3D Identification of clusters: a cancer cell case-study

Gonçalo Figueiró [1], Francisco M. Couto [2], André Moitinho [1]

[1] LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal
[2] SIM/CENTRA, Faculdade de Ciências de Universidade de Lisboa, Portugal

Whether in astronomy or bioinformatics, data driven scientific research is characterized by methods to handle large data sets. Biomedical datasets are no exception and are becoming increasingly complex, particularly because they're derived from multiple sources and hence heterogeneous, unstructured and high-dimensional [1]. It is expected that in coming years the number of biomedical datasets based in 3D point clouds will largely increase. These can be *natural* point clouds from nuclear medicine and scanners or *representative* point clouds in which 2D images are represented as 3D points.

The interactive 3D visualization of point clouds allows the discovery of patterns and relationships within the data, however if the dataset is large enough it might be impossible to visualize all points at once due to constraints in memory and computational power. But even if possible it often isn't the right approach since using traditional visualization techniques with these datasets leads to confusing and even useless representations due to cluttering and overplotting problems. Thus there is a need to represent the data without overbearing the user and allow the retrieval of information from the dataset.

We propose creating a method to identify regions of interest in a point cloud that takes into account not only the 3D positional coordinates but also all other dimensions.

In order to implement and evaluate our method we created a synthetic dataset to model healthy cells and cancer growth. An adapted version of UPMASK [2] was used to compute membership assignment. UPMASK is an unsupervised clustering method for membership assignment, that was previouly used to assess stellar clusters using only photometry and positions. It is data-driven and doesn't rely on models. It is particularly useful for situations where cluster members are buried in field stars. These characteristics make it a good candidate to use with our problem and case-study since cancer cells are also *buried* in healthy cells.

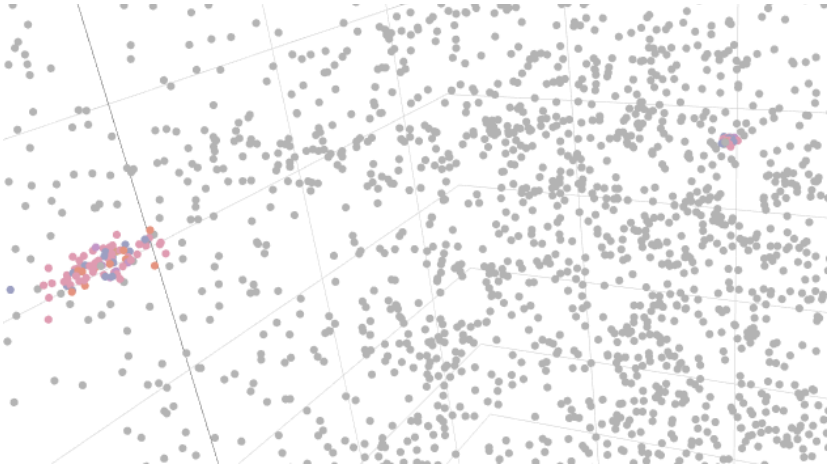

Figure 1– Cancer cells (red and blue), healthy cells (grey)

An initial version of this method allowed the correct identification of cancer cells (Fig. 1) even in situations where healthy cells where very close or even within the structure of the “tumor”.

Preference for presentation: Oral or Poster
Location: University of Minho
Author for Correspondence: gfigueiro@gmail.com

References

1. Andreas Holzinger, Igor Jurisica: Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. Springer (2014)
2. A. Krone-Martins, A. Moitinho: UPMASK: unsupervised photometric membership assignment in stellar clusters (2013)