3D visualization and analysis of HI in and around galaxies

Punzo, Davide

Chapter 5

Summary and future outlook

5.1 Synopsis of this work

Upcoming H I surveys, such as those envisaged with Apertif and ASKAP (and after that with SKA), will deliver big datasets leading radio astronomy into the regime of the so-called *Fourth Paradigm* (i.e., data-intensive scientific discovery).

Apertif is expected to begin its observing campaign of the northern sky in 2017. The daily Apertif data-cube will have dimensions of $2048 \times 2048 \times 16384 \sim 7 \times 10^{10}$ voxels and the expected number of H I detections is ~ 100 per day. WALLABY, a blind H I survey of the southern sky, will have similar characteristics. The large volume of data creates the need for new tools and algorithms that must exploit advanced ideas and solutions for storage, data reduction, visualization, and analysis to obtain scientific results.

Automated pipelines (Popping et al., 2012; Serra et al., 2012a, 2013, 2015; Di Teodoro and Fraternali, 2015b; Kamphuis et al., 2015a) will be responsible for finding the sources and creating catalogs. The pipelines will provides catalogs with hundreds of thousands of sources and masks. In addition, measuring parameters that give an indication of the properties of a source (i.e., source characterization) will be necessary to find and highlight specific cases of interest in the catalogs and it is under development (Giese et al., 2016). However, part of these datasets have a very complex nature (Sancisi et al., 2008) and they provide many challenges to automated source finders and modeling pipelines. For example, deep surveys will probe faint H I structures, typically located in the vicinity of the galaxies, such as tails, filaments and extra-planar gas. These structures are crucial for understanding galaxy evolution, particularly when they are studied in relation to the local environment. However, the signature of these datasets is very faint (i.e., signal-to-noise ratio $\sim$ 1) and difficult to interpret. Automated pipelines can miss such faint and complex emission. Therefore, powerful visual analytics tools will extremely needed to enhance the inspection to find such features in the results generated by the pipelines and will assist in their detailed analysis.

In this thesis we analyzed the role of 3D visualization and visual analytics for the manual inspection and analysis of H I in and around galaxies. The volume rendering of the H I emission from a galaxy provides an immediate overview of the spatial and velocity structures in the data (Chapter 2). These structures are usually coherent in all three dimensions (see Fig. 2.2): the 3D view is a mix of spatial and velocity coherent structures. Therefore, 3D visualization strongly enhances and accelerates the inspection of the data and masks. We concluded that 3D visualization (performed by using a ray casting algorithm; Roth, 1982) of extracted cubelets by automated source finder pipelines (e.g., `Duchamp` and `SoFiA`) from upcoming H I surveys will be a crucial tool to enhance the human assessment and quality control and to perform interactive analysis of complex cases. Although 3D visualization can enhance the manual inspection, the current astronomical 3D toolkits are inadequate. In Chapter 2 we, therefore, defined the requirements for a 3D visualization and analysis tool for H I data, and we reviewed the state-of-the-art of scientific visualization software packages.

In Chapter 3 we investigated state-of-the-art filtering techniques and their coupling with 3D visualization to boost the discovery and inspection of complex and faint H I structures. In order to find and enhance faint and extended features in H I data one usually lowers the resolution in right ascension, declination and/or velocity to increase the signal-to-noise ratio. The optimal shape of the smoothing kernel is of course the one that follows the 3D structure of the faint features in the data such that it is just resolved in each of the three dimensions. The exact shape is, however, not known *a priori*, hence one often uses a set of different kernels and then inspects the results to make a final decision about which one suits the data best. In Chapter 3 we showed that there is an optimal filtering algorithm that works very effectively on H I data. This filter is the intensity-driven gradient filter (also known as *anisotropic diffusion*). This filter, thanks to its adaptive characteristics, needs a minimal tuning of the input parameters and enhances the signal of the low signal-to-noise ratio emission without degrading the resolution of the high signal-to-noise ratio component.

Moreover, in Chapter 4 we introduced the `CloudLasso` selection tool. This is a 3D interactive selection tool (Yu et al., 2012). Volumetric data interaction tools (e.g., picking a voxel or selecting a region of interest in 3D) are the fundamental link between 3D visualization and data analysis. Volumetric selection tools are a key capability missing in traditional astronomical visualization tools. The `CloudLasso` is optimized for traditional 2D input/output hardware and can perform a 3D selection based on a 2D selection (made by the user) over the 3D view of the data. The `CloudLasso` requires that the user interactively navigates in the 3D space and chooses the optimal 3D view (which shows at the best the 3D structure of interest) to perform the 2D selection. Therefore, the combination of 3D visualization and the `CloudLasso` selection technique is a powerful tool, because it allows the users to select the data while he/she visualizes the full 3D signature of the data. The tool can be also used in the 2D views for a 2D selection or for refining the selection slice by slice. Our implementation, the `AstroCloudLasso` tool, is an intuitive and efficient 3D selection method, which can be used also for allowing manual modification of masks generated automatically by source finder pipelines (e.g., adding very faint signal missed by automated pipelines). A second application of the tool is to select a region of interest (ROI). The ROI can be successively used to perform calculations, such as tilted-ring model fitting, in the selection.

In Chapter 4, we also demonstrated that 3D is a powerful tool not only to provide a region of interest for the calculations, but also for the inspection of models (Fig. 4.11) and the data not fitted by models (e.g., extra-planar gas in NGC2403; Fig. 4.12). We showed that interactively visualizing kinematic models (e.g., tilted-ring models that describe the rotation and geometry of H I) of the symmetric cold thin disk as a segmentation makes it possible to locate quickly and efficiently any unusual features (e.g., asymmetries and extra-planar gas) in the data-cube of interest and already get an idea of their properties, thus directing further modeling. For example, a model of the extra-planar gas above or below the disk with a slower rotation and a vertical motion provides quantitative information about the rotation and the infall velocity of such gas. We chose to use the output model of the automated model-fitting algorithm, ^{3D} `Barolo`, which indeed fits the symmetric disk in Keplerian rotation of a galaxy, to visually highlight other components (i.e., extra-planar gas, high velocity gas complexes, etc.) in a data-cube. For example, in the case of the galaxy NGC2403 (Chapter 4), the 3D view in Fig. 4.12 clearly shows the presence of an extra gas component not fitted by the model. This component consists of extra-planar gas close to the disk of the galaxy that rotates at a lower velocity than the gas in the disk. 3D visualization of the data and model gives an immediate overview of the extra-planar gas. Therefore, 3D visualization, coupled to modeling, provides additional capabilities helping the discovery and analysis of subtle structures in the 3D domain.

We integrated all these techniques in an extension of an interactive 3D image viewer for medical research, `3DSlicer`[1]. Our visualization package, `SlicerAstro`[2], bundles together 2D and 3D interactive visualization with astronomical analysis capabilities. All functionality within `SlicerAstro` is geared toward handling 3D radio astronomical data. In addition, it uses built-in features of the core software of `3DSlicer`, fully exploiting the fast 3D rendering, linked display capabilities and segmentation or region-of-interest selection. It can handle 3D data-cubes of up to 10^9 voxels on modern desktop machines, which are usually equipped with *Graphic Processing Units* (GPUs). `SlicerAstro` has been designed with a strong and stable C++ core, and its classes are also accessible via `Python` scripting, allowing great flexibility for user-customized visualization and analysis tasks. Therefore, `SlicerAstro` offers new capabilities that enhance the inspection of 3D data-cubes and allows the analysis of complex sources faster than the traditional tools. The present implementation of `SlicerAstro` provides the following novel features that are of great interest for the exploration of radio astronomical spectral line (in particular H I) data:

[1] `https://www.slicer.org/`
[2] `https://github.com/Punzo/SlicerAstro`

(A) 3D Visualization of astronomical data-cubes using the FITS data format and active use of proper astronomical world coordinates and physical units.

In Chapter 4 we presented the main module, `AstroVolume`, of `SlicerAstro`. This module provides: a user-interface for handling the loading and writing of FITS files; display of astronomical World Coordinates; control of 2D and 3D color transfer functions; classes (MRML nodes) for storing the data; and data conversion tools for masks and `3DSlicer` segmentation objects.

Interactive 3D visualization available in `SlicerAstro` *enables an immediate overview of the coherent structures in the data, helping the inspection of complex datasets.*

(B) Interactive (semi)automatic smoothing in all three dimensions.

The intensity-driven gradient filter is available in `SlicerAstro`, in addition to general Gaussian smoothing. The default parameters are set to the optimal parameters researched and discussed in Chapter 3 for a variety of H I datasets. In addition, we provide parallelized implementations of the filters both on CPUs and GPUs. The last has interactive performance which makes feasible, for the user, to interactively search the optimal parameters of the filters for a larger variety of datasets.

Semi-automatic filtering, such as the intensity driven gradient filter, coupled with interactive 3D visualization greatly helps the discovery of very faint signals.

(C) Interactive 3D selection of H I sources.

In order to operate 3D selections of H I data, we adopted the `CloudLasso` algorithm. This technique is interactive and user-friendly. The `CloudLasso` is a lasso-constrained Marching Cubes (MC) method. Its implementation in `SlicerAstro` (`AstroCloudLasso` selection) has two applications: i) interactive modification of a mask; ii) selection of regions-of-interest (ROI) for further analysis.

The `CloudLasso` *selection technique is an essential tool to help analysis in the 3D space and it strongly enhances the efficiency and effectiveness of the analysis itself.*

(D) Interactive H I data modeling coupled to visualization.

`SlicerAstro` provides modeling capabilities using the automated 3D fitting routine ^{3D}`Barolo` for fitting the regularly rotating H I disks in observed galaxies. The data used for the model fitting can be preselected using the `AstroCloudLasso` tool. This module can be used for: i) generating a classical tilted ring model to describe the regularly rotating H I disk of an observed galaxy and visually separating, with different colors, the various components in the data-cube; ii) interactively refining the model while visualizing it overlaid on the data (both in the 2D and 3D views).

The combined 3D visualization of data and models obtained with ^{3D}`Barolo` *provides an immediate overview of the cold H I disk of galaxies and unusual gas components allowing an immediate discovery of the latter.*

5.2 Final remarks and prospects for future research

Interactive 3D quantitative and comparative visualization, 3D user interaction and analysis capabilities available in `SlicerAstro` form an effective new tool that will boost, in terms both of efficiency and quality, the analysis of complex sources.

In order to fulfill all the visualization requirements defined in Chapter 2, some quantitative (non 3D) features still have to be incorporated in `SlicerAstro`. For example, a tool displaying the histogram of the flux values of the data-cube will greatly help the user in setting the 2D color function. The capability to display flux density profiles (i.e. 1-D visualization linked to 2D and 3D views) is also necessary, especially when dealing with unresolved sources. Furthermore, a dedicated tool in `SlicerAstro` for easily displaying position-velocity (P-V) diagrams will improve the inspection and comparison of models. Specialized analysis tasks on 3D selections (e.g., calculating statistics, moment maps, etc.) can be performed by running scripts from the `3DSlicer Python` console. On the other hand, customized quantitative tasks can also be added, as core modules, in `SlicerAstro`, similar to the implementation of the `AstroModeling` module. These capabilities will be integrated in future updates of `SlicerAstro`. Moreover, the involvement of a broader user (and developers) community in `SlicerAstro` can bring additional expertise that can be translated in new analysis capabilities.

In addition, a future perspective is to add interoperability between `SlicerAstro` and the future H I data servers archives. One could imagine using high-dimensional data visualization software (e.g. `TOPCAT` Taylor, 2005) to explore the parameters of thousands of sources in catalogs. The user should be able to mark the data of interest and download the associated data-cube(s) from the catalog and then use `SlicerAstro` for further exploration of the 3D signatures. A comparison with one or more models can be performed for a detailed inspection and analysis of complex cases. Finally, the user should be able to upload easily the results of the analysis to the catalog. Guaranteeing such powerful connectivity between the on-line catalogs and visual analytics tools, such as `SlicerAstro`, will enhance the discovery of potential of the archives in the era of Big Data for radio astronomy. We will therefore investigate the possibility of using Virtual Observatory (VO) tools, in particular the astronomical communication protocol (SAMP), to realize such capabilities.

Finally, although our main focus for the development of `SlicerAstro` was the analysis of 3D H I data, it can, in principle, also be used to inspect and study any other type of 3D astronomical data such as mm/submm molecular line data and optical integral field spectroscopic data. Molecular line data and optical/NIR spectroscopic data often have the additional complication that more than one spectral line are present in a single spectral window. Specific techniques will be required to cleverly combine the known spectral lines and provide visual representations that highlight the specific 3D structure (i.e. spatial and kinematic information) of the data. In conclusion, `SlicerAstro` can be employed for the visualization of any other types of 3D astronomical data, but additional machinery needs to be added to `SlicerAstro` to properly visualize and, moreover, analyze such data in an effective manner.