Feature Extraction from Pioneer Venus OCPP Data

Freek Reinders [1] Frits H. Post [1] Hans J.W. Spoelder [2]

[1] Delft University of Technology
Faculty of Technical Mathematics and Informatics
PO Box 356, 2600 AJ Delft, The Netherlands

[2] Free University Amsterdam
Faculty of Physics and Astronomy
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

Abstract. Scientific visualization provides means to explore data and highlight interesting features in the data. In this paper we will discuss the visualization of astrophysical data. Light properties of sunlight scattered by the atmosphere of Venus were measured by the Pioneer Venus Orbiter. One of the objectives of this mission was to determine the properties of the clouds and haze in the atmosphere.
Given the amount and complexity of the data, it is important to be able to browse through the data and select maps with interesting features. We built a system that reads the raw data, prepares it and extracts cloud features. The feature extraction is achieved by the following steps: selection, clustering, attribute calculation and iconic mapping. After data exploration a number of consecutive images with coherent moving cloud features, is found. From the center position and the time between two frames, a qualitative measure for the cloud velocities is derived. The obtained velocities are well in correspondence with generally accepted results.
Thus we have showed that visualization techniques are powerful tools to browse through the data, recognize cloud features and determine the motions of the features in time.

Keywords: scientific visualization, feature extraction, data exploration, astrophysical data.

1 Introduction

On May 20th 1978 the Pioneer Venus Orbiter was launched by a Centaur launch vehicle from Kennedy Space Center, Florida USA. It arrived at Venus the 4th of December 1978 and was placed in orbit, where it stayed for almost twelve years. Among the 17 instruments aboard was the Orbiter Cloud Photo Polarimeter, or OCPP. The instrument measured the intensity and polarization of sunlight reflected by the atmosphere. Unpolarized sunlight is absorbed and scattered by cloud particles, the intensity and polarization characteristics of the reflected light depend on the structure of the atmosphere. The principal objective of the Pioneer investigation was to determine the properties of the clouds and haze, including the vertical and horizontal distribution of the particles, cloud particle size and refractive index. Also the variations in cloud morphology and the nature of the cloud motions are investigated.

The 2Gb of data is organized in 4000 so-called maps that contain collections of 2D spherical coordinates (the radius is assumed to be fixed) and the measured quantities

of the scattered light. Considering the large amount and the complexity of the data, visualization techniques are worthwhile. We developed a system to browse through the data, extract interesting features and visualize them. The combination of visual inspection and automatic feature extraction is a powerful tool to explore the data. The features extracted from the OCPP data are the cloud formations seen in the intensity and in the degree of polarization. For each cloud feature, a set of attributes is calculated and the features are visualized by ellipse icons. Displaying the ellipses of successive images gives us means to find coherent motion of cloud features. From the values of the center points in the attribute set and the time between two frames, we can give a qualitative measure for the cloud velocities. The obtained wind speeds are verified by results from astrophysical research.

These results are obtained in various ways (Rossow et al 1980 [3], Rossow et al 1990 [4], Toigo et al 1994 [7] and Smith and Gierasch 1996 [6]). The obtained velocities are divided in a zonal and a poleward component. It is generally accepted that the mean zonal wind speed corresponds to a solid rotation with an average equatorial speed of 93 m/s and the mean meridional wind is poleward, in the order of 10 m/s. This is proportional to an average global circulation periodicity of 4.5 days. The results are time dependent; they change over a period of weeks and they are a function of the latitude. These results will be used to verify the results obtained by our method.

This paper is organized as follows. In section 3 we will discuss the properties and some of the preprocessing of the OCPP data. Then we explain the feature extraction techniques in section 3, and in section 4 we present the results. Finally in section 5 the conclusions and future research topics are given.

2 The Pioneer Venus OCPP Data

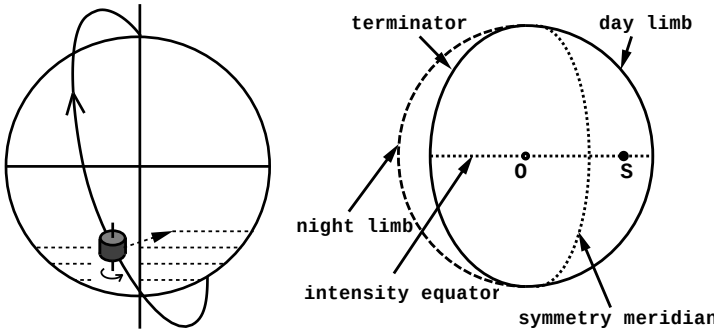
2.1 Data generation

The orbiter, rotating around his axis, measured the data in scan lines across the surface, see Figure 1a. During every scan line one of the four different wavelengths (270, 365, 550 and 935 nm) is measured. The surface is thus scanned by about 50 scanlines for every wavelength. One planet surface measured in this way is called a *map*. The measured properties of the reflected light are: the intensity I, degree of linear polarization $|P|$ and the direction of polarization χ relative to the local scattering plane. Furthermore, the positions of the scan points, the sub-solar point and the sub-orbiter point are given in spherical coordinates. The coordinate system is fixed in space, with the positive x-axis pointing in the direction of the constellation Aries, i.e. it does not rotate with Venus.

wavelength	270	365	550	935
data	I	$\|P\|$	χ	
position	longitude		latitude	
scattering properties	μ_0	μ	θ	$\phi - \phi_0$

Table 1: Overview of the data measured by the Pioneer Venus OCPP.

In Figure 1b) a diagram of the situation is given as seen from the orbiter's point of view. In this Figure O is the sub orbiter point and S is the sub solar point. The visible hemisphere has as its center the O point, while the illuminated hemisphere has S as its center. These two hemispheres overlap in a symmetric segment on the planet's surface, bounded at one side by the day limb and at the other side by the terminator. The meridian which cuts it into equal halves is called the *symmetry meridian*. The line through O and S is called the *intensity equator*. Both are important symmetry axes [1].



(a) Scanning over the surface

(b) Measurement situation

Figure 1: Diagram of Venus and the Pioneer Venus Orbiter.

With the positions of the sun and the orbiter, the local scattering geometry (Figure 2) is determined: μ_0 is the cosine of the angle between the incoming light and the normal, μ is the cosine of the angle between the outgoing light and the normal, the azimuth angle $\phi - \phi_0$ is the angle between the plane of incidence and the plane of reflection and the phase angle θ is the angle between the incoming and outgoing direction of the light. In Table 1 an overview of the data is given.

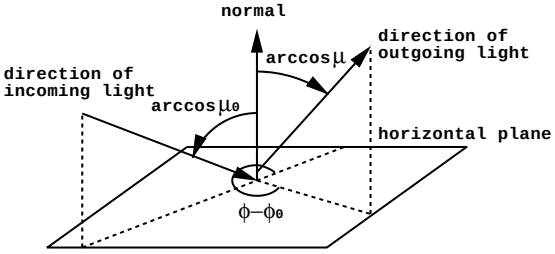


Figure 2: Local scattering geometry.

The scattering properties and the time between two measurements are different for each map. Since the polarization degree also depends on the phase angle, the data range in $|P|$ can be considerably different between two successive maps. Furthermore, the intensity and polarization quantities are not always related to each other. This causes the data to be inaccessible and hard to interpret. Therefore, it is important to be able to browse through the data: select specific map numbers, wavelengths and data quantities. Special routines to read the raw data were written in AVS. The visualization is realized by the following steps: resampling, limb darkening correction and data presentation.

2.2 Resampling

Since the data was measured in scan lines at non-uniform intervals, it is delivered as scattered data. In order to obtain the data on a regular grid, the data is resampled. This

is achieved by a nearest-neighbour interpolation algorithm (Figure 3). The area around each grid point is divided into a number of sectors. For each sector the nearest raw data point, within a certain *search radius*, is determined. If for all sectors the nearest neighbour is found, the interpolation will take place by distance weighting over these points.

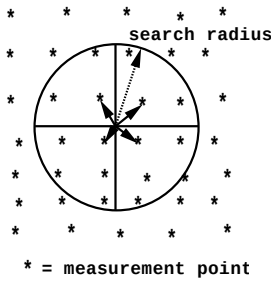


Figure 3: Resampling the raw data to a regular grid.

A good choice for the number of sectors is three or four. The search radius may not be too small: this means loss of data when the distance between raw data points is larger. However too large means searching in a large area, which means more computation time, while the effects are small. The best choice for the search radius is about two or three times the average distance between the raw data points.

2.3 Limb darkening correction

The geometry of a spherical object reflects the light with a certain *limb darkening*: the intensity is strongest in the middle and weaker near the edges. This is visible in the brightness distribution at wavelength $\lambda = 550$nm (Figure 4). For the higher wavelengths (550 and 935nm) scattering by gas molecules is less significant; therefore the influence of atmospheric properties on the intensity is small. However, the limb darkening also affects the lower wavelengths. This causes features near the limbs to be blurred or even become invisible.

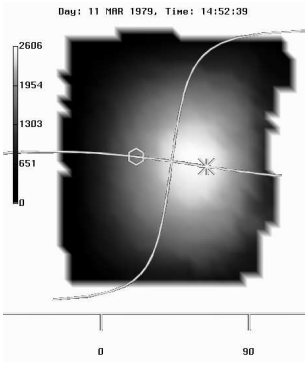


Figure 4: The limb darkening effect: the brightness distribution at wavelength $\lambda = 550$nm.

These features are enhanced by correction for the limb darkening based on Lambert's law. This law holds under the assumption that the light is reflected with equal intensity in all directions (perfectly diffuse reflection). This leads to the following correction procedure (Rossow et al [3]). First a disk integrated scaling factor I_d is calculated, then the appropriate zenith angle is subtracted from the raw intensity I^r.

$$I_d = \sum_n I_n^r [\sum_n \mu_{0n}]^{-1} \quad (1)$$

$$I_n^c = I_n^r - I_d \mu_{0n} \quad (2)$$

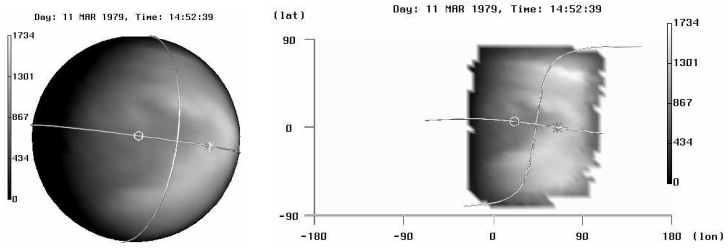
Thus the corrected intensities are an indication of intrinsic brightness deviations from the mean brightness over the disk. They can have positive or negative values, corresponding to relatively bright or dark features. This correction can be performed on the raw intensity data (e.g. before the resampling step). Two other correction methods have also been implemented; however, these do not give significantly better results [2]. Errors are estimated by calculating the Relative Difference Measure or RDM:

$$RDM = \sqrt{\frac{\sum_n (I_n^r - I_n^c)^2}{\sum_n (I_n^c)^2}} \quad (3)$$

where I_n^r is the raw intensity field in node n and I_n^c the correction field in node n. The accuracy of a method is measured by calculating the RDM between the brightness distribution at wavelength $\lambda = 550$nm and the correction field.

2.4 Data presentation

A colour presentation of the data can be given in two ways. Since the data is available in spherical coordinates, with the radius constant, it can be presented as a 2D map. The range of longitude is from -180 to 180 degrees and the latitude lies between -90 to 90 degrees. However, the data will be horizontally stretched out near the poles, which also deforms feature patterns. This can be avoided by projecting the data on a sphere (in Cartesian coordinates). In Figure 5 both presentations are displayed. The colours indicate the values of the scalar data. The time and date of the measurement are shown in the header. The positions of the sub orbiter and sub solar point are marked by o and * respectively, and the two symmetry axes are visualized by lines.



(a) Cartesian coordinates

(b) Spherical coordinates

Figure 5: Two types of data presentation, day 169, λ = 365 nm.

Thus we have created means to browse through the data, select a specific map, wavelength and data quantity and view it with user defined colour maps in one of the projections shown in Figure 5. The browser can be used to search significant patterns or features in the data by visual inspection. These features can than be detected automatically and visualized as described in the next section.

3 Feature Extraction

In this section we will discuss our technique of feature extraction. This method was described earlier by van Walsum [8]. The following steps are performed:

- Selection.
  The creation of selections is achieved by evaluation of a selection expression. The expression is specified by a language that consists of many operators and functions (unary, binary, statistical and gradient functions). The language uses a C-like syntax. With this, one can create virtually any selection that is based on combinations of different data values, or quantities derived from data values, and on the grid node positions.

- Clustering.
  In order to handle regions of selected nodes as entities it is necessary to perform clustering. Neighbouring selected nodes are grouped into regions based on a connectivity criterion. Each group is labeled for reference.

- Attribute calculation.
  For each feature an attribute set is calculated in order to characterize the shape, size, position and orientation of the feature. Integration over the selected nodes leads to the volume/surface area, center point and the second moments.

- Iconic visualization.
  The attribute set is mapped onto the parameters of a geometric object or icon (van Walsum et al [9]). The goal of iconic mapping is to visualize essential elements of a data set in a symbolic way. The icon must be clear, compact, meaningful, and it should be related to the physical concepts and visual languages of an application.

This is a process of *abstraction*: a region in a data set is viewed and treated as an entity, and is visualized at a higher level. The meaning of a selection is changed: previously a selection was viewed as a set of nodes that contains interesting data, while here a selection relates to a feature. Therefore the term feature applies to any region for which an abstract representation and visualization is worthwhile. That is why the process of selection, clustering, and attribute calculation is called feature extraction.

3.1 OCPP cloud features

In order to extract cloud features from the OCPP data, first a suitable selection criterion has to be found. The selection expression works with threshold values, and thus it is important to have a mechanism to determine a suitable value. It may consist of an absolute value or a function of certain statistical parameters. It must be invariant for different maps. Finding a proper threshold is a matter of exploration of the data by the user. The definition of features cannot be given by universal rules, but depends on the underlying physics.

Generally speaking, a cloud feature is defined as a deviation from the average value. A region with values below the surrounding is a “dark” feature and a region with higher values is a “light” feature. Obviously statistical parameters such as the mean or the

standard deviation (SD) are useful. If the distribution of the data is Gaussian, two thirds of the data values lie within one SD and 95% lies within two SD's from the mean. In both cases the criterion is invariant for different maps. Here another advantage of the limb darkening correction arises: the correction creates a Gaussian distribution. The relatively dark features have negative values and the bright features have positive values. In the expression below a selection of dark cloud features from the corrected intensity field at wavelength 365nm, labeled **I365**, is made.

ranged	=	(**I365** - avg(**I365**)) / sd(**I365**)
threshold	=	1.0
select	=	**ranged** $<$ -threshold

The simplest way to visualize the position and the size of a selected feature is with an ellipse icon. The average position of the cluster is mapped to the position and the second moments determine the length and the orientation of the axes. In case of the 2D ellipse a minimal attribute set of five parameters is needed. The surface area, center point and second moments are calculated by means of integration over the selected nodes. In Figure 6 a selection is made from intensity data with a wavelength $\lambda = 365$nm. The intensity data is corrected and the selection is made by the expression described above, e.g. dark features are extracted. The selected points are marked by the small white crossmarks and the ellipses visualize the cloud features.

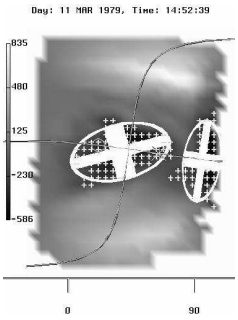


Figure 6: Iconic presentation of the selected features.

3.2 Cloud tracking

Now we have means to automatically extract features from the OCPP data. In order to find a series of maps with coherent moving features, we first make a rough selection from all maps. The following two restrictions are applied: the time between two successive images must be less than 12 hours and the phase angle θ must be less than 80^o. The first restriction is based on the expected speed of motion (4.5 days circulation) and the size of the visible part. Half the planet's surface is visible and this part becomes smaller with increasing phase angle; therefore the second criterion is introduced. These two criteria reduce the number of useful maps to a small number of series.

The remaining maps are then explored for coherent features using the browser. Intensity and polarization data may show different features, so both types of data are explored. Correspondence between two cloud features in two frames is established manually, based on the following criteria:

- Position.
  The feature in the second frame must be at the position expected from the position and velocity (known in advance) of the feature in the first frame.

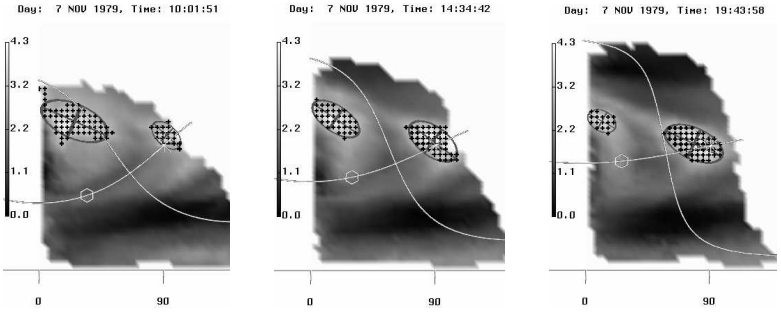
- Volume (or surface area).
  The two features must have approximately the same size.

From the calculated center positions and the time interval between two frames, the velocity of the features can be estimated.

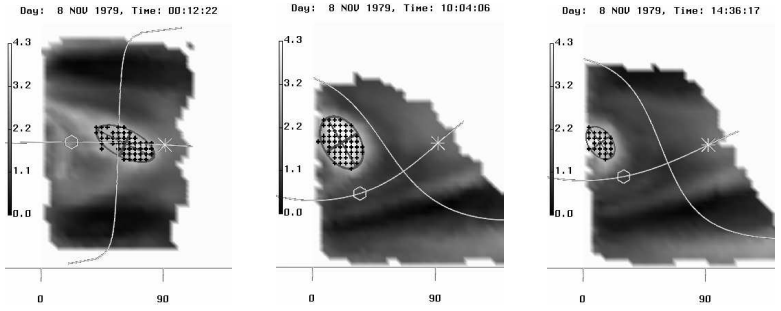
4 Results

4.1 Cloud tracking

The series shown in Figure 7 is an example of a series with coherent moving features. It covers about 28 hours and the phase angle changes between $63.2 < \theta < 68.8$. Since the phase angle is almost constant, the threshold can be taken as an absolute value. The selection is based on the polarization degree: $|P| > 3.0$. In the figure six frames are shown in which features move from right to left. In the first three frames two features are visible. Based on the positions and consistency we can say that the left feature moves out of the image and that only the right feature is still apparent in the fourth frame. In the frame f) it also almost disappears.



(a) t = 0 h (b) t = 4.5 h (c) t = 9.7 h



(d) t = 14.2 h (e) t = 24.0 h (f) t = 28.5 h

Figure 7: Determination of circulation velocity by means of feature tracking, day 507-512, $|P|$ data, λ = 365 nm.

In Figure 8 the longitude and the latitude of the second feature are plotted as a function of the time. In about 28.5 hours the feature has moved from longitude = 91.9 to 12.5 degrees. During this period the latitude was almost constant. This means that the circulation time is about 5.4 days. However in the beginning and at the end, part of the feature was outside the measurement area. This affects the position of the center point and the approximation of the circulation velocity. The effects of the borders may be avoided by checking the surface area of the feature: the surface of the feature in the first and the last frame are much smaller ($\pm$300 degrees^2) than the rest ($\pm$750 degrees^2), which is a reason to skip them. The second is also near the border, however it is almost as large as the others ($\pm$680 degrees^2) and therefore we can take it into account. The result is a circulation time of 4.4 days. This result is very well in correspondence with the expected average of 4.5 days. This supports the conclusion that the extracted feature is a real moving cloud formation.

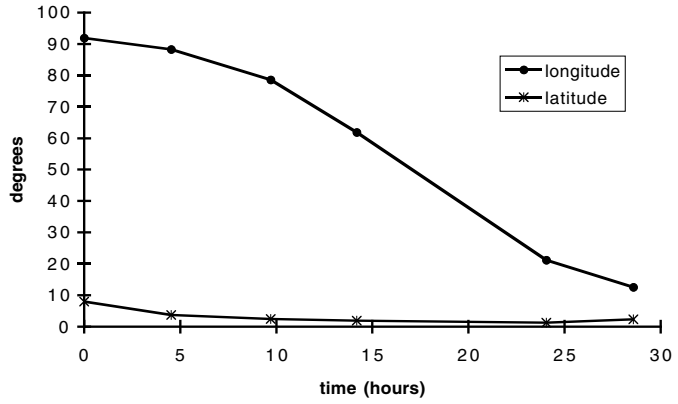


Figure 8: Movement of the center point of the feature.

5 Conclusions and Future Research

The data measured by the Pioneer OCPP experiment consists of a large, complex data set. The scattering geometries and the time between two measurements differs per map. This causes the data to be very inaccessible and hard to interpret. Therefore it is important to be able to browse through the data and select maps with interesting features. We built a system that reads the raw data, prepares it and extracts cloud features automatically. The correction of the intensity data for the limb darkening gives a good enhancement of features. A better distinction can be made between dark and light features.

Features are tracked by performing the feature extraction on consecutive images. However it takes some data exploration to find a suitable series of images with coherent moving features. The center positions of these features are used to approximate the velocity. Features near the border might influence the velocity, but the surface area can be used as an indication to include them or not. Tracking the large scale cloud features in the Pioneer Venus OCPP data results in a circulation time of 4.4 days. This is in correspondence with the average 4.5 days rotation found by Rossow at al [3].

We believe that these results justify the development of an algorithm for automatic feature tracking based on calculated attribute sets. The attribute set should hold the proper

information in order to determine feature translation, rotation and scaling. In addition, the algorithm should be able to recognize the events like: continuation, creation, dissipation, bifurcation and amalgamation [5]. In order to recognize these events, a number of test criteria for the attributes can be implemented.

Ellipses do not always give a good impression of size and position. For instance a long curved feature is not properly visualized by an ellipse. In order to visualize such features, icons must be used which also account for the shape of the feature. One way to generate shape attributes, is the determination of the medial axes or skeleton. This may be achieved by morphological operations on a cluster of selected nodes.

6 Acknowledgments

The authors wish to acknowledge the stimulating discussions with dr. J.F. de Haan and drs. R. Braak on the interpretation of the OCPP data.
This work is supported by the Netherlands Computer Science Research Foundation (SION), with financial support of the Netherlands Organization for Scientific Research.

References

[1] J.W. Hovenier. Principles of symmetry for polarization studies of planets. *Astron. & Astrophys.*, 7(1):86–91, 1970.

[2] Freek Reinders. Visualization of the pioneer venus ocpp data. Technical report, Delft University of Technology, The Netherlands, 1997.

[3] William B. Rossow, Anthony D. Del Genio, and Timothy Eichler. Cloud-tracked winds from pioneer venus ocpp images. *J. of Atmos. Sci.*, 47(17):2053–2082, 1990.

[4] William B. Rossow, Anthony D. Del Genio, Sanjay S. Limaye, Larry D. Travis, and Peter H. Stone. Cloud morphology and motions from pioneer venus images. *J. of Geophys. Res.*, 85(13):8107–8128, 1980.

[5] Deborah Silver and X. Wang. Volume tracking. *Proc. IEEE Visualization '96*, pages 157–164, 1996.

[6] Michael D. Smith and Peter J. Gierasch. Global-scale winds at the venus cloud-top inferred from cloud streak orientation. *Icarus*, 123:313–323, 1996.

[7] A.D. Toigo, P.J. Gierasch, , and M.D. Smith. High-resolution cloud feature tracking on venus by galileo. *Icarus*, 109:318–336, 1994.

[8] Theo van Walsum. Selective visualization on curvilinear grids. *PhD Thesis*, 1995. Delft University of Technology, The Netherlands.

[9] Theo van Walsum, Frits H. Post, Deborah Silver, and Frank J. Post. Feature extraction and iconic visualization. *IEEE Trans. on Visualization and Computer Graphics*, 2(2):111–119, 1996.