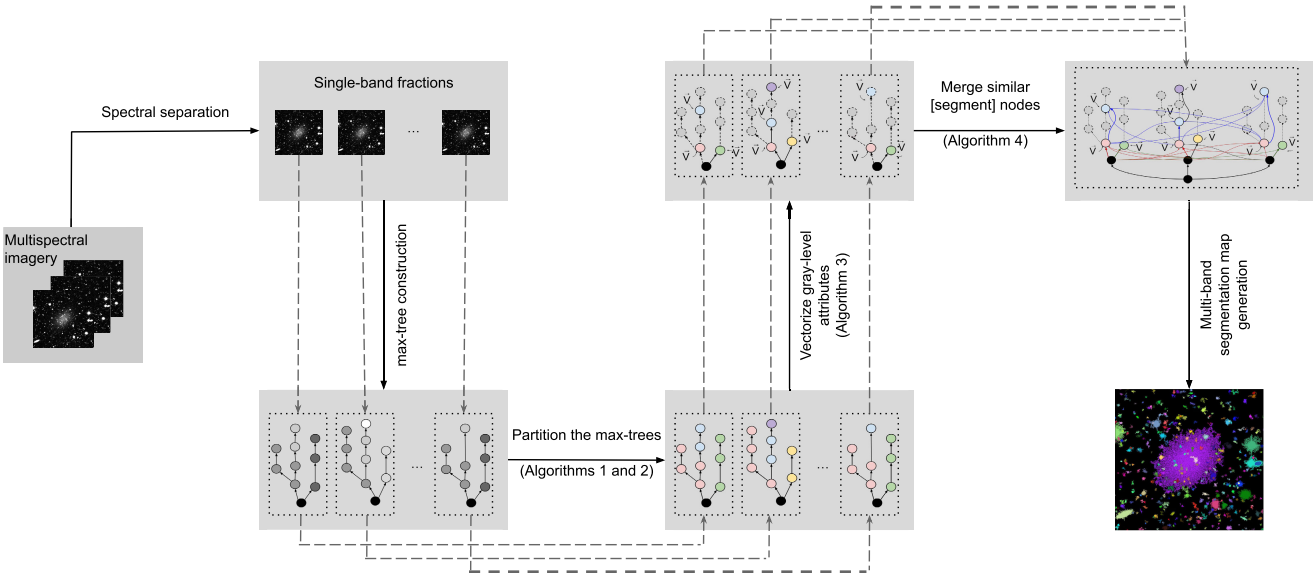


FIGURE 5. MMTO implementation overview. Solid arrows denote the algorithm sequence, while dashed arrows provide an overview of the data processing flow. Following the construction of max-trees on individual channels, grey-level attributes of significant nodes are transformed into a vector space, facilitating the generation of the final multi-band segmentation map. Correlated sources are identified within the graph structure to extract multi-band information.

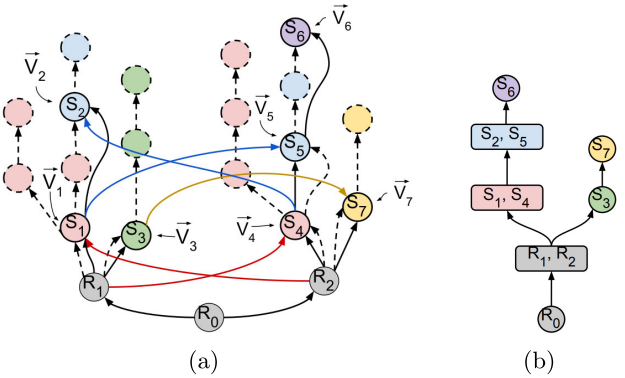


FIGURE 6. (a) Graph $\mathcal{G}$ derived by two semantic trees; solid lines correspond to semantic and dashed lines represent initial inclusion relations. Nodes with the same colors in different trees represent correlated sources; vectors are assigned to semantic nodes of each sub-tree. Cross-tree (colorful) edges are added to $\mathcal{G}$ to unify the path from the root to correlated sources. (b) Final segmentation map preserving the order of overlapping non-related sources.

Algorithm 4 Merge Similar Semantic Nodes

Require: Topologically ordered graph of semantics $\mathcal{G}$
Ensure: Unified graph $\mathcal{G}$
1: **for** $\mathcal{S}_n \in \mathcal{T}_i \in \mathcal{G}$ **do**
2: **for** $\mathcal{S}_m \in$ siblings$(\mathcal{S}_n, \mathcal{T}_j) \in \mathcal{T}_j \in \mathcal{G} - \{\mathcal{T}_i\}$ **do**
3: **if** $distance(\mathcal{S}_n, \mathcal{S}_m) \leq \delta$ **then**
4: **if** $similarity(\vec{V}_n, \vec{V}_m) \geq \beta$ **then**
5: $E \leftarrow E \cup \{(parent(\mathcal{S}_m, \mathcal{T}_j), \mathcal{S}_n))\}$
6: **end if**
7: **else if** $area(\mathcal{S}_n) \leq area(\mathcal{S}_m)$ **then**
8: $E \leftarrow E \cup \{(\mathcal{S}_m, \mathcal{S}_n)\}$
9: **else**
10: $E \leftarrow E \cup \{(\mathcal{S}_n, \mathcal{S}_m)\}$
11: **end if**
12: **end for**
13: **end for**
14: **for** $n \in \mathcal{G}$ **do**
15: $path(n) \leftarrow longestPath(n, \text{root})$
16: **end for**
17: **Return** $\mathcal{G}$

In graph $\mathcal{G}$, a pixel of the image may belong to several nodes. It means two irrelevant sources from different bands may overlap. To preserve the order of overlapping sources that do not merge as correlated semantic nodes, we add an edge from the node with the larger area to the smaller one, each one from a different semantic tree (line 7 of algorithm 4).

Figure 6 is a schematic illustration of the construction of graph $\mathcal{G}$ and merging cross-band correlated sub-trees by adding edges to unify their path (the longest path from the root is considered for each node in $\mathcal{G}$, to preserve the cross-band hierarchical order). Using this graph structure we simultaneously elevate multiple max-trees to a higher dimension, enabling the identification of cross-tree correlated sources. We then reintroduced these sources as a single segment, which led to the construction of the final multi-band segmentation map. The proposed method is implemented as shown in figure 5.

D. TIME COMPLEXITY ANALYSIS

This section analyzes the time complexity of the proposed method for constructing a segmentation map from b individual images. Let n_s and n_p be the number of semantically meaningful nodes and pixels in each image, respectively.

The initial step is to build the max-trees for each image, which has a time complexity of $\mathcal{O}(bn_p \log n_p)$. Then, for each semantic node in the max-trees, we need to find its *Parent* in the other trees, which has a time complexity of $\mathcal{O}(b^2 n_s^2)$, where $b \ll n_s \ll n_p$.

Next, we construct the directed acyclic graph $\mathcal{G}$. The first step is to add nodes to the graph, which has a time complexity of $\mathcal{O}(bn_s)$. Since there are b images, the total number of nodes in $\mathcal{G}$ is bn_s. Each node can have up to $2b$ edges. We then add edges to $\mathcal{G}$, with a time complexity of $\mathcal{O}(b^2 n_s m_s)$, where m_s is the number of direct children of a node ($m_s \ll n_s$).

After constructing the graph $\mathcal{G}$, we add edges between each node and the parents of related nodes in other max-trees based on cosine similarity over feature vectors. This process has a time complexity of $\mathcal{O}(n_s^2)$. Finally, we convert the graph $\mathcal{G}$ to a final unique tree using a distance map from the *root* to each component. Since $\mathcal{G}$ is a topologically ordered graph, this step can be accomplished with time complexity of $\mathcal{O}(|\mathcal{E}| + |\mathcal{V}|)$, where $|\mathcal{E}|$ and $|\mathcal{V}|$ are the number of edges and vertices in $\mathcal{G}$.

IV. EXPERIMENTS

In this study, we evaluate the effectiveness of the proposed method, referred to as MMTO, in detecting and merging correlated segments in multi-band image data. To assess its accuracy, experiments were carried out to compare the performance of MMTO with CGO, a framework for multi-band astronomical image segmentation. Our results demonstrate that MMTO surpasses CGO in terms of accuracy and computational efficiency. Furthermore, we apply the MMTO method to medical images and present a visual representation of the multi-band segmentation result to indicate its generality.

1) ASTRONOMICAL IMAGE SIMULATIONS

Evaluating the performance of astronomical image analysis algorithms can be challenging due to the lack of known ground truth in real image data. This is because many important features, such as very faint objects, may not be present in the captured data. To address this issue, simulated astronomical image data can be utilized as a substitute. The use of simulated data offers the advantage of providing a known ground truth, enabling the evaluation of algorithms based on the parameters of generated sources. Using the simulations, we can evaluate the performance of astronomical image analysis algorithms in a controlled environment, to ensure that the algorithms are robust and reliable when applied to real astronomical image data.

In this study, two types of three-channel data frames were generated to simulate astronomical images. We have restricted the simulations to this number of frames because, unlike MMTO which is able to handle any number of frames, CGO is limited to 3-band images. The first type mimics actual multi-band observations of a specific section of the sky and includes various sources such as stars, galaxies, and background galaxies, each with its own unique parameters, such as surface brightness, flux per unit solid angle of the image, effective radius, and axis ratio. To each frame, different Gaussian and Poissonian noises were added. The second group of simulations used the same parameters and noise distribution across different frames to evaluate the accuracy of semantic merging. Two three-channel datasets of these simulations are presented in figure 7, where the two types of simulations are shown in two separate columns.

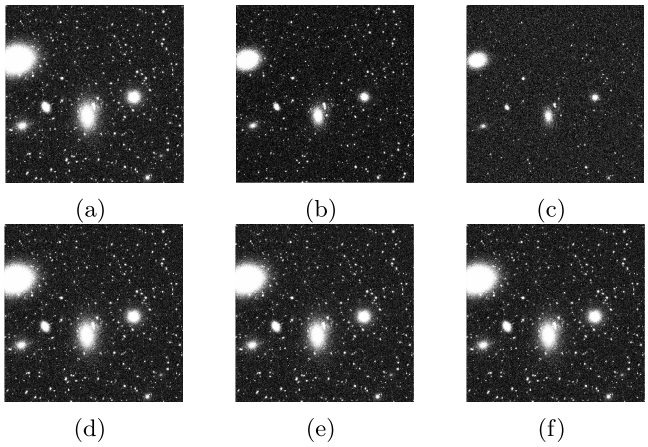


FIGURE 7. Three-band astronomical image simulation with different (a, b, and c) and same (d, e, and f) parameters of surface brightness, effective radius, axis ratio, and noise distribution. All the frames have a size of 4,000 pixels squared.

2) MMTO MERGING ACCURACY ON ASTRONOMICAL IMAGE SIMULATIONS

To test the accuracy of semantic merging, we used three-band simulations with sources generated using the same parameters and noise distribution in each frame, as shown in figure 7 (b, d, and f). It was expected that each semantic node would be merged into precisely two other segments representing the corresponding source in the other two channels. A total of 100 datasets of three-band simulations, comprising 300 frames, were generated with each having a size of 1,000 square pixels.

To build the semantic tree nodes, we utilized MTObjects, which has been shown to be an accurate single-band source detection and segmentation tool for astronomical image data [17]. The detection process performed by MTObjects involves subtracting the background and constructing a max-tree from the smoothed background subtracted image. This tree is then filtered using statistical significance tests to identify sources by determining nodes with an amount of flux, given their area (both flux and area are considered as elements of the vector representing the segment map) that are statistically distinct from the background.

The position of the center-of-mass of sources in three individual bands and the MMTO merger map is shown in figure 9 based on the ground truth of the simulations. The pale dots represent central position of the faint sources in the simulations that MTObjects does not mark their corresponding nodes as significant. MMTO merges the

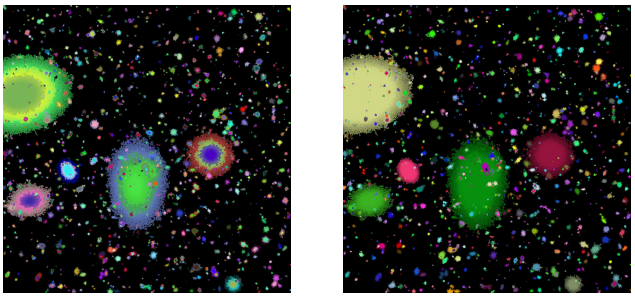


FIGURE 8. MMTO multi-band segmentation map output, where cosine similarity threshold (β in algorithm 4), is set to (a) a value greater than one or less than zero, (b) optimal value (.94 for this data set).

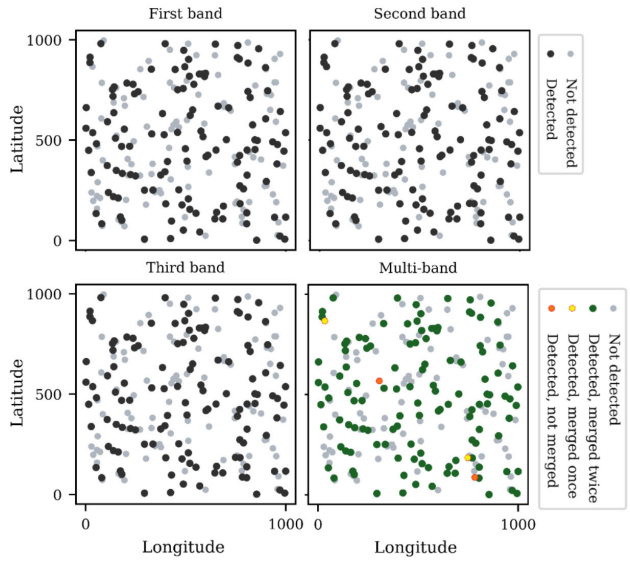


FIGURE 9. Accurate merging of ground truth sources in a three-band simulated dataset. Grey-level panels: bold dots represent detected sources, while pale dots represent non-detected sources in individual bands. Color points in the bottom-right panel indicate the position of sources that have been successfully merged, as shown by the number of mergers.

semantic nodes with their corresponding segments in other frames with an average accuracy of 98% ± 1 on 100 3-channel simulated image datasets. As shown in algorithm 4, MMTO has two adjustable parameters: the center-of-mass distance δ and the cosine similarity threshold β [of sources]. In this study, we set these parameters to 10 pixels and 0.94 (as illustrated in figure 4), respectively.

Figure 8 illustrates the multi-band segmentation map using the simulated image data with a different number of sources having different parameters on each channel (figure 7, a, c, and e) where the correlated segments have been merged (left panel) or not (right panel). To prevent merging similar segments, we simply set the cosine similarity threshold, β, to any value out of the [0, 1] range. Even if similar segments fail to merge, the algorithm preserves all the segments in the order of area, which is the property that we use to construct the DAG.

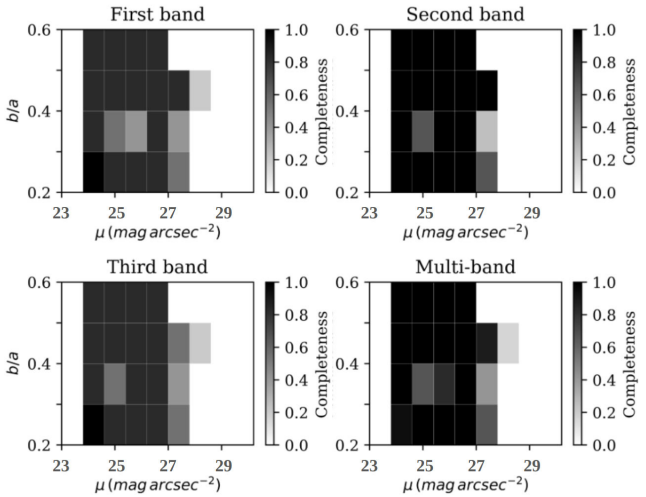


FIGURE 10. Multi-band detection completeness for MMTO (bottom-right panel) and single-band detection using max-tree-based method, MTObjects as a function of mean effective surface brightness and axis ratio of the sources. The grey scales correspond to the completeness indicated by the color bar.

3) DETECTION COMPLETENESS

To further evaluate the efficacy of MMTO, we conducted an empirical evaluation comparing the multi-band completeness against single-band detection using MTObjects. Our investigation utilized artificial images with known ground truth, incorporating a diverse array of sources appearing across multiple frames, each characterized by distinct parameters such as surface brightness. Additionally, certain sources exhibited emissions exclusively in one or two bands, contributing to the complexity of the assessment.

This analysis is depicted graphically in figure 10. Four plots were generated to illustrate the detection completeness of three single-band max-tree based and multi-band MMTO detection. Overall, MMTO exhibited superior completeness across a greater number of blocks compared to its counterpart operating on single bands. However, in specific instances, such as the second band, individual single-band displayed higher completeness. This phenomenon stems from the completeness metric capturing emissions distributed across multiple channels collectively for multi-band detection evaluation.

4) MMTO VS. CGO

The Component-Graph Object (CGO) framework, which is a multi-band astronomical object detection and segmentation approach, extends the single-band significance test concept of MTObjects into the multi-band domain using a component-graph structure. Similar to MTObjects, the node power attribute in CGO is calculated as the sum of the squared difference between the corresponding pixel values of the node and the level of its parent node. In the component-graph data structure, the supremum of the parent levels is used as a reference (alternative references such as the average, infimum, or maximum have also been proposed). The power

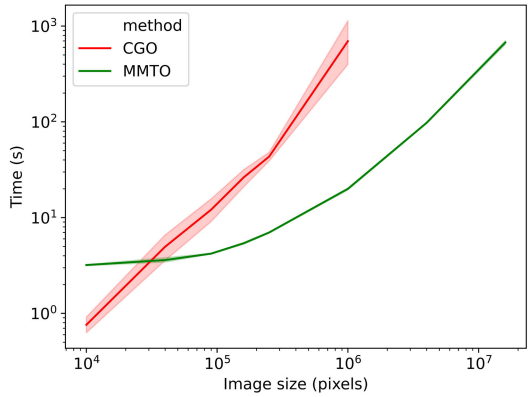


FIGURE 11. Execution time of MMTO and CGO as a function of image data size. Both axes are logarithmically scaled. The line chart illustrates the relative computational cost of MMTO and CGO for different image sizes measured in pixels. The point of intersection between the two lines indicates that the computational cost of attribute extraction for MMTO is higher than statistical tests on the component-graph structure for very small images. As the number of connected components increases with increasing image sizes, MMTO outperforms CGO in terms of execution time. These findings demonstrate that MMTO is a more efficient image segmentation technique for larger images with a significant number of connected components.

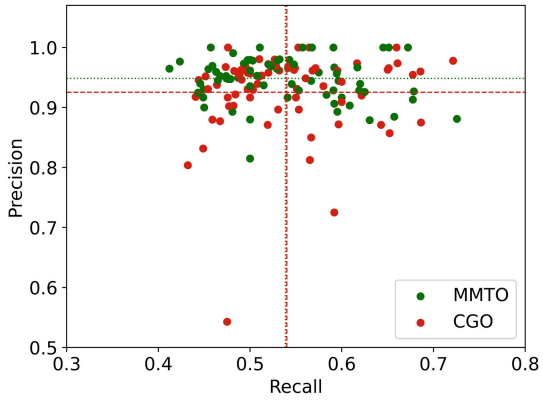


FIGURE 12. The scatter plot illustrates the comparison of precision and recall for MMTO and CGO on 3-band datasets. MMTO exhibits slightly better precision values than CGO, and the points are less scattered, indicating that the performance of MMTO is more consistent across the different test cases. These findings suggest that MMTO is a more reliable image segmentation technique for complex visual data with varying degrees of complexity.

attributes of MTObjects and CGO, are

$$power_{MTObjects}(N) = \sum_{x \in N} (f(x) - f(parent(N)))^2 \qquad (22)$$

and

$$power_{CGO}(N) = \sum_{x \in N} (f(x) - \sup_{y \in parent(N)} f(y))^{\circ 2} \qquad (23)$$

respectively, where $\circ$ is the element-wise power in CGO. For multi-band images, CGO uses a significance test combining separate bands and the synthesized band.

As a result of merging the full max-trees, grey-level by grey-level, CGO fails to generate the output in a reasonable time. The largest image we could use to evaluate CGO was 1,000 square pixels in dimensions and it crashed on some of these (in [18], sliced simulation tiles of 500 square pixels are used to test CGO); any larger image needs drastically more time, whereas MMTO can handle multi-band image data with the size of 20,000 square pixels per band. To have a decent comparison between these two methods, we limited the size of our simulations to 1,000 square pixel images or smaller and executed both methods on the same machine.

We generated 60 3-band datasets with the sizes of 100, 200, 300, 400, 500, and 1,000 square pixels (ten datasets for each size, 5 with the same parameters and 5 with different parameters) to compare the execution time of MMTO and CGO. Figure 11 shows the huge time difference between these two algorithms. Compared to CGO, MMTO has less dispersion from the mean value of each 10 3-band data for each size. The efficient and deterministic execution time of MMTO is due to it's selection of statistically significant nodes in each max-tree, including extracted grey-scale features mapped into a vector space instead of a grey-level by grey-level in an expensive non-hierarchical graph which is used by CGO.

In component trees and component graphs, objects are represented as sequences of significant nodes that appear differently at varying thresholding levels. Each node in MMTO and its primary sub-tree belong to the same object. However, in the context of the component graph with partial orders, multiple sub-trees containing incomparable nodes that belong to a single object may exist. To address this, CGO assumes that the sub-trees corresponding to an identical object will converge to similar centers. Despite the increased time complexity of identifying sub-trees in the component graph, CGO results in more false positives in assigning nodes to sub-trees. Figure 12 demonstrates the precision and recall of source-segmentation for both MMTO and CGO. While both methods have similar recall values, MMTO exhibits a less scattered distribution in terms of precision, improving detection precision from 0.92 to 0.95.

5) RESULTS ON INSERTED ACTUAL GALAXIES

We assessed the efficacy of the proposed method on a subset of three-band real galaxies that were embedded within a vacant part of the sky in three different *Fornax Deep Survey* (FDS) frames that do not contain large galaxies or stars. By evaluating the tool on these real galaxies, we aimed to validate the generalizability of its behavior to more complex galaxy structures like spiral galaxies. For this study, we extend the approach previously utilized by Haigh et al. [17] to multi-band. Specifically, we employed an empty region of the sky from a three-band FDS image as the template and inserted actual galaxies into this region such that the centers of the corresponding galaxies in different bands were cross-frame identical.

We selected a sample of eight galaxies from the EFIGI catalog [27], which contains multi-band galaxies from the

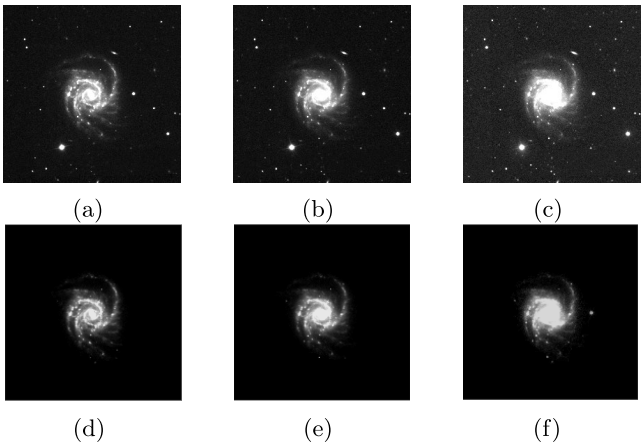


FIGURE 13. Top row: one of the inserted actual SDSS-galaxies in g, r and i bands (from left to right). Bottom row: The corresponding isolated galaxies determined through the use of k-flat zones [26].

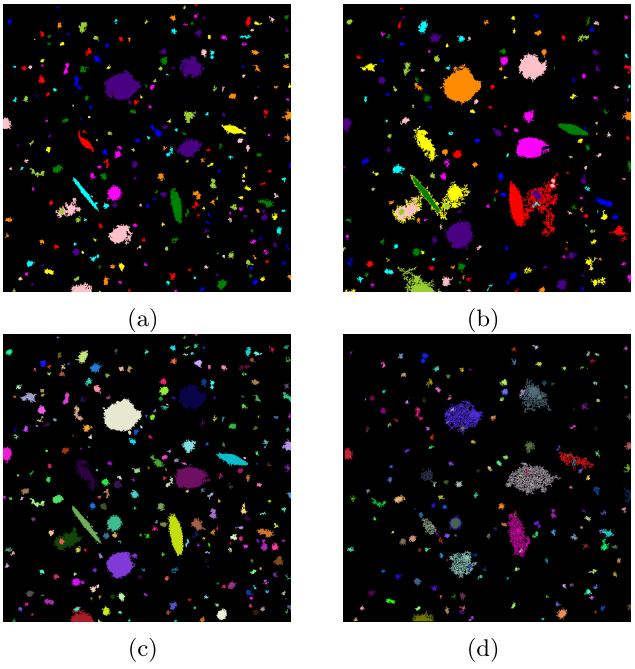


FIGURE 14. (a) Single-band segmentation map on the frame with the highest SNR. (b) Single-band segmentation map on the 3-band stacked image. (c) MMTO three-band segmentation map. (d) CGO three-band segmentation map.

fourth data release of the *Sloan Digital Sky Survey* (SDSS). The galaxy located at the center of each SDSS image was isolated through the application of k-flat filtering [26]. This method effectively removed regions of light that were not connected to the central pixel, while maintaining the intricate details internal to the galaxy. The first row of figure 13 shows one of the selected SDSS-galaxies to insert in three different g, r, and i bands. Their corresponding processed data are illustrated at the bottom row, in the same order.

To ensure that the inserted galaxies were consistent with the range of signals, we first transformed their pixel values to align with those of the template. We then identified and removed any pixels in the template that were masked by the isolated galaxies before adding the intensity of each, along with Poisson noise, to the template of each individual band. The normalized outcome of inserted galaxies have been rescaled to a factor of 10^{-10} to conform to the FDS frames.

Figure 14 depicts the multi-band segmentation map generated from three raw frames using MMTO and CGO, a single-band segmentation map on the best signal-to-noise frame, and the stacked weighted cross-band intensities. Image stacking is commonly used in astronomy to improve the signal-to-noise ratio and enhance the quality of the final image as single-band data. The normalized flat field characterizes the response of the telescope and instrument to uniform radiation, indicating the relative amount of light received by each pixel position. This weight is used to assess the relative importance of a pixel in comparison to other pixels within the image.

The presence of defect pixels, which may count too few or too many photons, is accounted for by storing them in hot- and cold-pixel maps. In addition, pixels may be deemed unusable for various reasons, including saturation, cosmic ray hits, or satellite tracks. Pixel maps are created to identify these issues, and weights are adjusted accordingly. Saturation occurs when a pixel's count exceeds a predefined threshold, which may also lead to neighboring pixels becoming "dead" with counts below a lower threshold. Cosmic ray events are detected using specialized source detection filters, such as Retina filters trained to identify cosmic rays using neighboring pixels as inputs. Satellite tracks can be identified through a line detection algorithm like the Hough transform, which produces a "peak" in the transformed image where there is a significant signal along a line. This peak can be clipped and transformed back into a pixel map that masks the track. The weight of a given pixel with x and y as the coordinates can be written as

$$\mathcal{W}_{x,y} = \frac{1}{\sigma^2} \cdot \frac{F_{x,y}}{\mathcal{I}_{x,y}} \cdot P^{hot}_{x,y} \cdot P^{cold}_{x,y} \cdot P^{saturated}_{x,y} \cdot P^{cosmic}_{x,y} \cdot P^{satellite}_{x,y} \tag{24}$$

where σ^2 is the variance of the raw image, $F_{x,y}$ is the flat field pixel value, $\mathcal{I}$ is the illumination correction, and P^{hot}, P^{cold}, $P^{saturated}$, P^{cosmic}, $P^{satellite}$ are the hot, cold, saturated, cosmic ray and satellite pixel maps. The weight images are used for masking unwanted pixels before stacking [28]. The pixel values of the stacked image of b bands are written as

$$stacked_{x,y} = \frac{\sum_b ({}^bW_{x,y} \times {}^bImage_{x,y})}{\sum_b {}^bW_{x,y}}. \tag{25}$$

To evaluate the quality of segmentation in figure 14, the ground truth segmentation map of inserted galaxies was employed and the Intersection over Union (IoU) metric was computed, as illustrated in figure 15. The results indicate that MMTO outperforms other methods in segmenting the inserted galaxies. While the single band segmentation of the stacked image and the frame with the highest signal-to-noise ratio (SNR) exhibit a similar IoU score, they also manifest

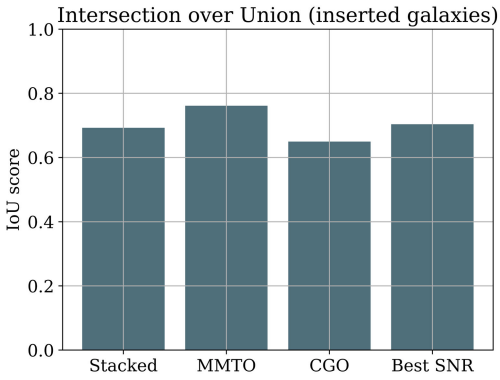


FIGURE 15. The Intersection over Union (IoU) plot illustrates the segmentation performance of different methods, where MMTO demonstrates the highest IoU score, while CGO exhibits the lowest due to the presence of hollow detections. The single band segmentation of the stacked image and the frame with the highest signal-to-noise ratio (SNR) show similar IoU scores but with some limitations in certain regions.

some drawbacks. Specifically, the stacked image displays [malformed] under-segmentations in certain regions. CGO demonstrates the lowest IoU score due to the presence of hollow detections for most of the sources.

6) MMTO VS. MASK R-CNN DEEP LEARNING

To underscore the effectiveness of our approach, we present an comparison between MMTO and a fairly recent deep learning methodology leveraging the Mask Region-based Convolutional Neural Network (Mask R-CNN) [29]. This method enables the execution of multi-band source detection and deblending within a unified machine learning framework. This work extends the implementation of Mask R-CNN by Abdulla,[4] enabling the utilization of multi-band FITS files for input during training and detection phases.

To ensure a fair comparison, we used the Photon Simulator (PhoSim) dataset for evaluations, which served for training the Mask R-CNN model (they generated 1000 simulated DECam images for the training set, with a total training set of approximately 150,000 astronomical sources; each image has around 150 object masks corresponding to a star or galaxy).

The results showcased in figure 17 depict precision and recall metrics for both MMTO and R-CNN. While both methods exhibit comparable recall rates, MMTO demonstrates superior precision. Notably, MMTO displays a more concentrated distribution, indicating robustness, particularly in datasets characterized by heightened complexity. An example of a typical PhoSim training set image with its detection mask corresponding to both galaxy and star classes, using Mask R-CNN and MMTO segmentation output is shown in figure 16 where MMTO preserves nested sources segmentation as a result of it's hierarchical structure and captures fainter outer parts of the sources missed by R-CNN.

In upcoming investigations, we will enhance our evaluation by training deep learning models on custom multi-band simulations, including segmentation ground truth. Notably, we will highlight the capacity of our MMTO method to accurately preserve nested sources, a critical aspect for effectively segmenting intricate astronomical images. This comprehensive analysis will yield deeper insights into the comparative performance across diverse astronomical datasets.

[4] *https : //github.com/matterport/Mask_RCNN*

7) RESULTS ON MULTI-COLOR FLUORESCENCE IMAGE

The proposed method has been primarily evaluated on astronomical image data. To demonstrate its applicability to other sensors, we applied it to a skin image acquired using a 3-channel fluorescence microscope. The visual showcase results demonstrate the potential for adaptation based on the image data characteristics, making it a significant contribution to the field and opening up new possibilities for efficient and accurate multi-band image segmentation, particularly in processing large-scale images.

Multi-color fluorescence imaging is a powerful tool in the field of molecular biology, allowing researchers to visualize the expression, localization, and interactions of multiple proteins in a single sample. This technique involves the use of fluorescent probes, each with a unique spectral signature, which binds to specific targets in the sample. When exposed to light of the appropriate wavelength, these probes emit fluorescence, which is then captured and imaged by a microscope equipped with filters. The simultaneous detection of multiple fluorescent signals can reveal complex patterns of protein distribution, providing valuable insight into cellular processes such as cell division, differentiation, and signaling pathways. Multi-color fluorescence imaging has become an indispensable tool in a wide range of biological research applications, including studies of development, disease, and drug discovery.

Figure 18(a) presents visual outcomes obtained from a fluorescence microscope-acquired 3-channel 640×390 pixels image and its multi-band segmentation map generated by MMTO (figure 18(b). The sample was stained by the procedure in [30], in which both the red and green fluorescent dyes are targeted collagen Type VII as geometric verification marker, and the blue dye targets DNA in the cellular nuclei. Thus, the red channel structures should correlate with the green, but not the blue.

As multi-color fluorescence has less complexity compared to astronomical images, we could mark the significant nodes representing the emissions using a simple attribute filtering (based on the intensity of connected components) and partition them into max-tree group nodes representing individual sources; we filtered the raw images based on intensities higher than 100, 100, and 50 for red, green, and blue channels, respectively. Figure 19 (a, c, and e) shows the color maps of the raw grey-level images of different bands, along with the reconstructed sources from the simplified max-trees of the raw images, holding significant nodes. We used the same feature set to apply MMTO on multi-color

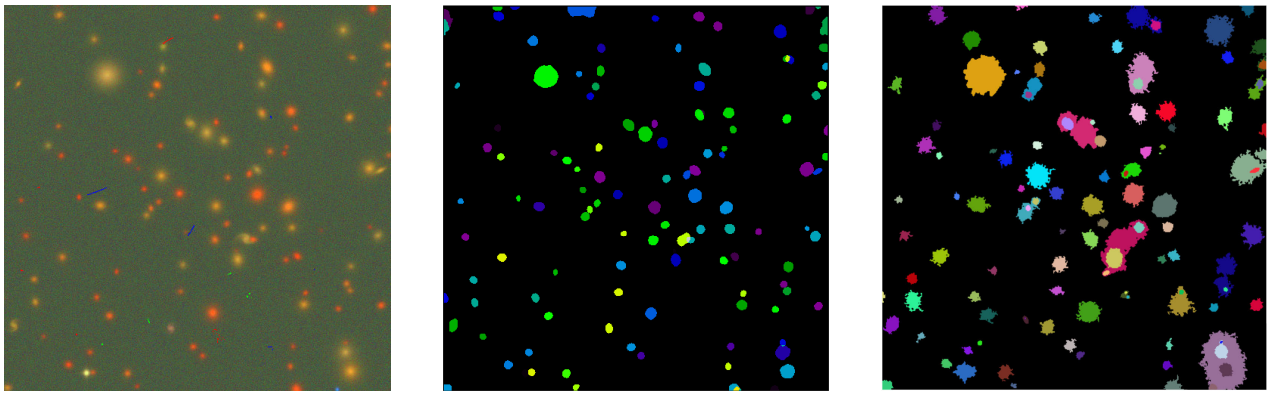


FIGURE 16. An example of PhoSim DECam image. (a) The simulated color image of a crowded field was used as Mask R-CNN training data. (b) Detection mask corresponding to both galaxy and star classes, of Mask R-CNN. (c) MMTO three-band segmentation map.

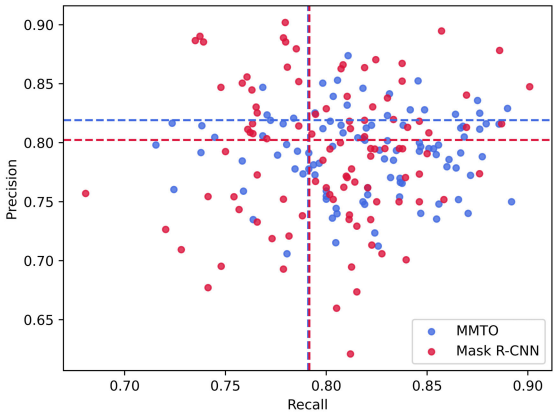


FIGURE 17. The scatter plot illustrates the comparison of precision and recall for MMTO and Mask R-CNN on one hundred 3-band PhoSim dataset. MMTO exhibits slightly better precision values than Mask R-CNN, and the points are less scattered, indicating that the performance of MMTO is more consistent across the different test cases.

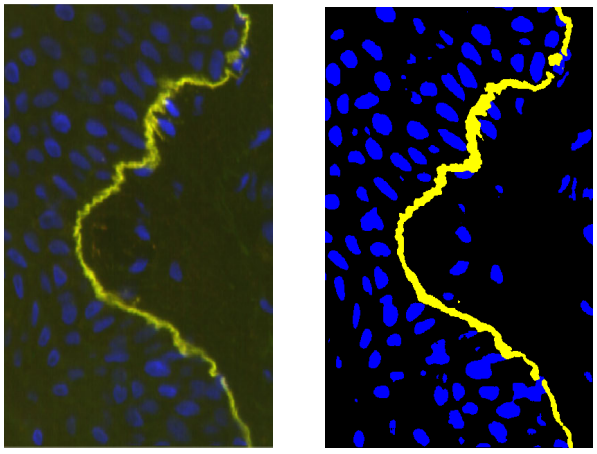


FIGURE 18. (a) A 640 × 390 pixel skin image acquired using a 3-channel fluorescence microscope [30], and (b) its multi-band segmentation map generated using MMTO.

fluorescence imaging, with the same parameters optimized for astronomical simulations (cosine similarity threshold set to .94 and the distance threshold set to 10 pixels).

We acknowledge the possible need for further examination of the feature selection process to represent the sources and measure their similarity for other applications. In astronomical contexts where we have extensively evaluated our approach to complex multi-band astronomical simulations with known ground truth, we found that the normalized flux and area, first Hu moment, and flux over the area were the most effective features (figure 4); adjusting these features based on specific sensor characteristics may be necessary for optimal accuracy.

V. DISCUSSION

We proposed an algorithm for multi-spectral source-segmentation. Our innovative approach involves constructing a topologically ordered graph from multiple max-tree node partitions, holding single-band precious grey-level features; these features are obtained from the max-tree node attributes, that are an accurate representative of the shape, size, position, and intensity of a partition of the image.

When compared to the component-graph-based approach, our method showed significant improvements in terms of time complexity due to the efficient data structure construction and process. Moreover, our method, which employs statistical tests on individual bands to designate significant max-tree nodes as partitions representing the sources, achieves superior detection accuracy.

The presented methodology showcased exceptional segmentation accuracy when applied to real galaxies with diverse characteristics. Comparative analysis was conducted against three approaches: a max-tree-based method applied on the single-band image with the best signal-to-noise ratio, the fusion of multi-band image data into a single-band representation of a higher resolution using the weighted combination (known as stacking in astronomy) and multi-band

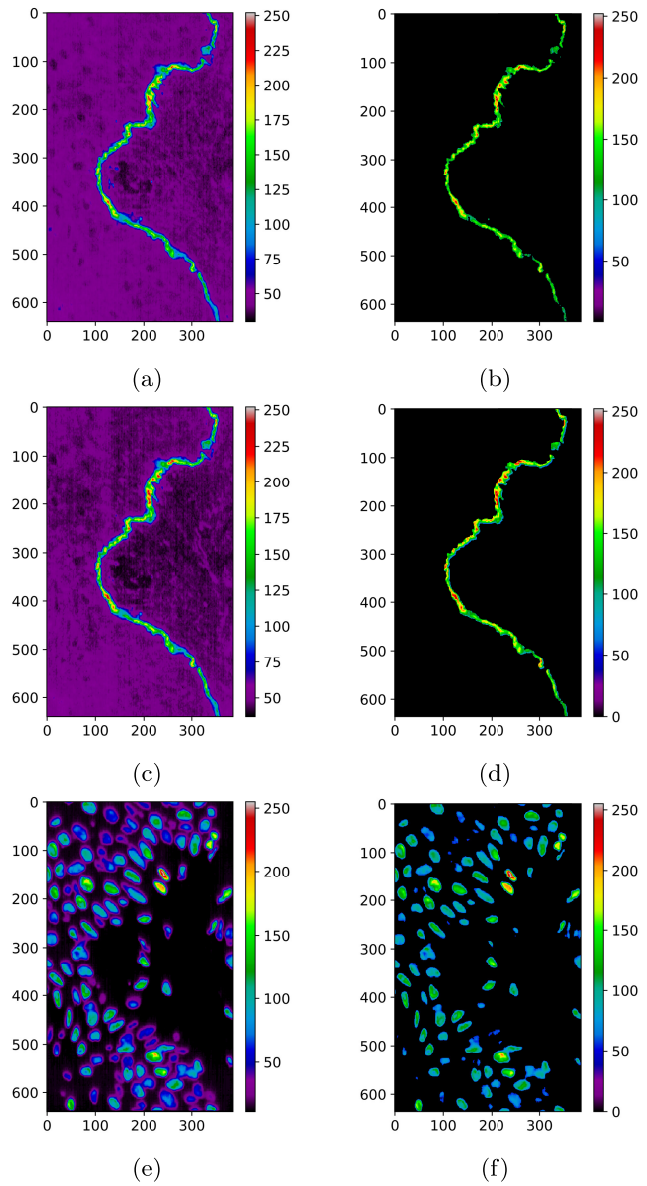


FIGURE 19. Left column: the color map of three channels of figure 18-a (red, green, and blue channels from top to bottom). Right column: significant nodes color map. The single-band max-trees have been simplified using attribute filtering with the intensity threshold set to 100, 100, and 50 from the top to the bottom.

detection using an state-of-the-art method named CGO. Across these comprehensive evaluations, MMTO consistently outperformed the comparatives, affirming its superior performance in accurately segmenting complex features and highlighting its robustness in handling diverse structures and imaging conditions.

The method's adaptability was tested using a multi-color fluorescence microscope image, showcasing its potential beyond astronomy. Our method effectively separated distinct emissions through max-tree attribute filtering and node partitioning based on intensities. Despite initially being optimized for astronomical simulations, this successful application hints at the method's broader utility in diverse scientific fields, underlining its potential for accurate multi-band image segmentation, by fine-tuning according to specific sensor characteristics to optimize accuracy in various domains.

VI. FUTURE WORK

The proposed algorithm has shown promising results in terms of accuracy and time efficiency; however, there is still room for improvement. Future research will aim to improve the statistical tests used to identify and partition significant nodes on initial max-tree structures and to deblend separate overlapping sources having short center-to-center distances that can result in malformed segments, allowing for more accurate cross-band correlated source detection.

The integration of machine learning algorithms into the hierarchical structures for source detection can further improve the accuracy and efficiency of the proposed algorithm. By focusing on these key areas, we can better understand the complex relationships between multi-spectral grey-level images.

REFERENCES

[1] C. Elachi and J. J. Van Zyl, *Introduction to the Physics and Techniques of Remote Sensing*. Hoboken, NJ, USA: Wiley, 2021.

[2] A. F. H. Goetz, B. N. Rock, and L. C. Rowan, "Remote sensing for exploration; an overview," *Econ. Geol.*, vol. 78, no. 4, pp. 573–590, Jul. 1983.

[3] R. Weissleder and M. J. Pittet, "Imaging in the era of molecular oncology," *Nature*, vol. 452, no. 7187, pp. 580–589, Apr. 2008.

[4] D. A. Agard, Y. Hiraoka, P. Shaw, and J. W. Sedat, "Fluorescence microscopy in three dimensions," *Methods Cell Biol.*, vol. 30, pp. 353–377, Jun. 1989.

[5] B. Herman, *Fluorescence Microscopy*. New York, NY, USA: Garland Science, 2020.

[6] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. Berlin, Germany: Springer, 2007.

[7] E. Bertin and S. Arnouts, "SExtractor: Software for source extraction," *Astron. Astrophys. Suppl. Ser.*, vol. 117, no. 2, pp. 393–404, Jun. 1996.

[8] D. J. Prole, J. I. Davies, O. C. Keenan, and L. J. M. Davies, "Automated detectionof very low surface brightness galaxiesin the virgo cluster," *Monthly Notices Roy. Astronomical Soc.*, vol. 478, no. 1, pp. 667–681, Jul. 2018.

[9] M. Akhlaghi and T. Ichikawa, "Noise-based detection and segmentation of nebulous objects," *Astrophysical J. Suppl. Ser.*, vol. 220, no. 1, p. 1, 2015.

[10] A. S. G. Robotham, L. J. M. Davies, S. P. Driver, S. Koushan, D. S. Taranu, S. Casura, and J. Liske, "ProFound: Source extraction and application to modern survey data," *Monthly Notices Roy. Astronomical Soc.*, vol. 476, no. 3, pp. 3137–3159, May 2018.

[11] P. Teeninga, U. Moschini, S. C. Trager, and M. H. F. Wilkinson, "Statistical attribute filtering to detect faint extended astronomical sources," *Math. Morphology-Theory Appl.*, vol. 1, no. 1, pp. 100–115, 2016.

[12] N. Passat and B. Naegel, "An extension of component-trees to partial orders," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 3981–3984.

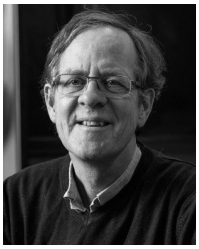
[13] F. Tushabe and M. H. F. Wilkinson, "Color processing using max-trees: A comparison on image compression," in *Proc. Int. Conf. Syst. Informat. (ICSAI)*, May 2012, pp. 1374–1380.

[14] C. Kurtz, B. Naegel, and N. Passat, "Connected filtering based on multivalued component-trees," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5152–5164, Dec. 2014.

[15] E. Carlinet and T. Géraud, "A color tree of shapes with illustrations on filtering, simplification, and segmentation," in *Proc. 12th Int. Symp. Math. Morphology Appl. Signal Image Process. (ISMM)*. Reykjavik, Iceland: Springer, May 2015, pp. 363–374.

[16] E. Carlinet and T. Géraud, "MToS: A tree of shapes for multivariate images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5330–5342, Dec. 2015.
[17] C. Haigh, N. Chamba, A. Venhola, R. Peletier, L. Doorenbos, M. Watkins, and M. H. F. Wilkinson, "Optimising and comparing source-extraction tools using objective segmentation quality criteria," *Astron. Astrophys.*, vol. 645, p. A107, Jan. 2021.
[18] T. X. Nguyen, G. Chierchia, L. Najman, A. Venhola, C. Haigh, R. Peletier, M. H. F. Wilkinson, H. Talbot, and B. Perret, "CGO: Multiband astronomical source detection with component-graphs," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 16–20.
[19] P. Salembier and M. Wilkinson, "Connected operators," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 136–157, Nov. 2009.
[20] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.
[21] Y. Xu, T. Géraud, and L. Najman, "Morphological filtering in shape spaces: Applications using tree-based image representations," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 485–488.
[22] C. Merino-Gracia, K. Lenc, and M. Mirmehdi, "A head-mounted device for recognizing text in natural scenes," in *Proc. 4th Int. Workshop Camera-Based Document Anal. Recognit. (CBDAR)*. Beijing, China: Springer, Sep. 2011, pp. 29–41.
[23] W. Li, Z. Wang, L. Li, and Q. Du, "Feature extraction for hyperspectral images using local contain profile," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5035–5046, Dec. 2019.
[24] R. Jones, "Connected filtering and segmentation using component trees," *Comput. Vis. Image Understand.*, vol. 75, no. 3, pp. 215–228, Sep. 1999.
[25] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 2, pp. 179–187, Feb. 1962.
[26] G. K. Ouzounis and M. H. F. Wilkinson, "Hyperconnected attribute filters based on K-flat zones," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 224–239, Feb. 2011.
[27] A. Baillard, E. Bertin, V. de Lapparent, P. Fouqué, S. Arnouts, Y. Mellier, R. Pelló, J.-F. Leborgne, P. Prugniel, D. Makarov, L. Makarova, H. J. McCracken, A. Bijaoui, and L. Tasca, "The EFIGI catalogue of 4458 nearby galaxies with detailed morphology," *Astron. Astrophys.*, vol. 532, p. A74, Aug. 2011.
[28] A. Venhola, R. Peletier, E. Laurikainen, H. Salo, T. Lisker, E. Iodice, M. Capaccioli, G. V. Kleijn, E. Valentijn, S. Mieske, M. Hilker, C. Wittmann, G. van de Ven, A. Grado, M. Spavone, M. Cantiello, N. Napolitano, M. Paolillo, and J. Falcón-Barroso, "The fornax deep survey with VST: III. low surface brightness dwarfs and ultra diffuse galaxies in the center of the fornax cluster," *Astron. Astrophys.*, vol. 608, p. A142, Dec. 2017.
[29] C. J. Burke, P. D. Aleo, Y.-C. Chen, X. Liu, J. R. Peterson, G. H. Sembroski, and J. Y.-Y. Lin, "Deblending and classifying astronomical sources with mask R-CNN deep learning," *Monthly Notices Roy. Astronomical Soc.*, vol. 490, no. 3, pp. 3952–3965, Dec. 2019.
[30] S. Bruins, M. C. de Jong, K. Heeres, M. H. Wilkinson, M. F. Jonkman, and J. B. van der Meer, "Fluorescence overlay antigen mapping of the epidermal basement membrane zone: I. Geometric errors," *J. Histochemistry Cytochemistry*, vol. 42, no. 4, pp. 555–560, Apr. 1994.

MOHAMMAD HASHEM FAEZI received the M.Sc. degree in computer science from Shahed University, Tehran, Iran. He is currently pursuing the Ph.D. degree in computer science with the University of Groningen, intelligent systems group, specializing in image processing and machine learning. His research interests include advancing image analysis techniques, particularly in the realm of mathematical morphology and hierarchical structures.



REYNIER PELETIER received the Ph.D. degree from the University of Groningen, in 1989. He is currently a Professor with the Kapteyn Astronomical Institute, University of Groningen, specializing in observational astronomy. He had a postdoctoral research position with Harvard University, European Southern Observatory, La Palma Observatory, and the University of Durham. He was an Assistant Professor with the University of Nottingham, U.K., from 1999 to 2003. In 2003, he moved to Groningen, where he was awarded a full professorship, in 2009. His research interests include the formation and evolution of galaxies and their subcomponents using novel quality instrumentation, evolution of dwarf galaxies in clusters, and their detectability in deep astronomical imaging. He was a member of the SAURON Team studying 3D spectroscopy of galaxies. He led the highly successful SUNDIAL Marie Slodowska Curie ITN, which developed several collaborations between astronomers and computing science.

MICHAEL H. F. WILKINSON (Senior Member, IEEE) received the M.Sc. degree in astronomy from the Kapteyn Astronomical Institute, University of Groningen, in 1993, and the Ph.D. degree from the Institute of Mathematics and Computing Science, University of Groningen, Groningen, in 1995. He worked on image analysis of intestinal bacteria with the Department of Medical Microbiology, University of Groningen. He was appointed as a Researcher with the Centre for High Performance Computing, Groningen, working on simulating the intestinal microbial ecosystem on parallel computers. After this, he was a Researcher with the Johann Bernoulli Institute on image analysis of diatoms. He is currently an Associate Professor with the Bernoulli Institute, working on morphological image analysis and especially connected morphology and models of perceptual grouping. An important research focus is on handling giga- and tera-scale images in remote sensing, astronomy, and other digital imaging modalities.