Efficient morphological tools for astronomical image processing
Moschini, Ugo

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

Introduction

When I was a child, I began to look at the night sky with my dad. He bought a pair of binoculars and a telescope, a really good one. In spite of having all that great equipment, the sentence I was frequently uttering to my father while browsing the sky seeking a particular galaxy was “I do not see it...” or “Which one is it?!”. Many of the nebulae, galaxies or star clusters are in fact very faint. Whenever I was able to spot a distant galaxy with feeble structures in it, I truly had the feeling of staring at something interesting, mysterious and fascinating. Or most likely, at my young age, I was considering that just cool. Apart from the objects that are difficult to spot, the field of view, even while looking through an amateur telescope, is literally crowded with stars: I could easily get lost and I did not know exactly where to look to find the desired object.

Years later, it turns out that such early experiences of mine readily translate into issues present in modern astronomy. In astronomical image analysis, a problem that arose in recent years is related to the huge amount of data collected by modern instrumentation. A vast quantity of data refers to images of nearby and distant galaxies and stars, and to remote sensing images of planet Earth, taken by satellites in orbit. With the advances of technology, images of resolution equal to a few gigapixels are not exceptional. Not only the image resolution is high, but also the information carried by each pixel has often a high bit-depth. An example are the radio surveys that study radio emission from distant galaxies. In general, both 2D and 3D images contain hundreds of objects (sources of light, in astronomers’ jargon): their extraction in an automatic way is essential. The Sloan Digital Sky Survey (SDSS) Data Release 7 (Abazajian et al., 2009), used throughout the chapters in this thesis, contains 357 million unique objects. Manual extraction of objects is not feasible. In astronomy, objects of interest are normally stars and galaxies. Stars have a point-like, compact structure while galaxies (or also nebulae, star clusters, ...) show more complex, diffused and extended structures. Their shape varies greatly, according to their type. Usually, objects are nested on top of each other: stars on top of galaxies, star clusters on top of a diffused halo, and so on. Often, precisely the faint objects, that I found cool as a child, are of great importance to astronomers to advance their knowledge of the Universe. For example, the evolution and morphology of galaxies are studied by analysing the faint signs of ongoing and past interactions. Faint and...
extended sources are often difficult to identify with state-of-the-art software. One of the most famous applications used for astronomical object detection is Source Extractor (Bertin and Arnouts, 1996). It became very popular due to its robustness to several datasets and its ability to deal well with large images (Starck and Murtaugh, 2002). Object detection is not the end of the story. Astronomers are also interested in understanding what the detected sources are, how they are related to each other and ultimately to give some sort of characterization or classification.

Among the several techniques available to perform image analysis and image processing, Mathematical Morphology (Serra, 1982; Matheron, 1975) represents a whole field of research based on set theory, lattice algebra and topology. It provides a powerful and diverse set of operators applicable to several fields. To name a few, application areas include medical image processing (Metzler et al., 2002; Ouzounis et al., 2009; Kiwanuka, 2013), image compression and simplification (Salembier et al., 1996; Tushabe and Wilkinson, 2007; Maragos and Evangelopoulos, 2007; Soille, 2008), industrial applications and material sciences (Beucher, 1995) and remote sensing imagery (Soille and Pesaresi, 2002; Benediktsson et al., 2011; Ouzounis and Soille, 2012). The use of Mathematical Morphology in the specific domain of astronomical object recognition is still limited. However, in the detailed review paper of Masias et al. (2012) on source detection approaches in astronomical images, references to methods that employ morphological solutions (Aptoula et al., 2006; Berger et al., 2007; Perret et al., 2009; Yang et al., 2008) are reported. In the last years, there is surely a growing interest in both the astronomy and image processing community to explore further applications and methodologies in this field, as in Perret et al. (2010, 2012); Ouzounis and Wilkinson (2011). The same goes for the use of morphological tools in classification tasks: pattern spectra (Maragos, 1989) are employed for pattern recognition and information mining tasks. In astronomy, they have been already used in distinguishing stars from galaxies (Candeas et al., 1997) and identifying different types of galaxies (Starkenburg et al., 2009).

1.1 Scope

In this thesis, I will be looking at ways to use the Mathematical Morphology framework to perform image segmentation and classification of objects in astronomical datasets. In connective morphology, all image operations are performed not on the pixels as isolated instances, but on connected components. In standard connectivity (Serra, 1998), a connected component is a set of pixels that are connected through a path (usually defined by 4 or 8 pixel-adjacency relations) and have the same intensity values. Operators that work at the level of connected components are called connected filters (Salembier and Serra, 1995; Breen and Jones, 1996). They are shape-preserving filters that operate by removing (or retaining) certain components
1.1. Scope

Figure 1.1: An example of connected filtering: (a) original image; (b) connected components whose shape is eccentric are preserved and stars are removed.

according to a given criterion. They can be used both for filtering or segmentation, simply labelling a component with an object identifier rather than deleting it from the image. They were extended and generalised to attribute filters (Breen and Jones, 1996; Salembier et al., 1998), where several kinds of criteria could be used. Within attribute filters, shape filters (Urbach and Wilkinson, 2002) have the distinctive feature of being shape-sensitive and scale-invariant. They are used in many fields, with important applications in medical image processing for the extraction of vessels, like in the recent work of Kiwanuka (2013). Fig. 1.1 shows an example of connected filtering. It shows the original and the output image yielded after applying a connected filter that preserves components whose shape is eccentric: the elongated galaxies are preserved and the most circular structures, such as stars crowding the field of view and some object halos, are removed. Efficient ways to perform connected filtering and object segmentation applied to astronomical image processing will be presented in this thesis. Different kinds of connectivities will be discussed, leading to different operators. Apart from standard connectivity, both second-generation connectivity (Serra, 1998; Braga-Neto and Goutsias, 2002; Ouzounis and Wilkinson, 2007a) and hyperconnectivity (Serra, 1988; Braga-Neto and Goutsias, 2003b; Wilkinson, 2011b) will be explored, in different contexts, from filtering to classification. More specifically, this thesis can be divided in the following four research topics.

Big data Given the high resolution and the dynamic range of astronomical and remote sensing images, the need of parallelizing the construction of the max-tree (Salemberi et al., 1998) structure is fundamental. The max-tree is a tree structure that represents the image connected components in a hierarchical way. Efficient implementation of connected filtering and segmentation is done on the nodes of such trees. State-of-the-art parallel algorithms (Wilkinson et al., 2008) do not deal well
1. Introduction

with bit depths higher than 16 bits per pixel and with floating point data (Carlinet and Géraud, 2013), often present in astronomical datasets. As a solution to this problem, a novel max-tree parallel building algorithm is proposed that works effectively also on double-precision floating point 2D images and 3D volumes.

Object segmentation At first, to preserve the faint structures in which astronomers are interested, a filtering method based on second-generation and mask-based connectivity (Ouzounis and Wilkinson, 2007a) is explored. The connectivity among pixels is defined via morphological viscous closings (Vachier and Meyer, 2005). Such operators allow for clustering together the faint, dark, structures close to the noise level, usually broken up in countless, tiny connected components, and at the same time keeping the bright, close-by structures (stars) separate.

Even if faint structures could be better clustered together as a single component, the problem of performing the actual object segmentation remained. The morphological approach is then combined with the statistical knowledge of the data sets used and tested on the Sloan Digital Sky Survey catalogue. Second-generation and mask-based clustering connectivities are abandoned in favour of designing a connected filter based on the statistical characteristics of the noise present in the images. Such statistical attribute filter identifies the sets of connected components that do not belong to the background (noise) and represent instead actual objects.

A collaboration with the group of astronomers led by professor van der Hulst, at the Kapteyn Astronomical Institute was started to study high-resolution floating point 3D volumes containing radio emission from galaxies. Within this project, both the novel parallel max-tree algorithm and the statistical attribute filter are used together and adapted to the features of the noise in the radio volumes. The results are compared with the methods currently used by astronomers (Serra et al., 2015).

Classification The use of pattern spectra for classification purposes is studied, as a follow-up from previous experiments (Starkenburg et al., 2009) carried out at the group of Intelligent Systems at the University of Groningen. The goal is to be able to classify merging and projected galaxies in the Sloan dataset. The former are galaxies showing an actual interaction with a companion galaxy, usually in the form of faint filamentous structures; the latter are galaxies that just look close to each other, due to a projection effect on the sky. A parallel method to compute pattern spectra on the whole image is adapted to the computation of local pattern spectra, one for every object identified. Two-dimensional pattern spectra of every segmented object, with area and image moment invariants on the two dimensions are used as feature vectors and fed to a C4.5 tree classifier, giving good classification performances.
Hyperconnected filtering In the last part of the thesis, a new algorithm is proposed to perform filtering of viscous-hyperconnected components. A particular hyperconnectivity class that shows similarities with the theory of viscous lattices in (Serra 2002) is analysed: it prevents image structures not related to each other from becoming connected due to noise or other undesired structures. Such problem is known as leakage and it arises frequently in standard connectivity. While this last topic is not directly connected to the analysis of astronomical images, it could open up different ways of performing object segmentation or, more in general, image filtering. It should be subject to further research, investigating all the possible application fields, including astronomy.

1.2 Thesis organization

The remaining chapters of this thesis are all journal or peer-reviewed conference papers published or under review. Some of the concepts might be repeated in multiple chapters, to make each chapter as self-contained as possible.

In Chapter 2, images generated through viscous mathematical morphology operators (Vachier and Meyer 2005) are evaluated as they are used as connectivity mask images in mask-based connectivity. The aim is to keep neighbouring stars separate, while clustering the speckles or grains that faint galactic structures are made of. A new second-generation connectivity is introduced, together with the concept of connectivity mask. Experiments varying the parameters used to generate the mask images are presented and discussed. The actual morphology of the objects is preserved and object segmentation techniques would benefit from that.

In Chapter 3, an automatic method to extract astronomical objects is proposed. It is based on attribute filtering on connected components, organised in a max-tree structure, that is used throughout the rest of the thesis. Connected attribute filtering is briefly introduced. It is shown that Source Extractor (Bertin and Arnouts 1996), a widely used state-of-the-art software, often fails to detect the faint extended sources. A filter based on the distribution of an attribute with respect to its distribution in the case of noise is proposed. Components are either considered parts of an object or noise. Object detection is tested on 2D optical images extracted from the Sloan Digital Sky Survey Data Release 7.

Chapter 4 presents a new parallel algorithm to build max-trees of high resolution and very high-dynamic range images. This novel parallel solution combines root-to-leaf flooding and leaf-to-root merging max-tree algorithms in a two-step algorithm, based on two methods in (Berger et al. 2007) and in (Wilkinson et al. 2008). State-of-the-art sequential and parallel algorithms are reported and their performance discussed. The proposed parallel algorithm is the first one capable of supporting efficiently bit-depths larger than 16 bits per pixel. An analysis of the algorithm and its
performance on real-world and simulated data is presented, including 2D images and 3D volumes.

In Chapter 5, the parallel max-tree algorithm presented in Chapter 4 is combined with the statistical attribute filtering proposed in Chapter 3 to perform object segmentation on high-resolution 3D radio volumes. Such cubes contain measurements of the emission of modelled galaxies, as they would be observed by the WSRT (Westerbork Synthesis Radio Telescope). The cubes show velocities versus projected spatial dimensions on the sky. The noise model used by the statistical filter is adapted to the kind of noise present in the cubes. The first results of this segmentation approach are limited but promising.

Chapter 6 and Chapter 7 introduce the concept of pattern spectra and its application to remote sensing imagery and galaxy classification, respectively. In particular, Chapter 6 is based on a first exploratory piece of work and it was intended for a workshop with “work in progress” type presentations. However, the concepts presented there are useful for a better understanding of Chapter 7. Pattern spectra are histograms that quantify the amount of information that is present in the image structures that satisfy a given range of size or shape attributes. The max-tree structure is used to compute pattern spectra efficiently. A pattern spectrum for every segmented object is built in parallel. Such pattern spectra are named local, to differentiate them from pattern spectra computed instead on the whole image, more common in literature. Galaxies are segmented with the algorithm proposed in Chapter 3. Size and Hu’s and Flusser’s image invariant moments (Hu, 1962; Flusser, 2000) are computed and a pattern spectrum is calculated for every galaxy. The goal of the classification is to distinguish mergers from projected galaxies. The C4.5 tree classifier (Quinlan, 1993) with bagging gives the best classification result. Mergers and projected galaxies are classified with a precision of about 80%.

In Chapter 8, the concept of hyperconnectivity is explained and a hyperconnectivity class that tries to address the leakage problem typical of connected filters is described. A novel algorithm is provided to perform attribute filtering on a tree of viscous-hyperconnected components. The algorithm works in two steps: first, the cores of the hyperconnected components are identified and their max-tree is built. Such max-tree is then updated with the contribution coming from the other pixels that belong to more than one component.

Finally, the last chapter provides a summary of the findings present in this thesis and further developments are proposed.