ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/269318854

Automatic detection and automatic classification of structures in astronomical images

Conference Paper *in* Proceedings of SPIE - The International Society for Optical Engineering · July 2014

Impact Factor: 0.2 · DOI: 10.1117/12.2055579

READS

12

6 authors, including:

Mauricio Solar

Universidad Técnica Federico Santa María

44 PUBLICATIONS **196** CITATIONS

SEE PROFILE

Diego Mardones

University of Chile

96 PUBLICATIONS **1,711** CITATIONS

SEE PROFILE

Karim Pichara

Pontifical Catholic University of Chile

18 PUBLICATIONS **29** CITATIONS

SEE PROFILE

Victor Parada

University of Santiago, Chile

50 PUBLICATIONS **388** CITATIONS

SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Mauricio Solar

Retrieved on: 07 May 2016

Automatic detection and automatic classification of structures in astronomical images

Rodrigo Gregorio[a], Mauricio Solar[a], Diego Mardones[b], Karim Pichara[c], Victor Parada[d], Ricardo Contreras[e]

[a]Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile;
[b]Universidad de Chile (UCh), Santiago, Chile;
[c]Pontificia Universidad Católica (PUC), Santiago, Chile;
[d]Universidad de Santiago de Chile (USACH), Santiago, Chile;
[e]Universidad de Concepción (UdeC), Concepción, Chile

ABSTRACT

The study of the astronomical structures is important to the astronomical community because it can help identify objects, which can be classified based on their internal structure and their relation to other objects. For this reason, we are developing an automated tool to analyze astronomical images into its components. First, the 2D images will be decomposed into different spatial scales based on wavelet transforms. Then we will implement detection algorithms to each spatial scale, such as Clumpfind, Gaussclump, or Dendrogram techniques. The goal is to build a tool that is available to the community and satisfy the requirements of the next Chilean virtual observatory (ChiVO).

Keywords: ALMA, Astronomical Images, Detection, Classification

1. INTRODUCTION

Chile has huge potential for astronomical observation, because it is the home of the majority of the most important observatories in the world, and the favorite place for further projects planned to be constructed during the next decade. With the clearest skies in the world, Chile is becoming an astronomical power. The currently largest astronomical project is located on the Chajnantor plateau in the Atacama desert: The Atacama Large Millimeter/Submillimeter Array (ALMA). ALMA began scientific observations in the second half of 2011 and the first images were released to the community on 3 October 2011.

ALMA will provide a large amount of data. This information will open up new windows to study the cold Universe. This will allow us to study the presence of physical structures over large spatial scales, with different chemical properties. For example, with this data we may simultaneously characterize the properties and spatial distribution of star clusters, or of spiral arms, or of bulge and disk in nearby spiral galaxies. Alternatively, we could study the spatial distribution and chemical properties of protostellar disks with their warps or spiral structures and bipolar winds, simultaneously with the dense gas cores, and filaments, within giant molecular clouds.

It propose to develop a tool for the detection and classification of structures at different spatial scales in astronomical images. Firstly, we will use filters from 2D wavelet transforms to generate sub-images at different scales. The tool will be focused initially to 2D images or a series of 3D images planes. Then, it will apply detection algorithms in each sub-image, such as Clumpfind, Gaussclump, Dendrograms, CSAR, SExtractor, generating an object catalog for each image input. The catalog includes an initial classification of the type of object or structure found and allows to study statistics and spatial relationships between them.

This proposal includes joint work between astronomy and computer engineering, and provides support for scientific research and development of new applications in the field of astro-informatic.

Further author information: (Send correspondence to Rodrigo Gregorio)
Rodrigo Gregorio: E-mail: rgregori@alumnos.inf.utfsm.cl, Telephone: +56 9 56393706

Section 2 gives a brief description of the problem. Section 3 is focused on exploring the current state of detection algorithm. Section 4 describes the proposed algorithm. Finally, section 5 concludes what has been achieved with this proposal.

2. PROBLEM DESCRIPTION

The internal structures of a molecular cloud observed in spectral lines are more easily visible in contour maps.[1] Some existing algorithm to find clumps are based on this approach (Clumpfind), each structure has a maximum peak and for determinate the amount of clumps is necessary only count the number of peak. Therefore, a problem to see is if this approach can be performed in other type of regions.

On the other hand, the existing algorithms have been developed for images within the visible spectrum. in ALMA the range of frequencies is extended and with this new data arises the problem of applying existing algorithms on them.

Existing algorithms include different types of approaches that given good results. Then, it appear the question about is possible to obtain better results if it mixes these approaches or when integrates these algorithms it can be find new results as well as new properties among the objects found. The implementation of these algorithms have been made in different programming languages, this makes difficult its application and also comparisons between them. It is necessary have the same implementation for them to be compatible with the tools used by astronomers such as CASA*.

3. CURRENT APPROACH

3.1 GaussClumps

It was proposed by Stutzki and Guesten,[2] consist on adjusting gaussian profiles to the brightest peaks in the data (this can be done in two o three dimensions). Adjustments are made progressively from the brightest to the lowest in the data. After adjusting each gaussian is added to the output catalogue and then it subtracted to adjust the next until decomposed the original image into a superposition if gaussian in different positions, orientation, size and intensity. The sum of gaussian components reproduces the original image plus a noise that ideally must be normal and homogeneous throughout the image.

In this algorithm, the clumps may be overlapped. Therefore, the input pixels can not be assigned to a single clump.

3.2 ClumpFind

It was developed by Williams et all.[1] This algorithm stars from the brightest peak of the image (first clump) and the all of pixels are associated to the same clump, gradually decreases in intensity until a second isolated peak appears. This process continues by assigning all pixels of the original image to the nearest clump. The clumps have arbitrary shapes, and they are separated by saddle points when they are connected.

With this algorithm the cloud structure is divided into a number of spatially disjoint clump, so that collectively reproduce the original image including sky noise. Thus, the algorithm is equivalent to taking a model of the original image and split it in the direction perpendicular to the sky, like a cookie cutter where every cookie has a different shape and all fit perfectly like pieces of a puzzle to reproduce the original model.

3.3 Dendrogramas

Rosolowsky et all[3] use dendrograms in the representation de the essential features of the hierarchical structure of the isosurfaces cube molecular line data. If clumpfind algorithm divides a contour map on pieces of a puzzle, the idea behind the dendrograms is to divide the puzzle into a collection of different layers as a mass of leaf, ie, as if every contour of a geological map represents a different object.

Clearly, the sum of the classify objects reproduce the original image. Unlike previous methods, dendrograms ordered structures hierarchically in tree, branches and leaves. This system defines physical relationships between different objects cataloged.

---

*CASA: the Common Astronomy Software Applications package, is being developed with the primary goal of supporting the data post-processing needs of the next generation of radio astronomical telescopes

3.4 Wavelets

Wavelets are a special kind of Fourier transform that represent an signal in terms of dilated and translated versions of a mother wavelet.

Alves et all[4] use the wavelet transform of an image to identify and then rebuild dense cores. It is mentioned that the methods traditional threshold based not have good results, that is the reason why they use an algorithm that consists in using Wavelet transforms on an image.

Identifying objects in the wavelet space is to isolate structures given one scale. The structures in consecutive scales are connected if the local maximum the first is within the structure of the following.

4. STRUCTURE DETECTION ALGORITHM

The proposal consist in develop an algorithm to classify objects by analyzing the same type of data delivered by the ALMA and the number of objects observed. Based on the above, it generates a solution that addresses issues that are of interest for the community and ChIVO. Figure 1 shows the outline. The algorithm identify and classify astronomical structures at different scales, using the Wavelet Transform. It is generated a subset of images at different scales and then apply a detection algorithm (gaussclump, clumpfind, dendrograms) to identify objects in each image. Putting all of the images generated was due to obtain the whole set of structures containing the image that is in FITS (Flexible Image Transport System) format.[5]

The structures found in the previous step are classified using a algorithm that allows to recognize at what object type belong the found structures. The users determine which criteria are considered to classify these objects.

Finally, it is expected to generate an application that can be used in a standalone or be included in the CASA software and/or be within Chilean Virtual Observatory.

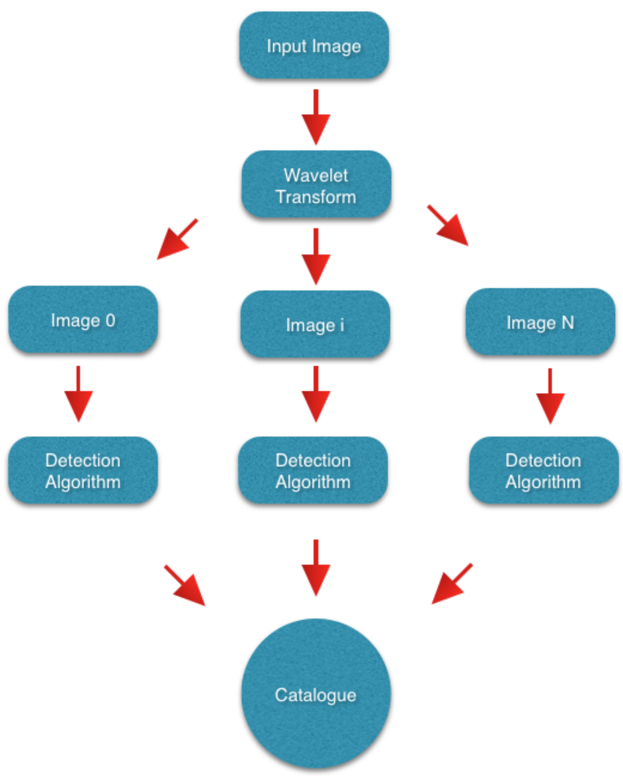


Figure 1: Creation process of structures catalog from a FITS image

4.1 Wavelet Transform for two dimensions

In the book of Starck and Murtagh,[6] it made a review of the various applications of wavelets in astronomy and provides a description of the algorithm for the implementation of a discrete wavelet transform known as A Trous. In addition, others examples of applications of wavelets in astronomy are given in the literature.[4,7]

The wavelet functions allow decompose hierarchically signals and their subsequent reconstruction, and extract certain types of information from them. The wavelets are tools for signal decomposition, such as images in a hierarchy of increasing resolutions what considered higher resolution levels where is possible achieve more and more details of the images.

A very important feature of the wavelet functions is to analyze a signal at different scales. In the wavelet analysis, the scale plays an important role, since the algorithms process data at different scales and spatial resolutions. If a signal on a small scale (low resolution) is only observed larger organizations. Similarly, if the same signal on a large scale (high resolution) is only observed small items or parts of the signal. Therefore, a signal analyzing $f(t)$ is decomposed into a series of scaled and translated versions in order to represent it as a superposition of a set of basis functions or scaled and translated wavelets.

In practice, it applies a function prototype wavelet called mother wavelet from which a family of scaled and translated versions is derived. Analysis of a time-dependent signal $f(t)$ is made from two viewpoints: (1) the temporal analysis, using a version of the collapsed mother function with high frequency; (2) the frequency analysis, it is developed with a dilated low-frequency version. The original signal $f(t)$ can be represented in terms of a wavelet expansion, operations on data can be performed using only the corresponding wavelet coefficients.

The simplest and most common way to implement the wavelet properties for the study of images, consist in applies convolutions on images using filters whose coefficients are derived from the wavelet functions. This set of filters, for both decomposition and reconstruction or synthesis, is called filter bank. A basic idea for filtering processes is that the energy distribution in the frequency domain identifies a structure. Therefore, if the frequency spectrum is broken down into a sufficient number of sub-bands, the energy of different structures are not equal. Taking advantage of this quality have been designed several types of filter banks and these include the separable filters Laws (1980), circular and wedge filters (Coggins and Jain, 1985).[8]

A digital filter is a sequence of values that is used to soften or emphasize certain aspects of a signal, either one or two dimensions. It is applied on a signal using a convolution and produce a different signal output. The filter is moved throughout the signal by calculating an inner product between the filter coefficients and those points of the signal where the filter is located. The digital representation of a filter is known as the impulse response and those which have a finite number of coefficients are called finite impulse response filters and the filters with an infinite number of coefficients are called infinite impulse filters.

Digital filters can be symmetric or asymmetric. Symmetric filters and especially those who have a form with peak in the middle, has a number of advantages: preserving the location of the sharp transitions in the signals and facilitate the processing of their edges. Symmetric filters are sometimes called linear phase, since if they are not their deviation is evaluated by the magnitude of the phase deviation from a linear function. Furthermore, exist two types of symmetric filters, symmetrical about the central value of the filter (with odd-dimensional) and hence their coefficients satisfy the relation $h(k) = h(-k)$; and symmetrical on half of the filter (dimension pair), satisfying the relation $h(k) = h(-k-1)$. Asymmetric filters satisfy $h(k) = -h(-k)$ or $h(k) = -h(-k-1)Ps$.

One drawback of the Discrete Wavelet Transform (DWT) is that it is not invariant to translations, i.e., an initial image and the other in which it has made a small shift, will present different wavelet transform coefficients. This setback is important in applications such as edge detection, determination of spatial patterns and recognition of images in general. It can avoid this effect by applying a redundant or undivided wavelet transform (undecimated wavelet transform): (1) not dividing the input image, just operating on it with the corresponding filter banks, such that the subset of images resulting from the wavelet transform have the same dimensions as the original image; or (2) by applying on it and shifted versions of their ordinary DWT (involving sub-sampling), getting a redundancy of information which is equivalent to not divide it.

4.2 A Trous Algorithm

It was developed by Holschneider et all. in 1989. Consists of a decomposition based on discrete wavelet transformation in which no sub-sampling of the images are produced, but they always have the same resolution. A convolution is performed with basic lowpass filters H and high pass filter G which are expanded by inserting an appropriate number of zeros between the coefficients (figure 2). The image is decomposed by two-dimensional low pass filter, thereby obtaining an image of approximations, whereas the coefficients of detail are result from the difference between two consecutive images filtered by a lowpass filter. To reconstruct the image the process is reversed.

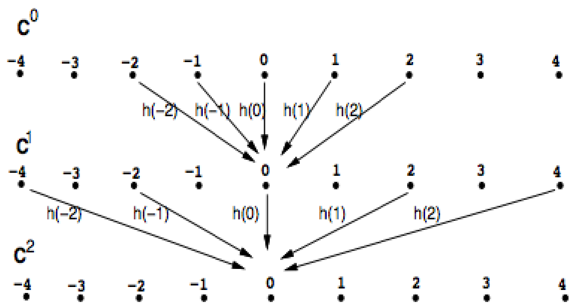


Figure 2: Construction of low pas filter

The two-dimensional low-pass filter generally consists of a bi-cubic spline filter associated with the scaling function, but it can apply other filters whose coefficients are corresponding to those used in the discrete wavelet decomposition . Figure **??** show the calculation routine for get images of approximations and details at different levels, with a generic low-pass filter of 5x5 pixels.

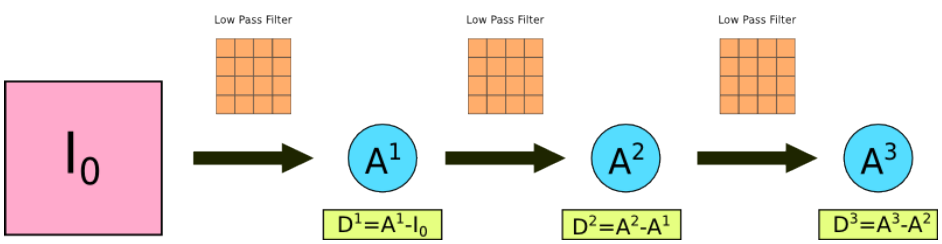


Figure 3: Decomposition process of an image with the a trous algorithm

The a trous algorithm implemented use a discrete b3-spline filter with size of 5x5 pixels. Table 1 show de defined coefficients that are multiplied by 1/256.

Table 1: Discrete B3-spline filter

1	4	6	4	1
4	16	24	16	4
6	24	36	24	6
4	16	24	16	4
1	4	6	4	1

4.3 Detection Algorithm

This algorithm is based in Clumpfind[1,9] that is an automatic algorithm for analyzing the structure in a spectral line data cube. The algorithm works by building contours, being multiples of the rms (root mean square)noise of the observations, then it seek the emission peaks as new clumps. Each emission is assigned a clump, and continues to the less intense. It is based on how the eye would analyze the maps: the data set is outlined, looking spikes, and continues with lower levels contours sequentially (figure 4).

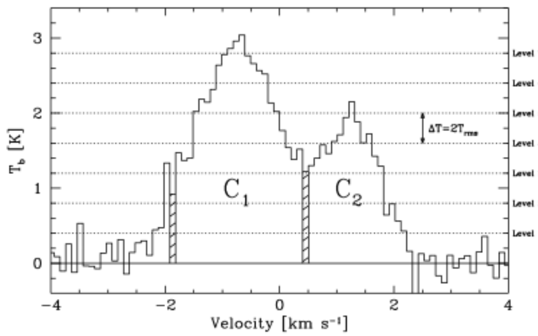
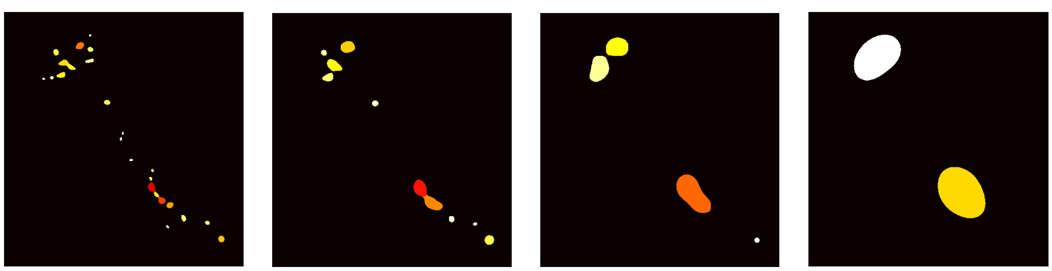


Figure 4: Spectrum with contour levels

The fundamental problems are: how to set contour levels for data and how to handle the case where two or more clumps are mixed.

The contours must be spaced, $T = \Delta T, 2\Delta T, 3\Delta T ...$, since noise is added linearly at each level. If ΔT is very small, the contour map filled structures appear difficult to differentiate between real features and noise spikes. On the other hand, when ΔT is very large the contour map will lack contrast and subtle features are lost. In tests conducted by Williams[1] established a suitable value for $\Delta T = 2T_{rms}$, where $2T_{rms}$ is the rms noise in the image.

The algorithm defined as a clump to the collection of pixels in which the highest contour is isolated from any other clump, i.e., are not connected. The clumps must be isolated in the same contour level, however, be mixed at lower levels. A vector that contain each clump in the contour is created, each new clump is a new item in the vector. A proof using the NGC 6334 region are showed in figure 5. For these cases the last level is discarded because is interpreted as noise.



(a) Level 5 (b) Level 6 (c) Level 7 (d) Level 8

Figure 5: Four last contour levels form NGC 6334 region. This it is obtained by mix wavelet transform algorithm and a detection algorithm (clumpfind)

4.3.1 Results

The detection algorithm implemented generate the position of the highest peaks in each clump, the position of its centroid, area, shape, and RA/DEC position. With these data we can construct a dendrogram mix each level. To find relationships between levels hierarchical structures only consider the location of the peak or centroid and not its value, thus the structures between related levels are necessary. It determined which clumps are within another between each level, this is similar to the dendrograms but only considered belonging to set the leafs.

The input image is in FITS format. the catalog generated is given in a table. Finally, all those implementation are included in a standalone software.

5. CONCLUSIONS

This paper presents the result of the design of an automatic detection and classification of structures. This implementation try probe what join different algorithm is possible and extend this result to ALMA data too. When many clump merge together, a significant portion of emission may originate from more than one clump. for this reason is proposed implement more detection algorithm to create a set of them and apply it in different images to generate catalogues with different structures and find different relationships.

The catalogue generated contain the structures of one region (FITS image). For create a global catalogue is necessary join the each output catalogue of collect them.

This implementation will be included as a tool in the Chilean Virtual Observatory.

ACKNOWLEDGMENTS

This work is being funded by project FONDEF D11I1060: "Development of a astro-informatics platform for management and intelligent analysis of large-scale data". The research team belong to LIRAE (Laboratory of Interdisciplinary Research on Astro-Engineering) at UTFSM.

REFERENCES

[1] Williams, J. P., de Geus, E. J., and Blitz, L., "Determining structure in molecular clouds," *The Astrophysical Journal* **428**, 693–712 (June 1994).

[2] Stutzki, J. and Guesten, R., "High spatial resolution isotopic CO and CS observations of M17 SW - The clumpy structure of the molecular cloud core," *The Astrophysical Journal, Part 1* **356**, 513–533 (June 1990).

[3] Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., and Goodman, A. A., "Structural Analysis of Molecular Clouds: Dendrograms," *The Astrophysical Journal* **679**, 1338–1351 (June 2008).

[4] Alves, J., Lombardi, M., and Lada, C. J., "The mass function of dense molecular cores and the origin of the IMF," *Astronomy & Astrophysics* **462**, 17–21 (Jan. 2007).

[5] R. J. Hanisch, A. Farris, E. W. Greisen, W. D. Pence, B. M. Schlesinger, P. J. Teuben, R. W. Thompson, and A. Warnock III, "Definition of the flexible image transport system (fits)," *Astronomy & Astrophysics* **376**(1), 359–380 (2001).

[6] Starck, J.-L. and Murtagh, F., [*Astronomical Image and Data Analysis*], Springer, 2nd ed. (2006).

[7] Jaffard, S., Meyer, Y., and Ryan, R. D., [*Wavelets and Astronomy*], ch. 12, 187–201, Society for Industrial and Applied Mathematics (2001).

[8] Coggins, J. M. and Jain, A. K., "A spatial filtering approach to texture analysis," *Pattern Recogn. Lett.* **3**, 195–203 (May 1985).

[9] Williams, J. and Teuben, P., "A miriad routine for analyzing structures in spectral line data cubes," (1998).