Visualization of Very Large High-Dimensional Astronomical Data Sets

B.J. Ferdosi[†], H. Buddelmeijer[‡], J.M. van der Hulst[‡], E.A.Valentijn[‡], J.B.T.M. Roerdink[†]

[†]Institute for Mathematics and Computing Science, University of Groningen
[‡]Kapteyn Astronomical Institute, University of Groningen
{B.J.Ferdosi, J.B.T.M.Roerdink}@rug.nl, {H.Buddelmeijer, J.M.van.der.Hulst, E.A.Valentijn}@astro.rug.nl

Abstract

Enormous size and dimensionality of astronomical data pose a challenge in exploration, analyzing and understanding them. New tools are needed that can not only handle large size (million data points) but also high dimensionality (dozens of parameters). To maintain applicability these tools have to be scalable.

1. High Dimensional Data Visualization Methods

- **Scatter Plot Matrix** shows the pairwise relationship of a set of variables in matrix format.

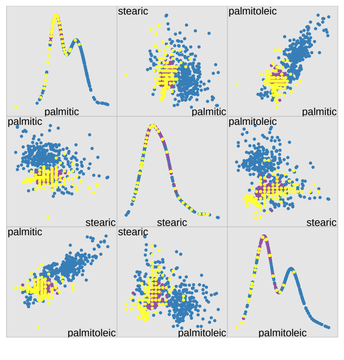

Figure 1: *Scatter Plot Matrix*

- **Parallel Coordinate Plot** [1] constructed with axes in parallel. Suitable for a large number of variables and hundreds of data entries.

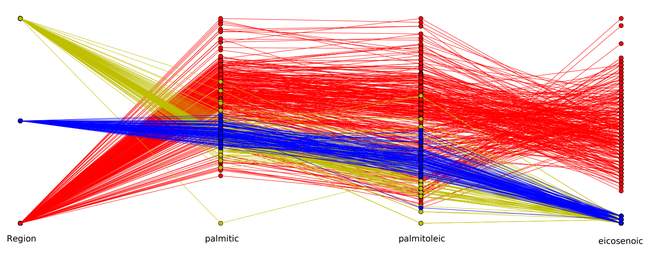

Figure 2: *Parallel Coordinate Plot*

- **Tours** [2] are basically motion graphics: to study joint distribution of non-orthogonal projection of data from a large number of projection directions.

2. Tools

- **GGobi** The visualization methods discussed above are nicely implemented in this tool. Performs better for relatively small datasets.
- **Xmdv** uses hierarchical approach. Data visualization through parallel coordinates. User interaction is not as smooth and intuitive as GGobi.
- **Limn** works well for large ($>$ million data points) data. Utilizes screen binning and multimedia techniques [3]. Data visualization method: scatterplot matrix, tours. Works well with large data but not with very high dimensionality.

3. Multiscale Morphological Method for Pattern Finding

- Application of multiscale methods work well for finding patterns and trends in data [4]. Our proposed method for finding patterns in data is shown in fig. 3.

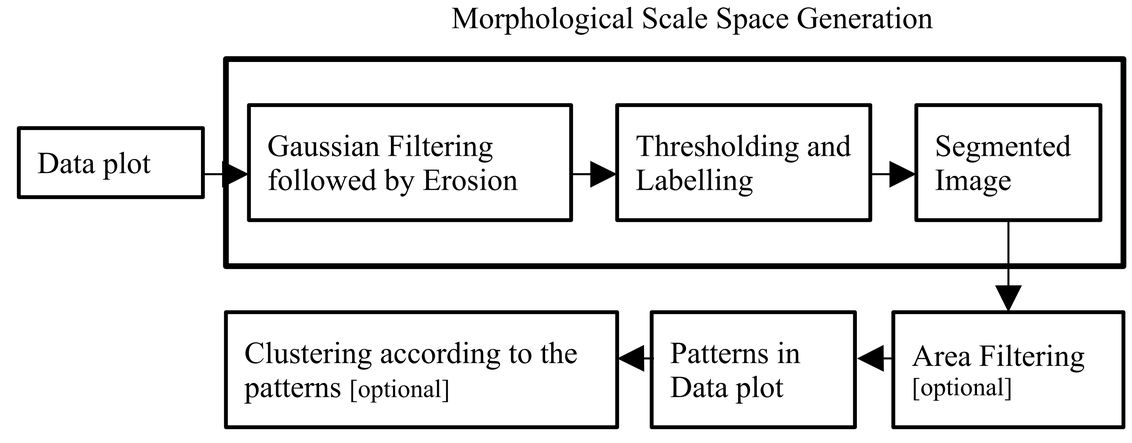

Figure 3: *Method for detecting patterns in data*

4. Datasets

- **Goal:** SDSS + UKIDSS: millions of sources and dozens of parameters.
- **Test Data:** 2dF South Population Photometric dataset [Data: G. Sikkema].
- **Test Question:** Discriminate and identify red and blue galaxy populations in the data using a color-magnitude diagram.

5. Test Results

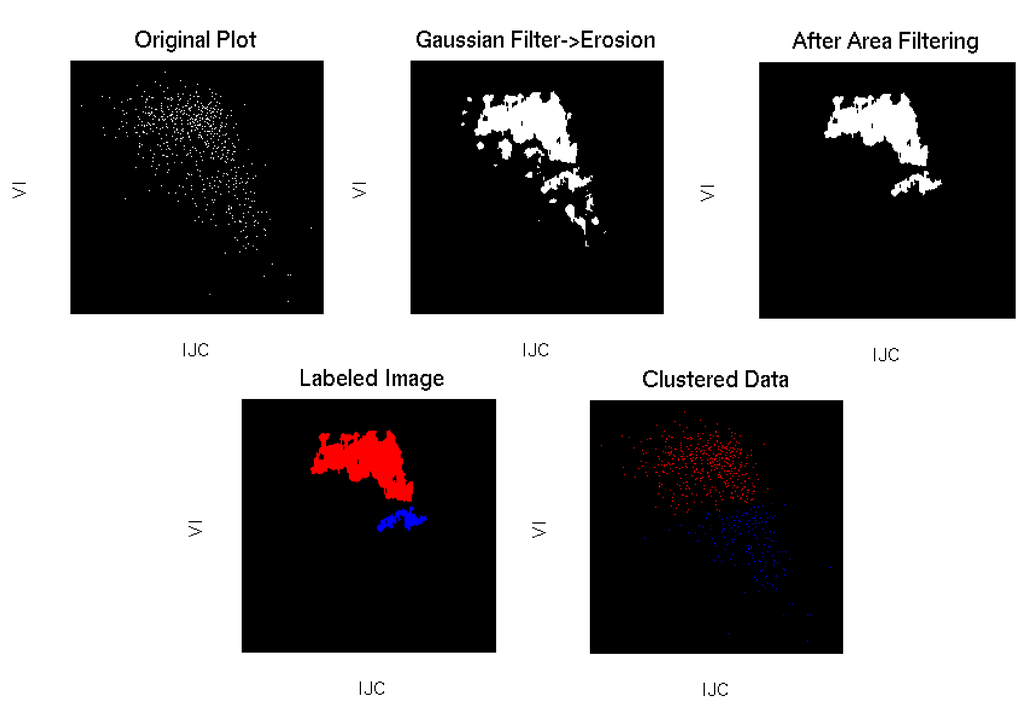

Figure 4: *Sequence of processing steps showing clear bimodality in the color-magnitude diagram*

6. Conclusions

- Multiscale morphological method can be very robust to find patterns in astronomical data
- However, results are very much dependent on the filter size and other parameters
- **Future work:** 1. Investigate the techniques for finding proper range of filter size to produce expected scale space. 2. Extend the method to higher dimensions.

References

[1] *A.Inselberg, B.Dimsdale,Parallel coordinates:A tool for visualizing multidimensional geometry. Proc. of Visualization 90, p. 361-78, 1990.*

[2] *D.Asimov,The grand tour: a tool for viewing multidimensional data,SIAM J. Sci. Statist. Comput., 6, 128-143, 1985.*

[3] *D.Cook,L.Miller,M.Suarez,P.Sutherland,J.Zhang,Using Multimedia Animation with Real-time Graphic Overlays for Visualizing a Million Cases of Multivariate Data, How to Visualize a Million workshop, Augsburg, Germany 2002.*

[4] *M. Wattenberg,A Model of Multi-Scale Perceptual Organization in Information Graphics,IEEE Sypposium on Information Visualization 2003, p. 23-30.*

This work was supported by **NWO**, STARE grant, no. 643.200.501.