ABSTRACT

Distant astrophysical objects like planetary nebulae can normally only be observed from a single point of view. Assuming a cylindrically symmetric geometry, one can nevertheless create 3D models of those objects using tomographic methods. We solve the resulting algebraic equations efficiently on graphics hardware. Small deviations from axial symmetry are then corrected using heuristic methods, because the arising 3D models are, in general, no longer unambiguously defined. We visualize the models using real-time volume rendering. Models for actual planetary nebulae created by this approach match the observational data acquired from the earth’s viewpoint, while also looking plausible from other viewpoints for which no experimental data is available.

Keywords: algebraic reconstruction, volumetric modelling, volume reconstruction, 3D modelling

1 INTRODUCTION

When stars not larger than a few sun masses die, they often eject part of their matter until only a small glowing nucleus is left in the center of a gaseous shell. These objects are called planetary nebulae. Their shape is often spherical or bipolar (i.e. cylindrically symmetric), but irregular shapes exist as well. Due to the radiation of the central star, the atoms in the shell get ionized and begin to emit light of characteristic wavelengths when the electrons recombine. Usually, there is not much absorption in the shell, so that only emissive effects have to be taken into account for reconstruction and visualization. A more comprehensive introduction to planetary nebulae can be found in [OF06].

In astrophysical research on planetary nebulae, being able to determine plausible models of their three-dimensional shape is an important precondition for a better understanding of the physical processes underlying their structure. For example, simulations of photoionization processes rely on a model of gas volume densities as input data and can in turn validate those models with respect to their physical realism. Interactive 3D visualizations of planetary nebulae can also be useful for scientific and educational purposes, such as digital planetariums.
While some algorithms already exist to reconstruct the 3D shapes of planetary nebulae using a single input image, we present a fast, GPU-based approach that not only outperforms existing solutions in terms of computing time, but also includes asymmetric features without requiring further input data.

2 RELATED WORK

While planetary nebulae are a common research object in astrophysics and astronomy, their three-dimensional visualization has been purely artistic work for a long time. One notable exception is the rendering of the Orion nebula by Nadeau et al. [NGN+01], who created a scientifically accurate fly-through animation of that nebula, but the 3D model they used had to be worked out by hand by astronomers. A large amount of astronomical research work is done on the subject of classifying [Cur18, KK68] and explaining [KPF78, CP83] the three-dimensional structure of planetary nebulae and to simulate the physical processes inside planetary nebulae for the proposed 3D geometries in order to confirm these observational findings [MF89, AKR00, EMB03].

However, the processes leading to the observed structures are still not well understood, and reliable determination of their 3D shape is an open problem.

Sabbadin et al. [Sah84, SCB00] as well as Saurer [Sau97] take a semi-automatic physics-based approach for the reconstruction of the 3D geometry of planetary nebulae. They assume that the velocity of a certain region of gas around the nebula is strongly correlated to its distance from the central star. Calculating the Doppler shift of some well-known emission lines allows to get the velocity component towards the observer. Combining these, depth information can be reconstructed. However, the relation between velocity and distance from the central star is generally unknown, and exact Doppler shift measurement requires elaborate experimental setups, while our reconstruction approach relies on easily available photographic images only.

Methods for tomographic reconstruction of axisymmetric objects based on a single image have been proposed by Hanson [Han93], who applies this approach to man-made objects with known and theoretically perfect axisymmetry.

Magnor et al. [MKHD04, MKHD05] present a hardware accelerated reconstruction method for planetary nebulae that works with a single photograph as input. Their analysis-by-synthesis approach is based on the assumption of axial symmetry and Constrained Inverse Volume Rendering (CIVR). While they also propose corrections for small deviations from axial symmetry, these corrections are not realized in the provided examples. Furthermore, the reconstruction using their approach is computationally very expensive.

Linju et al. have proposed a variant of the above algorithm that estimates absorption and scattering using an infrared image of the same object, allowing to reconstruct the dust density as well and thereby extending the range of reconstructible objects [LLM07b, LLM07a].

Another piece of work by Linju et al. [LHM+07] describes a method to reconstruct the volume density distribution of dust in reflection nebulae using an analysis-by-synthesis approach. The algorithm does not rely on symmetry assumptions but exploits the special properties of light transport in an environment dominated by scattering and absorption to produce non-exact but plausible 3D volumes. However, these properties are not present in planetary nebulae which are usually dominated by emission, and the method can therefore not be applied.

3 ALGEBRAIC RECONSTRUCTION

A common approach for getting three-dimensional volume models from two-dimensional images is tomographic reconstruction [KSS88]. This method is used, for example, in computed tomography (CT) to get volume densities out of multiple x-ray images of an object. While in the case of CT images the density of the object causes absorption, the contrary is the case for planetary nebulae, which emit light proportionally to the density of ionized gas. The intensity \( I_i \) of a certain pixel \( i \) in a discrete two-dimensional image of a planetary nebula is just the integral over all the emission densities along the incident light ray \( \text{Ray}(i) \) of this pixel. Using a discrete volume model, this can be written as a sum over all volume elements \( v_j \), where each summand is the length of the ray that lies within the volume element (denoted by \( |\text{Ray}(i) \cap v_j| \)), multiplied by its emission density \( \rho_j \):

\[
I_i = \sum_j |\text{Ray}(i) \cap v_j| \cdot \rho_j, \quad (1)
\]

This system of linear equations, usually written as \( Ax = b \) with \( x_j = \rho_j \), \( b_i = I_i \) and \( A_{ij} = |\text{Ray}(i) \cap v_j| \), can now be solved for the \( \rho_j \). Under certain preconditions, this gives a unique solution for the volume emission densities.

For the solution to be uniquely defined, the rank of the matrix \( A \) must be at least equal to the number of volume elements. This is usually achieved by using images from multiple viewpoints. In practice, the system will almost always be overdetermined, as one would rather use more pixels than necessary to get more stable results. An approximate solution can then be computed using iterative algorithms that minimize the 2-norm \( ||Ax - b||_2 \) of the residual error.

For most astrophysical objects, getting images from multiple viewpoints is impossible due to the large distance to those objects. This means that if a regular grid
of volume elements (or voxels) is used, the linear system is not uniquely defined and would in fact not yield any information about the three-dimensional structure of the object.

This problem can be solved by making additional assumptions about the geometry of the object, e.g., an axial symmetry that is common in planetary nebulae. Such symmetries can easily be reflected in the choice of voxels by combining all regions of space that are assumed to have the same emission density into one voxel (Fig. 2). This can reduce the three-dimensional complexity of the voxels \((x, y, z)\) to a two-dimensional one in the case of cylindrical voxels \((r, z)\) so that the solution of the linear system is unique.

The appearance of axisymmetric models obviously depends strongly on the choice of the symmetry axis. Several possibilities exist for determining a plausible solution of the linear system is unique.

4.1 Specifying the Linear System

In order to specify the system of linear equations (eq. 1), we need to calculate the matrix elements \(A_{ij}\) and the right-hand vector \(b\). While the \(b_i\) are already given by the intensities of the pixels, setting up the \(A_{ij}\) requires further calculations.

We recall that \(A_{ij} = |\text{Ray}(i) \cap v_j|\) is the length of the ray through pixel \(i\) that lies within the volume element \(j\). This means we have to calculate the intersection points of lines with voxels that have the form of a hollow cylinder (cf. Fig. 2). Since the viewing angle is usually very small due to the large distance to the object, we can assume that all the incident light rays are orthogonal to the image plane, so that they are defined as

\[
r(t) = \begin{pmatrix} p_x \\ p_y \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix},
\]

where \(p_x\) and \(p_y\) are the pixel \(x\) and \(y\) coordinates, respectively. This means we are using a three-dimensionally extended version of the image coordinate system for our calculations.

4.2 Solving the Linear System

The linear system \(Ax = b\) must now be solved for the voxel intensities \(x\). Since the system is usually overdetermined, in general only an approximate solution minimizing the residual norm \(||Ax - b||_2\) is possible. This solution can be determined by iterative algorithms such as the Conjugate Gradient Least Squares (CGLS) method [Han96].

The fundamental idea of the Conjugate Gradient algorithm is that solving \(Ax = b\) with \(A\) symmetric and positive definite is equivalent to minimizing \(f(x) = \frac{1}{2}x^TAx - b^Tx\). Starting from \(x = 0\), each iteration step \(k\) modifies the intermediate solution vector \(x_k\) by descending in the direction of the gradient of \(f(x)\), so that \(x_{k+1} = x_k + \epsilon_k \nabla f(x)|_{x=x_k}\) (where \(\nabla\) is the gradient operator). For fast convergence, it is important that \(\epsilon_k\) is computed such that \(x_{k+1}\) ends up close to the one-dimensional minimum in direction \(\nabla f(x)|_{x=x_k}\), and that the directions of descent are conjugate to each other, that means that for all directions \(d_i = \nabla f(x)|_{x=x_i}, d_j = \nabla f(x)|_{x=x_j}: d_i^TAd_j = 0\).

In our case, however, the matrix \(A\) does not fulfill the above conditions. But since for any matrix \(A\) the matrix product \(A^TA\) is symmetric and positive definite, we can multiply our system \(Ax = b\) by \(A^T\) and solve \(A^TAx = A^Tb\) instead. The CGLS implementation does this multiplication implicitly, preserving sparsity of the matrix \(A\), which allows for more efficient (memory saving) algorithms.
Since the norm $|x|_2$ of the intermediate solutions increases monotonically during the iteration, it is necessary to start the iteration with $x = 0$ in order to not exclude any possible solution. The residual norm, on the other hand, is guaranteed to decrease monotonically, so convergence is guaranteed if numerical errors can be neglected. In practice, the iteration can be stopped as soon as the convergence speed falls below a chosen minimal value.

In the original algorithm, the value range for the vector $x$ is not restricted in any way. Particularly, the entries of the solution vector can be negative. Since negative emission intensities are impossible (they cannot even be regarded as a physically valid model for absorption), the intermediate solutions have to be projected onto the subspace of positive solutions after each step so that the positivity of the solution is guaranteed [IM04].

After this step, we have obtained a radial map of the model, that is a 2D grid of densities whose axes are the $r$ and $z$ cylinder coordinates of the corresponding voxel. Rotating this map around the $z$ axis gives the full axisymmetric 3D model.

Implementations of iterative least-squares solvers for linear systems are widely available (Matlab’s \texttt{x=A} operation, for example), but normally do not allow for additional restrictions to the solution vector $x$. Since we need to guarantee $x_j \geq 0$ for all vector components, we adapt an existing implementation [han] to suit our needs.

For the matrix manipulations that are used in the algorithm we make use of the GPU’s parallel computing power using the nvidia CUBLAS library\footnote{http://www.nvidia.com/object/cuda_develop.html} library. This approach speeds up the reconstruction process by two orders of magnitude (Fig. 3) with respect to a purely CPU-based implementation of the same library\footnote{The UBLAS library from http://www.boost.org/doc/libs/1_35_0.html}. However, the matrix size in the GPU accelerated version is limited by the graphics card memory and the maximum texture size\footnote{On our setup with 768 MB of video RAM, the model is restricted to about 7000 entries in the radial map, or about 120 slices and 60 rings.}, but the resulting limit on the model size is usually not much smaller than the limit imposed by the quality of available input images.

### 4.3 Correcting for Asymmetries

While from a macroscopic point of view many planetary nebulae show axial symmetry, on a smaller scale there is always some deviation from perfect symmetry. This can be seen in the residual image that is left when the projection of the reconstructed model onto the image plane is subtracted from the real image. We distribute this residual intensity among the voxels of our model so that the projection equals the original image (cf. Fig. 5).

Since there is no depth information available for the asymmetric part of the intensity, this distribution will always be ambiguous, so we have to choose a “distribution function” that gives subjectively good results when viewed from different angles.

To break up the axial symmetry, the model of cylindrical voxels is first converted into a model of cubic voxels. For efficiency, the model is aligned to the image plane, so that its $x$ and $y$ resolution equal the image $x$ and $y$ resolution. Then $z$ is resolved is such that the whole model fits into the voxel volume. Since by that choice of the cubic voxel model every voxel only contributes to a single pixel, we do not need to take interdependencies between voxels into account. So the only decision that is left is which voxels that share common $x$ and $y$ coordinates will get how much of the residual intensity of the pixel $(x,y)$.

A convenient distribution function turns out to be the following: Each voxel gets an amount of residual intensity that is proportional to the amount of intensity it already contributes to the pixel intensity\footnote{Alternative approaches could use more assumptions about where asymmetries are likely to occur. For example, one might assume that the exploding star itself is perfectly symmetric and only when the expanding gas cloud hits small space debris, it is asymmetrically deformed. This would imply that the residual should preferably be applied to the outer regions of the shell.}. So if the reconstructed pixel intensity is $I_p = \sum_{v \in V} I_v$ where $V$ is the set of all voxels that contribute to the current pixel, the residual intensity is $I_r$ and the original pixel intensity is $I_o = I_p + I_r$, the new voxel intensities are

$$I'_v = I_v + \frac{I_r}{I_p} I_v = I_v \left(1 + \frac{I_r}{I_p} \right) = I_v \left(1 + \frac{I_o}{I_p} \right).$$

As this is just a multiplication of all voxels that are projected onto the same pixel with a common value

![Figure 3: Comparison of computing time for CPU and GPU based implementations, in logarithmic scale, for matrices of size 1x1 up to size 14000x14000.](image)
\( \frac{L_p}{r_p} \), this is efficient to calculate knowing the projected intensity \( I_p \) and the residual \( I_r \). The function is also optimal in the sense that it guarantees non-negative voxel intensities whenever possible, which in our case is always the case because the pixel intensities are always non-negative. This means that the rendered model will exactly reproduce the observed image when rendered from the original point of view, provided that the voxel resolution is large enough. It also preserves visual coherence between neighboring voxels which usually have similar intensities.

4.4 Visualizing the Results

The output of the reconstruction algorithm is a three-dimensional grid of cubic voxels, each of which has a certain emission density. To visualize this grid from an arbitrary viewpoint, volume ray casting can be used. Since the model is purely emissive, this means that from each screen pixel a ray is cast through the volume and intensities along the ray are summed up. This volume ray casting process is well suited for implementation on graphics hardware. The voxel model, for example, can easily be represented as a three-dimensional texture, and integrating the intensities along a given ray can be approximated by summing up the texture values at a number of fixed-distance sampling points along the ray using a fragment shader.

In order to get an impression of the chemical composition of a planetary nebula, reconstructed voxel models for different wavelengths can be shown simultaneously, assigning a color to each of them. Since the interesting spectral lines are often too close to each other (like the hydrogen and nitrogen lines at 656nm and 658nm, respectively) or even invisible to the human eye, false-color display is used. In the simplest case, when three different source images are to be displayed, these are naturally assigned to the red, green and blue color channels.

5 RESULTS

5.1 An Artificial Test Case

In order to verify the accurateness of our reconstruction, we first reconstruct an artificial model with perfect axisymmetry. For this, a radial map of intensities is drawn and projected into the image plane at different inclination angles. The resulting images are then reconstructed and the reconstructed intensity maps are compared to the original. We can show that for inclinations at least up to 45 degrees, the reconstructed radial map very closely resembles the original (see Fig. 4).

5.2 NGC7009 (Saturn Nebula)

The Saturn Nebula (Fig. 5), discovered by William Herschel in 1782, shows a bright, slightly S-shaped structure in the center, surrounded by a darker, barrel-shaped one. The S-shaped structure has noticeable reddish glowing tips. The original images were moderately disturbed by stars, so slight preprocessing had to be done.

5.3 Mz3 (Ant Nebula)

The Ant Nebula, discovered in 1922 by Donald Menzel, has a number of different gaseous outflows from its bright center. The most visible outflow consists of two approximately spherical lobes, but more subtle “rays” can also be observed outside these lobes. Interestingly, all these features share a common axis of symmetry, so in principle the simultaneous reconstruction of all important features would be possible (Fig. 6). However, due to the large difference in intensity and the linear scale chosen in the visualization, these structures cannot be observed together in one output image. Anyway, the interesting fine-grained asymmetries of the spherical lobes are visible in the asymmetry-corrected output (Fig. 7).

Due to the presence of many bright stars in the original images, heavy preprocessing needed to be done. This may have caused loss of fine-grained structures in some areas.

Magnor et al also reconstructed the Ant Nebula using their CIVR approach. The algebraic reconstruction algorithm that we implement outperforms CIVR – which leads to results that are optically very simi-
lar to the purely symmetric part of our reconstruction – by far: while using CIVR, “the reconstruction of a 128x32-pixel density map takes approximately one day on a 2.4 GHz PC in conjunction with an nVidia GeForce FX 3000 graphics card” [MKHD04], our algebraic approach can reconstruct the nebula with comparable resolution in a matter of seconds, including asymmetry correction.

5.4 NGC6543 (Cat’s Eye Nebula)
The Cat’s Eye Nebula (Fig. 1), discovered by William Herschel in 1786, has a very complex and not quite axisymmetric structure. Its reconstruction is nevertheless quite accurate due to the asymmetry correction. Due to its relatively large brightness compared to the surrounding stars, no preprocessing was needed to get clean results.

6 CONCLUSIONS
We have presented an efficient algebraic reconstruction approach to derive 3D information from single images of axisymmetric and purely emissive objects like planetary nebulae. Axial symmetry is broken in a controlled way in order to achieve closer resemblance between the model and observational data. The calculations are carried out efficiently by making use of the GPU’s parallel computing power, and the resulting models closely resemble the actual photographs.

There are, however, some limitations inherent to our approach. The asymmetry correction is not based on any physical measurement and can only be heuristically and ambiguously determined, as long as no additional data is provided. The reconstruction is also only possible for nebulae whose axis of symmetry is not too far inclined with respect to the image plane because no reliable 3D information can be derived if the axis is close to parallel to the viewing direction. For the same reason, objects without symmetry cannot be reconstructed at all.

However, for axisymmetric objects the results closely resemble the original. The CGLS algorithm is guaranteed to converge, and small asymmetric features can be included in such a way that the original image is exactly reconstructed. For objects with spherical symmetry, the algorithm is also applicable, though it could be optimized further to benefit from the stricter constraints.

The resulting three-dimensional models can be rendered from arbitrary perspectives. This allows for a wide field of applications, in scientific as well as artistic contexts. They may help understanding the complex structure of planetary nebulae and the physical mechanisms that underlie their formation.

In ongoing research we evaluate the possibility of allowing a wider variety of symmetry constraints. A user-supplied set of model parts such as cylinders, spheres, and other geometric primitives that may represent a physical explanation of the shape of the object could be used instead of just a single cylinder. This would allow to model more complex symmetries such as the helix that is present in the Saturn Nebula.

In order to assure physical realism, depth information from Doppler shift measurements could easily be incorporated into the error function, which would allow to loosen the symmetry constraints where better depth information is available.

ACKNOWLEDGEMENTS
This project is partially funded by the German science foundation DFG under grant 444 MEX-113/250-1 and the Mexican science foundation CONACyT under a bilateral cooperation grant.

All astronomical images taken from the Hubble Space Telescope [HST], partly acquired through the MAST search engine [MAS].

REFERENCES


[Han] netlib.org CGLS algorithm. file REGU/cgls.m from http://netlib.org/numerical/na4-matlab7.tgz.


Figure 5: The Saturn Nebula. From top to bottom: images taken using filters for 502nm, 555nm and 658nm that are assigned to the red, green and blue color channels, respectively; reconstruction without asymmetry correction; reconstruction with asymmetry correction, from earth; reconstruction with asymmetry correction, from outer space. Note how the S-shape of the original image is lost during reconstruction because it violates the axisymmetry constraint, and how the asymmetry correction restores this important feature.
Figure 6: Reconstruction of the Ant Nebula without asymmetry correction: radial map (top) and reconstructed view from earth (middle) and from outer space (bottom). The red, green and blue color channels are assigned to 673nm, 658nm and 487nm, respectively.

Figure 7: The Ant Nebula. Top: images taken using filters for 487nm, 658nm and 673nm that are assigned to the red, green and blue color channels, respectively. Middle and bottom: asymmetry-corrected reconstruction, view from earth and from outer space. In the bottom image, artifacts of the asymmetry correction due to overly bright regions in the input image can be observed as lines parallel to the viewing direction of the original image.