



**HAL**  
open science

## **CGO: Multiband Astronomical Source Detection With Component-Graphs**

Thanh Xuan Nguyen, Giovanni Chierchia, Laurent Najman, Aku Venhola, Caroline Haigh, Reynier Peletier, Michael H.F. Wilkinson, Hugues Talbot, Benjamin Perret

### ► To cite this version:

Thanh Xuan Nguyen, Giovanni Chierchia, Laurent Najman, Aku Venhola, Caroline Haigh, et al.. CGO: Multiband Astronomical Source Detection With Component-Graphs. ICIP 2020 - IEEE International Conference on Image Processing, Oct 2020, Abu Dhabi, United Arab Emirates. pp.16-20, <10.1109/ICIP40778.2020.9191276>. <hal-03132048>

**HAL Id: hal-03132048**

**<https://hal.science/hal-03132048v1>**

Submitted on 4 Feb 2021

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

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



HAL Authorization

# CGO: MULTIBAND ASTRONOMICAL SOURCE DETECTION WITH COMPONENT-GRAPHS

Thanh Xuan Nguyen<sup>1\*</sup>    Giovanni Chierchia<sup>1</sup>    Laurent Najman<sup>1</sup>  
 Aku Venhola<sup>2</sup>    Caroline Haigh<sup>3</sup>    Reynier Peletier<sup>4</sup>    Michael H.F. Wilkinson<sup>3</sup>  
 Hugues Talbot<sup>1,5</sup>    Benjamin Perret<sup>1</sup>

<sup>1</sup> LIGM, Univ Gustave Eiffel, CNRS, ESIEE Paris, F-77454 Marne-la-Vallée, France

<sup>2</sup> Astronomy Research Unit, University of Oulu, Oulu, Finland

<sup>3</sup> Bernoulli Institute, University of Groningen, Groningen, the Netherlands

<sup>4</sup> Kapteyn Institute, University of Groningen, Groningen, the Netherlands

<sup>5</sup> CentraleSupélec, INRIA, Université Paris-Saclay, France

## ABSTRACT

Component-graphs provide powerful and complex structures for multi-band image processing. We propose a multi-band astronomical source detection framework with the component-graphs relying on a new set of component attributes. We propose two modules to differentiate nodes belong to distinct objects and to detect partial object nodes. Experiments demonstrate an improved capacity at detecting faint objects on a multi-band astronomical dataset.

**Index Terms**— Morphology, component-graphs, multi-band image, astronomical source detection

## 1. INTRODUCTION

In the framework of mathematical morphology, component-trees (CT) and component-graphs (CG) are classical structures for image modeling and analysis. The CT (Min-Tree, Max-Tree [1] [2], Tree of Shape [3]) benefit from fast, efficient construction and varied filtering algorithms [4] [5] [6]. However, they are limited to single-band image processing [7]. Extension to multi-band image processing requires a total vectorial order that is arbitrary and application-dependent [8] [9]. On the other hand, the CG is designed to handle multi-band images by relying only on partial orderings. The CG is more powerful at the cost of a higher complexity [10] [11]. This work aims at exploring possibilities of the CG for source detection on multi-band astronomical data.

To detect astronomical sources, SExtractor [12] is the de-facto standard automated tool. MObject/Sourcerer [13] [14] was introduced to extend the SExtractor thresholding strategy by using the Max-Tree. MObject/Sourcerer has already shown its robustness. It has far fewer parameters than other faint source detection methods. However, both methods focus on single-band processing while multi-band information

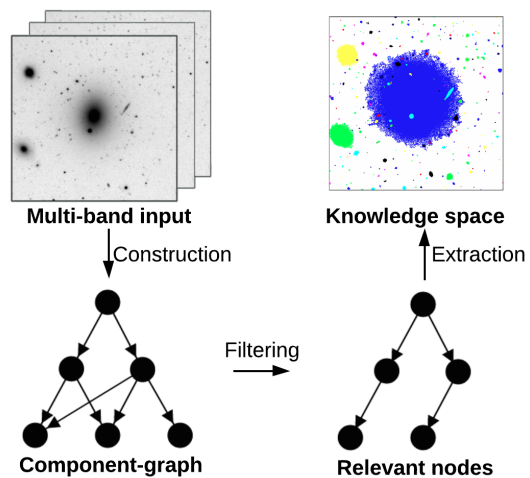
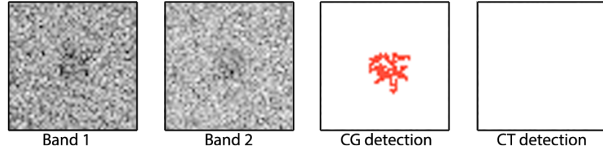


Fig. 1: General filtering framework using the CG.

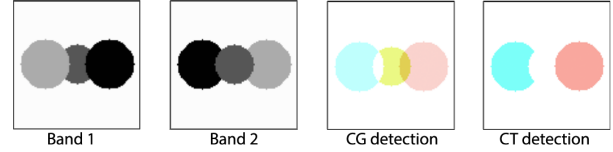
is available. Relying on the CG, the goal of this work is to extend astronomical source detection to the context of multi-band processing regardless the wavelengths. We thus improve detection sensitivity and deblending capacity. The proposed framework is illustrated in Fig. 1. Intuitively, Fig. 2a shows how the CG can improve the detection sensitivity by leveraging the multi-band information, whereas Fig. 2b shows how the richness of the CG structure can enhance the deblending capacity. Apart from these advantages, Fig. 3 reveals that the CG is no longer a tree, but a directed acyclic graph (DAG), which is significantly more challenging to filter than the classical CT structures [15].

**Contribution** Motivated by astronomical applications, we propose a CG-based detection and segmentation framework for multi-band images, named *CGO*. After some preliminary definitions in Sec. 2, we introduce in Sec. 3 a new set of multi-band component attributes, and two proposals for node differentiation and partial object detection. Experimental results in Sec. 4 show that our method is able to detect

\*Thanks to the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 721463 to the SUNDIAL ITN network.

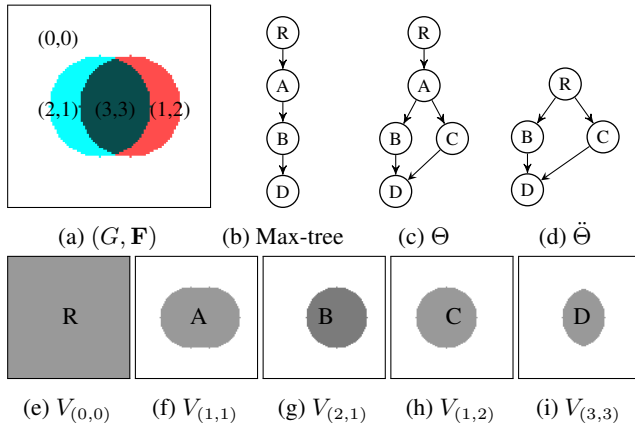


(a) A two-band single object input with Gaussian noises and its detected source in the CG and the CT. At the same level of confidence, the CT in separate bands could not capture the source. It shows that the CG is able to detect faint sources, i.e., sources at low signal-to-noise ratios.



(b) A two-band input containing three overlapping circles. Middle circle appears in the CG as an isolated component while it is merged with adjacent regions in the CT of separate bands: this explains the missed circle in the CT detection. It depicts that the CG is richer than the CT for capturing input structures.

**Fig. 2:** Illustrations of the CG features for (a) detection sensitivity and (b) deblending capacity.



**Fig. 3:** (a) A two-band image, viewed as a graph  $(G, \mathbf{F})$  valued on  $\mathbb{V} \subseteq \mathbb{N}^2$  equipped with the marginal order relation  $\leq_m$ ; (b) The Max-tree of the first band; (c) The CG  $\Theta$ ; (d) the CG  $\tilde{\Theta}$ ; and (e-i) Threshold sets  $V_v$  for  $v \in \mathbb{V}$ . The letters (R, A–D) refer to the connected components correspond to the nodes in the tree and graphs.

faint sources on simulated and real multi-band optical images (KIDS bands  $u, g, r, i$  [16]), with significantly better precision and recall scores than a state-of-the-art method [13].

## 2. COMPONENT-GRAPHS

We recall necessary notions on graphs and component-graphs. Fig. 3 shows an example of two CG variants and a Max-tree constructed from a two-band input.

**Connected Component** Given a graph  $G = (V, E)$ , where  $V$  is a finite set and  $E \subseteq \{(x, y) | x, y \in V \wedge x \neq y\}$ , we say that a sequence  $(x_0, \dots, x_n) \in V$  is a *path* in  $V$  from  $x_0$  to  $x_n$  if  $(x_{i-1}, x_i) \in E, \forall i \in \{1, \dots, n\}$ . A subset  $V' \subseteq V$  is said to be connected if for any two distinct elements  $x, y \in V'$ , there exists a path from  $x$  to  $y$ . *Connected components* of  $G$  are the maximal connected subsets of  $V$ . The set of all connected components of  $G$ , denoted as  $C[G]$ , is a partition of  $V$ .

**Valued Graph** Let  $\mathbf{F} : V \rightarrow \mathbb{V}$  be a function from  $V$  to a nonempty set  $\mathbb{V}$  which is equipped with an order relation  $\leq$ . We say that  $(G, \mathbf{F})$  is a vertex-valued graph (or valued graph).

**Component-Graph** Given a valued graph  $(G, \mathbf{F})$ , we define the threshold set  $V_v = \{x \in V | \mathbf{F}(x) \geq v\}$ , where  $v \in \mathbb{V}$ . The threshold set  $V_v$  induces a subset  $E_v = \{(x, y) \in E | x, y \in V_v\}$  and a sub-graph  $G_v = (V_v, E_v)$ . The set of connected components of the sub-graphs  $G_v$  of  $G$  is denoted as  $\Psi = \bigcup_{v \in \mathbb{V}} C[G_v]$ .

If  $(\mathbb{V}, \leq)$  is totally ordered, the Hasse diagram of the partially ordered set  $(\Psi, \subseteq)$  forms a Max-tree (see Fig. 3b) of the valued graph  $(G, \mathbf{F})$ . If  $(\mathbb{V}, \leq)$  is partially ordered, the Hasse diagram of the partially ordered set  $(\Psi, \subseteq)$  forms a CG  $\Theta$  of the valued graph  $(G, \mathbf{F})$  [17] (see Fig. 3c). This research works on a simplified version of the CG, denoted as  $\tilde{\Theta}$  (see Fig. 3d), where its set of connected components  $\tilde{\Psi} = \{X \in \Psi | \bigcup_{Y \in \{Z \in \Psi | Z \subsetneq X\}} Y \neq X\}$  contains only connected components that contribute to the visibility of the image  $\mathbf{F}$  [10]. The CG  $\Theta$  and  $\tilde{\Theta}$  are DAGs. In the remaining of this paper, we always use the simplified CG  $\tilde{\Theta}$ .

The *parents* of a node  $N$  are the smallest nodes of  $\tilde{\Theta}$  greater than the node  $N$ :  $\text{parents}(N) = \min\{X \in \tilde{\Theta} | N \subsetneq X\}$ . The *significance*  $\text{sn}(N)$  of a node  $N$  of  $\tilde{\Theta}$  is a predicate saying whether the node  $N$  is relevant for filtering, its design is upon applications. The *closest significant ancestors* of a node  $N$  are its significant ancestors having maximal levels

$$\text{sn}_{\text{anc}}(N) = \underset{X \in \{Y \in \tilde{\Theta} | N \subsetneq Y \text{ and } \text{sn}(Y)\}}{\text{argmax}} \quad \mathbf{L}(X), \quad (1)$$

where level  $\mathbf{L}(N) = \min\{\mathbf{F}(x), x \in N\}$  is the minimal pixel value in  $N$ . Because of the partial order  $\leq$ , a node  $N \in \tilde{\Theta}$  may have several parents and several closest significant ancestors.

## 3. FILTERING THE COMPONENT-GRAPH

We introduce CGO, a framework to handle multi-band source detection with the CG. Given the CG  $\tilde{\Theta}$  with node attributes introduced in Sec. 3.1, filtering is described in Algorithm 1.

First, two types of relevant nodes are extracted during a root to leave order traversal of the CG: significant nodes without significant ancestors that are associated to disjoint objects; and significant nodes different from its closest significant ancestors that are associated to extended-objects lying on top of their parents. The node differentiation is based on the dupli-

---

**Algorithm 1:** Filtering the component-graph  $\ddot{\Theta}$ .

---

**Input :**  $\ddot{\Theta}$ , a component-graph.  
**Output:** objs, a list of object nodes.  
 /\* Filter duplication from root to leave \*/  
 1 **foreach** significant node  $N \in \ddot{\Theta}$  **do**  
 2 | **if**  $\text{sn}_{\text{anc}}(N) = \emptyset$  or  $d(N, X) \forall X \in \text{sn}_{\text{anc}}(N)$   
 |     **then**  $\text{objs} \leftarrow \text{objs} \cup N$  ;  
 /\* Filter partial objects \*/  
 3 **foreach** node  $N \in \text{objs}$  **do**  
 4 | **if**  $\text{partial}(N, \ddot{\Theta})$  **then**  $\text{objs} \leftarrow \text{objs} \setminus \{N\}$  ;  
 5 **return** objs;

---

cated object assumption, described in Sec. 3.2.

Second, we check whether relevant nodes are parts of the same object based on partial object detection, described in Sec. 3.3. Due to partial ordering in the  $\ddot{\Theta}$ , an object can appear as several isolated components. Partial object detection is designed to deal with this issue on the fly. The second step guarantees that each detected node is only associated to a single object. In the end, nodes are marked as objects.

### 3.1. Generalized Significance Test

We extend the idea of MTOBJECT significance test in single-band [13] into the multi-band context with respect to a chi-square distribution of normalized node power and node area. **Node power** is the sum of the squared difference between node pixel values and the level of the parents. Since a node in the CG  $\ddot{\Theta}$  may have several parents, this definition use the supremum (average, infimum, max area node are other possibilities) of parent levels as reference.

$$\mathcal{E}(N) = \sum_{x \in N} \left( \mathbf{F}(x) - \bigvee_{y \in \text{parents}(N)} \mathbf{L}(y) \right)^{\circ 2}, \quad (2)$$

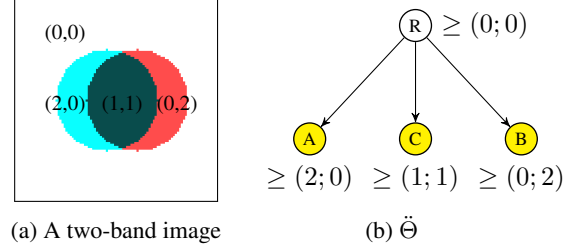
where  $\bigvee$  is supremum operator and  $\circ$  is element-wise power. **Node normalized power** normalizes the node power by local background variance.

$$\mathcal{E}'(N) = \mathcal{E}(N) \oslash (\hat{\sigma}_{\text{bg}}^2 + \mathbf{L}(\text{parents}(N))) \oslash \text{gain}, \quad (3)$$

where  $\text{gain} = (\hat{\mu}_{\text{bg}} - \bigwedge_{x \in V} \mathbf{F}(x)) \oslash \hat{\sigma}_{\text{bg}}^2$  refers to the CCD gain in astronomy;  $\oslash$  is element-wise division; and  $\hat{\mu}_{\text{bg}}, \hat{\sigma}_{\text{bg}} \in \mathbb{R}^n$  stand for mean and standard deviation of background of image  $\mathbf{F}$ .

**Node significance** relies on hypothesis testing. For multi-band images, we propose a significance test combining separate bands and synthesized band, formalized as

$$\text{sn}(N) = \left( \exists i \in [0, n] \mathcal{E}'(N)_i > \text{cdf}\chi^2(\alpha, a(N)) \right) \text{ or } \left( \sum_{i=0}^n \mathcal{E}'(N)_i > \text{cdf}\chi^2(\alpha, n a(N)) \right), \quad (4)$$



**Fig. 4:** (a) A two-band input and (b) its CG  $\ddot{\Theta}$ , where yellow nodes are significant. Considering the first band, node  $A$  and  $C$  should be considered as two parts of a single object, but they are isolated because of disagreement in the two-band space. Similarly for node  $B$  and  $C$  in the second band. Partial Object Detection is designed to mark  $C$  as a non-object node by checking whether  $C$  merges with its adjacent siblings.

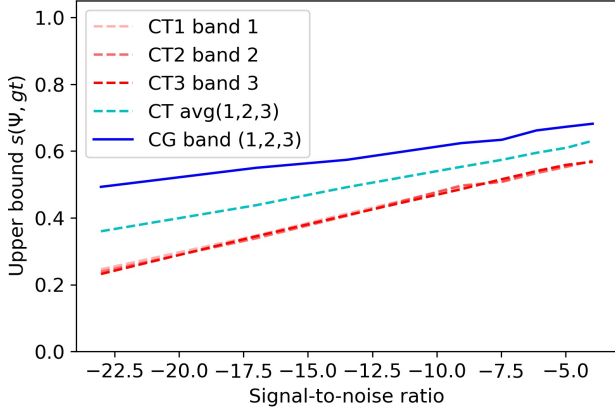
where  $n$  is number of band;  $\text{cdf}\chi^2()$  is the chi-square cumulative distribution function;  $\alpha$  is a significance level; and  $a(N) = |N|$  is the area of the node  $N$ .

### 3.2. Duplicated Object Assumption

In the CT and the CG, objects appear differently at different thresholding levels as sequences of significant nodes. For instance, the object in Fig. 3a is represented by one branch ( $A \rightarrow B \rightarrow D$ ) in the Max-tree (see Fig. 3b), or two branches ( $B \rightarrow D$ ) and ( $C \rightarrow D$ ) in the  $\ddot{\Theta}$  (see Fig. 3d). The key of object filtering on these hierarchies is to differentiate nodes belonging to distinct targets. In the context of the CT with a total order, the assumption of main branch (defined as the significant descendant of a node with the largest area) was used in MTOBJECT [13], where a node and its main branch reside in the same object. Then finding objects is finding their sequences of main branches. However, in the context of the CG with partial orders, there may exist several branches containing incomparable nodes belonging to a single object. So as to identify those nodes, we introduce a new assumption: branches corresponding to an identical object eventually grow to similar centers. We define a predicate saying whether two nodes belong to the same object as  $d(N_1, N_2) = \|\text{ml}(N_1) - \text{ml}(N_2)\| < r$ , where node  $N_1, N_2 \in \ddot{\Theta}$ , function  $\text{ml}(N_i)$  returns center pixel, precisely, the brightest pixel of the node  $N_i$ , and  $r$  is a thresholding radius. The center pixel can be possibly designed as the central mass or the center of the best eclipse fitting of the node covering region.

### 3.3. Partial Object Detection

In the CG, significant adjacent components may be incomparable, i.e., orders in separate bands disagree. Those components will be captured as isolated objects whereas they may belong to the same object in separate bands. An example is shown in Fig. 4, where multiple significant adjacent components are incomparable in a two-band input, but they are not isolated objects.



**Fig. 5:** Detection upper bounds of the CT and the CG.

To determine whether a significant node  $N$  is an object or a part of an existing object, its adjacent nodes  $\mathcal{ADJ}(N)$  have to be checked, but  $\mathcal{ADJ}(N)$  is costly to query. In this work, we pay attention to the adjacent sibling set which is a subset of  $\mathcal{ADJ}(N)$  but is available in the CG in constant time. Precisely, a significant node  $N$  is said to be part of another object in two possibilities: 1. If it exists an adjacent sibling belonging to an object, then  $N$  falls into the same object with the sibling; 2. If the union of  $N$  and the adjacent sibling forms a non-significant node, then  $N$  is a non-object node.

#### 4. EXPERIMENTS

We analyse CT and CG detection capacities on simulations and compare the CGO versus MTOBJECT on synthetic data.

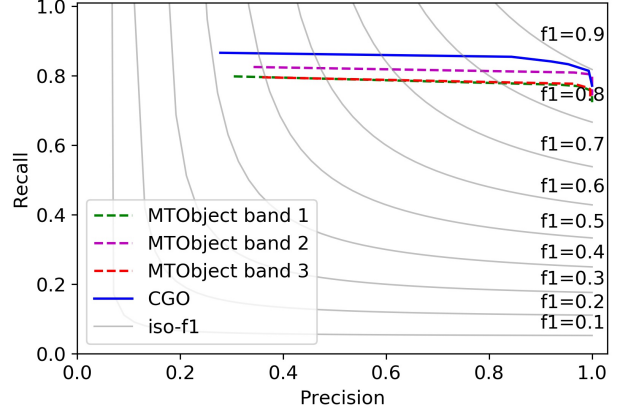
##### 4.1. Upper Bound Detection Capacity of CT and CG

In order to detect target objects, their morphological representation must capture them as components. We see how well objects are preserved in the CT and the CG by studying their component similarity upper bounds on synthetic data. For a set of components  $\Psi$  and a ground-truth  $gt$ , the upper bound similarity  $s(\Psi, gt)$  is equal to  $\max_{N \in \Psi} J(N, gt)$ , where  $J$  stands for the Jaccard distance between two components.

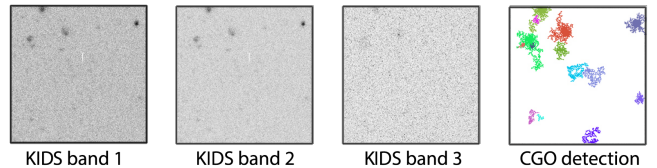
We simulate  $10^4$  three-band images each with a single source with Gaussian noises. Ground-truth is defined as the region covering 99% of the object brightness. The CT on separate bands, average band, and the CG on the three-band image are constructed. We analyse the similarity upper bound of each structure on the dataset, illustrated in Fig. 5. On this synthetic data, CG provides higher similarity upper bounds than CT, i.e., the CG is more likely to detect and segment target objects properly.

##### 4.2. Proposed Astronomical Source Detection Results

**Dataset** We use a three-band optical simulation with ground-truth [18], which imitates the Fornax Deep Survey. It contains 1500 stars as point sources, 4000 background galaxies and 50



**Fig. 6:** Precision-recall of CGO versus MTOBJECT.



**Fig. 7:** A KIDS image ( $u, g, r$ ) bands [16] and CGO detection.

background clusters. Because the CG construction is time consuming ( $\mathcal{O}(n^2)$ ), we slice the simulation into tiles (size (500, 500) pixels, overlapping 250 pixels).

**Metric** For evaluation, we use precision, recall, and F1-score. In order to match between detection map and target map, each target object is represented by its brightest pixel, as a consequence, each representative pixel is included in either background or one object in the detection map. In case of one detected object accommodates multiple representative pixels, the brightest will be chosen.

**Quantitative results** We have compared our CGO versus the state-of-the-art MTOBJECT [13], the only parameter for both is the significance level  $\alpha$ . Since the signal in border regions is less reliable, we do not take into account object centers lying within 100 pixels from the borders. Precision and recall curves for each method are shown in Fig. 6. It is clear that the proposed method improves on MTOBJECT at both precision and recall metrics. An illustration on a real astronomical KIDS image [16] is shown in Fig. 7.

#### 5. CONCLUSION

We have investigated how the CG structure can handle source detection on multi-band data. From the theoretical point of view, our studies have shown that the CG are better at preserving object structures comparing to the classical CT. Furthermore, we proposed CGO – a detection framework and a set of new component attributes on the CG to detect astronomical sources. Practical results show that CGO outperforms existing methods on the multi-band astronomical dataset at both precision and recall metrics.

## 6. REFERENCES

- [1] Philippe Salembier, Albert Oliveras, and Luis Garrido, “Antiextensive connected operators for image and sequence processing,” *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570, 1998.
- [2] Edmond J Breen and Ronald Jones, “Attribute openings, thinnings, and granulometries,” *Computer Vision and Image Understanding*, vol. 64, no. 3, pp. 377–389, 1996.
- [3] Pascal Monasse and Frederic Guichard, “Fast computation of a contrast-invariant image representation,” *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 860–872, 2000.
- [4] Christophe Berger, Thierry Géraud, Roland Levillain, Nicolas Widynski, Anthony Baillard, and Emmanuel Bertin, “Effective component tree computation with application to pattern recognition in astronomical imaging,” in *2007 IEEE International Conference on Image Processing*. IEEE, 2007, vol. 4, pp. IV–41.
- [5] Laurent Najman and Michel Couprie, “Building the component tree in quasi-linear time,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3531–3539, 2006.
- [6] Thierry Géraud, Edwin Carlinet, Sébastien Crozet, and Laurent Najman, “A quasi-linear algorithm to compute the tree of shapes of nd images,” in *International symposium on mathematical morphology and its applications to signal and image processing*. Springer, 2013, pp. 98–110.
- [7] Benoît Naegel, Nicolas Passat, Nicolas Boch, and Michel Kocher, “Segmentation using vector-attribute filters: methodology and application to dermatological imaging,” 2007.
- [8] Benjamin Perret, Sébastien Lefevre, Christophe Collet, and Éric Slezak, “Connected component trees for multivariate image processing and applications in astronomy,” in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 4089–4092.
- [9] Benjamin Perret and Ch Collet, “Connected image processing with multivariate attributes: An unsupervised markovian classification approach,” *Computer Vision and Image Understanding*, vol. 133, pp. 1–14, 2015.
- [10] Benoît Naegel and Nicolas Passat, “Colour image filtering with component-graphs,” in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 1621–1626.
- [11] Nicolas Passat, Benoit Naegel, and Camille Kurtz, “Component-graph construction,” *Journal of Mathematical Imaging and Vision*, pp. 1–26, 2019.
- [12] Emmanuel Bertin and Stephane Arnouts, “Sextractor: Software for source extraction,” *Astronomy and Astrophysics Supplement Series*, vol. 117, no. 2, pp. 393–404, 1996.
- [13] Paul Teeninga, Ugo Moschini, Scott C Trager, and Michael HF Wilkinson, “Statistical attribute filtering to detect faint extended astronomical sources,” *Mathematical Morphology-Theory and Applications*, vol. 1, no. 1, 2016.
- [14] Michael H. F. Wilkinson, Caroline Haigh, Simon Gagnes, Paul Teeninga, Nushkia Chamba, Thanh Xuan Nguyen, Hugues Talbot, Laurent Najman, Benjamin Perret, Giovanni Chierchia, Aku Venhola, and Reynier Peletier, “Sourcerer: A robust, multi-scale source extraction tool suitable for faint and diffuse objects,” in *The Realm of the Low-Surface-Brightness Universe Proceedings IAU Symposium*. 2019, number 355, International Astronomical Union.
- [15] Eloïse Grossiord, Benoît Naegel, Hugues Talbot, Laurent Najman, and Nicolas Passat, “Shape-based analysis on component-graphs for multivalued image processing,” *Mathematical Morphology-Theory and Applications*, vol. 3, no. 1, pp. 45–70, 2019.
- [16] Jelte TA de Jong, Gijs A Verdoes Kleijn, Konrad H Kuijken, Edwin A Valentijn, et al., “The kilo-degree survey,” *Experimental Astronomy*, vol. 35, no. 1-2, pp. 25–44, 2013.
- [17] Nicolas Passat and Benoît Naegel, “Component-trees and multivalued images: Structural properties,” *Journal of Mathematical Imaging and Vision*, vol. 49, no. 1, pp. 37–50, 2014.
- [18] Aku Venhola, *Evolution of dwarf galaxies in the Fornax cluster*, Ph.D. thesis, University of Groningen, 2019.