The basic nature of the data that Visual Data Mining (VDM) deals with is usually visual (images of all sort, satellite scenes, radar scenes, magnetic resonance images, time series of images, photos, movies etc.), but data can also be strictly numerical or literal (natural language texts). Visualizing the data must not be a simple graphical representation, but a technique to increase the informational entropy of a message [Chen 2008, Chen 2010].

Another impulse for developing VDM techniques is the ever growing volume of image databases. The specialists in remote sensing and image processing need tools, preferably automatic ones, for dealing with terabytes of images that need to be organized, understood and used [Datcu 2003, Datcu 2005].

DataVis3D application

- DataVis3D is a prototype of a VDM tool aiming at interactively and efficiently browsing and understanding of the structure of large data sets of EO imaging products.
- One of its main applications will be in EO Payload Data Ground Segment (PDGS) of the German satellite mission TerraSAR-X.
- However, the tool could operate with any type of EO imaging products, such as Sentinel 1 and 2 missions and other similar missions (Spot, Pleiades, Landsat, etc.).
- The current version has implemented the following features:
  - The EO images are partitioned in patches with the size adapted to capture meaningful contextual information; Experimentally the optimal size was found to be of ca. 200 pixels, independently on the image resolution.
  - Patches are indexed by extracting a relevant parameter set and this is represented in a descriptor; Generally the descriptors capture different properties, e.g. radiometric, phase, or geometry, thus will not describe the same information content. A library of specific descriptors for multispectral and SAR is used: Gabor features, Multispectral descriptors, and SIFT. Thus, the whole archive is represented in the n-dimensional space of the extracted features, each patch being a point.
  - For visualization, a multi-dimensional reduction of the n-dimensional descriptors space to a 3-dimensional space is performed;
  - Thumbnails of the images are visualized in a 3-dimensional space associated to their description projections; It is not the usual geographic overlapping of images, but a 3D spatial grouping, by means of their visual resemblance.
  - The GUI provides zoom and rotation navigation in the 3d space, and a selection tool based on a user-defined sphere;
  - The selected tiles may be projected and visualized on Google Earth.

![Screenshot of DataVis3D, running on a set of ~6000 tiles obtained by dividing a Landsat scene into 100 x 100 tiles. On the left main screen, their representation in the 3d space of the extracted features is displayed. In this space, zooming and rotation navigation is provided, together with a selection tool based on a user-defined sphere. The selected tiles may be projected and visualized on Google Earth.](image)

References

- C. Chen - An Information-Theoretic View of Visual Analytics, IEEE Comp. Graph. Appl. 28, issue 1, 18, 2008.
- M. Buican, C. Cucu-Dumitrescu, F. Serban, D. Teleaga - A 3D Visual Data Mining Application, Image Information Mining Conference: Knowledge Discovery from Earth Observation Data, DLR, Oberpfaffenhofen, Germany, 2012.