Auroral Feature Detection by combined Image Segmentation and Clustering Analysis

SM41A–1689

P. Guio[1], N. Achilleos[1], R. Prangé[2]

[1] University College London, UK [2] Observatoire de Paris, France

Email contacts: p.guio@ucl.ac.uk, nick.achilleos@ucl.ac.uk, renee.prange@obspm.fr

Download poster from `http://www.ucl.ac.uk/~ucappgu`

Abstract

We extend our work on image segmentation (VOISE algorithm, [9]) where we used VOISE to construct segmentation maps of auroral images from the Jovian polar regions. In this paper we transform such segmentation maps to a system of magnetic latitude and longitude. This transformation makes use of an eccentric dipole representation of the Jovian internal magnetic field that corresponds to the coefficients of the VIP4 model [3]. We comment on the efficiency and reliability of this approach for enabling the characterisation of arc-like features, approximately aligned along contours of constant magnetic latitude, and regions linked to the polar cap and the cusp. We also comment on the role of the 'coherence' of data in detection of features near the background levels.

Introduction

Planetary auroral emission is a particularly useful diagnostic tool for the study of magnetospheric processes. In the polar regions of the planet currents flow into and out of the upper atmosphere, and energetic particles precipitate down the magnetic field lines, impacting on the molecules in the atmosphere. At Jupiter and Saturn, the dominant atmospheric particle is hydrogen, which emits in the UV when excited by auroral electron impact. In order to observe this UV emission we need to observe from above the Earth's atmosphere, and so we use the Earth-orbiting HST [2].

The auroras of the outer planets have been studied relatively little in comparison to the Earth's, however we have advanced our understanding of the environment of Jupiter and Saturn. Jupiter's magnetosphere is the largest cavity in the solar wind flow, i.e. within the heliosphere, and the dynamics of it are largely dominated by the fast rotation of the planet (period $\sim 10\,\mathrm{hr}$). It has been established that the most significant component of Jupiter's auroras, the main auroral oval, is driven by planetary rotation.

Jupiter exhibits two other classes of auroral feature: the moon footprints and the polar auroras. The polar auroras are still somewhat of an enigma, as they are the most highly variable component of Jupiter's auroras in terms of brightness. The location and variation of these auroral emissions make it possible that at least some of them may be caused by the solar wind interaction, although this has yet to be proved conclusively. [11] shows good evidence that Jupiter's polar cap boundary, i.e. the perimeter of the region where planetary field lines open into the solar wind, is coincident at times with some of these polar emissions.

Magnetic coordinates and VOISE segmentation

In order to define a magnetic system of coordinates, we used the Schmidt coefficients of order $l \leq 2$ from the VIP4 model of Jupiter's magnetic field by [3]. These coefficients from spherical harmonic expansions can be manipulated in a standard manner to generate an eccentric dipole approximation to the full magnetic field model (e.g. [8]). For the VIP4 coefficients, the equivalent dipole is displaced $0.1\,\mathrm{R_J}$ from the planet centre.

The orientation of the dipole is defined by its angle of tilt with respect to the rotation axis $\theta_d = 9.5\,^\circ$ and its longitude $\lambda_d = 159.2\,^\circ = 200.8\,^\circ\,(\mathrm{SIII})$. Magnetic colatitude θ_μ and local time T_μ are defined using the eccentric dipole position and orientation, as well as the Jupiter-Sun direction.

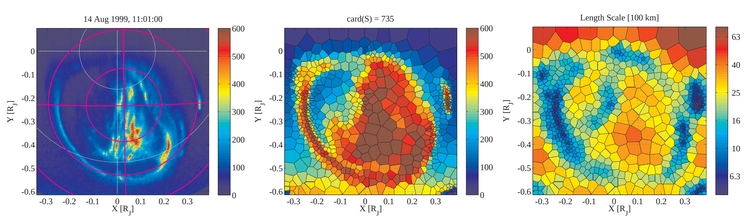

Figure 1

Fig. 1 presents the segmentation process. *Left:* image extracted from the polar projection of Jupiter's UV auroral emission computed from an HST STIS image (FUV-MAMA detector) taken on August 14, 1999 at 11:01 UT (291×301 pixels). The units are in Jupiter's equatorial radius ($1\,\mathrm{R_J} = 71492\,\mathrm{km}$). The origin corresponds to the projection of the northern rotational pole of Jupiter and the y-axis corresponds to the direction of the Sun. The dipole magnetic coordinate grid is shown as magenta solid lines for $10\,\mathrm{deg}$ intervals in magnetic colatitude. Note how arcs of the auroral oval are aligned with magnetic latitude on the dawn sector and the magnetospheric cusp is located around magnetic noon. The footprint of Io is seen at $\sim (0.35, -0.25)\mathrm{R_J}$. *Middle:* map of polygon median intensity I_{median} computed by VOISE [9] for the polar projection preprocessed by median filtering followed by histogram equalisation. The parameters are (i) divide phase $d_m^2 = 4\,\mathrm{pixels}^2$ and $p_\mathrm{D} = 85\,\%$ (ii) merge phase $p_\mathrm{M} = 50\,\%$, $\Delta\mu = 20\,\%$ and $\Delta\mathcal{H} = 30\,\%$ and (iii) regularisation phase: 2 relaxation iterations. *Right:* the length scale ℓ estimated from the geometry of the polygons of the Voronoi map.

Clustering Analysis

Cluster analysis is the art of finding groups in data. The aim of cluster analysis is to form groups in such a way that objects in the same group are similar to each other, whereas objects in different groups are as dissimilar as possible (see e.g. [10]).

Each object is characterised by p measurements and therefore n objects can be organised in an $n\times p$ matrix, where the rows correspond to the objects (seeds or polygons) and each columns corresponds to a property of that seed. The fth measurement of the ith object is denoted x_{if}.

The measurements x_{if} are converted into normalised variables $z_{if} = (x_{if} - m_f)/s_f$, where m_f and s_f are resp. the median and the mean absolute deviation of the fth property over all objects.

The next step is to compute the distance between objects, in order to quantify their degree of dissimilarity. For each pair of objects i and j, we choose the Euclidean distance in z-space

$$d(i,j) = \sqrt{(z_{i1} - z_{j1})^2 + (z_{i2} - z_{j2})^2 + \ldots + (z_{if} - z_{jf})^2}. \quad (1)$$

This collection of distances can be arranged in an $n\times n$ symmetric distance matrix, each entry corresponding to one pair of objects i and j. This is called a dissimilarity matrix. Note that there are many other possibilities to characterise dissimilarity, depending on the type of data [10].

The list of seeds generated by VOISE (each associated with a polygon of data) are the objects we choose to cluster. Let us consider a group of neighbouring seeds, i.e. where each polygon in the group shares at least one edge with another polygon. Such a group of seeds is defined as a cluster if the polygons in the group are similar with respect to a given combination of properties, e.g. seed coordinates, median intensity I_{median} within each polygon, length scale ℓ of the polygon.

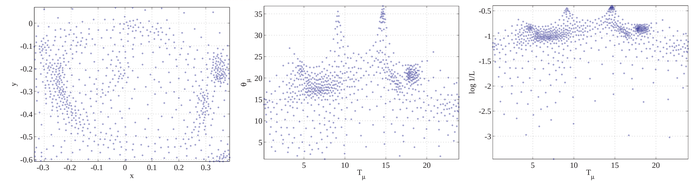

Figure 2

Fig 2 presents seed location in different coordinate systems for the Voronoi map in Fig 1. From left to right, Cartesian coordinates (x, y), magnetic coordinates (T_μ, θ_μ) and $(T_\mu, -\log(L))$. Note how the clustering 'compactness' of the seeds is dependent on the coordinate system.

- When partitioning a set of objects, the main objective is to find clusters whose objects show a high degree of similarity, while objects belonging to distinct clusters are as dissimilar as possible (k clusters in total).
- The Partitioning Around Medoids (PAM) algorithm [10, 12] is based on the search for k representative objects within the data set, called k-medoids or centrotypes of the clusters.
- When a set of k medoids is found, the k clusters are constructed by assigning each object of the data set to their nearest representative object using the dissimilarity matrix.
- The quality of the clusters can be inspected using the so-called silhouette $s(i)$ of each object i. Values of $s(i)$ are bounded in the interval $[-1, 1]$, and measure how well objects have been classified. Values close to -1 mean that the object is misclassified, while values close to 1 are well-classified and values around 0 are intermediate cases.
- A summary is provided by the following values:
  - The average of s for all objects in a cluster is called the average silhouette width of that cluster.
  - The average of s for all objects is called the average silhouette width for the entire data set $\overline{s}(k)$, for a partition into k clusters.
- $\overline{s}(k)$ can be computed for different numbers of clusters k and the silhouette coefficient $\mathrm{SC} = \max_k \overline{s}(k)$ provides a measure of the strength or 'compactness' of clustering structure for an 'optimum' number of clusters k.

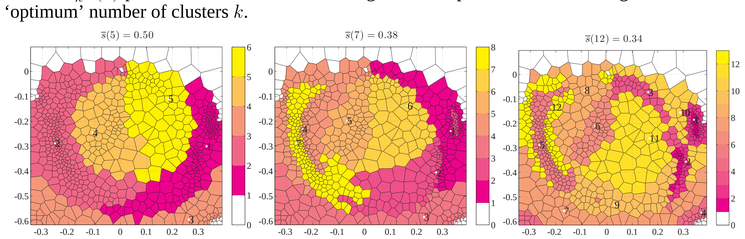

Figure 3

Fig. 3 shows results of Partitioning Around Medoids (PAM) clustering for different combination of properties. *Left:* seed position in the (T_μ, θ_μ)-coordinate system. *Middle:* seed position and the median intensity I_{median} within each polygon. *Right:* seed position, median intensity and length scale ℓ of the polygonal region. The clusters are coloured from largest to smallest (cluster) averaged silhouette width (the label 1 is largest width), and the medoids for each cluster are also labelled. The 'mist' of low intensity regions have been removed from the clustering analysis with a threshold of $100\,\mathrm{kR}$ (the sensitivity of the sensor is $\sim 1\,\mathrm{kR}$) and appear as transparent polygons. The optimum number k of clusters has been determined by running PAM with k varying from 2 to 15 and was found to give the largest silhouette coefficient SC for each combination of properties. Note how the intensity property has constrained clustering around regions of high intensity. Note also that the constraint produce clustering with larger number k of clusters and lower average silhouette width. Note how the scale length property has further constrained clustering around regions with small length scale

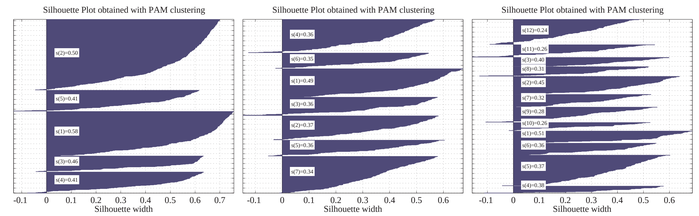

Figure 4

Fig. 4 shows silhouette width plots of all objects organised by cluster, and sorted within each cluster, for the three clustering in Fig. 3. The average silhouette width for each cluster $s(k)$ is labelled for each silhouette profile and the average silhouette value $\overline{s}(k)$ for the entire data are the ones given in Fig 3. Ideally we want the silhouettes to be as uniformly wide as the average of the $s(i)$ for all object i belonging to that cluster. The plot discerns clear-cut from weak clusters.

Conclusion / Future work

VOISE combined with a clustering analysis such as PAM is a promising technique for objective detection of features such as magnetically-organised auroral features in auroral images. The method could be further refined by using a Principal Component Analysis (PCA) with robust statistics as a preprocessing to the clustering analysis in order to reduce the dimensionality of the VOISE data [4]. Other clustering analysis techniques, such as DBSCAN [7, 5] and OPTICS [1, 6] based on density function, which contain explicitly the notion of noise and are able to find arbitrarily shaped clusters, would need to be investigated as well.

Acknowledgements

We acknowledge the support of the HST auroral imaging team at Boston University.

References

[1] M Ankerst, MM Breunig, HP Kriegel, and J Sander. Optics: Ordering points to identify the clustering structure. In *Proc. 1999 ACM SIGMOD - Int. Conf. on Management of Data*, pages 49–60, Philadelphia, USA, 1999. ISBN 1-58113-084-8.

[2] J. T. Clarke, J. Ajello, G. Ballester, L. Ben Jaffel, J. Connerney, J.-C. Gérard, G. R. Gladstone, D. Grodent, W. Pryor, J. Trauger, and J. H. Waite. Ultraviolet emissions from the magnetic footprints of Io, Ganymede and Europa on Jupiter. *Nature*, 415:997–1000, February 2002.

[3] J. E. P. Connerney, M. H. Acuña, N. F. Ness, and T. Satoh. New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *J. Geophys. Res.*, 103:11929–11940, June 1998.

[4] M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, and B. Walczak. Robust statistics in data analysis - a review basic concepts. *Chemometrics Intell. Lab. Syst.*, 85(2):203–219, February 2007.

[5] M Daszykowski, B Walczak, and D. .L. Massart. Looking for natural patterns in data - part 1. density-based approach. *Chemometrics Intell. Lab. Syst.*, 56(2):83–92, May 2001.

[6] M Daszykowski, B Walczak, and D. L. Massart. Looking for natural patterns in analytical data. 2. tracing local density with optics. *J. Chem. Inf. Comput. Sci.*, 42(3):500–507, May 2002.

[7] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. 2nd Int. Conf. Knowledge Discovery Data Mining*, Portland, USA, 1996.

[8] A. C. Fraser-Smith. Centered and eccentric geomagnetic dipoles and their poles, 1600 - 1985. *Rev. Geophys.*, 25:1–16, 1987.

[9] Guio, P. and Achilleos, N. The VOISE Algorithm: a Versatile Tool for Automatic Segmentation of Astronomical Images. *Mon. Not. R. Astron. Soc.*, pages 1051–+, August 2009. Available at http://arxiv.org/abs/0906.1905.

[10] L. Kaufman and P. J. Rousseeuw. *Finding groups in data. an introduction to cluster analysis*. Wiley-Interscience, New York, 2nd edition, 2005. ISBN 9780471735786.

[11] L. Pallier and R. Prangé. More about the structure of the high latitude Jovian aurorae. *Planet. Space Sci.*, 49:1159–1173, August 2001.

[12] S. Verboven and M. Hubert. Libra: a matlab library for robust analysis. *Chemometrics Intell. Lab. Syst.*, 75(2):127–136, February 2005.