COMPUTER SCIENCE HONOURS
FINAL PAPER
2016

Title: Overlay of Optical Data Catalogues on 3D Radio Data Cubes

Author: Bryce Billing

Project Abbreviation: AstroVis2

Supervisor(s): Associate Professor Michelle Kuttel, Dr. Sarah Blyth

Category	Min	Max	Chosen
Requirement Analysis and Design	0	20	5
Theoretical Analysis	0	25	0
Experiment Design and Execution	0	20	20
System Development and Implementation	0	15	5
Results, Findings and Conclusion	10	20	20
Aim Formulation and Background Work	10	15	10
Quality of Paper Writing and Presentation	10		10
Quality of Deliverables	10		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	0	10	
Total marks		80	

Overlay of Optical Data Catalogues on 3D Radio Data Cubes

Bryce Billing
University of Cape Town
bllbry001@myuct.ac.za

ABSTRACT

Radio data cubes are the typical data products from radio telescope interferometers and typically contain 3D information in two spatial directions and one frequency axis. This is different to the usual 2D images made with telescopes at other wavelengths. Astronomers usually cross-reference objects from different datasets, they can overlay images or information from catalogues, but it is not currently possible to do this in three dimensions in the existing radio data visualization software packages. Being able to do this is important in aiding analysis, identifying structure and finding sources in low signal-to-noise observations. The overlaying of optical data catalogues on 3D radio data cubes will help astronomers visualize information from two different wavelengths and use information about galaxies that we already know to make observations about the radio data. It will also be easier for astronomers to visualize galaxy clusters and radio sources. We aimed to develop the first effective method for the overlay of optical data catalogues on 3D radio data cubes. We approached this problem using AstroVis, which is an already existing visualization software package and extending it to incorporate the overlay functionality. We tested the overlay with simulated data so as to eliminate all the variables such as, resolution differences between radio data and optical data, that could affect the positioning differences of the overlay markers and the radio data cube galaxies. Our results showed that overlay is easy to use and the markers were accurate at marking the galaxies. The overlay functionality is likely to reduce the amount of time required to overlay optical data as well as help astronomers easily visualize points of interest such as galaxy clusters.

CCS Concepts

• **Human-centred computing →Scientific Visualization**

Keywords

Radio Astronomy; Visualization; Data Cubes; Overlay; Optical Data Catalogue.

1. INTRODUCTION

With construction of The Square Kilometre Array (largest radio telescope in the world) soon to begin and its pathfinder telescope, the MeerKat, currently under construction, it has led to an increase in the need to be able to efficiently analyse radio data cubes, due to vast amounts of data they produce. For example, the SKA will have a 40 gigabyte (GB)/second data volume [1]. These data volumes cause an interest in developing software to try and eliminate as many manual operations as possible and increase the need to improve the current visualization software available.

In astronomy researchers need to visualize their data in order to identify features of the objects they are studying and part of this analysis involves comparing data taken at different wavelengths by different telescopes (so-called multi-wavelength astronomy). This is how they can build up a better picture of the structures and processes of the astronomical object they are studying. One method that is extremely useful is to overlay data/images/catalogues taken in different wavebands. The overlays allow astronomers to use information that they already know (optical data catalogues) to make judgements about unknown data (radio data cube of the same area). For example, astronomers are able to use a catalogue with galaxy positions, to mark galaxies in the corresponding data cube being analysed. The overlay will have many applications; improving the number of radio detects, reducing manual operations, pattern recognition, identifying structure and finding sources in low signal-to-noise observations. A radio detection is when a galaxy or other astronomical object is discovered when using radio data from a radio telescope. It will be useful to mark galaxies that have very weak emissions (neutral atomic hydrogen - a normal hydrogen atom which is electronically neutral and has one electron and one proton, commonly called HI [2].) in regions of noisy data. For example, the overlaying in a radio data cube of an optical galaxy that has been found, may lead to the discovery of a radio source. Astronomers consider it useful to view information at different wavelengths as it allows them to learn more about the data i.e. radio data cubes tell you about the location of the neutral atomic hydrogen gas in the galaxy and optical wavelengths show you the location of stars overall. It can also help astronomers understand the nature of the data faster, therefore, speeding up analysis.

An optical data catalogue is a "long list of observations of astronomical objects." [3] It is also defined as a table, where each row contains information about an astronomical object [4]. The optical catalogue referred to in this project contains information about the Right Ascension (RA), Declination (Dec) and Redshift (z) of a list of galaxies. These correspond to the x, y and z coordinates respectively. The catalogue also stores an ID for each of the galaxies. The RA and Dec are measured in degrees and the Redshift is calculated from either the velocity (km/s) or frequency (MHz) of the galaxy.

A radio data cube is a 3D data structure that stores information about a specific volume of the universe obtained from a radio telescope. An interferometer correlates the signals between the antennas of the radio telescope, to create a dataset containing Fourier coordinates (u, v, w). A Fourier transform then needs to be applied to the dataset to get it into spatial coordinates (RA, Dec). This creates an image of the region usually called a map [5]. Many images stacked form a data cube. These multiple images are generated by the fact that telescopes observe information over a range of frequencies.

3D visualization involves the visualization of the entire data cube, allowing the user to rotate and zoom into certain areas of the data cube, whereas, 2D visualization only show a slice (see Figure 1 below) of a data cube. 3D views help for qualitative interpretation of data by the user and is useful for the interpretation of spectral data cubes [6]. 3D visualization is also useful in conveying the meaning of information to non-experts [6]. With a 2D visualization the user has to step through each slice to get deeper into the data cube, however, 3D visualization allows for enhanced pattern recognition and improved comprehension of global features [6].

Current tools such as DS9 [7], Aladin [4] and Karma [8] support 2D overlay of optical data catalogues on radio data cube slices (2D section of a 3D radio data cube along a specific axis) [9, 7]. However, 3D overlay has not been implemented in any astronomy tools, to our knowledge, that are currently in use. LivOlay has implemented a general 3D overlay tool [10], however, there is no information on whether the tool was released, as there is little information on it. We will be extending AstroVis [6] for this project which is an existing data cube visualization software that supports both 2D and 3D visualization.

We aim to effectively overlay optical data on a 3D radio data cube. By 'effectively', we mean that it is accurate in marking positions of objects in the data cube and easy for the user to use.

The general approach to the overlay will be to make sure the catalogue is in voxel positions and not degrees, and if it is in degrees perform the necessary conversions. These conversions will be done by developing a conversion tool. Once the conversions have been done, then the catalogue can be loaded and stored and using these voxel position as the centre of the galaxies. Markers will then be drawn at these voxel positions. The overlay will also be rendered on the 2D slice view. This will all be done as an extension to AstroVis [6].

2. BACKGROUND

2.1 Data Formats

2.1.1 Radio Data Cubes
A radio data cube is a 3D data structure of radio information obtained from a radio telescope from a specific volume of the universe, it has 2 spatial dimensions (usually RA and Dec) and a third axis corresponding to frequency or velocity. These respectively correspond to x, y, and z in a normal coordinate system respectively. We will use a simulated data cube for our project. A data cube is shown in Figure 1 below.
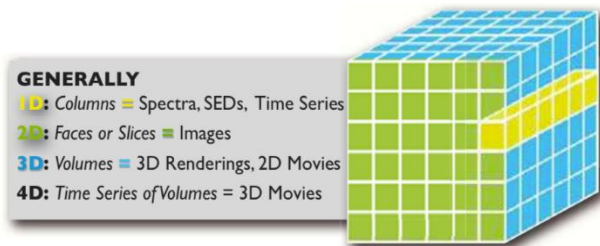

Figure 1. The Format of a data cube produced by a radio telescope [9]

Data cubes are stored in many different formats, the most common types being FITS [11] and HDF5 [12]. An HDF5 file is organized as a directed graph and data model objects are mapped onto the storage model. With the use of B-tree data structures, it allows for the data to be discontiguous allowing the resizing and insertions of datasets. It is a format for managing and storing complex scientific data [12]. It is portable, extensible and parallelizable, allows for an unlimited variety of datatypes and has built in compression [12]. A FITS file has a set of Header-Data Units (HDU) which are contiguous blocks of data following ASCII headers [11]. FITS has random groups, ASCII tables, and compression and has been the main data format for astronomy since the late 1970s [11]. The size of the data cube depends on the

field of view surveyed and the number of channels used. For example, the SKA will generate multi-terabyte data cubes [6].

2.1.2 Catalogues
Catalogues store text-based information about astronomical objects which might include properties such as spatial coordinates, redshifts, luminosities at different wavelengths and galaxy ID's. They are stored in a table format, where each row in the table represents an object. They can be stored in .txt, .csv and .tsv file formats. The catalogue stores the information in the order of Galaxy ID, RA, Dec (in degrees), frequency (Hz) or velocity (km/s), these will then be converted to voxel positions using the conversion tool built into the overlay (Section 4.2). Optical data catalogues are created by astronomers when analysing optical data from a specific section of the universe.

2.2 Visualization
Visualization is the process of transforming quantified data which is not visual into a visual representation [13]. It allows a large amount of information to be rapidly interpreted by a human [14]. It allows for problems with data to be immediately observable and it can help easily convey information to non-experts [6]. It is useful in astronomy as it allows for astronomers to observe the data collected from the large radio telescopes and make observations about it.

Visualization comprises of four basic stages, namely: collecting and storing data, converting the data into something that is easy to manipulate by pre-processing it, the graphics algorithms and display hardware that produce an image and the human cognitive and perceptual system [14]. Visual information processing consists of a three staged model. The stages consist of parallel processing to extract low level properties, pattern perception and visual working memory [14]. The pattern recognition stage is where the visual system divides up the view into regions and simple patterns, for example, regions of the same colour and texture. This will help the visual system determine the difference between the overlay and the radio data cube and the stage of the visual system that we aim to exploit with our design principles.

Ben Shneiderman's mantra, which is a set of capabilities that a visualization should have, for it to be more effective, of overview first, zoom and filter and then details on demand, will influence the decisions made with both the design and functionality of the overlay [15].

2.3 Visualization Software
Astronomy tools are often developed by institutions that specifically focus on analysing astronomical data. The aim and focus of these tools is to effectively visualize astronomical data, so that it is easier to make observations using the data. Few tools support overlay, however, the following tools support a form of overlay.

Karma is a visualization library that supports both 3D and 2D visualization. It was developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Australia. It was designed to provide a toolkit to help with developing visualization applications as well as a suitable environment for production tools [16]. Karma only supports 2D overlay, and it does so with the use of renzograms [8]. Renzograms display the velocity information in a cube. A renzogram overlays, for every channel in the cube, a single-level contour map over an image. Each channel's contour map is in a different colour [8]. A renzogram shows peaks in the velocity profile [8]. However, the final image is a 2D image.

LivOlay allows data rendered in different applications to be rapidly overlaid [10]. It was developed by Hao Jiang, Daniel Wigdor, Clifton Forlines, Michelle Borkin, Jens Kauffmann and Chia Shen. It visually overlaps and geometrically registers multiple remote desktops on a shared display [10]. LivOlay supports 3D overlay of data, it is not, however, specific to astronomy.

Aladin is an interactive software sky atlas, which allows the user to visualize sky images [4]. It was developed by the Centre de Données astronomiques de Strasbourg (CDS) [4]. It supports the overlaying of optical data catalogues on 2D images. It is mainly used to visualize and check catalogues and images, search and browse available astronomical data, prepare observations and create field charts [4].

DS9 is a 2D and 3D astronomy visualization tool. It was developed by the Smithsonian Astrophysical Observatory in the United States of America. It was developed to view FITS files and binary tables [17]. DS9 supports 2D overlay of symbols from a catalogue on the current image displayed [7]. DS9 also supports the overlay of coordinate grids and contours on top of an image [18]. It allows for remote analysis and region annotation with text and simple shapes [19].

AstroVis was developed for viewing very large FITS files on a personal computer or laptop [6]. AstroVis loads FITS files and it gives the user both a 3D view of the cube as well as a 2D slice view of the cube. AstroVis allows the user to get a quick general overview of the cube, allows the user to rotate and scale the 3D data cube, translate and zoom into the 2D slice view and step though each of the slices.

2.4 User Centred Design
User centred design involves the user from the early stages of the design process to ensure an understandable and usable end product [20]. It allows for an early understanding of the requirements needed by the user. It is especially useful for scientific software, because you are developing for a specific end user you can incorporate functionality that they are used to i.e. do not have to develop it for the general public. This process comprises of four main steps: identify end users, identify requirements, implement and evaluate the design [20]. Identifying end users involves deciding who is using the software. Identifying the requirements involves determining what the end users expect from the system. Implementation can be done in stages, starting with a rough concept and ending with a complete solution and finally evaluating the design involves testing the system, preferably with end users to determine its viability [20].

To evaluate the system, we aimed for expert evaluation through a combination of interviews and observations. Interviews allow for an in-depth reflection and it is good to understand the experiences of different people [21]. Observations allow us to study and gather information on an activity and uncover what the user actually does, rather than what we think they should do [21].

3. DESIGN
For the design of the 2D and 3D overlay we followed a user centred design approach. We developed each feature of the overview based on the requirements specified by the end users (astronomy experts).

3.1 Design Goals
We aim to overlay optical data catalogues on radio data cubes, to allow for easier discovery of galaxies and better utilization of known information i.e. using the information from optical data catalogues (known information) to mark galaxies in the radio data cube. Our overlay should be easy to distinguish from the surrounding noise in the data cube and at the same time not obscure important information.

Our goal is to be able to clearly identify galaxies that have been listed in catalogues on a 3D radio data cube. The overlay markers must be placed as close to the centre of each of the galaxies marked as possible.

The overlay functionality should be usable and intuitive to improve productivity. Usability is defined as the ability of the software product to be understood, used, learned and attractive to the user, when used under detailed conditions. [22]

The design provides an overview of the information as well as the ability to zoom into the data and filtering of markers by galaxy ID incorporated.

3.2 Design Principles
We followed basic principles to guide our design, such as Shneiderman's mantra of overview first, zoom and filter and then details on demand [15]. Another design principle we used is comparisons, as the purpose of the overlay is to compare the optical and radio data [23].

We aim to maximise information density, to convey as much information to the user at one time. We followed the principle of interconnection to achieve this, as it emphasizes the use of multiple visualizations to convey the data to the user [23].

The principle of dynamic scaling allows the overlay to scale and transform with the data cube, making it easy to track the markers no matter what angle the user wishes to view the data from [23].

Visual metaphors are also used in the design to help the user use information that they already possess to improve the overlay [23]. We also use preattentive attributes to help distinguish the overlay from the background data was used. We chose to use the design principle of dominance/emphasis by changing the preattentive attributes of the markers so they differ from the surrounding data [24]. The preattentive attributes incorporate form, colour and spatial position. For the overlay specifically, we will be focusing on colour and form (shape) of the markers, and more specifically the hue of the colour and the curvature of its form i.e. the hue of the colour of the markers will be a different colour to that of the surrounding data and the curvature of the markers will also be different to the curvature of the data cube elements [25].

3.3 Design Process
We followed a user centred design approach and completed the steps as follows: the identification of the user was a simple process as the project was requested by the astronomy department and therefore, the end user would be astronomers. The identification of the requirements was done by organising interviews with the astronomy experts and finding out what they expected from the system, in terms of its functionality. The evaluation of the system was done using expert evaluation with the use of observations and interviews to determine what their thoughts on the system were.

3.3.1 Prototype Design
Our overlay will be rendered on both the 3D data cube and the 2D slice view in AstroVis to present an overview of the galaxies present (3D overlay) as well as more detail (2D overlay). This incorporates overview and details on demand of Shneiderman's mantra. The use of the two different overlay displays encourages

comparison, as the user is able to compare the information from the 3D and 2D overlays.

3D pins are used as markers for the 3D overlay, as it is a symbol often used for a marker and follows the design principle of visual metaphors. The markers for both the 2D and 3D overlay will transform and scale with the radio data cube and at the same time will always face the camera, maintaining the principle of dynamic scaling (zoom).

The 2D overlay allows the user to select between two options of markers: a cross (which marks the centre of the galaxy (Figure 5(a))) and a square (which surrounds the galaxies (Figure 5(b))). These shapes are angular and due to the number of curves present in the data, they will help the overlay stand out as it makes the preattentive attributes of the overlay different from that of the radio data.

Red, blue and yellow appear frequently in a noisy data cube (Figure 2) and therefore, to make the markers stand out, colours not in this spectrum, or of different luminosity must be used. Green is not in the colour spectrum of AstroVis and therefore, was selected as the main colour for the overlay. The option to change the overlay colour to cyan for users who are red-green colour blind is also available. The overlay markers also have black outlines, which separate the markers from one another and the surrounding environment.
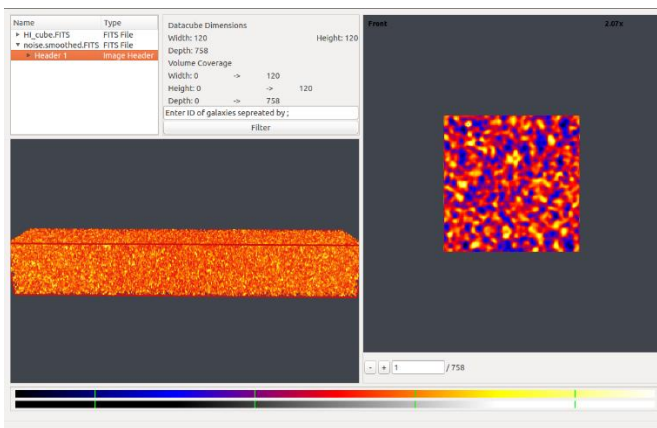

Figure 2. AstroVis colour spectrum.

The overlay functionality is separate from any of AstroVis' functions i.e. the overlay has its own menu so all its functions are under one menu to try and improve the usability of the system, so as to not confuse the user between the functions for the data cube and the functions for the overlay.

3.3.2 Evaluation
We evaluated the accuracy of the markers as well as the usability of the overlay system. The overlay was also tested to see what the biggest file size it could load was before the system slowed down considerably.

A synthetic data set is used to test the application, comprising of a data cube and the corresponding data catalogue. This helped to avoid issues with the position of the markers being due to redshift issues and positional differences in the optical and radio data cubes, as this is not in the scope of this project. Positional differences are due to the resolution differences between optical

and radio data: optical images have much better angular resolution and therefore, the centre of an astronomical object is easily determined, however, with radio telescopes, a galaxy has very poor angular resolution and, therefore, it is difficult to get the exact same centre as that of the optical image. The redshift issues we refer to, are caused by the uncertainties in the redshift values of both optical catalogues and radio data cubes i.e. when a galaxy's redshift is determined in both the optical and radio data there will always be a related observational uncertainty on the measurement.

We used expert evaluation, with three domain experts from the Astronomy department at the University of Cape Town. We also used two non-experts who were doing postgraduate research in the field of Astronomy. The participants were given an explanation of the system and features and then a set of tasks were presented for them to achieve. The tasks consisted of converting a catalogue using the built in conversion tool, loading a catalogue, filtering specific galaxies given to the user, toggling on and off both the 3D and 2D overlay, changing the colour of the overlay and the shape of the 2D overlay.

After all the tasks have been accomplished the user is interviewed as to what they liked and what they did not like, what would be a useful addition to the application in terms of overlay and whether they think that it is useful for astronomy.

The accuracy of the galaxy markers is tested by reducing the number of galaxies visible to one or two and seeing whether the visible galaxies accurately mark a galaxy in the data cube and this is evaluated on a scale of 1 to 5.

The overlay is tested by loading catalogues of multiple sizes until the overlay takes a considerable amount of time to render the overlay. This was to test the maximum file size it can load in a reasonable time.

The data was analysed by taking the averages of the time taken to render the multiple size catalogues. The average value of the accuracy of the galaxy will be recorded. Low values will mean the information is not accurate or clear, however, high averages will mean that it is clear and accurate. The catalogue loading data was plotted in a table with the file size, number of objects present and the average loading time.

4. IMPLEMENTATION
The prototype (Figure 3) uses AstroVis to load and render the 3D data cube and the 2D slice view. The overlay component loads, processes and displays the catalogue data. The stored catalogue is then used to render the overlays.
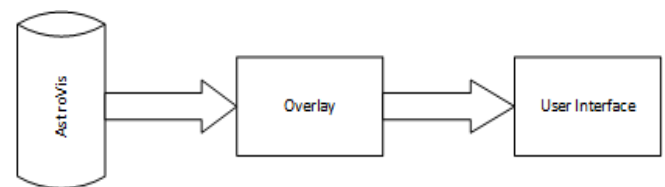

Figure 3. System architecture diagram.

4.1 Methodology
The prototype was developed with the Waterfall software engineering methodology. The stages in the Waterfall methodology are: System Engineering, Requirements Analysis,

Design, Construction, Testing, Installation, Maintenance [26]. For the overlay, the installation and maintenance phases of the methodology are not applicable as this is not software that is being released immediately.

The system engineering phase involves determining the high-level system architecture. This was done in the original prototype of the AstroVis, so for the overlay we just had to extend the functionality of AstroVis.

Requirements analysis involved developing a list of the necessary requirements for the project, in consultation with experts in the Astronomy department at the University of Cape Town. Requirements were: a catalogue loader which loads the optical data catalogue and displays the markers on the radio data cube, the filtering of the overlay to be able to only see certain galaxies, the need for a conversion tool to convert a catalogue in spatial coordinates into voxel positions and the toggling of the overlay on and off.

The design phase involved determining the design specification, software architecture specification and system test specification.

The construction phase involved editing the source code of the original AstroVis visualization prototype.

Testing consisted of getting expert evaluators to test certain aspects of the prototype mentioned earlier in the Design section of the report.

4.2 Catalogue Conversion
The optical data catalogue stores the information in degrees. However, to overlay the information it must be converted to voxel positions. The redshift also needs to be converted to either frequency or velocity before its conversion to voxel position, using the following equations:

$$RA_{pix} = \frac{RA - RA_{reference}}{RA_{delta}}$$

$$Dec_{pix} = \frac{Dec - Dec_{reference}}{Dec_{delta}}$$

If it is a velocity data cube, then these equations must be used to convert the Redshift:

$$V = cz$$

Where c = speed of light in km/s

$$V_{pix} = \frac{V - V_{reference}}{V_{delta}}$$

If it is a frequency data cube, then these equations must be used to convert the Redshift:

$$F' = \frac{F_{HI}}{1 + z}$$

Where F_{HI} = Rest frequency of HI (neutral hydrogen) in Hz

$$F_{pix} = \frac{F' - F'_{reference}}{F'_{delta}}$$

The Channel span for frequency also needs to be calculated:

$$Freq = \frac{V}{c * \frac{1 + z}{F_{HI}}}$$

Where c = speed of light in km/s

V = velocity

F_{HI} = Rest frequency of HI (neutral hydrogen) in Hz

The data after the conversion is saved in the format: RA, RA span (Number of voxels the galaxy spans in the RA direction, often 1), Dec, Dec span (Number of voxels the galaxy spans in the Dec direction, often 1), Channel (z direction) and Channel span (Number of voxels the galaxy spans in the z direction, for AstroVis it has been set to 100 km/s long) in a .txt file.

4.3 Data Loading
Data is loaded into AstroVis using C++ file I/O. The data in the catalogue is structured in that each line in the catalogue is a galaxy and each galaxy has tab separated columns. The columns do not have headings. The first column is the ID of the galaxy, the second column has the RA voxel position, the third column has the RA voxel span, the fourth column has the Dec voxel position, the fifth column has the Dec voxel span, the sixth column has the Channel voxel position and then the seventh column has the Channel span. This catalogue is generated in this format after using the converter discussed in earlier sections.

The catalogue is read in line by line. Each line is split on the tab character, so that an array of the values is created. This array is then used to populate a new Catalog struct which has a variable for every column in the catalogue text file. These Catalog structs are then stored in a QVector<Catalog> which is then passed to the DetailRender and VolRender classes to be used in the rendering of the overlay.

4.4 Alterations to AstroVis
New classes were added for the catalogue loading and 2D overlay, make the overlay as independent as possible from AstroVis' already existing structure.

The following classes were added to AstroVis: Catalog struct, CatalogueLoader which handles the reading and storing of the catalogue itself, CatalogueHelper which handles the drawing of the 2D overlay as well as checking whether a specific x, y and z coordinate is in the catalogue data and CatalogWidget which brings up the dialog allowing the user to select the file they wish to load.

We added a converter to convert a data catalogue in degrees into the necessary voxel positions. ConverterDialog which allows the user to specify the CDELTA and CRVAL values and the converterwidget allows the user to select the file that they wish to convert.

Changes made to AstroVis classes, include the addition of drawing functions to the VolRender class for the 3D overlay to be drawn. The DetailRender class was also changed to call the necessary classes to draw the 2D overlay. New menu items were added to AstroVis' menu bar to allow the user to access the overlay functionality, such as toggling on and off the overlay, filtering, changing the colour of the overlay, changing the type of 2D overlay and converting and loading the catalogue files. A filter text box was also added to the graphical user interface, where the user would be able to type in galaxy IDs that they want to only see.

4.5 Challenges
As AstroVis was not updated since 2012, changes had to be made in order for AstroVis to compile. Determining the required changes and libraries needed was probably our main difficulty.

Understanding the transformations that AstroVis applied to the data cube and slice, was also a challenge, as these same transformations needed to be applied to the overlay.

It is difficult to tell with the 3D overlay and the quantity of galaxies even in a small data cube, whether the markers are in the correct position. AstroVis also culls the galaxies as you zoom in closer making this task more challenging. This can lead to difficulties in the implementation, as you are not sure whether the transformations being applied are exactly correct.

4.6 Evaluation

We tested the software both qualitatively and quantitatively. We used experts and non-experts to do the qualitative testing through interviewing and observations. We then tested the capabilities of the software by testing it with different file sizes and measuring time taken for various file sizes. All tests were done in Ubuntu using a MacBook Pro 13" 9,2 2.50GHz i5 processor, 4GB DDR3 RAM.

For the quantitative testing we tested with various file sizes ranging from 2.5 MB to 30.8 MB. We stopped testing at 30.8 MB because the loading time had slowed down substantially. Although the file size of the 30.8 MB may seem like an inadequate size for the software to start struggling with, however, when considering the fact that a 30.8 MB file contains 629466 astronomical objects, then it is more understandable.

5. RESULTS AND DISCUSSION

Our discussion of the final prototype begins with an overview of the system and the overlay features before proceeding to discuss the results of the evaluation.

5.1 System Overview

The overlay (Figure 6) reads in the catalogue and places markers at the corresponding voxel positions. These markers are rendered on both the 2D (Figure 5) and 3D overlay (Figure 4). To load a catalogue into the overlay, select the Overlay drop down menu and then select load catalogue. This will then prompt you to select your catalogue. Once you have done so and pressed open, the catalogue will be loaded and rendered immediately.

A screenshot of the final 3D overlay in a noise free data cube is displayed below (Figure 4). The markers are shown by the green pins and the galaxies are shown by the orange objects in the cube. The tip of the pin marks the centre of the galaxy as this is the easiest place to mark, as you know its RA, Dec and Velocity/Frequency but you do not always know how long the galaxy is.
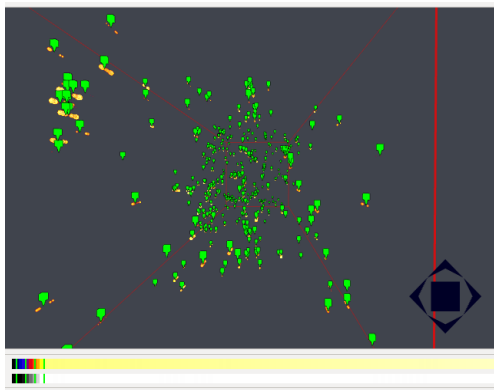

Figure 4. Final 3D overlay in noise free data cube.

The 2D overlay (Figure 5(a) and Figure 5(b)) shows the different types of overlay available. The square overlay surrounds the galaxy and the cross marks the centre of the galaxy obtained from the catalogue. The option of having two separate styles of 2D overlay was so that the users have greater flexibility over what they wanted to see.
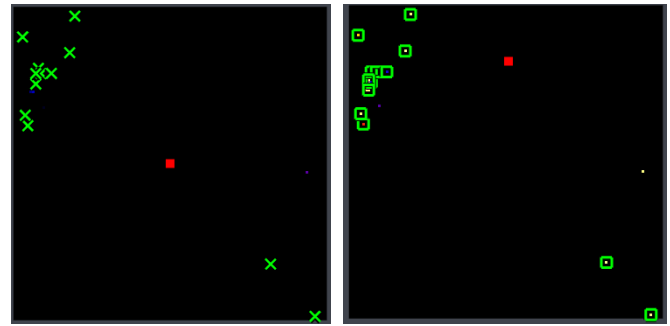
(a) Cross markers   **(b) Square markers**

Figure 5. Final 2D overlay markers.

The overlay functionality is a separate menu (Figure 6). However, the filter box was incorporated in with AstroVis' current display as it became too many steps for the user if we made it a separate dialogue pop up.
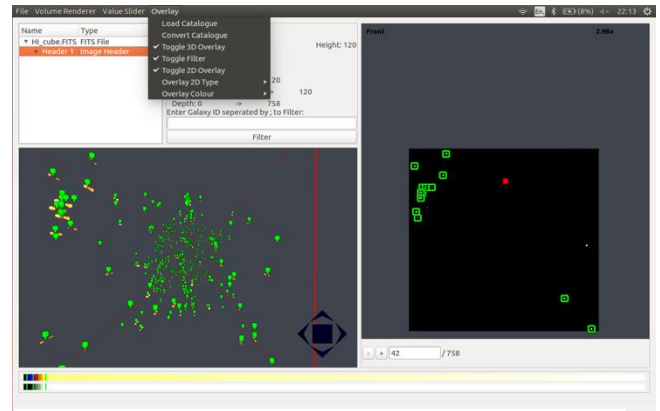

Figure 6. Final overlay prototype with loaded catalogue.

5.2 Evaluation

We were satisfied with the loading capability of the overlay. The graph (Figure 8) shows that there was a linear increase in time taken to load the data with an increase in size of the catalogue being loaded.

Table 1. Table of Testing the loading time of a catalogue for various catalogue file sizes.

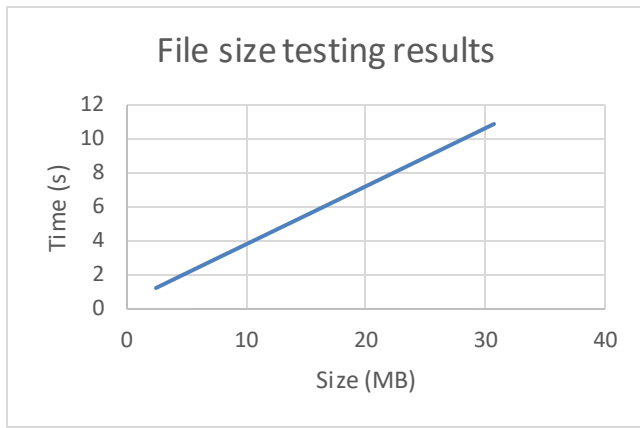| Size (MB)	Number of Objects	Average Time Taken to Load (s)
2.5	52717	1.22
7.7	159862	3.00
15.4	319739	5.62
30.8	629466	10.89

Figure 8. Graph of testing the loading time of a catalogue for various catalogue file sizes.

The testing yielded varied results. They all had both negative and positive comments about the overlay. The participants all struggled with the order in which the events of the conversion tool took place i.e. they would have preferred to have selected the file they were converting first before entering the values necessary for the conversion. They also mentioned how there was no progress bar, so there was no way of telling when the conversion had completed.

The participants suggested some improvements to the overlay functionality: allowing only a subset of galaxies to be shown, have a key to show what an overlay marker looks like and show the actual RA, Dec and Frequency/Velocity value of the galaxy in the 2D overlay. The subset of galaxies was suggested to give greater control when analysing the data, to try and avoid overcrowding the visualization and avoid noise spikes. The key was suggested as a user found it difficult to understand initially what colour represent overlay and what colour represented the galaxies. The suggestion to show the RA, Dec and Frequency/Velocity was to try and increase the information that the overlay could show.

The adjacent 2D and 3D display was found to be useful by the evaluators. They also liked the fact that all the overlays functionality was under one menu. They found the filtering function a useful tool and thought the symbol choice for the markers to be very appropriate as it is often used in other visualizations as a marker.

All the participants thought that the overlay would be useful for research. Their reasons were: it is a simpler way to identify items in the data set without having to do it manually, it will allow astronomers to give insights into the data that would be otherwise difficult to make, it can be used to help identify clusters, the human eye is good at picking out patterns and it, therefore, will help develop an intuition about what to search for quantitatively and there was nothing similar to it available in any other software packages.

The accuracy of the galaxies was found to be a difficult question to answer, due to the functionality of AstroVis not allowing you to pan or zoom in to close to a galaxy, as it culls the galaxies that are close to it. The two non-experts gave the accuracy of the markers 3.5 and 5 out of 5. Two of the experts gave the accuracy of the markers a rating of 4 out of 5 and the third expert explained that they do not deem it necessary to rate the accuracy as specified, because in actual data there will always be a physical offset and matching positions is very difficult at different wavelengths, as

they will never be perfectly on top of one another, however, they did say that the markers looked accurate from what they could see.

The process for the evaluation was effective, because we were present during the evaluation we were able to take notes while the user was performing specific functions, however, the problem that we found was that the evaluators seemed to have rushed into trying to perform the tasks, without taking time to go over the interface as advised. This caused a lot of confusion at the beginning of the testing. We were satisfied with the results of the evaluation as there were a lot of positive comments about the overlay functionality. A summary of the positives, negatives and impressions from our evaluation is shown below.

Table 2. Summary of the positive and negative results from evaluation

Participant	Negative	Positive	Useful for Research
A	Not immediately obvious what everything means.	2D and 3D effective. Effective symbol choice.	Yes.
B	Order of conversion tricky.	Simpler way to identify items without having to do it manually. Very obvious overlay.	Yes.
C	Order of conversion tricky.	Pinpointing centre is effective. Easier to use and more modern interface. Filtering function effective.	Yes.
D (non-expert)	2D and 3D redundant.	Helps with quick glance detection. Easy visualization of everything in 3D space.	Yes
E (non-expert)	Filter not right word for filtering of galaxies.	Will help with multi-wavelength astronomy.	Yes

6. CONCLUSIONS AND FUTURE WORK

This project produced an effective tool for overlaying optical catalogue data on 3D radio data cubes. The use of the user-centred design process and the waterfall methodology allowed us to develop an overlay that met the specific needs of the targeted users and also made sure that the tool was developed on time. The conversion tool was not as easy as was intended, however, it was a minimal problem that could be fixed in future iterations.

Future work will involve extending the overlay to allow for more than one catalogue to be overlaid at one time and the placing of custom markers in the data cube that would then be added to the catalogue.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Wu, X. et al. Data Mining with Big Data. *IEEE transactions on knowledge and data engineering. 26*(1), 97–107, 2014.

[2] Swinburne University. Neutral Hydrogen | COSMOS. *Astronomy.swin.edu.au*. Retrieved October 26, 2016 from http://astronomy.swin.edu.au/cosmos/N/neutral+hydrogen.

[3] Jaschek, C. Astronomical Catalogues: Definition, Elements, and Afterlife. *Quarterly Journal of the Royal Astronomical Society. 25*, 3, 1984.

[4] Fernique, P. et al. Aladin User manual. 1–91. Retrieved October 26, 2016 from http://aladin.u-strasbg.fr/java/AladinManual6.pdf, 2009.

[5] Chengalur, J., Gupta, Y. and Dwarakanath, K. *Low Frequency Radio Astronomy* 3rd ed., National Centre for Radio Astronomy, Pune, 2007.

[6] Perkins, S, Questiaux, J, Finnis, S, Tyler, R, Blyth, S and Michelle, K. Scalable desktop visualisation of very large radio astronomy data cubes. *New Astronomy. 30*, 1–7, 2014.

[7] Joye, W.A. New features of SAOImage DS9. **Proceedings of * Astronomical Data Analysis Software and Systems (ADASS) XV, ASP Conference Series, C. Gabriel, C. Arviset, D. Ponz, and E. Solano, eds. 351*, 574-576, 2006.

[8] Gooch, R. 3 Viewing and Comparing Multiple Images and Cubes. *Atnf.csiro.au.* Retrieved October 1, 2016 from http://www.atnf.csiro.au/computing/software/karma/user-manual/node4.html#SECTION00400000000000000000, 2006.

[9] Goodman, A.A. Principles of high-dimensional data visualization in astronomy. *Astronomische Nachrichten. 333*(5-6), 505-514, 2012.

[10] H. Jiang, D. Wigdor, C. Forlines, M. Borkin, J. Kauffmann, and C. Shen. LivOlay: Interactive Ad-hoc Registration and Overlapping of Applications for Collaborative Visual Exploration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1357–1360, 2008.

[11] Alexov, A., Schellart, P., ter Veen, S., van der Akker, M., Bähren, L., Greissmeier, J.M., Hessels, J.W.T., Mol, J.D., Renting, G.A., Swinbank, J. and Wise, M. Status of LOFAR data in HDF5 format. *Astronomical Data Analysis Software and Systems XXI, 461*, 283, 2012.

[12] Price, D.C., Barsdell, B.R. and Greenhill, L.J. HDFITS: Porting the FITS data model to HDF5. *Astronomy and Computing*, 12, 212-220, 2015.

[13] Manovich, L. What is Visualization? 1. *The Journal of the Initiative of Digital Humanities, Media, and Culture, 1*(2), 1-32, 2010.

[14] Ware, C. *Information Visualization: Perception for Design* 4th ed., Netherlands: Elsevier, 2013.

[15] Shneiderman, B., September. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* 336-343, 1996.

[16] Gooch, R. KARMA: a visualization test-bed. **Proceedings of * Astronomical Data Analysis Software and Systems (ADASS) V, ASP Conference Series*, G. H. Jacoby, and J. Barnes, eds. *101*, 80-83, 1996.

[17] Joye, W.A. and Mandel, E. 2003. New features of **Proceedings of * Astronomical Data Analysis Software and Systems (ADASS) XII, ASP Conference Series, H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds. 295*,489-492, 2003.

[18] Smithsonian Astrophysical Observatory. SAOImage DS9 Reference Manual. 1-129. Retrieved September 28, 2016 from http://www.astro.louisville.edu/software/sbig/archive/xmccd-4.1/xmccd-4.1e/docs/ds9/ds9_user.pdf, 2010.

[19] Mohan, P, Hawkins, C, Klapaukh, R and Johnsron-Hollit, M. Three Tools to Aid Visualisation of FITS Files for Astronomy. 2003–2006, 2016.

[20] Silubonde, S, Warner, D and Kuttel, M. Effective Visulaization of Tuberculosis Three-Drug Assays: A Design Study. *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists.* ACM, 38-48, 2016.

[21] University College London. Evaluation methods for public engagement projects. 1–43. Retrieved October 26, 2016 from https://www.ucl.ac.uk/public-engagement/documents/evaluationtoolkits/evaluationmethods/100831_Methods_for_evaluation.pdf.

[22] Abran, A., Khelifi, A., Suryn, W. and Seffah, A. Usability meanings and interpretations in ISO standards. *Software Quality Journal, 11*(4), 325-338, 2003.

[23] Jeffery, Clinton, and J.A.-G. Visualization Principles and Techniques. *Writing Virtual Environments for Software Visualization.* Springer, New York, 2015.

[24] Paul, J. Principles of Design. Retrieved October 26, 2016 from http://www.getty.edu/education/teachers/building_lessons/principles_design.pdf, 2011.

[25] Mazza, R. *Introduction to information visualization.* Springer Science & Business Media, Berlin, 34, 2009.

[26] Bennet, S., McRobb, S. and Farmer, R. *Object-oriented systems analysis and design using UML*. McGraw-Hill Education, New York, 2010.